Assessing the welfare of extensively managed sheep: an evaluation of animal-based welfare indicators

Susan Emily Richmond

Thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

The University of Edinburgh

2016

Declaration

I declare that I have composed this thesis. This is my own work, and any assistance received during the project has been duly acknowledged. The work described has not been submitted for any other degree or professional qualification.

Susan Emily Richmond

September 2016

Abstract

The increased public interest in the welfare of animals used in food production has led to the emergence of welfare assessment schemes for a range of livestock species. There are currently over 100 million sheep in the EU which are primarily bred for milk, meat and wool production and the majority of these animals are managed extensively. The general perception of sheep in extensive systems living "natural lives" with few welfare compromises, along with the practical challenges of adequately assessing their welfare, has caused them to be largely ignored in comparison to other species. It was only relatively recently that the lack of animal-based welfare assessments for extensively kept small ruminants was recognised. Thus, the aim of this project was to evaluate potential animal-based welfare indicators for use during on-farm welfare assessments of extensively managed sheep. The current project used the Welfare Quality® 4 principles and 12 criteria as a foundation for selecting indicators for the assessment of extensively managed sheep. Following a comprehensive review of the scientific literature and a meeting attended by experts from across the EU, 16 indicators were selected for evaluation. Each principle and criteria were covered by at least one of these 16 indicators ensuring the main aspects of sheep welfare were addressed. The indicators selected for investigation could either be applied without handling or gathering the animals during an `Assessment in the Field´, or during a more thorough hands-on `Assessment at Gather´. The reliability, feasibility and validity of some indicators (e.g. body condition scoring) are already established. However for others (e.g. Qualitative Behavioural Assessment approach), at least one of these criteria required further investigation before the indicator could be accepted. The reliability of selected measures was evaluated by assessing their repeatability and inter-observer agreement. Face validity was assumed for the indicators selected during the expert

meeting, and further cross validation was performed when appropriate using additional information collected on the animal's physical health status. During the Assessments at Gather blood samples were collected for the assessment of metabolic profiles, and faecal samples provided gastro-intestinal parasite counts.

The Assessments at Gather were performed on the same 100-135 Scottish Blackface ewes across a two year period (2011-2013) on a Scottish hill farm. The Assessments at Gather occurred five times a year coinciding with key points in the ewes' reproductive cycle: pre-mating, mid-pregnancy, late pregnancy, mid lactation and weaning. During the assessments data were collected on the ewe's body condition, coat cleanliness, faecal soiling score, respiratory conditions, anaemia, lameness and udder condition (udders assessed during lactation only). Current stage in the reproductive cycle and seasonality were found to have significant impacts upon the long-term reliability of the following measures: body condition score, tooth loss, nasal discharge and anaemia scores ($P<0.001$) with variation both within, and between years. On commercial farms older and less productive ewes tend to be removed from the flock once a year prior to mating. Of the indicators applied to the ewes during the Assessments at Gathers, tooth loss and body condition score were the best predictors for the ewe's exclusion from the flock, predicting the future removal of a ewe from the flock 12 months in advance of the shepherd's decision. For the Assessments in the Field, indicators which did not necessitate close contact were required. A whole-animal method (Qualitative Behavioural Assessment (QBA) was therefore particularly useful as it can be performed with minimal disturbance. Rather than quantitatively scoring the behaviour patterns of an animal the assessor focusses on how the animal interacts with their environment. This information is translated in to qualitative descriptors such as "calm" or "agitated". Good inter-observer reliability was found when three observers assessed 49 individual ewes on two occasions ($W=0.77$, $P<0.001$). When QBA was applied 13 times to 50 individual ewes over a six-month time period (spanning late pregnancy to post-weaning) four meaningful Principal Components were identified; the first two accounted for more than half of the explained variation between sheep. The two main components were 'General Mood' (PC1), describing the overall affective state of the ewe, and 'Arousal' (PC2) reflecting energy levels. General Mood scores significantly increased in the post-lambing period compared to pre-lambing observations, and significantly increased again post-weaning ($P<0.001$). Ewes were significantly experiencing significantly higher Arousal in post-lambing and post-weaning conditions compared to pre-lambing ($P<0.001$), but there was no difference between post-lambing and post-weaning. During the Assessments in the Field data were also collected on: the ewe's response to human approach, a surprise test, the ewe's social group size, group demographics and behavioural synchrony. Ewes with lower mood scores tended to have larger distances between them and other ewes ($P=0.023$). The distance to which a human could approach before the ewe fled was significantly related to Arousal ($P=0.05$), as ewes in a higher energy state fled from the approaching human sooner than those who were in lower Arousal states. Ewes in social groups with higher numbers of ewe and lamb vocalisations tended to have lower General Mood scores ($P=0.014$), and lower Arousal scores ($P<0.001$) than those in smaller groups.

Indicators which met the conditions of feasibility, reliability and validity (for example, those reported above) proved to be suitable for use when assessing the welfare of extensively managed sheep. The effect of time on the reliability of the indicators applied during the assessments have important implications for understanding temporary fluctuations in the animal's welfare caused by either internal (reproductive state) or external (environmental) factors. These fluctuations may not be representative of a farm's overall welfare levels in the long term and therefore further careful consideration of the most appropriate time to apply the selected indicators is required.

Lay Summary

The increased public interest in the welfare of animals used in food production has led to the emergence of welfare assessment schemes for a range of livestock species. There are currently over 100 million sheep in the EU which are primarily bred for meat, milk and wool production. The majority of these animals are managed extensively, i.e. in systems which require relatively low human labour input and the animals spend most of their lives outside. The general perception of sheep in extensive systems living "natural lives" with few welfare compromises, along with the practical challenges of adequately assessing their welfare, has caused them to be largely ignored in comparison to other species. Until relatively recently the only welfare assessments available for these animals focussed on the resources available to the animals (resource-based assessments), however the need for animal-based assessments which provide information on how animals cope with their situation has been recognised. Thus, the aim of this project was to evaluate potential animal-based welfare indicators for use during on-farm welfare assessments of extensively managed sheep.

Previously, the Welfare Quality® project developed welfare assessment protocols for intensively managed pigs, cattle and poultry, and devised four welfare principles: Good Feeding, Good Housing, Good Health and Appropriate Behaviour. Within these four principles they highlighted 12 criteria which underpin the welfare assessments and ensure the animals experience good physical and psychological welfare. The current project used these principles and criteria as a foundation for selecting indicators for the assessment of extensively managed sheep. Following a comprehensive review of the scientific literature and a meeting attended by experts from across the EU, 16 indicators were selected for evaluation. Every principle and

criteria were covered by at least one of these 16 indicators ensuring the main aspects of sheep welfare were addressed. It is important that welfare indicators are: reliable (good consistency over time or agreement between observers), feasible (practical on farm) and valid (i.e. it measures what we intend). For some selected indicators these were already well established, for example body condition scoring (how much fat the animal is carrying), however for others such as the Qualitative Behavioural Assessment approach, at least one of these criteria required further investigation before the indicator could be accepted. The reliability of the selected indicators was tested both by assessing the agreement between observers, and the consistency of results over time. The validity of specific indicators was tested by collecting additional information on the animal's physical health status. Collecting blood samples from the animals allowed us to assess their metabolic and immune system function, and gastro-intestinal parasite counts were conducted on their faeces.

The indicators selected for investigation could either be applied without handling or gathering the animals during an `Assessment in the Field´, or during a more thorough hands-on `Assessment at Gather´. During the Assessments at Gather the same 100-130 ewes were individually assessed across a two year period (2011-2013) on a Scottish hill farm. The Assessments at Gather occurred five times a year at key points in the ewes' reproductive cycle: pre mating, mid pregnancy, late pregnancy, mid lactation and weaning. The stage in the reproductive cycle and current calendar season were found to have significant impacts upon the long-term reliability of the following measures: body condition score, tooth loss, nasal discharge and anaemia scores with variation both within, and between years. On commercial farms older and less productive ewes tend to be removed from the flock once a year prior to mating. Of the indicators applied to the ewes during the Assessments at Gather, tooth loss and body condition score were the best predictors for the ewe's exclusion from the flock, predicting the future removal of a ewe from the flock 12 months in advance of the shepherd's decision.

For the Assessments in the Field, indicators which did not necessitate close contact were required. A whole-animal method (Qualitative Behavioural Assessment (QBA)) was therefore particularly useful as it can be performed with minimal disturbance. Rather than scoring what behaviour an animal is carrying out (e.g grazing or resting) QBA focusses on how the animal interacts with their environment. This information is translated into qualitative descriptors such as "calm" or "agitated". Good agreement between observers for QBA was found when three observers assessed 49 individual ewes on two occasions. When QBA was applied to 50 individual ewes over a six-month time period (spanning late pregnancy to post-weaning) two main meaningful emotional dimensions were identified; 'General Mood' and 'Arousal'. Changes were observed in both the ewes General Mood and Arousal over the six-month time period. General Mood increased in the post-lambing period compared to pre-lambing observations, and increased again post-weaning. Ewes were in significantly higher Arousal states in post-lambing and post-weaning conditions compared to pre-lambing, but there was no difference between post-lambing and post-weaning. Relationships were also found between the QBA results and other indicators applied during the Assessments in the Field. For example the ewes' General Mood and nearest neighbour distance were related, indicating that ewes with lower General Mood scores tended to have larger distances between them and other ewes. The distance to which a human could approach before the ewe fled was related to Arousal, as ewes in a higher energy state fled from the approaching human sooner than those who were relaxed. Ewes in groups with higher numbers of ewe and lamb vocalisations tended to have lower General Moods but were in higher states of Arousal.

Indicators which met the conditions of feasibility, reliability and validity (for example, those reported above) proved to be suitable for use when assessing the welfare of extensively managed sheep. The effect of time on the reliability of the indicators applied during the assessments have important implications for understanding temporary fluctuations in the animal's welfare caused by either internal (reproductive state) or external (environmental) factors. These fluctuations may not be representative of a farm's overall welfare levels in the long term and therefore further careful consideration of the most appropriate time of year to apply the selected indicators is required.

List of conference proceedings and presentations

The work in this thesis has been presented as oral and poster presentations.

<u>Oral presentations at international meetings</u>

Beltran de Heredia, I., Arranaz, J., **Richmond, S.E.,** Dwyer, C.M., and Ruiz R. (2015) Animal-based welfare indicators in dairy sheep: preliminary results from on-farm monitoring . Second Dairycare Conference (Cordoba, Spain).

Richmond, S.E., Wemelsfelder, F., and Dwyer, C.M. (2014) The Development of a Welfare Assessment Protocol for Extensively Managed Sheep AWIN Annual Meeting (Prague, Czech Republic).

Richmond, S.E., Dwyer, C.M., Wemelsfelder, F., Beltran, I., Ruiz, R. (2014) Sheep Welfare Indicators. Welfare Assessment Protocol. Copa Cogeca sheep working group (Brussels).

Richmond, S.E., Georges, C., Baxter, E.M., Wemelsfelder, F., and Dwyer, C.M. (2013) Does handling experience alter the response of sheep to the presence of an unfamiliar human? 47th Congress of International Society for Applied Ethology (Florianopolis, Brazil).

Dwyer, C.M., Lesage, S., and **Richmond, S.E.** (2013) Neonatal lamb rectal temperature, but not behaviour at handling, predicts lamb survival. 48th Congress of International Society for Applied Ethology (Vitoria, Spain).

Stilwell, G., Dwyer, C.M., Ruiz, R., Vieira, A., Ajuda, I., Battini, M., Mattiello, S, and **Richmond, S.E.** (2013) Body condition score in small ruminants: Developing an indicator for hunger. AWIN 2nd Annual Meeting (Vitoria, Spain).

<u>Oral presentations at regional meetings</u>

Richmond, S.E., Wemelsfelder, F., and Dwyer, C.M. (2014) The inter-observer reliability of fixed list QBA assessments during on-farm welfare assessments of hill sheep ISAE UK & Ireland Regional Meeting, 2014 (Edinburgh, UK).

Poster presentations at international meetings

Richmond, S.E., Dwyer, C.M., and Wemelsfelder, F. (2014) Qualitative Behavioural Assessment and variation in the demeanour of individual hill sheep in a longitudinal study. 6th International Conference on the Assessment of Animal Welfare at Farm and Group Level (WAFL) (Clermont-Ferrand, France).

Dwyer, C.M., **Richmond, S.E.,** Wemelsfelder, F., Beltran, I., Ruiz, R. (2014) Assessing seasonal variation in welfare indicators in extensively managed sheep. WAFL (Clermont-Ferrand, France).

Beltran De Heredia, I., **Richmond, S.E.,** Wemelsfelder, F., Ruiz, R., Arranz, J., Canali, E., and Dwyer, C.M. (2014) WAFL (Clermont-Ferrand, France).

Richmond, S., Georges, C., Baxter, E.M., Wemelsfelder, F., and Dwyer, C.M. (2013) Does handling experience alter the response of sheep to the presence of an unfamiliar human? AWIN, 2ND Annual Conference (Vitoria, Spain).

Poster presentations at regional meetings

Richmond, S.E., Wemelsfelder, F., and Dwyer, C.M. (2014) Assessing sheep demeanour. Scot Sheep, National Sheep Association. (Berwickshire, UK).

Richmond, S.E, and Dwyer, C.M. (2014) How long does a bad winter last? Impacts on sheep health and production of a poor winter. Scot Sheep, National Sheep Association. (Berwickshire, UK).

Acknowledgements

I would like to thank sincerely my principal supervisor, Cathy Dwyer for all her advice and guidance throughout my project. I genuinely could not have had a better advisor, or one so willing to read numerous drafts of chapters, even whilst on holiday.

I also wish to thank my secondary supervisor, Francoise Wemelsfelder. It was her QBA approach which first led me to SRUC for my month's placement in 2011, and this was the stepping stone for my transition from working with primates to livestock. I have thoroughly enjoyed this aspect of the project. Francoise's energy and enthusiasm for my efforts helped me to keep going, even when I was knee and waist deep in snow on the hills!

I have benefitted greatly from the help of BIOSS statisticians, Iain and Javier. Their advice with this particularly troublesome data set was invaluable. They were incredibly patient with me and their optimism and excitement throughout our meetings certainly made the tasks of analysis much easier.

Thank you to Sarah Hall, our lab manager, for all her advice and support with the lab work – especially when it seemed, for such a long time, that nothing was working for so long! I also want to say thank you to Jenny Coe for her help in lab and the AWB tech team: Marianne Farish, Matt Turner, Eilidh Baker, Agnieszka Futro, Jo Donbavand and Mhairi Jack for their assistance on farm.

I really enjoyed working with the farm staff, Keith, Malcolm and Maureen and would like to thank them for their help with the ewes and not laughing too much when I arrived at the farm with rather odd items for surprising the sheep! The sheep were also incredibly patient and put up with my antics for two years, and giving me only a few nasty bruises in the process!

Thanks to visiting French student Charlotte Georges who assisted me and kept me company while lambing and watching the videos collected for Chapter 7. Also, thanks to MSc student Katie MacDonald who kept me sane on the visits up the hill. I really appreciated our silly moments when it seemed like we would never find those missing ewes! I would also like to mention the other PhD students, specifically Jo Pollock, Suzanne Desire, Kay Russel, Carol Thompson, Sarah Ison, Lucille Bellegarde, and Claire Donaldson. I am very grateful for their friendship and support during the roller-coaster of emotions we have endured over the last few months.

Big thanks to Emma Baxter and Fritha Langford for all their academic and general life advice. I definitely owe you both a beer!

And last, but certainly not least I would like to thank my family. Massive thanks to my little mum for all her moral support, listening to me chat about sheep and for doing my washing when I got too busy. Thank you to dad and my sister Karen for your help in improving my use of the English language. Now that I will have more free time I will love to be nearer to my little nephew Angus as he grows up and can't wait to teach him all about sheep – whether he likes it or not!

List of abbreviations used throughout this thesis

Abbreviation	Definition
BCS	Body Condition Score
BHB	ßeta- hydroxybutyrate
MCHC	Mean Corpuscular Haemoglobin Concentration
NEFA	Non-esterified *fatty acids*
PCV	Packed Cell Volume
QBA	Qualitative Behavioural Assessment
RBCC	Red Blood Cell Count
SEM	Standard Error of the Mean
QBA	Qualitative Behavioural Assessment

List of Tables

List of Figures

Table of Contents

Chapter 1 General Introduction

1 Sheep production in the European Union

There are currently over 100 million sheep in the EU (Eurostat, 2015) which are primarily bred for milk, meat and wool production. Since their domestication, approximately 11 000 years ago (Chessa et al., 2009), sheep management systems have evolved in order to optimise product yield (Goddard, 2008). The specific management style adopted by farmers depends upon the commodities produced and location (Dwyer and Lawrence, 2008). Although still present in some areas of the world, the traditional pastoral approach of nomadic shepherds has generally given way to the modern intensive and extensive management systems.

1.1 Intensive production

The concept of intensive production is perhaps more readily associated with other livestock species such as pigs and poultry, but sheep may also be managed in a similar, albeit less extreme, system. In intensive systems animals are typically housed both day and night with little or no access to pasture, fed on feed supplementation instead of natural grazing, and have frequent contact with their stock workers (EFSA Panel on Animal Health and Welfare, 2014). The relatively recent increase in the intensification of the sheep industry has led to the permanent housing of dairy ewes and fattening lambs. For example, intensive sheep production is popular Australia where lambs are regularly finished in feedlots (Duddy et al., 2007).

1.1.1 Welfare problems associated with intensive sheep production

Although the sheep in intensive systems typically have freedom of movement than other species, for example the confinement of pigs in sow stalls or crated veal calves, there are specific welfare concerns associated with the intensive sheep production industry. The reduction or elimination of natural grazing and decrease in space allowance impacts upon the ability of the sheep to perform normal behaviours (Fraser, A, 1983). Close confinement with conspecifics can further the spread of disease such as respiratory disorders, lameness or mastitis in dairy sheep (Fitzpatrick et al., 2006; Kilgour et al., 2008). The reproductive behaviour of sheep in intensive systems has also been modified to increase milk yield. Ewes are first mated in intensive dairy systems almost a year younger than their traditionally or extensively managed counterparts (Kilgour et al., 2008).

In some countries in Europe and North America sheep may be kept in a part-indoor, part-out door system in which they spend specific portions of their lives, such as lambing and lactation, housed in intensive conditions but other times on pasture (Kilgour et al., 2008). For some of these animals their time spent on pasture may be more akin to that of an extensive system.

1.2 Extensive production

The concept of extensive often used inconsistently in the literature as variations may occur in the quantity of improved grazing available to the animals, stocking density, ratio between animals and farm employees and restrictions imposed upon the animals movements (Dwyer and Lawrence, 2005). The European Food Safety Authority (EFSA Panel on Animal Health and Welfare, 2014) proposed that farms operating under extensive conditions could be grouped into three classifications: Semi-Extensive, Extensive, and Very Extensive. The first category to be defined was Semi-Extensive in which animals are moved to improved pasture areas (which may or may not be fenced) in which they may stay for days or weeks without housing. If these areas of improved pasture do not meet the nutritional demands of the sheep they may be supplied with supplementary feed. The stock person is somewhat involved with the flock and may check their animals daily however physical contact is not routine. Animals living in the Extensive category defined by EFSA remain unhoused year round, and range on pasture which may, or may not be considered as improved grazing. These animals are also provided with supplementation as necessary however the stock-person does not spend time with the animals other than to move them and carry out essential husbandry tasks when they are gathered. In Very Extensive systems the animals range on unfenced non-improved pastures and do not receive any supplements. The stock-person only visits the sheep rarely and the animals are never housed. Regardless of the degree of the extensiveness, or the purpose for which the sheep are bred, the sheep living in these systems are likely to have the same basic needs and therefore have the potential to suffer similar welfare compromises. In this thesis, the term “extensive” is used to encompass aspects relating to all these categories. This refers to animals which are unhoused for most, if not all, of the year and may or may not receive supplementary feed when foraging on unimproved grazing areas. These animals will typically have low levels of human input and interactions other than when relocating between grazing land or performing essential husbandry tasks.

The free-range aspect of these animals’ lives may be influencing the general public’s perception of these management systems. The public tend to believe that as the animals live a “natural life” they are therefore free of welfare constraints (Caroprese and Casamassima, 2009; Matthews, 1996; Sørensen and Fraser, 2010). These beliefs however may not be accurate. Contrary to popular belief the unrestricted aspects of extensive systems do not automatically guarantee high welfare standards and these systems often pose unique and complex problems for the animals (P Goddard et al., 2006; Matthews, 1996; Waterhouse, 1996). These welfare concerns have received much less attention in comparison to intensively reared species e.g. pigs, perhaps due to this due to this naive public perception (Dwyer and Lawrence, 2008).

The concept of Animal welfare is multi-dimensional with many definitions encompassing various aspects of the animals' life experiences. The Brambell committee (1969) stated that welfare is a wide term that embraces both the physical and mental wellbeing of an animal and FAWC used this to develop the Five Freedoms for domestic and captive animals (Farm Animal Welfare Council, 1992). It was later acknowledged that the sheer absence of negative experiences is inadequate and animals should be able to both reach a high level of biological functionality and have positive experiences (Botreau et al., 2009; Fraser, 1993). Specific welfare issues for extensively managed sheep are described in detail below.

1.2.1 Potential welfare issues for extensively managed sheep relating to feeding

Imbalances in nutrition are common in extensively managed animals (Humann-Ziehank et al., 2008) and poor nutrition and metabolic disease have been identified as a major welfare concern for adult sheep (C. J. Phythian et al., 2011). When animals live outside as in extensive systems they are dependent upon the natural availability and quality of forage and are subject to the uncontrolled effects of season and climate (Turner and Dwyer, 2007). Fluctuations in either of these factors may result in the dietary needs of the animals not being met. Grass and other forage material may be unavailable to the animals either due to the plants' failure to thrive in the environment and provide sufficient grazing, or due to adverse weather conditions, such as heavy rain or snowfall, which may lead to the animals being unable to obtain the grass below (EFSA Panel on Animal Health and Welfare, 2014;

Turner and Dwyer, 2007). Even when forage is immediately available to the animals the nutritional content may be insufficient. The problems of malnutrition (in which nutritional conditions are limiting by being improperly balanced) and under nutrition (volume of food is insufficient) are welfare concerns for extensively managed sheep (Caldeira et al., 2007a; Kyriazakis et al., 1998)

Malnutrition and specific mineral imbalances can cause various health issues, for example copper deficiency can result in problems such as anaemia, bone disorders and an increased susceptibility to infections (Underwood and Suttle, 1999) and low levels of vitamin E in pregnant ewes can lead to a high mortality rate (Lui et al., 2014). The chronic hunger resulting from frequent prolonged periods of under nutrition is likely to be a negative psychological experience and may therefore lead to suffering (Dawkins, 2006a; Verbeek et al., 2011). In the worst cases severe food restriction can ultimately end in death. Under nutrition may be a consequence of ineffective husbandry, serious neglect or circumstance. If the famer does not adequately assess the quality and quantity of herbage available to their animals, and provide supplementation where necessary the animals will become undernourished.

Under nutrition of some individuals can also occur even if adequate feed for the group is provided. If social competition for food is high, or spatial constraints prevent the subordinate animals from feeding the less dominant animals may suffer (Bøe, 2012; Thogerson et al., 2009). Improperly designed or installed feeding equipment such as feeder rings or creep feeders can also physically impair the accessibility for the ewes (Bøe, 2012; EFSA Panel on Animal Health and Welfare, 2014). Health problems such as lameness or tooth damage and loss can also prevent animals from feeding normally (McGregor, 2011). To prevent welfare problems for both adult sheep and lambs the farmer should ensure all animals are able to obtain adequate nutrition (McGregor 2011). Efficiently functioning teeth are essential for maintaining good sheep health as incisor wear, damage and loss have been shown to affect the feed intake of sheep leading to reductions in body weight and milk production (Dove and Milne, 1991; McGregor, 2011).

1.2.2 Potential welfare issues for extensively managed sheep relating to environmental challenges

Sheep living outdoors in extensively managed systems may be exposed to environmental hazards. Such hazards may be natural, for example topological features for example unfenced cliffs or rivers, or poisonous plants. Man made products can also be damaging to the animals welfare as they could injure themselves on sharp barbed wire or broken fencing if it is discarded carelessly in their environment.

In extensive production most animals will spend much of their lives outside, however there may be times, usually at key points in their reproductive cycle, when they are housed. Dairy sheep are particularly likely to be housed and although the traditional methods of wide-ranging grazing are still true for much of their lives (Boyazoglu and Morand-fehr, 2001), more intensive husbandry procedures and housing during bad weather is becoming more common (Averós et al., 2014a). The housing provided to these animals during this time may not be suited to the needs of the animals. Although many factors must be considered when evaluating the adequacy of housing conditions, of primary relevance is the availability of space (Averós et al., 2014a; Petherick, 2007). Space limitations can have major impacts in terms of welfare and performance of animals (Estevez et al., 2007). Space is important because ultimately it determines the types of behaviour the animals are able to perform and their duration. Confinement of animals in areas of low spatial availability is known to be stressful (Horton et al., 1991; Sevi et al., 2007) and can have consequences on their behaviour (Averós et al., 2014b; Bøe et al., 2006), physical health (Sevi et al., 1999) and production traits (Gonyou et al., 1985). The behaviour patterns such as those associated with feeding and drinking, excretion, and resting are critical for immediate survival whilst other behaviours such as locomotion, exercise, self-grooming and social behaviours are essential for longer term welfare (Petherick and Phillips, 2009; Petherick, 2007). The movement and space use depends upon the amount of feeding and lying space which are important resources as covered below (Asher and Collins, 2012). Usable space may be influenced by environmental complexity, the complexity and location of resources and the behaviour of conspecifics (Leone and Estevez, 2008). Social animals, such as sheep, are motivated to perform in behavioural synchrony with group members but when unable to do so due to resource and space limitations, their welfare can be compromised (Asher and Collins 2012). These environmental inadequacies lead to competition for space and overcrowding of resources such as feeding areas (Asher and Collins 2012).

Unhoused sheep may not have adequate shelter from extreme weather. The nature of these climactic conditions depends upon their location. Sheep in northern European countries are likely to be exposed to cold conditions, however those in warmer climates, or housed while maintaining a full fleece are more likely to experience heat stress (C. M. Dwyer, 2008). When natural shelter is lacking, it is important that the farmer provides alternatives in order to warrant high levels of welfare and the survival of his or her flock (Deag, 1996). In extensive environments in which the landscape does not offer natural protection the shelter provided to the animals must be suitable and large enough for the whole flock. If space is limited subordinate sheep are often displaced from the shelter (C. M. Dwyer, 2008; Sherwin and Johnson, 1987). If this happens on a regular basis the physical discomfort and aggression levels accompanied with this displacement may lead to these animals becoming chronically stressed.

In wet weather the shelters should provide protection from both the precipitation and wet ground caused by rainfall and improper drainage. Animals having to lie on wet ground can compromise more than one aspect of their wellbeing. Being forced to lie on wet or dirty ground affects the comfort of the sheep. Sheep prefer to rest on soft dry flooring and such conditions depending on the type and amount of flooring materials (Færevik et al., 2005). If provided, bedding materials should be sufficiently comfortable to ensure the animals have enough resting time and space to maintain good welfare and productivity (Norring et al., 2008; Tuyttens, 2005). When bedding is used properly it can provide thermal insulation and absorb excrement, prevent drafts and improve skid and slip resistance and protect animal from hard surfaces (Færevik et al., 2005) all of which can improve comfort and welfare (Teixeira et al., 2014). Competition between sheep for clean lying areas results in the higher ranking individuals monopolising the attractive space (Bøe et al., 2006). Marsden and Woodgush (1986) found that lying space was the second most important resource involved in displacement incidents in sheep. Inadequate lying space, which results in some animals being unable to maintain behavioural synchrony with the flock, can also impact upon the ewe's stress levels (Asher and Collins, 2012). Group behavioural synchrony has been defined as "the observed degree of conforming of behaviour between individuals within a group where conforming of behaviour is performing the same behaviour as the other individuals at the same time point" (Asher and Collins 2012). Behavioural cohesion is a well-known phenomenon in gregarious species such as sheep that show a consistent and synchronous pattern of activity and resting and the inability to coordinate their behaviour can be detrimental to their welfare (Raussi et al., 2011; Rook and Penning, 1991).

Dirty fleeces caused by insufficient clean and dry lying space can also cause subsequent physical welfare problems as it can cause the skin to become irritated and create prime conditions for pathogens and ecto-parasites (Stubsjøen et al., 2011). Dirt on animals' coats pose additional risks to welfare. Damp and dirty environments lead to the spread of specific bacteria which cause painful health problems such as mastitis and lameness (Schreiner and Ruegg, 2003; Winter, 2008). These physical health problems are discussed further in "potential welfare issues for extensively managed sheep related to Good Health".

1.2.3 Potential welfare issues for extensively managed sheep relating to health

Disease and pain caused by husbandry procedures are believed to be the biggest welfare issues faced by sheep (Phillips, 2009). Various health problems have been identified as causes of concern with regards to sheep welfare. At a flock level, lameness has been identified by sheep farmers in Great Britain as their highest cause of concern for poor health (P Goddard et al., 2006; Kaler et al., 2008). Some particular major health issues for sheep depend upon the animal's age and sex (C. J. Phythian et al., 2011). Specifically for young lambs, husbandry procedures such as tail docking and castration were identified as key welfare issues along with gastro-intestinal parasites (Phythian et al 2011). Key concerns regarding infectious diseases, parasitic infection and routine husbandry procedures are described below.

Infectious disease

In 2005 Nieuwhof and Bishop estimated that lameness in sheep cost the UK economy £23.9 million annually. In the UK it is estimated that approximately 10%

of the national flock is lame and the most common infectious causes are inter-digital dermatitis, foot rot and more recently contagious ovine digital dermatitis (Kaler et al., 2008). These infectious causes can quickly spread throughout a flock (Winter, 2008) and over 90% of farmers report lameness in their flocks with a within flock prevalence of 8-10% (Kaler et al., 2010, 2008). It is estimated that 90% of lameness has been caused by foot rot, an infection of *Fusobacterium necrophorum* and *Dichelobacter nodosus* bacteria (Kaler et al 2010). Regardless of the cause, lameness is a major welfare concern as it is a painful condition which if left untreated can be debilitating for the animal (Winter et al 2008). Due to the nature of extensive systems, in which a farmer may go weeks or months without seeing individual animals, it is often not possible for the shepherd to identify and treat health problems such as lameness as easily and quickly as in an intensive system (Dwyer and Lawrence, 2008; P Goddard et al., 2006). This can lead to animals suffering problems for weeks or month which would have otherwise been relatively straightforward to treat (Turner and Dwyer 2007).

Mastitis was also identified by Phythian et al (2011) as an important welfare concern for sheep. Mastitis is an inflammatory condition of the udder usually due to bacterial infection caused by dirty facilities or equipment (Bergonier and Berthelot, 2003; Caroprese and Casamassima, 2009; Olechnowicz and Jaśkowski, 2014; Weary et al., 2006). Poor maintenance of the milking machine, and insufficient washing of hands and equipment are known to increase the risk of the disease (Albenzio et al., 2002; Caroprese and Casamassima, 2009; Olechnowicz and Jaśkowski, 2014). The disease can vary in its severity however it is acknowledged as an important disease in sheep as it limits milk production of the ewe leading to lower lamb growth rates and can even result in ewe and lamb mortality (Blagitz et al., 2014; Clements et al., 2003; Fragkou et al., 2014; Legarra et al., 2007). Mastitis also impairs the welfare of the infected ewe as it is a moderately painful condition (Fitzpatrick et al., 2006), ranked second to the perceived painfulness of dystocia (Thompson et al., 2015). Bergonier and Bethelot (2003) found annual incidence of clinical mastitis is generally lower than 5% of flocks but sub clinical mastitis ranges from lower than 10% to over 50% in some flocks. Effective treatment of the disease requires an early diagnosis and prompt treatment (Fragkou et al., 2014). In dairy flocks udder problems are usually first noticed during milking which occurs at least once a day. However, in extensive meat production systems the udders of ewes do not receive such regular and close attention. This delay in diagnosis and treatment means the early stages of the disease may be missed, potentially leading to severe welfare implications to the ewes (Fragkou et al 2014).

Parasitic infections: Ecto-parasites

Ecto-parasites may represent significant, and possibly fatal challenges to sheep in extensive systems and cause chronic stress and pain (Colebrook and Wall, 2004; Dwyer and Bornett, 2004; Pete Goddard et al., 2006). It is estimated that almost 30% of the global spending on animal health is on the prevention and management of parasitic diseases (Jackson et al., 2012).

Ecto-parasite arthropods, such as the scab mite, live on the skin. As a result of this, direct damage might be caused to the skin and other tissues (Colebrook and Walll 2004). The feeding behaviour of these parasites may also lead to significant blood loss, secondary infestation, itching, and alopecia, and in some cases it results in death of the sheep (Colebrook and Wall 2004). The welfare implications of ecto-parasites may extend past the physical damage. When present in high numbers they can cause alterations in the behaviour of the infected sheep by increasing levels of behaviour such as rubbing and leading to reduced time spend grazing or ruminating, or in some cases leading to self-wounding (Colebrook and Wall 2004).

Blowfly, *Lucilia sericata*, is the primary cause of cutaneous myasis of sheep in northern Europe where the condition is commonly termed “blow fly strike” (Colebrook and Wall 2004). An infection is established when batches of eggs are deposited on the wool and after hatching the larvae feed on the animals skin using specially adapted mouth hooks. This infestation causes extensive tissue damage resulting in inflammation (Colebrook and Wall 2004). The accumulation of faeces in the wool of the breech area (tail, perineum and anus) of sheep has been shown to be a major precursor to fly strike (Broughan and Wall, 2007; Scholtz et al., 2012). Soft and moist faecal material adheres to the wool and builds up around the tail and crutch of the sheep (Reid and Cottle, 1999; Scholtz et al., 2012; Waghorn et al., 1999). As well as loose faeces increasing the risk of fly strike, low faecal viscosity can be an indication of intestinal parasites (Bath and van Wyk, 2009; Broughan and Wall, 2007; Wall, 2007).

Parasitic infections: endo-parasites

Gastro-intestinal parasites are a major problem for both animal health, productivity and welfare (Maia et al., 2014). *Coccidia* are microscopic, single celled parasites which live in the intestines. Coccidiosis, which is a disease of the intestines caused by *Coccidia*, has been observed in almost all sheep rearing countries of the world and it is estimated that most, if not all domestic ruminants will become infected at least once in their lives (Platzer et al., 2005). Although housed animals are mainly at risk (Platzer et al 2005), it is also be a problem for those grazing as it remains unknown how long the parasites can survive on pasture (Henderson, 2010). The number of coccidiosis diagnoses typically increase during the spring and early summer (Elsheikha, 2009). Although these outbreaks generally affect younger animals all age groups are susceptible to infection (Taylor and Catchpole, 1994). Older animals may also be at risk as their resistance decreases due to stress, dietary changes, or concurrent infection (Taylor, 2012). From a welfare perspective, animals with high oocyte counts may not necessarily suffer, although coccidiosis is a welfare concern as symptoms include, diarrhoea, fever, weight loss and in extreme cases, death.

Gastro-intestinal nematode worms are reported to be the major cause of production loss in sheep (Roeber et al., 2013). Nematode worms of the *Strongyle* type, for example *Haemonchus contortus,* infect the surface of the abomasum and intestines of grazing animals such as sheep (Roeber et al., 2013). *Haemonchus* burrow into the mucosal layer of the stomach and consume the animal's blood (Bath and van Wyk, 2009). Until relatively recently this parasite only posed significant concern for animals living in tropical countries, however it has spread through southern Europe becoming increasingly common (Rinaldi et al., 2015). Its discovery in the UK makes it an increasing concern for British farmers (O'Connor et al., 2006; Rinaldi et al., 2015). These worms are a serious concern for animal welfare as the animal experiences severe anaemia, diarrhoea, dehydration and if left untreated the animal may die (Bath and Wyk 2009). When provided to animals with gastro-intestinal worm burdens, anthelmintic treatment has been shown to successfully reduce the incidence of faecal soiling and subsequent fly strike (Broughtan and Wall 2007). However for animals which are infrequently gathered or handled, this treatment may not be readily available.

Management procedures

Diseases and parasitic infections are not the only health concerns for extensively managed sheep. The policy and management decisions made by a sheep farmer can have a profound effect on the health and welfare of the flock. Lambs routinely undergo procedures which are known to be painful such as tail docking, castration and ear tagging (Guesgen et al., 2011; Jongman et al., 2000; Mellor and Stafford, 2000a; Molony and Kent, 1997). In many countries specific husbandry procedures are a legal requirement, e.g. ear tagging (DEFRA, 2003), however even if these procedures are known to be painful there is no legislation regarding the provision of analgesia or anaesthesia. There are steps that a stock person can take however to minimise the adverse effects of these practices on their animals.

Ear tags are required as an effective way to identify individual animals following birth, however care must be taken to ensure the lambs welfare is not compromised as a result (Edwards and Johnston, 1999). There are various metal and plastic tags available on the market but through careful consideration of the tag type and proficiency in inserting the tags unnecessary trauma can be avoided (Edwards and Johnson 1999). Commonly used sheep tags can cause major lesions to the integrity of the ears even when applied properly (Edwards and Johnson 1999). Metal loop tags caused greater damage to the ear at insertion and are also significantly associated with greater lesions than pliable plastic ear tags (Edwards and Johnson 1999). The positioning of the tags is an important consideration too as both metal and plastic tags can become too tight and rip the ear following growth if placed incorrectly at insertion (Edwards et al., 2001).

Tail docking is procedure routinely performed on lambs to prevent the risk of faecal soiling and the subsequent fly strike discussed above (Clark et al., 2011). Although there are benefits to tail docking, the procedure itself is known to compromise the lambs health and welfare in the short term (Bateson, 1991; Guesgen et al., 2011) and also how they perceive pain in the long term (Clark et al., 2014). There are additional welfare concerns associated with tail docking as if the tail is docked too short (not covering anus and vulva) it can lead to an increase in rectal prolapses (Fisher et al., 2004; Thomas et al., 2003) and increase in carcinoma of vulva (Swan et al., 1984). Castration is another management procedure which is done with the aim of preventing welfare problems in the future (e.g. unwanted pregnancies or aggression in males) however the process itself has been shown to cause acute pain in lambs (e.g. Mellor & Stafford, 2000b). Despite the benefits of using analgesics to alleviate this pain it is still estimated that this does not occur frequently (Mellor and Stafford 2000).

1.2.4 Potential welfare issues for extensively managed sheep relating to behaviour

The behavioural needs of sheep may not be met in all extensive environments and the lack of environmental diversity may be detrimental to sheep welfare (Dwyer, 2009). Environmental complexity is important for animals, not only due to the increase in choice of forage material available to grazing animals, but it has also been shown to affect the expression of behaviour (Boissy and Dumont, 2002; Sibbald and Hooper, 2004; Sibbald et al., 2008).

Vigilance is considered to be an indicator of fear, especially for prey animals and it offers insight in to the animal's perception of predation risk (Hopewell et al., 2005; Rieucau and G. A. Martin, 2008). Predation is a particular problem for extensively managed animals. In many European countries large predators such as lynx, wolves, foxes and eagles prey on sheep. The perceived threat of predators has significant impact upon the social behaviour expressed by sheep (Dwyer, 2004). Predators appear to prey most heavily on the periphery of the social group and the best way for a sheep to avoid predation is to be inconspicuous by being synchronous and to aggregate with other individuals who are showing similar behavioural patterns to themselves (Dwyer, 2004; K. E. Ruckstuhl and Neuhaus, 2001). An increase in group size also allows animals to reduce individual vigilance behaviours, leaving the animal free to devote more time to other behaviours such as feeding and social interactions (Roberts, 1996). Prolonged increase in vigilance behaviours have been shown to be detrimental to the animals' nervous system and concentration, resulting in reduction in ability to detect predators and decision making abilities (Dukas and Clark, 1995). An increase in vigilance behaviour may also happen at the expense of other behaviours such as rest or feeding (Rushen and Depassille, 2005). Decreased levels of vigilance have been related to environmental diversity along with changes in grazing behaviour (K. Ruckstuhl and Neuhaus, 2001). Barren environments (such as a flat featureless paddock) which offer sheep no protection from predators have been associated with an increase in alarm behaviours in comparison to hilly areas with trees (Dwyer and Lawrence, 2008).

The relationship between animals and farmers is a complex. Although in extensive environments the role of the human is more remote than during intensive production systems, the relationship between a stockperson and their flock can have considerable consequences on their welfare. Despite the generations of selective breeding since sheep were first domesticated approximately 11 000 years ago (Chessa et al., 2009), one of the most frightening stimuli they experience is humans (Boissy and Bouissou, 1995; Hemsworth and Coleman, 1994; S Waiblinger et al., 2006).

This fear response of animals towards humans is either founded on an absence of habituation to human contact, or a learned negative association (Boivin et al., 2003;

Rushen et al., 1999a). A stable human-animal relationship does not fully develop after one or two interactions but is built up from a series of interactions over weeks or months (S Waiblinger et al., 2006; Windschnurer et al., 2009). Sheep living in extensive management systems typically receive only neutral or aversive contact with people, e.g. restraint, shearing or medication administration. Even before animals are physically handled for these procedures, the working of the flock by people or with dogs is a known stressor for sheep (Dwyer, 2009). The movement of sheep in this way exploits the innate anti-predator tactics of sheep such as flocking together and running. During the domestication process humans learned to take advantage of this response to fearful stimuli and use dogs tactically to manoeuvre their flocks (Dwyer 2009). Although at a group level such a fear response may aid the control of the animals, at an individual level animals that are fearful of people are generally more difficult to work with as they are more likely to attempt to escape and act aggressively, making the situation more dangerous for both parties (Boivin et al., 2003, 1994; D'Souza et al., 1998; Hemsworth, 2003; Rousing et al., 2005; Rushen et al., 1999a; Sorge et al., 2014; Tallet et al., 2006) .

Fear associated with humans is likely to be one of the most detrimental things to an animal's welfare as it can lead to acute or chronic stress (Boivin et al., 2003; Hemsworth, 2003; Jones and Waddington, 1992; Rushen and Depassille, 2005).

1.3 Chronic stress as a welfare concern

Many of the welfare concerns listed above (e.g. fear of humans, competition for resources and prolonged health problems) contribute to the animal experiencing chronic stress. The experience of chronic stress in itself is a welfare concern. When an animal experiences acute stress, the autonomic nervous system is activated to facilitate a typical "fight or flight" response. The autonomic nervous system is part of the peripheral nervous system that controls physiological functioning below the level of consciousness. For example stimulation of the autonomic nervous system promotes an increase in heart rate, digestion, respiratory rate, salivation, perspiration and pupillary dilation (Donkelaar et al., 2011). This is followed by the activation of the hypothalamic pituitary adrenal axis which mediates an endocrine response and leads to the release of hormones such as glucocorticoids and catecholamines, and the mobilisation of the immune system to provide the animal with sufficient energy to combat the stressor (Möstl and Palme, 2002). These adaptations are beneficial when the animal is reacting to an immediate stressor such as the detection of a predator however, long-term activation of the stress response to chronic stressors is maladaptive and a cause of poor welfare. Chronically elevated glucocorticoids for example have been associated with a number of physiological consequences indicative of poor welfare such as hyperglycemia, neuronal cell death, and suppression of the immune and reproductive systems (Sapolsky, 1992; Wingfield and Romero, 2001). Chronic stress has been found to be associated with greater parasite burdens, as chronically stressed sheep are unable to mount as efficient and effective responses to pathogen challenges as unstressed animals (Dwyer and Bornett, 2004).

1.4 The need for livestock welfare assessments

Increased consumer awareness and concern regarding the animals used in food production has led to an upsurge in "ethical consumerism" (Webster et al., 2015) as the public become more selective in the animal products they buy (Blokhuis et al., 2003; Kılıç and Bozkurt, 2013; Martelli, 2007). This increased interest in farm animal welfare has led to the emergence of many different programmes and assessment schemes designed to ensure certain levels of animal welfare in food production systems. On-farm welfare assessments have been developed for numerous livestock species, and for various purposes e.g. advising farmers how to improve the welfare of their animals, inspecting compliance with legislative requirements, implementing welfare certification schemes and comparing systems to refine legislation (Botreau et al., 2009). In order for these schemes to be effective and provide a meaningful assessment of the animals' welfare it is critical that sufficient effort is invested during the development process. Established scientific knowledge regarding the animals' physical and behavioural needs must be paired with expert opinion and beliefs. The inclusion of producers and industry members is vital in order to ensure economic and practical aspects are remembered. It also allows people working in the sector to voice their opinions on what constitutes the important concerns of the industry and can offer insight to problems of which scientists may be unaware (Blokhuis et al., 2003; Fraser, 2006; Hemsworth et al., 2015; Napolitano et al., 2009; Stubsjøen et al., 2011).

1.5 Assessing animal welfare

As stated previously the term "welfare" represents a state within the animals, including their emotions and recognising that fact is critical when devising a comprehensive welfare assessment protocol or scheme (Hemsworth et al., 2015). Challenges are presented when trying to evaluate positive or negative experiences and arrive at conclusions from the animals' perspective as their experience of a situation, including the emotional component, cannot be measured directly (Roger, 2013). However, there are informative physiological, behavioural and qualitative indices that can be used (Hemsworth et al., 2015).

Behaviour is widely used in clinical assessments of animals and is also very well suited to on-farm welfare assessments (Lovatt, 2010; Rutherford, 2002). Behaviour can be seen as a result of an animal's decision making processes and it can be used to assess how well, or badly, an animal is coping within its situation (Dawkins, 2006b). The advantages of using behaviour to assess welfare include that is non-invasive and mostly a non-intrusive measure (Dawkins 2006b). For species such as sheep their social behaviours can be used to derive information about their welfare. Sheep are gregarious animals and choose to associate in social groups (Dwyer 2008). Alterations to their social grouping or behavioural patterns may indicate potential welfare problems (Gougoulis et al., 2010) and therefore aspects of social and group behaviour may prove to be useful indicators during a welfare assessment. For example, human-animal relationships is an important component of a domesticated animal's welfare and testing the animal's reaction to the presence of humans should be used during on-farm welfare assessments (Waiblinger et al., 2001; Winckler et al., 2003; Windschnurer et al., 2008). A potential conflict arises here as for social animals it may be beneficial to test them with other conspecifics to minimise isolation distress (Cattle: Lensink et al 2000 a, b, 2001b, c. Waiblinger and Menke 1999, Waiblinger et al 2003b, Sambraus 1974, Waiblinger et al 2002, 2003b, Rousing and Waiblinger 2004, Boissy and Bouissou 1988, Goats: Lyons et al 1988a, Lambs: Markowitz et al 1998). But, there is potential for disruption caused by the behaviour of the other group members as fear may be transmitted through vocalisations or behaviours (Susanne Waiblinger et al., 2006). It may therefore be beneficial to assess groups of animals, rather than individuals in some instances.

A measure, which covers an aspect of physiology or behaviour that is used to gauge the past, present or future welfare status of an animal, can be regarded as an animal welfare indicator (Sørensen and Fraser, 2010). Relatively recently, the concept of "iceberg indicators" has come to the attention of researchers developing welfare assessment protocols. Iceberg indicators typically provide information on an outcome measure such as production or physical injury following an event, such as the completion of a breeding cycle, or at the end of the animal's life (EFSA Panel on Animal Health and Welfare, 2014; Farm Animal Welfare Council, 1992). More than just an indicator of one particular aspect of the animal's life, e.g. production traits or physical injuries, these indicators may be able to offer more. Comparable to the tip of an iceberg signalling the mass below the surface of the water, these indicators may predict further valuable information regarding other aspects of the animal's welfare (Heath et al., 2014). These indicators by themselves however do not provide sufficient information but they can potentially add a great deal when included as part of a comprehensive assessment (Heath et al., 2014).

The data for production traits such as growth rate, survival or reason for culling may be collected by the assessor contacting the farmer and obtaining records. The advancement of Precision Livestock Farming (PFL) tools is making the collection and monitoring of such data faster and easier. Using linked technologies such as Electronic ID ear tags or boluses, weigh crates and cameras, information on the animals and their environment can be recorded automatically using specialised software (Berckmans, 2014; DEFRA, 2014). This data can be used by farmers to monitor their animals continuously and make adjustments to management practices as necessary to improve productivity and welfare (Berckmans, 2014). Although publications investigating the concept are scarce and the work that has been performed concentrates on intensive systems, the concept of Iceberg Indicators does appear to have potential for use during welfare assessments of extensively manged animals. In addition it is possible that data collected using PFL tools could also be used to keep tabs on farms after an initial welfare assessment as data can be analysed and potential problems, or improvements, identified at any given time.

On-farm animal welfare assessments typically fall into two categories: resource- or management-based, and animal-based. Although resource-based welfare assessments have their merits (for example, they require relatively little training and are time efficient), animal-based measures focus on the animal and are considered to have a more direct view of the animal's welfare (Barnett and Hemsworth, 2009; Mench, 2003; Mollenhorst et al., 2005; Rushen et al., 2011a; Sørensen and Fraser, 2010). Currently, on-farm welfare inspections for small ruminants, such as the Animal Needs Index or the RSPCA 'welfare standards for sheep', typically rely on the assessment of management- and resource-based measures focussing on structural and technical elements and management- related factors (Napolitano et al., 2009; C. J. Phythian et al., 2011).

Although sub-optimal management facilities and environmental factors may impose restrictions on an animals' welfare, assessing these measures alone is insufficient when gauging the animals' experiences and internal state. Specific links between environmental factors and the welfare as experienced by animals remain largely unknown (Capdeville and Veissier, 2001). Thus, the relative impact of each environmental and management factor on the appearance of a given welfare problem remains impossible to predict, particularly on a long term basis (Spoolder et al., 2009). Animals themselves also vary in terms of their genetics backgrounds, previous experiences and temperament and therefore may perceive, and react to, the same environment and situations differently (Sørensen and Fraser, 2010). Consequently the combination of animal- and resource/management-based indicators give the most valid assessment of animal welfare as perceived by the animal (EFSA Panel on Animal Health and Welfare, 2012a, 2012b). Together these approaches can be used to assess the current welfare state of the animal and also predict future experiences.

An example of a welfare assessment scheme in which the animal's experiences remained at the core, is the Welfare Quality® project. During this project, animal scientists created a multi-criteria evaluation model for animal welfare assessment at the farm or slaughterhouse. Using the FAWC Five Freedoms they defined four principles: Good Feeding, Good Housing, Good Health and Good Behaviour (Blokhuis et al., 2003). Within each of these principles, two to four criteria were specified creating a total of 12 key animal welfare criteria with each criterion representing a specific area of concern as shown in Table 1.1. As with many other welfare assessments, the protocols devised in the Welfare Quality project needed to be usable on various farms across the EU, as well as being sensitive enough to detect fluctuations in the welfare states of animals on the farm, and to reflect the welfare state of the flock or herd as a whole.

Table 1.1 The principles and criteria of animal welfare as developed by the Welfare Quality® project

Welfare Principles	Welfare Criteria
Good feeding	1 Absence of prolonged hunger
	2 Absence of prolonged thirst
Good housing	3 Comfort around resting
	4 Thermal comfort
	5 Ease of movement
Good health	6 Absence of injuries
	7 Absence of disease
	8 Absence of pain induced by management procedures
Appropriate behaviour	9 Expression of social behaviours
	10 Expression of other behaviours
	11 Good human-animal relationship
	12 Positive emotional state

The species included in the Welfare Quality® project were pigs, poultry and cattle. When considering the potential welfare problems for extensively managed sheep in the context of the four principles above, the need for such an animal-based welfare assessment is highlighted. This need has been acknowledged by researchers in Europe (Napolitano et al., 2009).

1.6 Current welfare assessment protocols for sheep

Protocols designed to assess sheep welfare have been developed mainly for use with housed or lowland animals and tend to focus on resource- and management-based measures. In Italy, Napolitano et al (2009) developed a welfare assessment protocol for use with non-extensively managed dairy sheep using a modified version of the Animal Needs Index which focusses on resource-based parameters such as space and water availability. The additional animal-based indicators included by Napolitano et al only assessed the physical aspect of welfare using measure such as lesions and cleanliness. Stubsjoen et al (2011) later developed an on-farm welfare assessment protocol using housed sheep in Norway comprising of resource-, management-, and animal-based measures. The animal-based measures selected by these authors also primarily focussed on physical health with only two behavioural components; fear and human animal relationships. The protocol also did not include the assessment of aggression, social and resting behaviours, or any positive emotional states. Phythian et al (2011) identified a number of potential sheep welfare indicators including resource-, management- and animal- based measures, again with a primary focus on physical health, which provided the first step in identifying valid and reliable indicators for welfare of lowland sheep in the UK. The assessments listed here have been developed for use with lowland sheep, housed animals, or those managed in non-extensive dairy systems. Although the underlying biology of sheep in intensive systems is similar to those managed extensively, there may be subtle differences between the animals typically kept in either system. Selective breeding over many generations has led to a divergence in animals typically kept in intensive and extensive systems. When selecting animals for an extensive system particular traits are highly valued: survival traits (Conington et al., 2001), good mothering capability (Dwyer and Lawrence, 2000), and the ability to withstand climactic extremes (Lawrence and Conington, 2008). These traits may be favoured over high production which is the primary driver in intensive systems (Conington et al 2001). Intensively managed animals, including sheep, have been bred with the aim of achieving maximum production. This has led to highly selected commercial animal breed lines e.g. Texel or Suffolk sheep, replacing rustic local breeds across the world (Carneiro et al., 2010; Groeneveld et al., 2010; Mcmanus et al., 2010) .Animals in extensive systems on the other hand tend to have been derived from local breeds giving rise to much higher diversity compared to intensive systems (Carneiro et al 2010). There is high genetic diversity both between and within these local breed populations (Groenveld et al 2010). As well as variation in the animals, the welfare issues experienced by animals in extensive systems may not be directly comparable with others. Some problems extensively managed animals suffer may be unique to the animal's ability to cope with their specific environment, and may thus require specific indicators not relevant to intensive systems. As described at the beginning of the chapter, the experiences and problems encountered by animals in intensive and extensive systems differ. Animals in intensive systems typically suffer much higher behavioural restriction but welfare of outdoor animals is linked to environmental factors such as lack of shelter or sufficient forage, high predation risk and the infrequency of human supervision can lead to insufficient medical intervention. This is not to say that the assessments produced by Napolitano et al, Stubjoen et al and Phythian et al should be rejected entirely, but modification may be required before they can be successfully applied to animals in extensive systems in order to ensure they are assessing true areas of concern for extensively managed sheep.

Another reason the welfare assessments are not readily transferrable between intensive and extensive systems lies in the feasibility of the methodologies. The protocols developed through previous research have been specifically created around the facilities and routines found in intensive systems. When attempting to apply welfare assessment protocols directly to extensive systems they may prove to be impractical. The first challenge faced by assessors in extensive systems, which is not so problematic in intensive systems, is locating and identifying the animals in order to perform the assessment. Specific methodologies, for example the dropping of a red ball from the ceiling of a shed to test fear levels have been well established for housed cattle, pigs and sheep (Forkman et al., 2007). It is simply not possible to conduct this test with animals living outside. Novel approaches are therefore required to collect similar data on extensive animals. Getting close to animals to inspect them physically is also problematic in an extensive environment as these animals typically have a large flight zone (Turner and Dwyer, 2007). The selection and development of measures which don't demand close physical contact is required.

To conclude; a more specific welfare assessment for extensively managed sheep is needed. This should comprise primarily of animal-based welfare indicators with additional resource- and environment-based measures included in the protocol. This assessment must cover the many aspects of welfare with consideration of these animals' specific issues and needs (Broom and Corke, 2002; Caroprese and Casamassima, 2009; Hemsworth et al., 2015; C. J. Phythian et al., 2011). Welfare indicators which have been developed for use with intensively managed animals can potentially be modified for use in extensive conditions and additional novel indicators may be required. Therefore the work described in this thesis was conducted in order to identify and evaluate suitable welfare indicators for use with extensively managed sheep.

In **Chapter 2**, specific welfare indicators which have potential for use with extensively managed sheep are identified. Evidence of their validity, feasibility and reliability is presented and the need for additional testing is discussed. In **Chapter 4**, selected indicators ae applied to Scottish Blackface hill sheep using the methodologies described in **Chapter 3**. In **Chapter 5** the inter-observer reliability and longitudinal variation in sheep expression (as measured by QBA) are assessed. These data are investigated in further detail in **Chapter 6** along with the convergent validity, feasibility and reliability of additional indicators applied during field observations. Finally, a discussion of the main findings and potential implications of the work reported in this thesis are presented in **Chapter 7**.

Chapter 2 The selection of potential animal-based welfare indicators for use with extensively managed sheep

2.1 Introduction

Any indicators selected for inclusion in a welfare assessment for extensive managed sheep would have to be applicable to sheep kept in a variety of environments; from outdoor paddocks and mountainous terrain where they spend much of their lives, to sheds in which they may spend some time (P Goddard et al., 2006; Waterhouse, 1996). Ease of use is also important, as measures which can be easily understood by assessors, producers and stock workers are more likely to be implemented and applied on-farm (Marchewka and Watanabe, 2013). It is essential that the indicators selected for an on-farm welfare assessment are valid (relevant to sheep welfare), reliable (produce consistent results when performed at different time points or by different assessors) and feasible (time and labour efficient), in order to deliver a relevant tool.

Validity is the main priority when selecting indicators for use in an on-farm welfare assessment (C. J. Phythian et al., 2011). There are a number of forms of validity on which to make this judgement. The concept of **validity** consists of **accuracy,** meaning the measure is free from systematic errors, **specificity** in answering the questions asked**,** and **scientific validity** (Martin and Bateson, 1993)**.** The **specificity** of a measure relates to the extent to which it is associated with its intended application. Within this specificity, the convergent and discriminant validation are important (Chronbach and Meehl, 1955). **Convergent validation** asks whether theoretically related measures are empirically associated with one another. This can be performed by testing for correlations between measures that are expected to be related. **Discriminant validation** however ensures that measures which are considered to be unrelated are, in fact, independent of one another. The **scientific validity** assesses whether the method provides scientifically relevant information and answers the appropriate research question (S Waiblinger et al., 2006). If the measure answers the research question it is considered to have **internal validity**. If this validity and relevance can be extended to other situations, experiments or animal populations its **external validity** is accepted (Waiblinger et al 2006). **Face** and **consensual validity** can be established by a consensus of opinion, for example during an expert panel or focus group (C. J. Phythian et al., 2011). In the absence of a "gold standard" or reference test for welfare assessment (C. J. Phythian et al., 2011; Rushen and Depassille, 2005) previous welfare research has also based the selection and initial validation of welfare indicators on expert opinion (Bracke et al., 2008; Cronin et al., 2002; C. J. Phythian et al., 2011; Whay et al., 2003). This is because expert opinion is considered to provide both face and consensual validity to the welfare indicators (C. J. Phythian et al., 2011; Scott et al., 2001).

It is also vital to assess the potential indicators' **reliability**. **Inter-observer reliability** is particularly important to establish in situations with different observers and animals, as is often the case in on-farm welfare assessments (Bokkers et al., 2012; Kaler et al., 2009; Martin and Bateson, 1993). This is due to the potential risk that differences recorded between groups of animals may in fact be observer differences (Martin and Bateson 1993).

Consistency in measures over time (repeatability) is another important aspect to be considered when assessing potential welfare indicators. Welfare assessments repeated on the same farm at different times may not always yield identical results. If discrepancies are found between the first visit and subsequent assessments, it may be difficult to identify whether the differences are due to poor intra-observer reliability (poor repeatability), a new welfare problem, or if it is merely linked to normal fluctuations in the animal's state (Phythian et al., 2015; Temple et al., 2013).When making comparisons between the scores received by the same farm over time, or comparing between farms, there are potential factors to consider; e.g. the animals' ages, reproductive history and their current position in the reproductive cycle may differ between farms or even at the same farm during different seasons (Rushen et al., 2011b). As animal-based welfare assessments focus directly on the animals' experience of a situation, it is critical to remember the potential impact these factors may have upon the reliability of a welfare assessment.

Finally, the feasibility of measures is key for welfare indicators. In order for welfare indicators to be adopted by farmers, or be included in welfare assessment schemes, they have to be user friendly and not impose excessive stress on the animals. The time taken to complete an assessment is critical and other welfare assessment schemes such as Welfare Quality have received criticism regarding the time taken to complete an assessment (Stubsjøen et al., 2011).

A clinical assessment of physical health is an essential component for a welfare assessment (Lovatt, 2010). However, when assessing the welfare of extensively managed sheep, logistical challenges can arise and such an assessment may not be possible: extensively managed animals may be difficult to identify individually and a large flight distance may prevent assessors approaching and handling the animals. Animals may range across a large area or number of locations making gathering animals for inspection expensive in terms of both time and labour. Gathering may also alter the welfare of the animals being assessed and might be unsuitable at particular times of year e.g. during mating and when lambs are at foot (Turner and Dwyer, 2007). Only performing a physical health check on the animals also results in other aspects of welfare being overlooked, for example assessing the behaviour of undisturbed animals can offer insight in to the psychological aspects of their welfare.

Thus, for these reasons, additional behavioural welfare indicators which do not necessitate close contact, and can be performed with minimal disturbance, are also required when performing a welfare assessment on extensively managed sheep (Fitzpatrick et al., 2006; Phythian et al., 2015; Turner and Dwyer, 2007). The aim of the work presented in this experimental chapter was to identify and select potential welfare indicators for use with extensively managed sheep. This was achieved by performing literature search followed by an expert panel meeting. The validity, reliability and feasibility of the potential welfare indicators were considered before a list of candidate indicators was agreed upon.

2.2 Materials and Methods

The methodology of the work described in this chapter comprised of a two stage approach. Firstly, a literature search was conducted to collate a database of academic papers in which potential welfare indicators were described. Secondly, the papers were reviewed and evidence for the indicators' feasibility, reliability and validity were discussed by an expert panel.

2.2.1 Stage 1 – Literature search

The 4 principles and 12 criteria outlined in the Welfare Quality project (Canali and Keeling, 2009) were the starting point for developing the list of potential sheep welfare indicators to be evaluated in this project. The principle of 'Good Housing' was broadened and renamed "Good Environment". This allowed the principle to be applicable to animals in both housed and non-housed conditions (e.g. 50% of UK sheep flocks are never housed, and most sheep production systems involve at least some outdoor management). The criterion "ease of movement", however, was considered to be only applicable to housed sheep.

A literature search was performed using the online database Web of Knowledge (http://apps.webofknowledge.com/). The primary goal of this review was to identify preliminary welfare indicators for use with extensively managed sheep. All data bases were included in the search with no limits set and the timespan was set to include the earliest possible year (1864) to the present and the language filtered to English. The search terms 'sheep', 'welfare' and 'indicator' were initially used and in order to capture as many potential indicators as possible additional searches were conducted using the terms 'assessment'; in place of indicator, and 'pain' in place of welfare as well as additional searches for each criteria. If no suitable indicators were yielded from these searches the terms were also widened to include other ruminant species (goats and cattle). Initial searches were conducted in 2011 (prior to my involvement in the study), and I conducted a later search in 2015 to account for new developments in the literature. Papers which were deemed to be irrelevant based on title or key words were removed. The information found during this literature search resulted in the formation of a list of candidate animal-based welfare indicators for each welfare principle and criterion. The next step was to assess the evidence of the candidate indicators' validity, feasibility and reliability.

2.2.2 Stage 2 - Expert panel meeting

The literature obtained was catalogued based on their applicability to the four Welfare Quality principles and twelve criteria. Evidence in support of the validity and type of validity available, reliability and feasibility for on farm assessment were gathered from the literature where available (shown in Tables 2.1 – 2.4). To refine this list, an expert meeting was conducted during which five international animal welfare and production scientists (whose experience ranged from 3 to 20 years) discussed each indicator in detail. These participants were from the UK, Spain and

Italy, countries which, when combined, represent more than 50% of sheep production in the EU (Eurostat). On the basis of whether the indicators had proven validity, reliability and feasibility indicators were accepted, rejected or selected for further evaluation and development.

2.3 Results

2.3.1 Stage 1 - Results from the literature search

A list of 28 potential animal-based sheep welfare assessment measures from 30 papers were accumulated from the literature search as shown in Tables 2.1-2.4. Information regarding the established validity, feasibility and reliability of the indicators was drawn from the literature found during the search. For some of these indicators there were numerous studies confirming aspects of the validity, feasibility and reliability e.g. body condition scoring. For others, such as behavioural synchrony, this information was lacking and therefore discussion and consultation with an expert panel was required before a decision could be made regarding its use as a welfare indicator for extensively managed sheep.

Table 2.1 Potential welfare indicators obtained from the literature regarding Good Feeding. Under the headings of Feasibility, Validity and Reliability "Yes" indicates this has been tested and confirmed, NT means Not Tested.

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivity	Specificity	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
Good Feeding	Absence of prolonged hunger	Body Condition Score	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Caroprese et al 2009,Napolitano et al 2009, Morgan-Davies et al 2008, Phythian et al 2011, Stubsjoen et al 2011, Caldeira et al 2007, Phythian et al 2010, Maitland et al 1995
		Lamb birth weights and % lambs weaned	Farm records	Yes	Yes	NT	NT	NT	NT	NT	Stott et al 2012
	Absence of prolonged thirst	Provision of clean water	Resource	Yes	Yes	NT	Yes	NT	Yes	NT	Napolitano et al 2009, Verksler et al 2008

Table 2.2 Potential welfare indicators selected from the literature regarding Good Environment. Under the headings of Feasibility, Validity and Reliability "Yes" indicates this has been tested and confirmed, NT means Not Tested.

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivity	Specificity	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
Good Environment	Comfort around resting	Coat cleanliness	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Caroprese et al 2009, Napolitano et al 2009, Stubsjoen et al 2011, Phythian et al 2011
		Space allowance and stocking density	Resource	Yes	Yes	NT	NT	NT	NT	NT	Caroprese et al 2009, Pines et al 2007
		Flooring slipperiness	Resource	Yes	Yes	NT	NT	NT	NT	NT	Napolitano et al 2009
	Thermal comfort	Rectal temp.	Animal	NT	Yes	NT	NT	NT	NT	NT	Lowe et al 2002
		Respiration rate	Animal	NT	Yes	NT	NT	NT	NT	NT	Caroprese et al 2000, Lowe et al 2002
		Provision of shelter	Resource	Yes	Yes	NT	NT	NT	NT	NT	Caroprese et al 2009, Napolitano et al 2009,Pines et al 2007

Table 2.3 Potential welfare indicators selected from the literature regarding Good Health. Under the headings of Feasibility, Validity and Reliability "yes" indicates this has been tested and confirmed, NT means Not Tested.

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivity	Specificity	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
Good Health	Absence of physical injury	Presence/ absence of coughing	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Phythian et al 2010, Stubsjoen et al 2011
		Presence/ absence of nasal discharge	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Phythian et al 2011
	Ease of movement	Hoof overgrowth	Animal	Yes	NT	NT	NT	NT	Yes	NT	Napolitano et al 2009, Caroprese et al 2009
	Absence of pain caused by management procedure	Management procedures	Farm records	Yes	Yes	NT	NT	NT	Yes	Yes	Caroprese et al 2009, Napolitano et al 2009, Phythian et al 2011, Stubsjoen et al 2011
	Absence of disease	Eye condition / abnormality	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Phythian et al 2011, Stubjoen et al 2011

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivity	Specificity	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
		Skin lesions and wounds	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Phythian et al 2012, Stubsjoen et al 2012
		Faecal soiling	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Caroprese et al 2009, Phythian et al 2011, Stubsjoen et al 2011
		Lameness	Animal	Yes	Yes	NT	NT	NT	NT	NT	Phythian et al 2010, Kaler et al 2009, Kaler et al 2011, Napolitano et al 2009, Stubsjoen et al 2011, Phythian et al 2011, Caroprese et al 2009
		Ewe mortality rates	Farm records	Yes	Yes	NT	NT	NT	NT	NT	Pines et al 2007, Phythian et al 2011

Table 2.4 Potential welfare indicators selected from the literature regarding Appropriate Behaviour. Under the headings of Feasibility, Validity and Reliability "Yes" indicates this has been tested and confirmed, NT means Not Tested.

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivity	Specificity	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
Appropriate Behaviour	Social Behaviour	Behavioural Synchrony	Animal	NT	NT	NT	NT	NT	NT	NT	Dwyer and Lawrence 2005, Boe et al 2006
		Vocalisation of ewes and lambs	Animal	NT	Yes	NT	NT	NT	NT	NT	Dwyer and Lawrence 2005
		Proximity of ewes and lambs	Animal	NT	Yes	NT	NT	NT	NT	NT	Dwyer and Lawrence 2006
		Separation from flock	Animal	NT	Yes	NT	NT	NT	NT	NT	Phythian et al., 2011
		Vigilance behaviour	Animal	NT	Yes	NT	NT	NT	NT	NT	Dwyer 2004

Principle	Criterion	Indicator	Resource-, management-, or animal based. Or information obtained from farm records.	Feasibility	Validity	Sensitivit y	Specificit y	Test-retest reliability	Inter-observer reliability	Intra-observer reliability	References
	Positive emotional state	QBA	Animal	Yes	Yes	NT	NT	NT	Yes	Yes	Phythian et al, 2011, Wickham et al 2011, Wemeslfelder and Farish 2004
	Good Human-animal relationship	Response to human approach	Animal	Yes	Yes	NT	NT	NT	Yes	NT	Caroprese et al 2009, Stubsjoen et al 2011, Phythian et al 2011, Lankin and Hutson 1982
		Response to humans at milking	Animal	Yes	Yes	NT	NT	NT	NT	NT	Lyons 1989
	Expression of other behaviours	Atypical behaviour /Stereotypie s	Animal	Yes	NT	NT	NT	NT	NT	NT	Caroprese et al 2009, Phythian et al 2011
		Maternal behaviour	Animal	Yes	NT	NT	NT	NT	NT	NT	O'Conner et al 1985, O'Conner et al 1989, Le Neidre et al 1998

2.3.2 Stage 2 - Results from the expert panel meeting

As mentioned previously feasibility is one of the key factors for indicators used during an on farm welfare assessment. The first task for the expert panel was to consider the feasibility of each indicator found during the literature search. Those which were deemed unfeasible for use during a welfare assessment were automatically discounted. The indicators rejected due to practical constraints were floor slipperiness and rectal temperatures. Play behaviour was removed as the welfare indicators and resulting assessment produced by the AWIN project specifically focuses on the welfare of adult ewes, for whom this behaviour was not expected to occur frequently enough, if at all, for it to be a meaningful welfare indicator. The expected low occurrences of atypical and stereotypy type behaviour also led to it being removed from the list of indicators. Maternal behaviour was also removed as the farm-assessments would not regularly occur during lambing periods. Ewe and lamb proximity was also removed as during an on farm assessment it may not be possible to correctly identify and pair the correct ewes and lambs together or know how many lambs would be expected in the social group and so it would not be a meaningful indicator. The distance between a ewe and her nearest neighbour was suggested as an indicator of social behaviour.

For the indicators lacking formal validity testing, (behavioural synchrony, separation from flock and vigilance), the inclusion by the expert panel offered face validity. At least one indicator was required for the assessment of each criterion listed in Table 2.5. For some criteria such as social behaviour there were several proposed indicators, but for others e.g. positive emotional state, only one indicator remained. An additional criterion of “absence of general fear” was suggested and included by the expert panel in order to account for the effects of predation on the welfare of extensively managed sheep. To assess Absence of General Fear, two indicators were selected; the animal's response to, and recovery from a surprise test. Shivering was added as a potential indicator for the criterion of Thermal Comfort in the Good Environment principle.

During the expert meeting it was noted that some aspects of sheep health were not adequately covered by the indicators found during the literature review and so additional measures were proposed. Due to the frequency and impact of mastitis on the welfare of ewes, it was suggested that health issues specific to the condition should receive further consideration. These problems are particularly relevant for dairy sheep which spend a vast proportion of their lives in various stages of lactation. It was decided that a full assessment of the udders should be included. Other animal-based indicators added to the list during the expert meeting were tooth loss to be included in Good Feeding, and the use of the FAMACHA scale to assess anaemia under Good Health. For two indicators, body condition scoring and lameness additional novel simplified versions of current established scales were suggested. These simplified scales were proposed with the aim of improving feasibility of the measure while retaining validity.

For the reasons given earlier in Chapter 1 animal-based indicators were preferred over resource- and management-based measures. Although when no feasible animal-based indicators were available in the literature, resource-based measures were proposed by the experts. For example, the best indicator to assess "Absence of prolonged thirst" was determined to be "the provision of clean water and evidence of use" (Table 2.5).

The applicability of the indicators was also discussed and whether the animals would require gathering and handling to assess specific indicators, or whether information could be gained from the animals with minimal disturbance while they were still on pasture. The indicators were allocated to two categories as shown in Table 2.5. The first was titled "Assessment in the Field" and included the indicators which did not require the animals to be handled, or for which disturbance would invalidate the measure e.g. behavioural synchrony. The second approach was an "Assessment at Gather" in which the animals could be in assessed in close proximity and the assessment of some measures e.g. body condition scoring and tooth loss require physical contact. Some indicators were suited to both approaches and were therefore included in both categories e.g. coat cleanliness.

This re-consideration resulted in a list of 32 potential indicators (Table 2.5), which were expected to be feasible to measure during an on-farm welfare assessment of either housed or unhoused animals. The indicators shown in regular non-bold font are those which were identified during the literature search and remained following the expert review. Indicators in bold text are those which were suggested at the expert meeting and subsequently retained.

Table 2.5 Potential welfare indicators selected by the expert panel

Principle	Criterion	Indicator	Assessment type
Good Feeding	Absence of prolonged hunger	Body Condition Score	Assessment at Gather
		Tooth loss	Assessment at Gather
		Lamb birth weights and percentage lambs weaned	Records
	Absence of prolonged thirst	Provision of clean water	Resource
Good Environment	Comfort around resting	Space allowance and stocking density	Assessment in the Field
		Aggression and displacement behaviour	Assessment in the Field
		Hoof overgrowth	Assessment in the Field
		Coat cleanliness	Assessment at Gather and in Field
	Thermal comfort	Panting	Assessment at Gather and in Field
		Shivering	Assessment at Gather and in Field
		Provision of shelter	Resource
Good Health	Absence of disease	Presence of coughing	Assessment at Gather
		Presence of nasal discharge	Assessment at Gather
		Udder condition	Assessment at Gather
		Faecal soiling	Assessment at Gather and in Field
		FAMACHA anaemia chart	Assessment at Gather
Good Health	Absence of disease	Lameness	Assessment at Gather and in Field

Principle	Criterion	Indicator	Assessment type
Good Health	Absence of disease	Eye condition	Assessment at Gather
		Fleece and skin condition	Assessment at Gather
		Ewe mortality rates	Records
	Absence of injuries	Skin lesions and wounds	Assessment at Gather
	Absence of pain caused by management procedures	Appropriate tail length	Assessment at Gather and in Field
		Ear damage	Assessment at Gather
Appropriate Behaviour	Social Behaviour	Behavioural Synchrony	Assessment in the Field
		Nearest neighbour	Assessment in the Field
		Vocalisation of ewes and lambs	Assessment in the Field
	Expression of other behaviours	Vigilance behaviour	Assessment in the Field
	Positive emotional state	Qualitative Behavioural Assessment	Assessment in the Field
	Good Human-animal relationship	Response to human approach	Assessment in the Field
		Response to humans at milking (dairy ewes only)	Assessment at Gather
	Absence of General Fear	**Response to surprise test**	Assessment in the Field
		Recovery from surprise test	Assessment in the Field

2.4 Discussion

The evidence of validity, feasibility and reliability for each of the welfare indicators provisionally accepted by the expert panel and literature search are discussed below. As mentioned above for some measures at least one of these criteria were not previously established and therefore these indicators required further evaluation before they could be accepted for use with extensively managed sheep in the final AWIN protocol.

2.4 1 Principle: Good Feeding

The terminology of "prolonged" is a key concept in Good Feeding. It is impossible that no animal will ever experience short term hunger or thirst as without these triggers the animal would not be inclined to eat or drink (Forbes, 2006; McKiernan et al., 2008; Verbeek et al., 2012).

Criterion: Absence of Prolonged Hunger

The indicators selected as most relevant for assessing prolonged hunger in sheep were body condition scoring (BCS), tooth loss and lamb survival.

Indicator: Body Condition Scoring

Body condition scoring is used in many species to assess the volume of fat and muscle an animal is carrying. The scale developed by Russel et al in 1969 is widely used by farmers and vets on farm. In some species, e.g. the horse, it is possible to perform condition scoring visually; however the heavy fleece of the sheep does not allow this. To obtain body condition scores for individual sheep the recorder palpates the lumbar region of the loin area behind the last rib over the kidney feeling the prominence of the spine and presence of muscular and fatty tissue (Russel et al., 1969). This is a semi quantitative method which allows the recorder to score the animal between 0 (emaciated) to 5 (very fat) with quarter point intervals.

Validity: The validity of body condition scoring has been demonstrated in many studies. Convergent validity has been demonstrated as BCS has a close relationship with indicators of biological function such as health, fertility and milk production (Roche et al., 2009; Webster et al., 2015). Low body condition score is indicative of increased energy output and reduced intake whereas high BCS is a sign of over feeding, or excessive confinement of animals (Caroprese and Casamassima, 2009). Verbeek et al (2012) found that high feeding motivation and negative energy balance of ewe with lower body condition scores suggested an increased risk of a compromised welfare state. Ewes with low (mean=2) or medium (mean=2.9) body condition scores were more motivated to walk farther to receive a food reward than those with high (mean=3.7) condition scores. Ewes with low and medium scores also consumed significantly more reward feed than ewes with High BCS. BCS has been shown to be an important welfare indicator in determining risks to ewe mortality as ewes with higher scores tended to have increased survival rates (Morgan-Davis et al., 2008). It has also been validated through comparisons with blood metabolites. Ewes with low BCS scores below 2.5 were shown to have higher concentrations of non-esterified fatty acids (NEFA) and lower concentrations of blood glucose than those with higher BCS >3 (Caldeira et al., 2007a). Body condition scoring has been shown to be a better predictor of body composition than live weight (Russel et al 1969), although breed differences must be taken in to account when interpreting results. There is a general consensus however that an animal scoring above 4 is overweight, and below 2 is emaciated. Animals scoring at either of these extremes are at risk of metabolic disturbance (Caldeira et al., 2007a; Carol and Huntingdon, 1988).

Feasibility: When using the protocol correctly is it easy to learn (Burkholder WJ, 2000) and the time taken to assess the animals has been shown to be reasonable for an on-farm assessment(Morgan-Davis et al., 2008; Napolitano et al., 2008; C. J. J. Phythian et al., 2011; Stubsjøen et al., 2011). Although it requires animals to be gathered and handled it is a recognised on farm method which suggests good feasibility and acceptance for use as an on-farm welfare indicator. It was proposed that this indicator could be applied during the Assessments at Gather.

Reliability: The inter-observer reliability of various body condition scales have been assessed with conflicting results. In 1962 Everitt et al reported poor inter observer reliability when four assessors scored 74 merino ewes using a scale between 1 and 10 as their scores for the same sheep significantly differed from one another ($p<0.001$). Their poor agreement may have been due to the scale used as it lacked specific categorisation, or because the assessors had never performed body condition scoring previously and only practiced on two ewes prior to the data being collected. Results using the Russel scale however prove to be more promising. Russel et al found high agreement both between observers (80%) and within when repeated over a three year time period (70%) (Russel et al 1968 reported in Russel 1969). Phythian et al (2011) found that both inter- and intra- observer reliability improved considerably when the assessors used a half point version of Russel's scale compared to a full point scale (intra-observer using a half point scale $W=0.7$ and full point scale $W=0.6$. Inter-observer using half point scale $W=0.7$, full point scale $W=0.4$). This suggests that the intervals of the scales used and consistency in methodology are extremely important to consider if this indicator is to be used during an on-farm welfare assessment.

The seasonal repeatability of this measure also requires further investigation. The body condition of ewes should change relative to season and reproductive condition.

Although breeds may differ it is advised that the highest body condition score should occur pre-mating so the ewe has sufficient reserves to account for the additional energy expenditures of pregnancy and lactation (Henderson 2010). The repeatability of the Russel scale throughout the reproductive cycle of the ewe is not reported in the available literature and requires further assessment.

A secondary body condition score was also proposed by the expert panel group in which animals are categorised in one of four categories: “emaciated”, “thin”, “fit” or “fat”. Although lacking the accuracy of the Russel scale, this novel measure may still provide assessors with sufficient information as the sheep which are of concern are those at the extremes of the scale i.e. emaciated and fat.

Indicator: Tooth damage and loss

Using tooth loss as a welfare indicator may allow at risk animals to be identified sooner and prevent nutritional imbalances before it becomes detrimental to their welfare. The mouth of a sheep is well adapted to the grazing nature of the animal. An adult sheep has 32 permanent teeth, eight of which are incisors at the front of the mouth, used during grazing, and 24 molars at the back of the mouth (Tatara et al., 2014). When assessing the teeth of a sheep the presence and condition of the incisors is typically performed by lowering the bottom jaw and visually assessing the number and condition of teeth present.

Validity: Damage to the incisors can dramatically reduce the intake of feed which ultimately results in a lack of nutrition (Erjavec and Crossley, 2010; Tatara et al., 2014). Tooth loss or damage can be caused by injury, intake of hard materials (e.g. rocks) while feeding or due to a chemical imbalance of the soil which can erode the enamel (Bloxham and Purton, 1991; Healy et al., 1967). A deficiency of calcium can also have negative impacts on the teeth, especially during lactation, and can lead to premature loss of the incisor teeth of ewes (Gunn, 1969). The visual assessment of dental damage and disorders is a common farm practice by shepherds when deciding whether to breed from a ewe. Missing teeth or poor dentition can lead to the premature culling of ewes before they have reached the end of their otherwise healthy reproductive life (Ridler and West, 2010; Tatara et al., 2014).

Feasibility: As with body condition score (BCS) assessing the dentition of an animal requires handling and close inspection which is not feasible when the animals are on pasture. Assessing the tooth loss of an animal which has already been gathered is quick, simple and informative supporting its use during the Assessments at Gather.

Reliability: The repeatability of tooth damage and loss does not appear to have been formally tested and reported in the available literature. For this reason the effects of calendar season and reproductive cycle on ewe dentition and consistency of the measure will be considered in this thesis. Descriptions of the inter-observer reliability of this indicator also appears to be lacking in the available literature.

Indicator: Lamb birth and weaning weights

In addition to her own welfare, issues affecting a breeding ewe also play a large part in the survival and growth of her lamb(s) both during pregnancy and post-lambing prior to weaning. For this reason lamb birth and weaning weights and lamb survival were chosen as potential iceberg indicators.

Validity: Maternal under nutrition during pregnancy has been shown to result in low lamb birth weights and impaired post natal survival in sheep (Binns et al., 2002; C. Dwyer, 2008; Dwyer et al., 2003).

Heavier lambs are much more likely to survive for both the first few days of life, and to weaning than lighter lambs, particularly when outside (C. Dwyer, 2008; Everett-Hincks and Dodds, 2008; Oldham and Thompson, 2011). Lamb survival, as determined as lambs weaned per ewe, and productivity has been significantly positively correlated to the overall welfare of breeding ewes on farm (Stott et al., 2012).

Although a higher welfare score has been linked to an increase in lamb production, using high productivity is generally not regarded to be a good indicator of welfare (Kilgour et al., 2008). The breed of sheep has significant impact upon fertility and fecundity and even in poor welfare conditions it is possible for some highly selected animals to maintain high production rates (Dwyer and Lawrence, 2005; Stott et al., 2012). Poor productivity on the other hand, may be indicative of health and welfare problems experienced by the ewe (Dwyer and Bornett, 2004). The convergent validity of this measure as an iceberg indicator required further investigation before it can be confirmed as a welfare indicator for extensively managed sheep.

Feasibility: The perceived simplicity of using farm records to obtain information regarding the number of lambs weaned per ewe, and the lamb market or slaughter weight initially implies good feasibility. Not all farms, however, will keep even such basic records and so the feasibility of this indicator requires further work. In extensive systems, where sheep lamb without human assistance, some farmers will not know their birth rates and not know how many lambs they have until the sheep are gathered for weaning. Some farmers may gather their animals for lambing and assist where necessary but still not record the number of live births. Farmers may view record keeping as something they have to do in order to meet the demands of a scheme rather than a tool for their own use (Escobar and Demeritt, 2015). To many farmers their daily practice of looking after their animals is their priority, and recording information such as lambing rates comes second to the animal's needs (Escobar and Demeritt 2015). In some European countries, however, such as

Norway, farmers are financially incentivised by the government and NGOs to participate in specific recording schemes for the main livestock species. For sheep the farmers are encouraged to record details on the genetic lineages of their animals along with other production traits (Saether, 2002).

Reliability: Evidence supporting the reliability of this measure is lacking. Along with the problems of farmers not collecting this data (as discussed above) there is also the possibility that farmers may fail to provide correct information. During the lambing period the farmer may also be too busy to record events as they happen and after it is too late to record this information accurately (Escobar and Demeritt 2014). Another challenge in establishing this as a welfare indicator is the difficulty when comparing results between different breeds of sheep. The preference for using more primitive breeds in extensive systems means productivity is typically lower in these systems than for heavily selected lowland breed flocks regardless of welfare state (Dwyer and Bornett 2004). The season in which lambs are born may also impact upon lamb survival as winter lambing tends to lead to higher lamb losses than a spring lambing (Fisher, 2004) and so comparing these results between farms may not provide reliable results. The repeatability of this measure also appears be good Binns et al (2002) found lamb mortality rates on farms were significantly positively correlated with that of the previous year $r=0.43$, $P<0.001$).

Criterion: Absence of prolonged thirst

At the expert panel meeting it was acknowledged that physiological or physical measures of dehydration, which would be suitable for an on-farm welfare assessment, were lacking. The skin pinch test, in which the time taken for the skin to return to normal following pinching with thumb and fore finger, is commonly used by vets to assess dehydration however the validity of this measure has been questioned (Blockhuis et al., 2013) as the time taken for skin to return to its normal contour was not linked to physiological indicators of dehydration (serum osmolality or packed cell volume), or behavioural (drinking behaviour) (Pritchard et al., 2005).The appearance of sunken eyes was also a potential animal based welfare indicator however this only appears to be symptomatic of extreme cases which do not occur under normal farm practices rather only from neglect or serious diarrhoea (Blockhuis et al., 2013). Thus the measure is not appropriate for the purpose of detecting thirst due to suboptimal provision of water on farm. It was suggested by the expert panel that the provision and evidence of use of clean water could be used as a potential resource-based welfare indicator

Indicator: provision of clean water and evidence of use

Water availability can be assessed by counting and looking at cleanliness and accessibility of the water troughs or natural resources available to sheep on-farm. Signs of use include footprints, and wear at the easiest access point. Cleanliness is measured by presence of contamination and accessibility by assessing whether any hindrance or barriers exist which may prevent sheep from using the water points

Validity: Access to drinking water is a simple and important resource for the survival of livestock. It is essential that the water provided to sheep is clean and easily obtainable (Markwick, 2007).

The feasibility and reliability of water cleanliness and availability by checking resources in the environment should be suited to an on-farm welfare assessment.

2.4.2 Principle: Good Environment

Sheep and goat houses are often inadequate in terms of design, materials and size (Caroprese et al 2009). In previous sheep welfare assessments, "Good Housing" focussed on resource measures, for example available substrate, space and ventilation (Napolitano et al 2009, Caroprese et al 2009). However, in the Welfare

Quality assessments, comfort around resting, thermal comfort and the provision of shelter were considered to be the most important in relation to animal welfare and animal based indicators were selected to examine how the animals are coping with their environments (Canali and Keeling, 2009).

Criterion: Ease of Movement (applicable to housed ewes only)

The implications of space restrictions on the sheep health and behaviour support its inclusion in a welfare assessment. Three indicators were selected to assess the ease of movement: space allowance, aggression and displacement type interactions between animals, and hoof overgrowth.

Indicator: Space allowance - resource

Space allowance refers to the average area available to the animal (Petherick and Phillips, 2009; Petherick, 2007). The space available should be sufficiently large in order to allow the normal expression of behaviours. The volume of space recommended for ewes varies with regard to the animals' age, reproductive status, presence of fleece and size. Ewes with lambs at foot should have at least 2.0 square meters of floor space, whereas non-pregnant small ewes could be kept with 1.0 square meters of floor space (DEFRA, 2003). The assessment of space available to the sheep during a farm assessment can be performed by counting the number of sheep in an enclosed area and measuring the floor space available.

Validity: This measure is considered to be valid as confinement of animals under conditions of high spatial density is known to be stressful (Horton et al., 1991; Sevi et al., 1999) and as such this is a valid welfare indicator. Research by Black et al (1994) suggests that high stocking density exacerbates the effects of heat stress, pneumonia and for resources. Stocking densities of 1m2/ewe during pregnancy was found to cause alterations to ewe movement patterns and use of space in comparison with pens providing 2 and 3m²/ewe (Averós et al., 2014a).

Feasibility: The speed and ease of calculation of space availability by performing a head count and measuring the pen makes this assessment feasible for an on-farm welfare assessment.

Reliability: The reliability of this measure does not appear to have been assessed in the literature.

Indicator: Aggression and displacement behaviours

An animal-based indicator: the presence of aggression and displacement behaviours was also suggested by the expert panel as an assessment of space availability. In confinement sheep production lying space is an important resource and competition for this resource can lead to aggression and social stress (Færevik et al., 2005). Single lying space is an attractive resource to housed ewes and (Bøe et al., (2006) found that a large number of displacements appeared to increase the overall restlessness of the group. (Marsden and Wood-gush (1986) found that limited lying space was the second biggest cause of displacements, second to food provision.

Validity: Abnormal behaviours including displacement activity, stereotypies and high levels of aggression have been shown to be valid indicators of stress and poor welfare (Lauber et al., 2012).

Feasibility: Observing housed animals in their pens and counting the number of displacement or aggressive behaviours observed within a set time frame is typically conducted under experimental conditions in which cameras can be set up to continuously record the behaviour of animals. However, the application of assessing aggressive behaviours during an on farm assessment of sheep does not appear to have been assessed in the literature.

Reliability: By combining the agonistic and displacement behaviours of cattle the Welfare Quality project found good inter reliability (W=0.83) and acceptable consistency over time (W=0.74). They conclude that the total agonistic behaviours observed may be used as a welfare indicator for dairy cattle. Further work is required however in order to determine whether this is true for housed sheep.

Indicator: Hoof over growth

In an experimental setting hoof growth rate tends to be measured in sheep by marking the hoof in two places and on subsequent measurements noting the distance between the lines (Shelton et al., 2011). However, this is impractical for on-farm welfare assessments. The proportion of animals in a pen with at least one over grown hoof was selected as a potential indicator for housed ewes.

Validity: A higher stocking density ($1m^2$/ewe) leads to ewes stride length decreasing, lower distances between ewes and generally higher activity and restlessness levels in comparison with 2 or $3m^2$/ewe (Averos et al 2014a). It has been suggested that this reduction in movement will result in changes to the sheep's hoof growth and wear as the hoof tends to be worn by large mammals when walking on hard or rocky surfaces (Shelton et al., 2011; Smith et al., 2014; Vokey et al., 2001). If wear is slower than growth, then hoof horn grows distal to the sole of the foot (Smith et al 2104). Claw overgrowth may therefore be a potential indicator of ease of movement in housed sheep. Another factor affecting hoof wear is lameness which prevents animals placing their foot firmly on the ground and eroding the hoof (Azizi et al., 2011). The prevalence of lameness and claw overgrowth is known to increase in the winter housed season in comparison with outdoor grazing in the summer months (Azizi et al., 2011).

Feasible: For on farm welfare assessment it has been shown to be feasible to measure hoof over growth by assessing the prevalence of over grown hooves (Caroprese et al 2009, Napolitano et al 2009)

Reliability: The prevalence of grown hooves in sheep has been has been shown to have very good inter-observer reliability when applied on farm (Napolitano et al., 2009).

Criterion: Comfort around resting

The indicator selected for this criterion is coat cleanliness. A large proportion of the sheep's daily activity budget is used for activities in which the animal lies down such as ruminating, resting and sleeping. On average a sheep may spend over 11 hours a day lying down (Das, 2001).

Indicator: Coat cleanliness

Due to this large portion of time sheep spend lying down it is important that they have sufficiently well drained ground on which to lie to prevent their fleece becoming wet and soiled. The cleanliness of the fleece can be used to obtain information on the availability of clean ground available to the animals on which to lie. Coat cleanliness was included in the Welfare Quality project for use with cattle (Forkman and Keeling, 2009a). It has also been suggested as a potential animal based welfare indicator for use with sheep (Stubsjoen et al 2011, Napolitano 2009) and has been included in the Bristol Welfare Assurance Programme (2004) and the AssureWel assessment protocols (2013). The cleanliness of the fleece can be done by visual assessment in which the amount and extent of soiling is assessed and whether the dirt is wet or dry.

Valdity: Consensual and face validity for coat cleanliness has been shown by agreement of sheep experts that cleanliness of belly it was an important indicator for sheep by Phythian et al (2011). Convergent validity of fleece cleanliness and environmental conditions however is lacking. Stubsjoen et al (2011) assessed coat cleanliness of housed sheep and the hygiene of the lying area, although did not report on the relationships between these measures.

Reliability: The inter- and intra- observer reliability of a binary coat cleanliness scale has been shown to be high (inter-observer α=0.97-1, and intra-observer reliability α=0.97) in work by Phythian et al (2012), and a four point scale based on the Animal Needs Index scale was also found to have good inter-observer reliability when applied to housed sheep (r=0.88, (Napolitano et al., 2009)).

Feasibility: As this measure does not require the animals to be gathered and handled it is feasible for this measure to be performed both during the Assessment in the Field and during the Assessment at Gather. The measure itself is quick and simple when using a scale in which clear categories are defined

Criterion: Thermal comfort

The indicators selected for the assessment of this criterion were panting, shivering and provision of shelter. Panting and heat loss from respiratory tract seem to be the main heat loss mechanisms in sheep (Silanikove, 2000).

Indicator: Panting

A rise in ambient temperature brings an increase in heart rate, respiration rate, and panting accompanied by reduced food intake and reduced water loss from urine and faeces (Silanikove 2000). An animal pants in order to increase evaporation from the respiratory tract, and is a clear sign of severe thermal distress (EFSA Panel on Animal Health and Welfare, 2014).

Validity: In sheep, the respiratory response to an increase in ambient temperature involves an initial escalation of breathing pace, and as the temperature continues to rise it is followed by slower heavy panting (Caulfield et al., 2014). Panting rate has been suggested as a potential welfare indicator for use in sheep experiencing heat stress (McCarthy, 2005; Pines et al., 2007)

Reliability: The reliability of using panting as an indicator of heat stress was attempted by Phythian et al (2012) however the incidence of panting was extremely low and they were unable to perform analyses. The low frequency of panting observed by Phythian et al was seen to be representative of the incidences observed in the UK during an on farm welfare assessment but panting is likely to be more important in other countries/climates and may also be a useful indicator of heat stress in housed sheep even in the UK.

Feasibility: Due to the close proximity required between the assessor and the animal this indicator would only be feasible for use with housed sheep.

Indicator: shivering

Due to the insulation offered by their fleece sheep are generally considered to be well adapted to cope with cold environmental conditions, this potentially means they are more likely to be exposed to cold conditions compared to other livestock species (C. M. Dwyer, 2008). The main physical adaptation for sheep to respond to cold stress is shivering (EFSA 2014, Dwyer 2008).

Validity: Shivering has been shown to be a good indicator of cold stress in some livestock species such as cattle (Tucker et al., 2007), however as the fleece of sheep provides insulation a very low thermal threshold of below freezing is required in order to elicit shivering in a fleeced sheep (EFSA 2014). Other on farm welfare assessments for sheep do not appear to have included it as an indicator.

Reliability: The reliability of visible shivering does not appear to have been assessed for sheep. The reliability of this method may be affected by the presence of the fleece which could hide visible tremors. The reliability of this method requires further assessment before it could be considered as a potential welfare indicator for adult sheep.

Feasibility: In lambs this may prove to be a useful measure although the feasibility of this measure in adult sheep is questionable as it takes such low temperatures for it to occur, and the presence of a fleece may prevent an accurate detection.

Indicator: Provision and use of shelter

The provision of shelter is very important for extensively managed sheep as unhoused animals are likely to experience adverse weather. Adequate provision of shelter and shade are important to allow the animal to cope with environmental conditions, for example sheep provided with sufficient shade are able to maintain a healthy body temperature when ambient temperatures are above 50°C (Sherwin and Johnson, 1987).

Validity: Sheep actively seek shelter when outside their thermo-neutral or comfort zone due to cold or wet weather, even if they have their fleece, indicating it is an important resource to maintain their welfare (Alexander, 1974; McBride et al., 1967; Nowak et al., 2008). They may shelter in naturally occurring areas such as trees, rocks or hollowed areas of hillside or use man-made structures (Deag, 1996; Sibbald and Hooper, 2003). In a study in Turkey with sheep farmers, shelter provision was perceived to be more important to an extensively managed animal's welfare than supplementary feeding or veterinary inspection (Kılıç and Bozkurt 2013).

Using a resource measures such as the presence and size of shelter available to the animals would be feasible and reliable for use during an on farm welfare assessment.

2.4.3 Principle: Good Health

It is well acknowledged that health and disease are important aspects of welfare (Cockram and Hughes, 2011). When considering the context of the sheep flock health management the detailed physical examination of the individual sheep is essential and highly relevant. A full physical evaluation is needed to ensure nothing is left out. A number of scoring systems for these examinations have been developed in experimental settings and with some modification it is believed they could be developed for use on farm in clinical practice, or for feasible welfare assessments (Lovatt 2010). Suffering is known to occur both with acute conditions such as physical injury and chronic conditions such as lameness (Cockram and Hughes 2011).

Criterion: Absence of disease

The important diseases for sheep welfare have been described in Chapter 1. The indicators selected to cover this criterion were a respiratory assessment, an assessment of the udder, lameness, eye mucosa colour (FAMACHA chart), faecal soiling and ewe mortality.

Indicator: Respiratory assessment: presence of coughing and nasal discharge

As discussed in Chapter 1 respiratory diseases represent a significant welfare issue for sheep and can stem from a variety of diseases and infections. Regardless of the cause of the problem, a respiratory infection typically consists of similar recognised symptoms including coughing, discharge, and sneezing (Bell et al 2008). Alterations in breathing patterns are also an indication that the animal is struggling to breathe and may have compromised lung space (Henderson, 2010). The presence or absence of coughing and nasal discharge was selected as a potential welfare indicator for extensively managed sheep as they would be able to pick up on the majority of respiratory problems faced by sheep.

Validity: The presence of coughing and discharge are well established indicators of poor respiratory health and are recognised as vets when performing a clinical examination (Lovatt 2010). These indicators were also included in Phythian et al's 2011 indicator validation study and have been included in sheep welfare assessments (Bath and van Wyk, 2009). In many cases the welfare implications of poor health are not directly related to the infection or disease, but rather the animal's experience of the situation (Cockram and Hughes, 2011; Kirkwood, 2007). The exact relationships between some physical health problems and the internal welfare state of an animal may be more complex than first assumed. It is possible that an animal diagnosed with a disease such as subclinical mastitis may be entirely unaware of the situation and thus their welfare remains uncompromised. Similarly, high numbers of parasites identified in a faecal sample from sheep do not necessarily reflect a state of poor welfare. Some individuals naturally have a high tolerance of gastro-intestinal parasites and may therefore be perfectly able to cope with higher numbers of parasites in their system. i.e. they are resilient (Sargison, 2013). For these reasons additional validity testing may be required for some indicators listed above. Specifically, cross-validation of indicators which are designed to assess physical health and those covering other aspects of welfare (e.g. affective state) are recommended.

Reliability: The reliability of these measures does require further evaluation however as agreement and consistency has not been well established in previous studies.

Binary scales in which the presence or absence of the condition are most commonly used although low prevalence has prevented statistical analysis (Phythian et al., 2012; Stubsjøen et al., 2011).

Feasibility: Assessing the presence or absence of coughing and nasal discharge during an Assessment at Gather is fast and simple, which makes it an attractive welfare indicator. The feasibility of performing this method with sheep has been further confirmed by Phythian et al (2015) and Stubjoen (2011) and it was also selected for inclusion in the Welfare Quality protocols for cattle (Forkman and Keeling, 2009a).

Indicator: Udder condition

Mastitis can be an acute (sudden-onset) or chronic condition. The condition can occur at either a clinical level in which inflammation of the udders is visible, or subclinical level when inflammation is not visibly detectable. As discussed in Chapter One, the disease is problematic for both animal welfare and production and therefore it must be examined during an on-farm welfare assessment.

The physical indicators selected by the expert panel to detect animals currently or previously infected with the disease were included to develop an indicator called "udder condition". Abnormalities in skin colour of the udder (indicating reduction in blood supply), shape (swelling caused by inflammation), consistency, hardness (fibrosis) and presence of lesions and scar tissue on the udder are indicative that a ewe is suffering from the condition. Some of these indicators can indicate current mastitis, or for others (fibroids) a past problem. Acute and chronic clinical mastitis is characterised by palpable changes in the consistency of the glandular tissue and the presence of fibroids are apparent. Subclinical mastitis cannot be detected by the clinical assessment and may need to be detected using the somatic cell counts.

Validity: Using a physical clinical assessment of an animal to detect the presence of acute clinical mastitis is well validated (Blagitz et al., 2014; Fragkou et al., 2014). These indicators have been validated using additional bacteriological tests and somatic cell counts which are also able to detect subclinical presence of the disease (Blagitz et al., 2014; Fragkou et al., 2014; Fthenakis and Jones, 1990). Scales have been developed to assess physical characteristics of the udder listed above. For example, Blagitz et al (2014) confirmed the validity of binary scales to score the presence or absence of injuries and inflammation, and three or four point scales to indicate the presence, and severity of asymmetry, discolouration, nodules and udder consistency, and size of the udders.

Feasibility: Performing bacteriology or somatic cell count analysis may be feasible for use with dairy sheep however the cost of analysis may be high. It also would not be feasible under farm assessment conditions to collect this data for meat sheep. Performing a clinical assessment on animals which have been gathered may be feasible for an on-farm welfare assessment (Lovatt 2010) but this requires further evaluation.

Reliability: The reliability of this measure also requires further validation before it can be accepted as a welfare indicator for use with extensively managed sheep. Currently there does not appear to be any literature regarding the consistency of udder scoring.

Indicator: Faecal soiling

Faecal soiling can be assessed by visually quantifying the volume of faeces adhered to the wool of the sheep, and its placement on the body or hind legs. Scales have been developed to allow farmers and assessors to categorise the extent of soiling (Broughton and Wall 2007, Larsen et al 1994).

Validity: The validity of faecal soiling as an indicator of infection has been well validated. Broughton et al (Broughan and Wall, 2007) suggested that "dag" scoring could be used an indicator to recognise animals in the field which are suffering from high gastro-intestinal parasite burdens. Faecal soiling present on an animal has been shown to be associated with higher gastro-intestinal parasite burden such as fluke and nematodes which cause diarrhoea (Bath and Wyk 2009). A longer duration in time since the last administration of anthelmintic drugs, lower faecal consistency, poorer pasture and lower live weights have also been associated with dirtier rears (Allerton et al., 1998; Broughan and Wall, 2007).

Feasibility: Assessing faecal soiling on farm using a visual scale has been shown to be feasible (Broughton and Wall 2007, Stubsjoen 2011). The speed at which the animals can be assessed and its non-invasive nature mean it is a potential welfare indicator (Broughan and wall 2007) and is particularly suitable for use during on-farm assessments. As this measure does not necessitate physical contact with the animals and may be performed while they are grazing undisturbed by human presence it could be applied during either the Assessment at Gather or Assessment in the Field.

Reliability: The inter-observer reliability and consistency of this measure over time does not appear to have been judged and still requires further evaluation.

Indicator: Eye mucosa colour – FAMACHA anaemia scale

A five point scale (FAMACHA scale) which can be used to identify anaemia in ruminants by assessing the colour of their lower eye lid mucous membrane was developed by South African researchers Bath et al (1996) to facilitate the clinical identification of sheep infected with barber pole worms (*Haemonchus contortus).* The colour of the eye membrane is compared to a five point chart. Each of the five points represents a different shade of red ranging from a dark red indicating that the animal is not anaemic, to almost white for animals which require immediate attention.

Validity: The FAMACHA scale has been well validated for use in sheep and goats in South Africa and other tropical regions of the world (Leask et al., 2013; Mederos et al., 2014; Papadopoulos et al., 2013). The scale however has not been validated in countries with colder climates in which *H. contortus* is not a common parasite. It is possible that this scale could be used to indicate anaemia caused by other reasons such as blood loss or nutritional imbalances however this requires further validation.

Feasibility: The scale is currently in use in many countries around the world, and due to the low cost and ease of use it would be very feasible to include in an on-farm welfare assessment. Another advantage is the speed of assessment per animal and its availability to be used by people of varying education and literacy levels (Maia et al., 2014).

Reliability: The reliability of the FAMACHA chart requires further work. Inter-rater agreement and test-retest evaluation of the scale has been tested although found to be moderate (W=0.66 and W=0.62 respectively, (Grace et al., 2007)). Moors and Gualy (2009) found differences in sheep of two different breeds with the same levels of parasitic infection rate. This implies that a modification of the scale may be required if it is to be used with sheep of different breeds as varying levels of pigmentation in the animals skin could confound welfare assessment results.

Indicator: Lameness

The proposed indicator to assess lameness in sheep is the seven point lameness scale developed by Kaler and Green (Kaler et al., 2009). This gait score looks at head nodding, stride shortening and weight bearing on all four limbs and allows lameness on multiple limbs to be taken in to consideration. A novel simplified version of the Kaler and Green scale with four categories was also proposed by the expert panel.

Validation: The validity of the lameness as a welfare indicator was discussed and presented in Phythian et al (2013). They reported that lameness was consistently considered to be a major welfare issue for sheep at all production stages including ewes in the breeding flock. This corresponds with the information found in other studies in which with 90% of farmers reporting lameness and the 10% prevalence in the national flock (Kaler and Green, 2008). The scale developed by Kaler and Green (2009) is an objective measurement based on visual observations of an animal. An additional benefit this scale has over others is that it allows the assessor to record the presence of lameness which affects more than one limb (Lovatt 2010).

Reliability: The seven point locomotion scale has been shown to have very high >0.9 agreement between observers (Kaler and Green 2009) and intra-observer reliability (a=0.99 Phythian et al 2012).

Feasibility: Although the Kaler and Green (2009) scale was developed for use in an experimental setting with training and practice it could be used for clinical examinations (Lovatt 2010) and during on-farm welfare assessments. As this scale was developed for use in experimental settings it was possible to ensure the ground walked on by the sheep was a solid flat surface, however this may not always be possible when performing a welfare assessment on farm, especially when using this measure in the Assessment in the Field conditions. A modified, less complex, version of this scale may prove to speed up the time taken for the assessment and although some small details may be lost it may still prove to identify sound and lame individuals. Broadening the scale might actually prove to be beneficial when assessing on uneven ground. The Kaler and Green scale may prove too sensitive for use with animals on hilly ground and errors may be made as changes in stride caused by the terrain may be attributed to lameness. By condensing the 7 point scale in to a four point scale it may be possible to distinguish between sound, mildly lame, moderately lame and severely lame animals without compromising the integrity of the measure however further work is required before this can be confirmed.

Indicator: Eye condition

Eye conditions in sheep may be a result of a disease or an injury and both are likely to cause inflammation, impair vision and even lead to permanent blindness. The presence or absence of twitching, excessive tear production or guarding of the eye when handled can all be indicators of an eye condition (Williams, 2010). Eye condition can be assessed by restraining the animal during the Assessment at Gather and recording any evidence of injury (current, recent or healed), and any evidence of inflammation or discharge which could indicate infection.

Validity: Assessing the eyes during a clinical assessment has been shown to be a valid welfare indicator for sheep (Phythian et al 2011, Lovatt 2010).

Feasibility: An assessment of the eyes of the sheep has been shown to be feasible during an on-farm welfare assessment (Phythian et al 2011, Stubjoen et al 2011).

Reliability: Phythian et al (2013) found presence/absence to have excellent inter-observer reliability when assessing the presence or absence of eye abnormalities (K=0.72).

Indicator: Fleece and skin condition

Fleece and skin conditions have been identified as important welfare indicators for sheep (Phythian et al 2011, Stubsjoen et al 2011). Wool loss and skin irritation can be used to identify ecto-parasite infestation, such as fly strike, and nutritional imbalances.

Validity: Sheep may experience wool loss following a period of psychological stress (Morgan et al., 1986), or malnutrition as inadequate intake of copper, calcium and cobolt can cause the wool fibres become thin and eventually break (Winter, 1995). Even if welfare problems are not present in the animals at the time of the on-farm assessment past conditions may be identified as even after treatment the loss of wool may remain (Plant, 2006) and as such it may be a valuable welfare indicator to gauge historical welfare compromises. The presence of flies and maggots or evidence of myasis can also be used to assess the presence of ecto-parasitic infections.

Feasibility: During the Assessment at Gather an assessment of the skin and wool is fast as described above, and during the Assessment in the Field it is feasible to assess wool loss from a distance. In cattle the physical examination of an animal took approximately three minutes to assess coat condition and wounds (Krebs et al., 2001).

Reliability: Phythian et al 2012 found good reliability when assessing fleece loss on-farm recording the proportion of animals with wool loss. Stubsjoen et al (2011) scored the presence and severity of skin lesions and skin irritations from 1 – 4 ranging from normal, loss of wool, redness and swelling and the presence of parasites or flies however these incidences were too low to assess reliability properly.

Indicator: Ewe mortality rates

Prior to mating farmers typically remove ewes which are suffering from chronic health problems such as mastits, or those that s/he fears will not be able to sustain another pregnancy due to low body condition score or tooth loss. These animals are either sold to lowland production systems for another breeding season (draughted) or are culled. If a farm has high levels of annual culling or draughting it may indicate they have high incidence of these welfare problems on farm. It is also possible that high cull rates are not representative of a farm overall welfare status. A farmer may have a high stock turn over due to production demands and may replace animals whose welfare are not compromised with other which will give him better returns. Assessing the proportion of ewes which are sold or culled at the end of each production year, along with the reasons for this decision may be a potential welfare indicator.

Validity: The use of cull records has been suggested as a welfare indicator in previous studies (C. J. Phythian et al., 2011; Science et al., 2001) giving at least face and consensual validity. Further validation however is required to understand the link between culling rates and welfare on farms during a welfare assessment, and whether the welfare of the animals removed from the flock is different from those which are retained.

Feasibility: As with lamb mortality the evidence supporting the reliability of this measure is lacking. Farmers are required by law (in the EU) to document the number and identities of animals sold at markets and those which are killed at slaughterhouses (DEFRA, 2003), thus suitable records should be readily available.

Repeatability: The repeatability of this measure appears to require further investigation as the consistency of extensive sheep farms' lamb production rates across consecutive years has not been reported in the literature.

Criterion: Absence of injuries

Injuries may occur due to improper handling, misuse of equipment or hazards in the environment, or they may have been caused by other animals.

Indicator: Skin lesions

Skin lesions and scarring can be used to identify current or past injuries which the animal has received. Skin lesions have been included in protocols for use when assessing the welfare of sheep (Stubsjoen et al 2011), however low incidence prevented conclusions to be made regarding the validity, feasibility and reliability of the measure.

Validity: Using the presence and severity of skin lesions to asses welfare has been performed in previous studies of other species such as cattle (Livesey et al., 2002).

The relationship between these measures and other aspects of welfare has not received much attention. In cattle, the percentage of animals with skin damage was used to provide estimates of thresholds when evaluating a herd as having either very good or very bad welfare (Bartussek et al., 2000). Little information is provided by the literature, however, to evaluate these findings fully. Further investigation is required in order to further validate this indicator for use with sheep.

Feasibility: Assessing the presence or absence of skin lesions is a simple and quick method for assessing injuries sustained by the animal. The fleece of the sheep may make the assessment of the body more troublesome however the assessment should be easy for the face, neck and legs and may occur at the same time as the wool and skin condition assessment.

Reliability: The inter-observer reliability of lesion scoring has been found to be good in cattle (80%, (Zubrigg et al., 2005)) but this information has not been reported for sheep.

Criterion: Absence of Pain Caused by Management Procedure.

Livestock are often subject to management procedures which are potentially painful. Procedures such as ear tagging are required by law, and other management procedures such as tail docking or castration may be performed to prevent welfare problems in the future.

The policy and management decisions made by a sheep farmer can have profound effect on the health and welfare of the flock. These procedures are typically performed on young lambs, and castration is only performed on male lambs. As the focus of this project is breeding ewes, the experts selected indicators which could be used to obtain information on historical and current welfare problems. The indicator selected were ear damage and length of docked tails.

Indicator: Ear damage

Ear tagging is required by law as a means of identifying and tracking animals throughout their lives. Although the procedure is mandatory damage to the ear caused by improper application, or tearing are welfare concerns (EFSA 2014).

Validity: Ear tags which are applied incorrectly can cause pain both at time of application and as the lamb grows the tag may cause the ear to rip (Edwards and Johnston, 1999). Tearing of the ear may leave an open wound which, in hot weather, is susceptible to fly strike or infection. The use of torn out ear tags has been used as a welfare indicator for sheep by Stubjoen et al (2011) however its relationship with other welfare indicators remains unknown.

Feasibility: Assessing the ears of the animal during the Assessment at Gather should be feasible and fast, noting the presence of any injuries, scars or open wounds.

Reliability: The inter-observer reliability and repeatability of this measure does not appear to have been assessed and therefore required further work.

Indicator: Appropriate length of docked tails

Sheep are frequently tailed docked. The tail of a sheep is considered to be too short if it is not long enough to fully cover the genitalia. A scoring system was proposed to indicate whether the tail of the sheep was too short, docked but acceptable length or not docked.

Validity: Tail docking is known to be a painful procedure (Fitzpatrick et al., 2006; Kilgour et al., 2008). However, there are additional welfare concerns associated with tail docking as if the tail is docked too short (not covering anus and vulva) it can lead to an increase in rectal prolapses (Fisher et al., 2004; Thomas et al., 2003) and increase in carcinoma of vulva (Swan et al., 1984).

Feasibility: The tail can be quickly and easily assessed visually either during handling as part of the Assessment at Gather or during the Assessment in the Field.

Reliability: The reliability of this measure does not appear to have been assessed and so further reliability testing is required. The assessors have to be aware, however, that the adult animals observed in the field may have been purchased after docking by a previous owner so additional cross-referencing with the farmer's account of procedures is necessary.

2.4.4 Principle: Appropriate Behaviour

Behaviour is widely used in clinical assessments of animals (Lovatt, 2010; Rutherford, 2002) and is also very well suited to on-farm welfare assessments. Behaviour can be seen as result of an animal's decision making processes and it can be used to assess how well, or badly, an animal is coping within its situation

(Dawkins, 2006b). The advantages of using behaviour to assess welfare include that is a non-invasive and mostly non-intrusive measure (Dawkins 2006).

Criterion: Social behaviour

Sheep are gregarious animals and choose to associate in social groups (Dwyer and Lawrence, 2008). Sheep therefore find isolation from conspecifics a stressful experience and will seek contact with other sheep (Barnard et al., 2015). Alterations to their social grouping or behavioural patterns may indicate potential welfare problems (Gougoulis et al., 2010) and therefore aspects of social behaviour may prove to be useful indicators during a welfare assessment.

Indicator: Behavioural synchrony

The behavioural synchrony in sheep has received much attention (e.g. Jorgensen et al 2011, Boe et al 2006, Rook and Penning 1991). Individual animals in a group are known to synchronise their behaviour with others during resting and grazing for a number of benefits including decreased likelihood of predation and maintaining the group structure (Jorgensen et al 2011, Rook and Penning et al 1991). A reduction in behavioural synchrony may be used as a welfare indicator to ensure all animals have the space and opportunity to perform behaviours consistent with others (Boe et al 2005). This may be performed during an on-farm welfare assessment by performing a visual scan of undisturbed groups of animals and assessing the proportion of animals engaged in specific behaviours or postures.

Validity: In cattle, synchronous lying and feeding have been used as positive welfare indicators (Forkman and Keeling, 2009a; Fregonesi and Leaver, 2002; Napolitano et al., 2009). Groups of animals performing lying or feeding behaviour synchronously have adequate space and access to resources without the need for competition (Napolitano 2009) and also have low levels of disturbance behaviour

(Fregonesi and Leaver 2002). A high degree of synchrony of resting or grazing behaviour within a herd or flock is indicative of a positive welfare state, in particular for subordinate animals (Napolitano et al., 2008). Further testing of convergent validity of this measure is still required as the relationship between behaviour synchrony and other welfare indicators remains unclear.

Reliability: The reliability of this indicator does not appear to have been tested in sheep. However, in the Welfare Quality project using cattle, observer agreement was found to be high when assessing the postures of the animals (r=0.80 to r=0.99, Forkman and Keeling 2009).

Feasibility: Synchrony has been used as a feasible welfare indicator using scan sampling methods in cattle (Napolitano et al 2009b, O'Driscoll et al 2008). Further testing is required before it can be confirmed as a welfare indicator for sheep as Stubsjeon et al (2011) were unable to include a measure of behavioural synchrony in their welfare assessments due to time constraints. As the animals can be observed undisturbed from a reasonable distance this furthers the potential feasibility of using the measure during the Assessment in the Field.

Indicator: Nearest neighbour

The flocking behaviour of sheep is thought to have evolved as a result of predation pressure by diluting the risk of attack (Dwyer, 2004). This behaviour is observed in both wild and domesticated sheep. Ewes with lambs at foot may naturally become more withdrawn from the flock as there is a temporary weakening of the social bond with other flock members but for ewes without lambs intentional separation from the flock is rare (Hinch et al., 1987). If a sheep is withdrawn from the social group or is behaving independently, this may indicate that the animal's welfare is compromised (Phythian 2015, 2011, EFSA 2014). The distance between a ewe and her nearest neighbour when grazing undisturbed could be used as a welfare indicator.

Validation: Separation from flock members has been suggested by sheep experts as a potential welfare indicator for sheep in Phythian et al (2011). However, the convergent validity between nearest neighbour distance and other welfare indicators requires further investigation (EFSA 2014).

Feasibility: In their 2012 study Phythian et al used “dull demeanour” as a measure when assessing sheep on farm. This term encompassed “behavioural separation from group, appearing dull with lowered head carriage and unresponsive to presence of observer or other sheep”. In this study the number of animals meeting these criteria was counted during a farm visit. This work suggests the feasibility of detecting lone animals during a welfare assessment is possible and supports good feasibility.

Reliability: The inter-observer agreement of the proportion of animals displaying “dull demeanour” by Phythian et al (2012) was found to be high (α =1). The reliability of assessors estimating the distance between these animals and the rest of the flock, however, was not tested. The consistency of the measure also required further work before it can be accepted as a welfare indicator.

Indicator: Vocalisations of ewes and lambs

Vocalisations are signals which have evolved as a social function with the aim of eliciting a response from other animals (Dawkins, 1990; Weary and Fraser, 1995). The use of ewe and lamb vocalisations has been selected for evaluation as a potential animal-based welfare indicator.

Validity: The use of vocalisations as a means to assess welfare has been suggested for various species including pigs (Weary and Fraser 1995), cattle (Grandin, 2001), poultry (Zimmerman et al., 2000) and rodents (Weary and Fraser 1995). Vocalisations in farm animals are generally considered as an indicator of negative feelings and an increase in vocalisation has been shown to be a valid indicator of poor welfare in slaughterhouses (Grandin 2001). Increased vocalisation may also be an indication of increased fear in sheep as lambs which are distressed also exhibit an increase in vocalisation (Greiveldinger et al., 2007; Manteuffel et al., 2004). Vocalisations may have other social functions though and not directly related to welfare at all (Boissy et al., 2007; Manteuffel et al., 2004). It must also be acknowledged that the absence of vocalisations also does also not guarantee good or bad welfare (Dwyer and Lawrence, 2008; Weary and Fraser, 1995). The validity of vocalisation rates as a welfare indicator for sheep requires further evaluation.

Feasibility: Vocalisations are a non-invasive measure which can be easily measured by humans without the need to expensive equipment. The number of animals vocalising was found to be feasible for cattle at slaughterhouses (Grandin 2001). When assessing the feasibility of vocalisation rates for chickens during the Welfare Quality project assessors found it difficult to distinguish between the noise of the animals under observation and others in the environment, and the noise of background machinery also interfered (Forkman and Keeling, 2009b). These problems caused the indicator to be excluded from Welfare Quality, although it has been suggested that with revision they may be included in an improved version of their protocols (Moura et al., 2008) . These issues however may not be so important when assessing extensively managed animals although the feasibility of using vocalisations as a welfare indicator for sheep requires further testing.

Reliability: Both the inter-observer and repeatability of vocalisation rates appear to be lacking for animals in extensive systems. Further investigation of these is essential in order for vocalisation to be considered a welfare indicator for extensively managed sheep.

Criterion: Expression of other behaviours

The complexity of the environment can be assessed either by assessing the environment itself, or using an animal-based measure to assess how the animals are responding to their environment (Edgar et al., 2013). The welfare indicator selected for this was an assessment of vigilance behaviour.

Indicator: Vigilance behaviour

During the assessment of behavioural synchrony (see above in "social behaviour") the vigilance patterns of social groups can also be assessed in order to evaluate vigilance levels.

Validity: As discussed in Chapter One vigilance is considered to be an indicator of fear (Riecau and Martin 2008, Hopewell et al 2005) and high levels of vigilance are detrimental to other aspects of the animals' welfare (Dukas and Clark, 1995). The assessment of vigilance behaviours does not appear to have been suggested in previous literature on on-farm sheep welfare assessments however it was considered for use with cattle during feeding in the Welfare Quality project.

Reliability: The inter-observer reliability of vigilance reported by the Welfare Quality project was high ($r=0.95$ to 1.00), however repeatability was found to be low ($r<0.2$). The consistency of vigilance levels of extensively managed sheep is apparently lacking and requires further evaluation before it can be accepted as a welfare indicator.

Feasibility: Assessing the vigilance levels of undisturbed sheep offers the same feasibility advantages as behavioural synchrony as using scan samples to assess the frequency of vigilance behaviours in a flock is quick and can be performed from a reasonable distance.

Criterion: Positive emotional state

Indicator: Qualitative Behavioural Assessment

A measure which may be particularly well suited to the assessment of the welfare of extensively managed sheep is Qualitative Behavioural Assessment (QBA). This measure is able to capture and assess both the positive and negative aspects of welfare (Wemelsfelder, 2007).

Unlike other animal-based welfare indicators, which focus on specific individual aspects of physical health and behaviour of an animal, QBA does not isolate these facets, since in doing so the concept of the "whole-animal" is lost. This "whole-animal" information cannot be regained at a later stage and potentially leads to important information being omitted when using quantitative scoring (Napolitano et al., 2012; Walker et al., 2010; Wemelsfelder and Lawrence, 2001).

The information gained when using QBA enhances and complements quantitative scoring measures used for assessing animal welfare (Andreasen et al., 2013; Wemelsfelder et al., 2001, 2000). During conventional quantitative behaviour scoring, the observer records which behaviours are present, but the QBA approach asks the observer how behaviours are being executed. When performing QBA the observer assimilates many pieces of information about the animal's body language and the way in which it interacts with the environment and translates this into qualitative descriptors such as "calm" or "agitated" (Wemelsfelder et al 2000). These descriptors may be developed using Free-choice profiling in which observers select their own terms, or a fixed list in which all assessors use the same descriptive terms.

Validity: Numerous studies have found good evidence supporting the validity of QBA both when using free choice profiling and a fixed list approach. Convergent validity has been demonstrated with good associations found between QBA and physiology and behaviour (Rousing and Wemelsfelder, 2006; Rutherford, 2002; Stockman et al., 2011; Wickham et al., 2012). Consensual validity has been established by Phythian et al (2011).

Reliability: Live on-farm QBA assessments using a fixed list approach have previously resulted in good levels of observer agreement in various livestock species. The Welfare Quality project reported good inter-observer reliability when applying QBA to pigs (W=0.82,(Forkman and Keeling, 2009c); poultry (W=0.83, (Forkman and Keeling, 2009b); and beef cattle (W=0.73), but not to dairy cattle (W=0.38, (Forkman and Keeling, 2009a). In addition, applying the same fixed list of terms to videos of dairy cattle Bokkers et al (2012) found observer agreement to range from W=0.24 to W=0.68. Yet, when the Welfare Quality QBA protocol was applied on 43 dairy cattle farms in a more recent study, inter-observer agreement was found to be good (W=0.72, Andresen et al 2013). A study evaluating the inter-observer reliability of observers viewing sheep video clips reported good agreement (W=0.78,(C. J. Phythian et al., 2013)) however the assessment of the inter-observer reliability and repeatability of a fixed list of terms applied to individual sheep on farm required further investigation.

Feasibility: The practicality of performing QBA strengthens its potential position as a welfare indicator for extensive animals. Animals can be observed in their normal environment and QBA can be performed from a reasonable distance with minimal disturbance. The feasibility of applying QBA fixed lists on farm has been tested and verified both in experimental situations and as part of the Welfare Quality protocol (Forkman and Keeling, 2009a, 2009b, 2009c; Muri et al., 2013; Sant'Anna and Paranhos da Costa, 2013).

The inter-observer reliability and repeatability of QBA will be evaluated in this project along with feasibility of this measure for use with extensively managed sheep.

Criterion: Good human-animal relationship

A domestic animal's fear of humans can have substantial negative effects on their welfare(Rushen et al., 1999b). The assessment of an animal's reaction to humans has been recommended for inclusion during on farm welfare-assessments (Rushen and Depassille, 2005; S Waiblinger et al., 2006; Winckler et al., 2003; Windschnurer et al., 2009). The indicator selected for evaluation to assess this criteria was the animal's response to a human approach and response to handling during milking (only applicable to dairy sheep).

Indicator: Response to human approach

Tests measuring the animal's reaction to humans fall into three categories: reaction to a stationary human, reaction to a moving human, and their response to being restrained or handled (Waiblinger et al 2006, de Passille and Rushen 2005). The reaction to a stationary human and response to being restrained have been validated, however, are more suitable for use with housed or at least temporarily confined animals (de Passille´ and Rushen, 2005; Waiblinger et al., 2006). The reaction to an approaching human may be the best suited for use when assessing extensively managed animals as it most closely resembles the situations the animals experience on a regular basis (Waiblinger et al 2006).

Validity: By measuring the animals latency to approach or the distance the human can approach to before the animals flee gives an indication of the animals comfort around humans (Boivin et al., 2003; Breuer et al., 2000; Fisher et al., 2000). When assessing the animal's response to a moving human the experimenter typically approaches the animal at a slow but steady pace and may have their arm outstretched at a 45° angle (Forkman and Keeling, 2009a). The tests reported in the literature vary in their protocols; the test may be terminated after a set time, or when the animal retreats from the person. A number of different parameters may be measured during these tests, for example the closest distance between the experimenter and the animal, the time the animal spent in a specific area of a pen, or whether physical contact was achieved (Waiblinger et al 2006). Studies assessing the discriminant validity of these tests with both cattle and sheep are lacking (Napolitano et al 2011, Waiblinger et al 2006) and therefore further validity testing is required in order to gain further insight in to its use as a welfare indicator.

Feasibility: The feasibility of the human-approach method has been demonstrated both during experimental and on-farm assessments (e.g. Waiblinger et al 2006, Forkman & Keeling, 2009a). As this test does not require a great amount of time or any equipment it appears to be feasible for use with extensively managed animals.

Reliably: The repeatability of individual flight distance is much lower when measured in a group in a paddock than when animals are individually tested in a yard (Fisher et al., 2000). Testing animals in a group remains one of the biggest problems in this area, especially under farm conditions where animals are generally reared in groups, especially on large commercial units (Boivin et al., 2003). The overall lack of consistency and standardisation between studies using these tests on farms has led to criticism and claims that it should not be used during an on-farm welfare assessment (Waiblinger et al 2006). Others disagree however and feel it offers valuable information when performed consistently (Waiblinger et al 2006, Napolitano et al 2011, Boivin et al 2003). In order for this test to be used during on-farm welfare assessments clear, standardised protocols such as those developed in the Welfare Quality project would have to be developed and stuck to in order to make comparisons between farm and between assessors.

Further development and evaluation is required of a human-approach test for use with extensively managed sheep. The validity and reliability will be investigated during this project.

Indicator: Response during milking

In most dairy systems animals are gathered and milked at least once a day during the lactation period. It is therefore very important that the handling does not cause stress to the animals. Using the animals' behavioural reactions to the handler during milking was selected as a potential welfare indicator for dairy sheep to assess the human-animal relationship.

Validity – An animal's previous experience with a handler has been shown to affect their milk production and handling during milking (Breuer et al., 2003; Hemsworth, 2003). Animals which have been handled aversively typically maintain a further distance from the handler, and show increased aggression e.g. kicking (Hemsworth 2003, Rushen et al 1999). Using the animal's reaction to the handler during milking can be used to indicate previous experiences and can inform an assessor of the handling styles adopted on the farm (Hemsworth et al 2000, Hemsworth et al 2002).

Reliability: The repeatability of kicking behaviour during milking has been shown to be variable in a study by Napolitano et al (2005). The consistency of the measure varied between farms with values ranging between W=0.35 to W=0.63. The inter-observer agreement of this measure however appears to be lacking and has not been reported for sheep or other dairy species. It is possible that the presence of observers during milking could affect the behaviour of the animals and reduce reliability of the measure, although this did not significantly affect the behaviour of dairy cattle in comparison to the animals' behaviour recorded using a camera (Munksgaard et al., 1999).

An issue which may confound this work with extensively managed animals is the irregularity of the handling in these systems. Sheep that do not receive frequent handling may find the experience in itself very stressful, not due to improper handling techniques, but due to the infrequency of the handling. Animals in their first lactation in particular may make this measure particularly unreliable. (Battini et al., 2011) found the reaction of extensively managed dairy cattle significantly differed throughout the year with the animals showing more fearful responses following their summer grazing period in the Alps. Further reliability testing is needed before this measure could be accepted as a welfare indicator for use with dairy sheep.

Feasibility: Behaviour during milking has been included in on-farm welfare assessments of dairy cattle and buffalo which implies the measure would also be feasible for the assessment of sheep systems. The length of time taken for the measure to be conducted may depend on the size of the flock however and as such decisions would need to be made regarding the proportion of animals assessed on each farm.

Criterion: Absence of general fear

Fear is a negative emotional state and is highly relevant to animal welfare impacting upon behaviour and physiology and as such assessing it is an important aspect of a welfare assessment (de Passille and Rushen 2005, Boivin et al 2003). In many studies a physiological indicator such as cortisol may be selected to assess an animal's fear levels however there are validity concerns with such measures (Dwyer and Bornett 2004) and they are not feasible for use during an on-farm welfare assessment (Dwyer and Bornett 2004, Forkman & Keeling, 2009a). Behavioural indicators were therefore selected for the assessment of this criterion: the animal's response to, and recovery from, a surprise test.

Indicator: Response to surprise test

Novelty is known to evoke a fearful response in livestock (Forkman & Keeling, 2009a, Greiveldinger et al 2007). In studies with housed animals, the preferred stimulus during a surprises test is a red ball or other inanimate object suddenly appearing in to the pen (Greiveldinger et al., 2007; Vierin et al., 2002); however this is not feasible when assessing animals in the field and so an original test was proposed. Rather than an object such as a ball dropping or being thrown towards the animals the use of an automatically opening umbrella was suggested for evaluation. This test was designed not to necessarily scare the animal but surprise.

Validity: In experiments assessing animals' reactions to unexpected events good correlations can be found between this response and reactions to other fear inducing stimuli (Romeyer and Bouissou, 1992). These findings may not be generalizable however as the breed of the sheep may influence fear reactions (Romeyer and Bouissou 1992). Although there have been relatively few studies, associations between unpredictable or surprising events (such as the appearance of a novel object) and physiological parameters, (such as heart rate) help to support the validity of this measure (Greiveldinger et al 2007). Further validation of this test with extensively managed animals is required.

Feasibility: Welfare Quality rejected the use of a surprise test (a sudden blow of air) in their on-farm welfare assessments due to lack of feasibility (Forkman & Keeling, 2009a). There do not appear to have been any previous studies assessing the feasibility of surprising extensively managed animals with a visual startle test such as this and therefore further assessment is required.

Reliability: The repeatability of this measure does not appear to have been tested in sheep and therefore requires further testing.

Indicator: Recovery from a surprise test

Following a surprising event an animal must assess whether the potential danger has passed it is safe to resume their previous behaviour (Dwyer, 2004). Assessing the length of time taken for a ewe to return to her previous behaviour following the surprise test was selected as a potential indicator of general fear levels.

The validity, feasibility and reliability of this measure have not been explicitly reported in the literature and as such all three aspects require further investigation.

2.4.5 Limitations of small expert panel

A fair criticism of the focus group approach would be that the number of experts attending the panel meeting was relatively small. It is possible that this may have led to some bias in the indicator selection. However it must be remembered that the countries represented at the expert meeting: Italy, Spain and the UK account for at least 50% of sheep production across the EU. The sheep industry these countries includes animals bred for both meat and milk, and the experts each had experience of at least one of these systems. If this approach were to be repeated it may be beneficial to increase the number of participants or countries represented in the panel, although this may not be feasible.

2.5 Conclusion

Following a literature review and expert panel meeting potential animal welfare indicators for use with extensively managed sheep were identified. For some indicators such as body condition scoring the validity, reliability and feasibility were well-established. For others however, at least one of these criteria were not met and therefore these indicators required further evaluation before they could be considered for use with extensively managed sheep. This further evaluation of these specific indicators is the objective of the work presented in this thesis. The work presented here focusses on the indicators applicability to unhoused ewes which are bred for meat production in the UK. The work presented in this thesis primarily focuses on the validity and feasibility of the indicators. The inter-observer reliability and long term repeatability of one measure, QBA, is discussed in Chapter 5. Short term repeatability (test-retest) of the other indicators was not possible during this PhD project due to time constraints for the researchers and the farm facilities. The farm used during data collection operated as a commercial farm, and the flock managed extensively. As such it was not feasible to repeat the assessments more frequently. Test-retest repeatability of the welfare indicators was performed by an AWIN partner research group based at Neiker Tecnalia in Spain. As their work was not part of this PhD project it is not explicitly discussed in this thesis.

Chapter 3 General materials methods

3.1 Study design/longitudinal project

The experimental work presented in this thesis (Chapters 4, 5 and 6) took place during a longitudinal data study spanning two years from November 2011 to August 2013. Throughout this thesis, the time between November 2011 and August 2012 is referred to as Year 1, and September 2012 until August 2013 as Year 2. The data collection time points and farm husbandry procedures conducted during these two years are shown in Figures 3.1 and 3.2

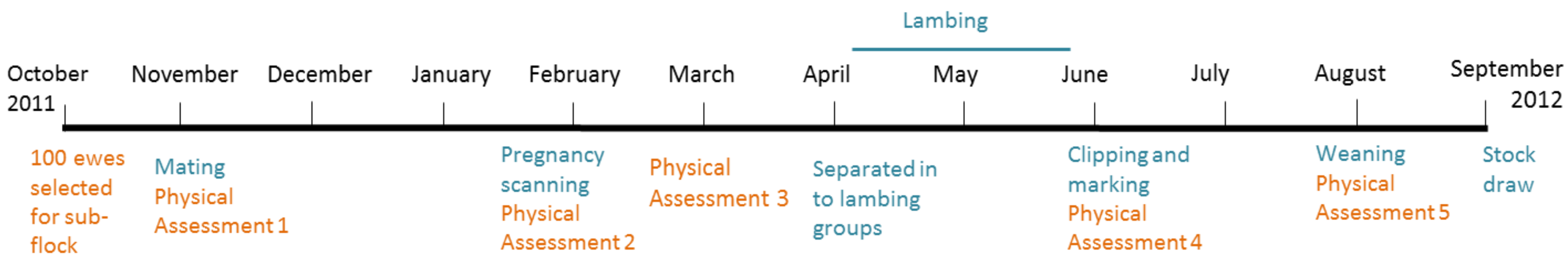


Figure 3.1 Timeline showing farm husbandry procedures and data collection performed on Castlelaw hill farm in Year 1. Blue text indicates farm husbandry, orange text indicates data collection

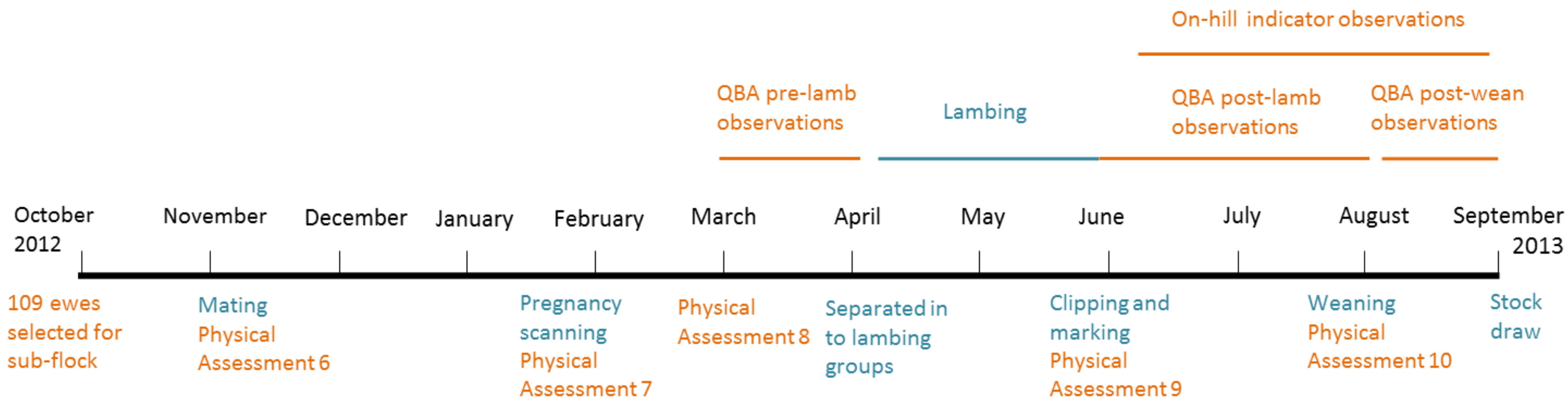
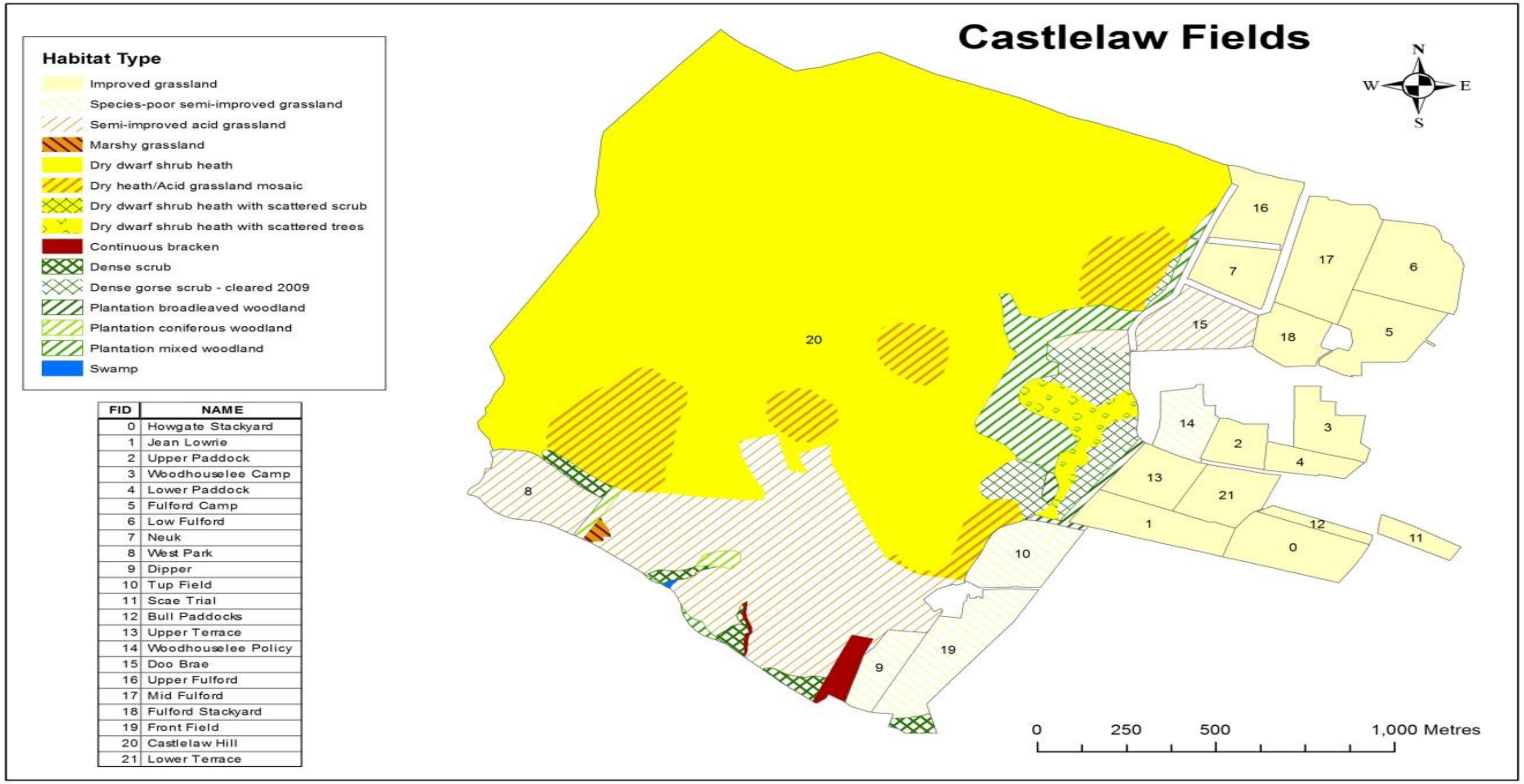


Figure 3.2 Timeline showing farm husbandry procedures and data collection performed on Castlelaw hill farm in Year 2. Blue text indicates farm husbandry, orange text indicates data collection

3.2 Animal location and farm husbandry

The sheep used in the indicator evaluation study were located on the SRUC Castlelaw hill farm in the Pentland hills south of Edinburgh. The farm flock consists of approximately 700 breeding Scottish Blackface ewes and their lambs have been studied for more than 20 years. All ewes and lambs on farm are covered by a Project Licence granted by the UK Home Office (PPL 60/3624 licence holder Lutz Bunger). The grazing land is owned by the UK Ministry of Defence and is located approximately 305 to 488m above sea level (Latitude: 55.859904, Longitude: -3.23292). The ewes were managed outdoors on natural pasture all year round and moved between free range grazing on the hill (unfenced, unimproved pastures of low nutritional quality), fenced areas of unimproved hill grazing (250 hectares unimproved/less favoured grazing), or improved fields near the farmstead (22 hectares of fenced improved grazing). These relocations were subject to management decisions and the position in the reproductive cycle and calendar season. On this farm the flock is closed and the sheep are managed to achieve commercial growth rates, using commercially relevant management. Decisions influencing sheep husbandry were made independently from this study.

The un-improved grazing land is dominated by heather moorland (*Callunas vulgaris*), together with a range of other grass species as shown in Figure 3.3.



© Crown Copyright/database right 2015. An Ordnance Survey/EDINA supplied service.

Figure 3.3 Plant diversity present on Castlelaw hill farm. Map produced by Sandra Stewart (SAC Consulting), image obtained from Dr John Holland

During normal management practices in late gestation, and for a brief period post-lambing, the sheep were moved by the shepherd to different grazing areas of the hill. For lambing the ewes were located at the bottom of the mapped areas (fields 9, 10 and 19 on map, and one unmarked area). They had unlimited access to grazing and were provided with nutritional supplementation approximately six weeks prior to lambing (16% Protein Ewe concentrate, East Coast Viners, UK, approx. 250-500g/ewe per day). The exact volume provided to the sheep is dependent on weather and grass availability and the average body condition scores recorded. The supplementation was provided to the ewes by the shepherd using a quad bike to transport it to specific areas on the hill in which the sheep tend to congregate.

Castlelaw farm operates under typical extensive farm management conditions and the animals spend the majority of their lives outside although they are gathered at least five times a year for routine husbandry procedures. The ewes are gathered in early September during which the shepherd performs the annual stock draw. Ewes experiencing recurring lameness, mastitis, or other health problems are removed from the flock at this time along with the oldest animals. If there is doubt regarding a ewe's ability to thrive throughout another breeding cycle due to low body condition scores and poor dentition, they are also removed from the flock. The removed animals are either relocated to an SRUC lowland farm in which living conditions are gentler than in the extensive system (draughted), sold or culled. At this time, 18month old ewes are added back in to the breeding flock for their first mating after wintering in the Scottish Borders.

In November, the ewes which have been retained for breeding are gathered, body condition scored by the shepherd, and assigned to one of 15 mating groups. The ewes aged between 1.5 and 6.5 years are naturally mated each year to 15 rams (13 Scottish Blackface and 2 Lleyns) over the course of 4 weeks. At the end of the four weeks the ewes are returned to the hill with three Bluefaced Leicester rams for an additional two weeks. The ewes remain on the unfenced hill ground until February.

In February, the ewes are gathered again for ultrasound scanning for the determination of pregnancy and litter size. Their body condition score is also re-assessed before they are returned to the hill ground. At the beginning of April the ewes are re-gathered and separated in to three groups: those which have not conceived (~10%), those expecting singles (~53%), and those scanned with twins (~35%) and triplets (~2%).

Lambing begins in mid-April and typically continues until the end of May. The ewes are not housed during lambing as on lowland farms, but are kept in fields near to the farmstead. Approximately 16 hectares of the improved grazing is used to hold the ewes at this time along with 11 hectares of unimproved hill ground (approximate average stocking density 25 ewes per hectare). The ewes remain in the lambing fields until the lambs are considered strong enough to return to the hill. This decision is made by experienced shepherds and generally lambs will go to the hill with their mother around 24-48 hours after birth. The decision is based upon the lamb's ability to maintain pace with the ewe and forecast weather conditions.

Before the ewes and lambs leave the lambing fields the lambs are caught and information regarding their sex, weight and mother's identity is recorded. At this time the lambs' ears are also tagged using Allflex lamb tags (Babe tags, Allflex, Europe). The lambs also receive preventative medication at this time for watery mouth (Spectam Scour Halt oral solution 1ml, Ceva Animal Health LTD, Buckinghamshire, England) and orf (0.02ml Via skin scarification inguinal groin area. Scabivax Forte, Schering-Plough LTD, Uxbridge, England). The lambs are also marked using spray paint (Richy stockmarker, Ripon, England) to later aid in the identifying the incorrect pairing of ewes and lambs.

The ewes and lambs are gathered in June. The ewes are shorn and body condition scored again, and the lambs are weighed and fleeces are re-marked with spray. All lambs are typically weaned in August; however in 2013 (Year 2) only the male lambs were weaned at this time due to space restrictions on the farm, and the females stayed with the ewes until the annual stock draw in September (after the end of this data collection period). In Year 1 all lambs were weaned from the ewes and relocated to an area in which they no longer had visual or auditory contact with their dams. Details on the animals, general husbandry and study design for this longitudinal study are given in this chapter. Additional details for specific methodologies are provided in their relevant chapters.

3.3 Study animal selection

Ewes were selected for inclusion in this project from the 700 head farm flock. The ewes were carefully selected in order for them to be representative of the breeding flock in terms of age, body condition scores and split evenly between mating groups. This selection was made prior to mating in the autumns of 2011 and 2012. Although the ewes selected for this study were subject to additional Home-Office licenced procedures, physical health assessments and behavioural observations they remained as an integral part of the breeding flock and received normal management. The ewes in the sub-flock, and their lambs, were used during the data collection reported in Chapters 4, 5, and 6.

3.3.1 Animal selection for Year 1 (November 2011 – August 2012)

Ewes

In November 2011, prior to mating, one hundred ewes were selected from the breeding population for inclusion in the sub-flock during the first year of the study. The demographics of the breeding flock and sub-flock using the information obtained in September 2011 are shown in Tables 3.1 and 3.2. These animals remained in the sentinel sub-flock until stock draw in September 2012.

Table 3.1 Percentage of ewes in Castlelaw breeding flock, and the Year 1 sub-flock, born between 2005 and 2010

Birth Year	Percentage of breeding flock	Percentage of Year 1 Sub-flock
2005	2	2
2006	8	8
2007	17	15
2008	20	15
2009	25	27
2010	28	33

Table 3.2 Weight and body condition scores obtained for the Castlelaw breeding flock and the Year 1 sub-flock in September 2011

	Breeding Flock	Year 1 sub-flock
Mean weight (kg)	52.30	51.4
Median BCS	2.5	2.5
Lightest Ewe (kg)	36.10	36.40
Heaviest Ewe (kg)	72.56	69.00
Lowest BCS	1.75	1.75
Highest BCS	4	4

Lambs

The ewes in the sub-flock gave birth to 115 live lambs in April 2012. The lambs were also included in the indicator development project. The lambs remained in the project until they were weaned and removed from the farm at approximately four months of age in August 2012. Although indicator data were collected on the lambs in June (marking) and August (weaning) these results are not reported in this thesis due to time constraints. The relationship between lamb birth, marking and weaning weights and ewe welfare indicator scores are explored in Chapter 4. Further lamb demographic information can be found in Table 3.3.

Table 3.3 Lamb produced by sub-flock in Year 1

N. ewes	N. live lambs born	Sex of lambs		Litter size		Breed of lambs		Lamb birth weight (kg)	
		Male	Female	Single	Twin	Scottish Blackface	Cross	Total	Individual lamb mean
100	115	63	52	54	61	103	12	479.60	4.28

3.3.2 Animal selection for Year 2 (September 2012 – August 2013)

Ewes

In the second year of the study 110 ewes were used for data collection. Of these 110 sheep, 75 ewes were retained from Year 1 and an additional 35 ewes were selected to replace older animals which were draughted or culled at the stock draw in September 2012. Figure 3.4 (below) illustrates this process. Prior to the annual stock draw, two Year 1 sub-flock ewes had died; one from mastitis and one of unknown reasons. Twenty three of the ewes from Year 1 study group were removed from the farm at the stock draw: 12 were removed due to tooth loss, the two oldest were automatically taken out due to old age, four were also removed to another farm for further breeding, three were removed due to low body condition scores, and two were removed due to chronic udder problems. The reasons listed here are the primary reason given by the shepherd for their removal from the flock; some animals may have had multiple health issues leading to their removal, for example, poor dentition and a low body condition score.

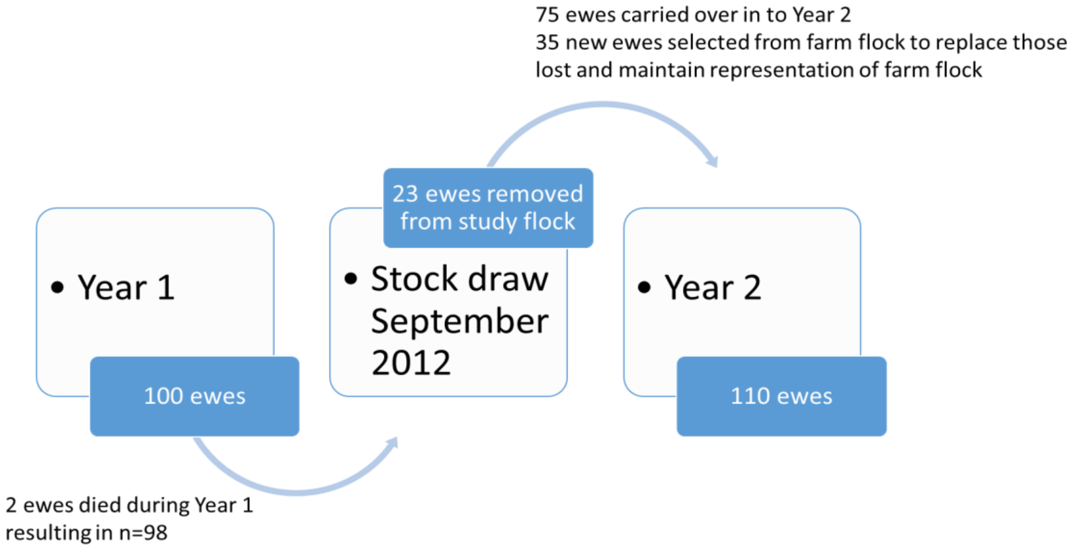


Figure 3.4 Diagram illustrating ewes' inclusion or removal in the sub-flock at the end of Year 1

The ewes selected for the Year 2 study group were again representative of the flock demographic in terms of their age, body condition and live weights in the autumn of 2012. These animals stayed in the sub-flock until the end of the data collection period in September 2013. The demographics of the breeding flock and sub-flock using the information obtained in September 2012 are shown in Tables 3.4 and 3.5.

Table 3.4 Percentage of ewes in Castlelaw breeding flock, and the Year 1 sub-flock, born between 2005 and 2010

Birth year	Percentage of Breeding flock	Percentage of Year 2 sub-flock
2006	2	3
2007	7	7
2008	13	14
2009	19	20
2010	26	26
2011	31	30

Table 3.5 Weight and body condition scores obtained for the Castlelaw breeding flock and the Year 2 sub-flock in September 2012

Measure	Breeding Flock	Year 2 sub-flock
Mean weight (kg)	52.48	52.86
Median BCS	2.75	2.75
Lightest Ewe (kg)	28.8	34.5
Heaviest Ewe (kg)	81.4	76.4
Lowest BCS	1.5 (one ewe)	1.75
Highest BCS	4	4

Lambs

The ewes in the second years sentinel sub-flock gave birth to 125 live lambs in April 2013 as shown in Table 3.6. As before the lambs were also included in the indicator development project. The lambs remained in this project until weaning. The male lambs were weaned and relocated in August and the female lambs remained with the ewes until stock draw in September 2013. Although indicator data were collected on the lambs in June (marking) and August (weaning) as above these results are not reported in this thesis due to time constraints. The relationships between lamb birth, marking and weaning weights, and ewe welfare indicator scores were explored in chapter 4.

Table 3.6 Lambs produced by sub-flock in Year 2

N. ewes	N. live lambs	Sex of lambs		Litter size		Breed of lambs		Lamb birth weight (kg)	
		Male	Female	Single	Twin	Scottish Blackface	Cross	Total	Individual lamb mean
110	125	63	62	53	72	120	5	472.00	3.78

3.4 Data collection

3.4.1 Indicator refinement

Some of the 32 indicators selected in Chapter 2 were seen to be only applicable to dairy or housed sheep: hoof overgrowth, space allowance and stocking density, and response to humans at milking. As the animals focussed upon in this thesis are unhoused meat sheep these indicators were therefore not evaluated or reported on here. The work conducted in this project occurred on an SRUC owned farm and therefore resource based measures were not included in the assessment: provision of clean water and provision of shelter. Panting and shivering were not assessed due to feasibility, along with aggression and displacement behaviour. Ear damage and appropriate tail length were also not assessed. No lambs on Castlelaw farm are tailed docked and ear tagging is performed by trained staff. Eye condition, fleece and skin condition and skin lesions were not scored using scales during the Assessments at Gather however if any evidence of infection or damage were observed during the assessment this was recorded in a comment section of the score sheet and the farm staff were immediately informed. The dismissal of the above 13 indicators from the

32 proposed in Chapter 2 resulted in the application of 19 indicators throughout this study as summarised in Figure 3.5.

Figure 3.5 Dismissal of 13 unsuitable indicators for this project resulting in 19 indicators receiving evaluation

The 19 indicators which provisionally deemed to be the most relevant for unhoused animals bred for meat production in Scotland, were applied and evaluated throughout this PhD project. As shown in Table 3.7 the data collected on the sub-flock ewes occurred during the Assessments at Gather in which the animals were gathered and handled, and during Assessments in the Field when the animals were observed on the hill. Information on the lamb birth, marking and weaning weights were obtained from the farm records, along with ewe mortality records.

Table 3.7 Indicators applied to ewes on Castlelaw farm during longitudinal data collection period over Years 1 and 2

Principle	Criterion	Indicator	Assessment	Aspect to be evaluated
Good Feeding	Absence of prolonged hunger	Body Condition Score (2 scales)	Assessment at Gather	Repeatability
		Tooth loss	Assessment at Gather	Repeatability, convergent validity
		Lamb birth and weaning weights and percentage lambs weaned	Records - Iceberg Indicator	Validity, repeatability
Good Environment	Comfort around resting	Coat cleanliness	Assessment at Gather and Field	Repeatability
Good Health	Absence of disease	Presence of coughing	Assessment at Gather	Repeatability
		Presence of nasal discharge	Assessment at Gather	Repeatability
		Udder condition: symmetry, colour and presence of fibroids	Assessment at Gather	Repeatability, feasibility
		Faecal soiling	Assessment at Gather and Field	Repeatability, validity, feasibility
		FAMACHA anaemia chart	Assessment at Gather	Repeatability, validity, feasibility
		Lameness (2 scales)	Assessment at Gather and Field	Repeatability, validity, feasibility
		Ewe mortality rates and removal from flock	Records	Validity, repeatability
Appropriate Behaviour	Social Behaviour	Behavioural Synchrony	Assessment in the Field	Validity, feasibility
		Nearest neighbour	Assessment in the Field	Validity, feasibility
		Vocalisations by ewes and lambs	Assessment in the Field	Validity, feasibility, repeatability

Principle	Criterion	Indicator	Assessment	Aspect to be evaluated
Appropriate Behaviour	Expression of other behaviours	Vigilance behaviour	Assessment in the Field	Validity, feasibility, repeatability
	Positive emotional state	Qualitative Behavioural Assessment (QBA)	Assessment in the Field	Validity, feasibility, inter-observer reliability, repeatability
	Good Human-animal relationship	Response to human approach	Assessment in the Field	Validity, feasibility, repeatability
Appropriate Behaviour	Absence of General Fear	Response to surprise test	Assessment in the Field	Validity, feasibility, repeatability
		Recovery from surprise test	Assessment in the Field	Validity, feasibility, repeatability

3.4.2 Procedure for data collection during for Assessments at Gather

On five occasions each year (at key points in the reproductive cycle of the ewe as shown in Table 3.8), when the sheep were gathered for routine husbandry procedures, the ewes in the sentinel sub-flock were given a "Nose to Tail" physical assessment comprising of the indicators in listed Table 3.7 During the Assessments at Gather the ewes were either inside the shed or in outdoor handling pens, where they were gently restrained. The Assessments at Gather were performed on one animal at a time. The results from these assessments are presented and discussed in Chapter 4.

Table 3.8 Distribution of ten Assessments at Gather over the two year period

Assessment at Gather	Year of study	Month	Point in reproductive cycle
1	1	November	Pre-mating
2	1	February	Mid-pregnancy
3	1	March	Late-pregnancy
4	1	June	Mid-Lactation
5	1	August	Weaning
6	2	November	Pre-mating
7	2	February	Mid-pregnancy
8	2	March	Late-pregnancy
9	2	June	Mid-Lactation
10	2	August	Weaning

The ewes were gathered from the hill by the shepherd using between one and three dogs at least one day before the data collection. The sentinel sub-flock were separated from the rest of the farm flock and were body condition scored by the shepherd. The sub-flock were situated an in-bye field overnight prior to the

Assessment at Gather. On the morning of data collections the ewes were gathered in pens. When lambs were "at foot" (June and August), they were separated from the ewes to improve efficiency and prevent injury during the data collection but reunited as quickly as possible following the assessment. In order to ensure the data collection was time efficient and feasible a team of research assistants and students (between 1 and 5) assisted in the data collection. The Welfare Quality principles covered by the Assessment at Gather were Good Feeding, Good Environment, and Good Health. The animal-based indicator data collection is described below.

Data collection for indicators relating to Good Feeding

Indicator: Body condition scoring

Assessment: Russel et al 1969 and novel simplified scale.

Body condition scoring was performed on the ewes by the shepherd as the sheep in the sub-flock were separated from the rest of the breeding flock. He used the six point scale (with sub divisions of half and quarter points) developed by Russel et al in 1969. Scores were assigned to animals based up on the prominence of their spine and hook bones during palpation of the lower back, descriptions of the categories as described by Russel et al are shown in Table 3.9. A simplified version of this scale was also applied to the animals at this time which resulted in the animals being classified as "emaciated", "thin", "fit" or "fat".

Table 3.9 Body condition score criteria (Russel et al 1969) applied to sheep during Assessments at Gather

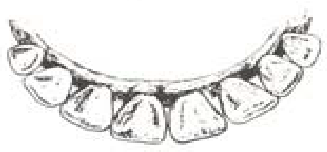
Body Condition Score	Description (as in Russel et al 1969)
0	Extremely emaciated and on the point of death
1	Spinous processes prominent and sharp. Transverse processes also short, the fingers pass easily under the ends, and it is possible to feel between each process. Mm. longissimus dorsi shallow with virtually no subcutaneous fat cover.
2	Spinous processes prominent but smooth, and individual processes can be felt only as fine corrugations; transverse processes smooth and rounded, and fingers can be passed under ends with little pressure; Mm. longissimus dorsi of moderate depth with little subcutaneous fat cover.
3	spinous processes have only a small elevation, are smooth and rounded, and individual processes can be felt only with pressure; transverse processes smooth and well covered, and firm pressure is required to feel over ends; Mm. longissimus dorsi full with moderate subcutaneous fat cover
4	Spinous processes can be detected with pressure as hard line between ends; Mm longissimus and associated subcutaneous fat; transverse processes cannot be felt; Mm. longissimus dorsi full with thick subcutaneous fat cover
5	Spinous processes cannot be felt even with firm pressure and there is a depression in sub-cutaneous fat spinous processes would normally be felt; transverse processes cannot be felt; Mm. longissimus dorsi very full with very thick subcutaneous fat cover; there may be large deposits of fat over tump and tail.

Indicator: Tooth loss

Assessment: Three point scale describing tooth loss and damage

The dentition of the animal was assessed by lowering the bottom jaw and lip to allow a visual inspection of the incisors. The presence of damaged teeth, or loss of teeth was scored using a three point scale shown in Table 3.10.

Table 3.10 Scale used to assign tooth loss score when assessing the incisors of ewes during the Assessment at Gather

Tooth loss score	Description	Diagram showing location of missing teeth (original diagram from NSW Agriculture, Australia)
0	No tooth loss. Animal has all eight incisors with no evidence of damage.	
1	Minor tooth loss. Some missing teeth detected. Up to three teeth broken or missing from the middle, lateral or corner incisors (highlighted in diagram).	
2	Substantial tooth loss. Four or more middle, lateral or corner incisors missing or damaged. Or at least one central incisor missing (highlighted in diagram).	

Data collection for indicators relating to Good Environment

Indicator: Coat cleanliness

Assessment: Five point scale based on the Food Standards Agency body cleanliness scale.

Coat cleanliness was assessed by way of a visual and tactile assessment of the fleece. A clean fleece resulted in the animal receiving a score of 0 however the presence of dirt, dung or other matter contaminating the fleece resulted in the animal receiving a score between one and four as shown in Table 3.11.

Table 3.11 Coat cleanliness scale used to assess severity of fleece soiling

Coat cleanliness score	Description	Pictorial scale
0	Clean and dry. *Image from FSA Cleanliness classification of livestock*	
1	Dry or slightly damp due to current weather conditions. If damp the moisture is only on the most outer layer of the fleece. Small amounts of mud or dirt may be on the body however this is due to the environment the sheep is in for the handling assessment and is not representative of normal conditions. *Image from FSA Cleanliness classification of livestock*	
2	Fleece is very damp or wet to the touch. Dirt and dung have contaminated the fleece prior to their presence in the handling area. *Image from FSA Cleanliness classification of livestock*	
3	Fleece is visibly wet and heavily contaminated with dirt and dung which may extend below the outer layer of the fleece. *Image from FSA Cleanliness classification of livestock*	
4	Coat is saturated with water and very heavily contaminated by dirt and dung. The dirt may extend to the inner layer of the fleece or touch the skin. Face and legs may also be coated in mud or faeces. *Image from AWIN Sheep assessment protocol*	

Data collection for indicators relating to Good Health

Indicator: Coughing

Assessment: Binary scale: present or absent

The assessment of coughing was conducted by listening throughout the assessment procedure and any incidence of coughing was recorded. If an animal coughed while another animal was assessed the identity of the coughing ewe was recorded and this information was included in her assessment results. The absence of coughing was assigned a score of 0, and the presence 1.

Indicator: Nasal discharge

Assessment: Binary scale: present or absent

The presence of nasal discharge was assessed by gently tilting the head of the sheep towards the assessor and examining the nostrils and philtrum of the ewe. If no nasal discharge was present the ewe received a score of 0, however if nasal discharge was seen a score of 1 was assigned.

Indicator: Udder condition (during lactation only)

Assessment: Udder symmetry, discolouration, presence of fibrosis and orf lesions

During the Assessments at Gather which took place during lactation (June and August) the ewes' udders were assessed for symmetry, discolouration, and presence of fibrosis and orf leisons. The scale and criteria for these measures are shown below in Table 3.12. Orf is caused by a parapox virus and causes papules which spread from the lamb's mouth to the ewe's udder during suckling. The assessment of orf leisons was performed using a binary scale: presence or absence (not listed in Table). This assessment was conducted visually when the animals were standing in a raceway, and during the assessment of fibrosis both sides of the udder were gently palpated by the assessor's hand. Any other injuries or abnormalities discovered during this assessment were recorded as notes.

Table 3.12 Scale used to assess udder health of lactating sheep during the Assessments at Gather

Indicator	Score	Criteria
Udder symmetry	0	Both sides of udder are even in size and shape
	1	One side of the udder is clearly smaller than the other when assessed visually however no signs of inflammation are present. Most likely caused by having a single lamb
	2	Udder is asymmetrical due to inflammation or infection.
Udder discolouration	0	No discolouration (redness, darkening of the skin) is present.
	1	Slight discolouration present but only affects less than 25% the udder.
	2	Moderate discolouration is present affecting up to 50% of one side
	3	Severe discolouration (at least 50% of one sideand/or both sides affected)
Udder fibrosis	0	Both sides are soft with no lumps or hardness detected
	1	One or two lumps felt or hard patches affecting less than 50% of one side
	2	Multiple lumps felt in one side of the udder, or small lumps in both sides
	3	Lumps or hardness making up a significant part of the udder at least one side

Indicator: Faecal soiling

Assessment: Five point “dag” scale

The extent of faecal soiling was assessed by eye and compared to a five point faecal soiling scale shown below (Figure 3.6). This scale was originally developed by researchers at AgResearch,Wallaceville in New Zealand for use with Merino lambs. The scale ranges from 0 indicating no faecal soiling to 4 demonstrating extensive soiling to hocks.

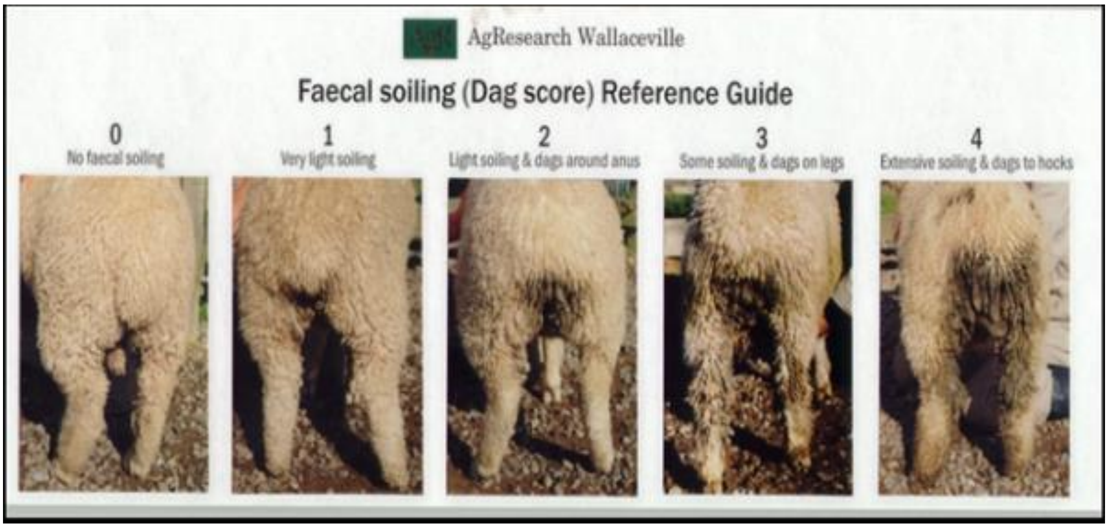


Figure 3.6 Dag score chart used to assess faecal soiling of ewes. Score developed by AgReasearch, New Zealand

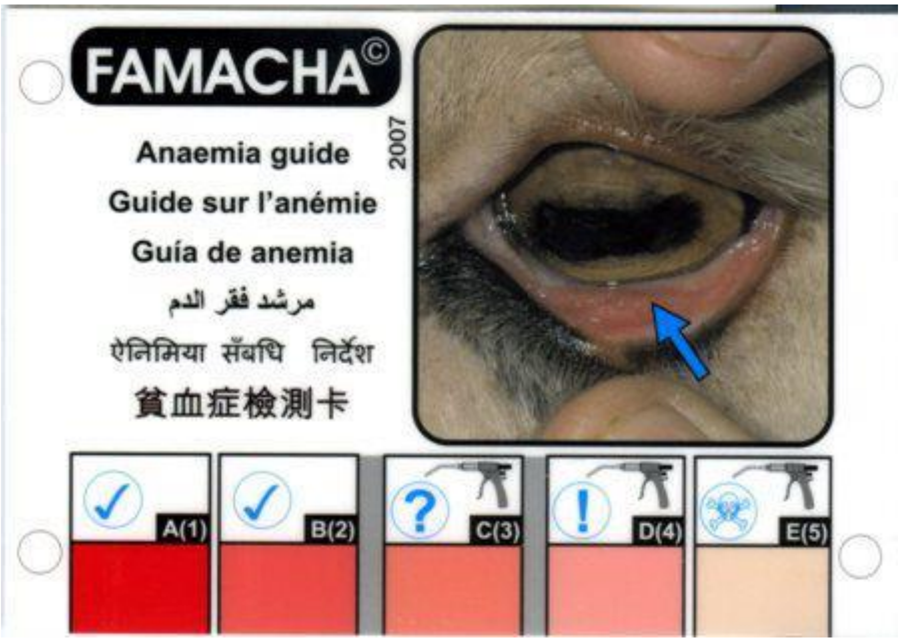


Figure 3.7 The FAMACHA chart developed by Bath et al (1996). To assess anaemia the colour of the animal's lower eye lid is compared to the scale

Indicator: Anaemia

Assessment: FAMACHA scale

To assess anaemia using the FAMACHA chart the lower eye lid was gently retracted and the colour of the facies posterious palperbrae compared to the five point FAMACHA colour scale as in Bath and Wyk (2009) shown in Figure 3.7.

Indicator: Lameness

Assessment: Kaler and Green (2009) lameness score and novel simplified scale.

Lameness was the last indicator to be applied to the sheep at the end of the Assessments at Gather. The ewes were allowed to regroup following their individual assessments and moved to the concrete area in between the shed and outdoor handling pens. As it was not feasible to assess the lameness of ewes individually they remained as a group on the concrete ground. One assessor (SER) walked around the perimeter of the area causing the animals to retreat. The assessor then walked through the middle of the group to ensure a good view of each ewe's gait. From this distance individual ewe ear tags could be read and ewes with an uneven gait were listed. The lameness scale used during this assessment was the Kaler and Green (2009) numerical scale shown below in Figure 3.8. This scale uses the chart to assign a score to a sheep to describe her movements. The numerical scale is read vertically and when using the chart, the assessor judges the sheep using the criteria and assigns a score for which the grey boxes are filled. For example a sheep which "bears weight evenly on all four feet" receives a score of 0, and a sheep which displays "uneven posture, short stride length and visible nodding of head" is allocated a score of 2.

A simplified scale was also applied to the sheep during this assessment. The letters below the chart in Figure 3.6 are the four categories. Ewes which would receive a score of 0 or 1 using Kaler and Green's chart were considered to be an "A", those scoring a 2 or 3 were now a "B", ewes which classified as a 4 or 5 on the numerical scale were a "C", and ewes which were unable to "stand or move" would be classed as a D. Following this assessment on the concrete ground the ewes were then moved to an area of flat grass behind the shed and were re-assessed using the same technique and scales.

The locomotion scoring scale, shaded area = all required for score

Score	0	1	2	3	4	5	6
Posture and locomotion							
Bears weight evenly on all four feet	■						
Uneven posture, but no clear shortening of stride		■	■	■	■	■	
Short stride on one leg compared with others		■	■	■	■	■	
Visible nodding of head in time with short stride			■				
Excessive flicking of head, more than nodding, in time with short stride				■	■	■	
Not weight bearing on affected limb when standing				■	■	■	
Discomfort when moving				■	■	■	
Not weight bearing on affected limb when moving					■	■	
Extreme difficulty rising						■	
Reluctant to move once standing						■	
More than one limb affected						■	
Will not stand or move							■

A B C D

Figure 3.8 Gait score comprising of Kaler and Green's numerical lameness scale, and complementary novel four point scale below

3.4.3 Procedures for data collection during the Assessments in the Field

In Year 2, data collection was also performed when the animals were in fenced and unfenced areas of the hill during Assessments in the Field. The animals were not relocated or moved for these observations, they were assessed in their normal environment. The Assessments in the Field took place between March and August 2013. The observations and tests conducted during these assessments were largely concentrated on behavioural measures but some physical health traits were also assessed. The indicators evaluated during Assessments in the Field cover Good Health, Good Environment and Appropriate Behaviour. The dates of assessments and indicators applied are shown in Table 3.13. During the first eight observations only the Qualitative Behavioural Assessment (QBA) was conducted due to a coinciding MSc project. The results from these observations are reported and discussed in Chapters 5 and 6.

To aid the feasibility of collecting these measures, this data collection focussed on a subset of the sub-flock, and 57 of the ewes were selected as shown in Figure 3.9 below. The sample group were selected following ultrasound scanning for pregnancy determination and litter size in February 2013. All selected ewes were carrying at least one lamb when scanned, however at lambing four ewes did not produce any offspring. These ewes remained with the rest of the sub-flock and farm breeding flock throughout this time period. The period following the removal of the male lambs is referred to as “post-weaning” even though the female lambs remained with their dams.

Table 3.13 The indicators performed during each Assessment in the Field, and ewes' stage in reproductive cycle during assessment

Assessment Number	Date	Indicator/test performed	Point in reproductive cycle
1	5/4/13	QBA	Late-pregnancy
2	8/4/13	QBA	Late-pregnancy
3	11/4/13	QBA	Late-pregnancy
4	15/4/13	QBA	Late-pregnancy
5	19/5/13	QBA	Post-lambing
6	22/5/13	QBA	Post-lambing
7	27/5/13	QBA	Post-lambing
8	30/5/13	QBA	Post-lambing
9	11/7/13	QBA, behavioural synchrony, vigilance, nearest neighbour, faecal soiling, ewe and lamb vocalisations, human approach, response to, and recovery from, surprise test.	Post-lambing
10	22/7/13	QBA, behavioural synchrony, vigilance, nearest neighbour, faecal soiling, ewe and lamb vocalisations, human approach, response to, and recovery from, surprise test.	Post-lambing
11	8/8/13	QBA, behavioural synchrony, vigilance, nearest neighbour, faecal soiling, ewe and lamb vocalisations, human approach, response to, and recovery from, surprise test.	Post-lambing
12	8/8/13	QBA, behavioural synchrony, vigilance, nearest neighbour, faecal soiling, ewe and lamb vocalisations, human approach, response to, and recovery from, surprise test.	Post-weaning
13	30/8/13	QBA, behavioural synchrony, vigilance, nearest neighbour, faecal soiling, ewe and lamb vocalisations, human approach, response to, and recovery from, surprise test.	Post-weaning

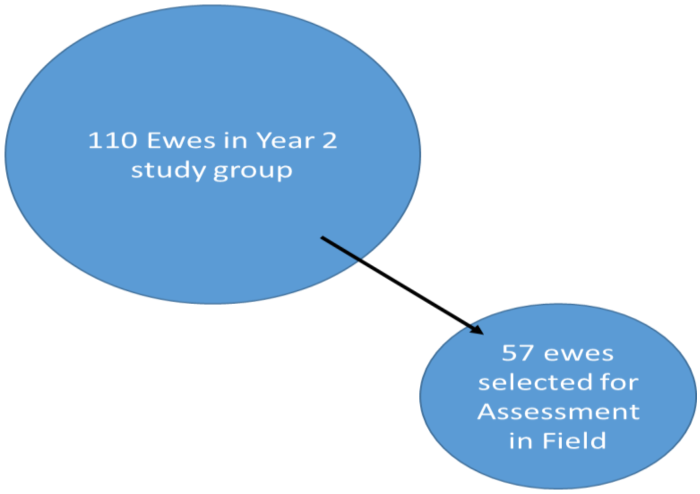


Figure 3.9 Ewes included in sub-set which were included in the Assessments in the Field

The ewes were primarily identified using their ear tags. However, to increase the distance from which the animals could be individually identified a green stripe was applied with marker spray across their shoulders (Ritchey super sprayline stockmarker, England), and uniquely coloured and numbered tags (Alflex maxi female cattle tag, Alflex Europe) attached around their neck with string, shown in Figure 3.10. The ewes were assessed in the order in which they were found.

Figure 3.10 Ewe from sub-flock in lambing field. The green stripe across shoulder and tag around neck aided in the identification of individuals

Each individual ewe was identified and observed from a distance of between 5 and 20 metres. In poor weather, or at distances greater than 10 metres binoculars (Olympus 8.40 DPS I) were used to ensure a good view of the animal.

Data collection for indicators relating to Good Environment

Indicator: Coat cleanliness

Assessment: Five point scale based on the Food Standards Agency body cleanliness scale.

During the Assessment on the Hill the cleanliness of the fleece was assessed using the same scale described above.

Data collection for indicators relating to Good Health

Indicator: Lameness

Assessment: Kaler and Green (2009) lameness score and novel simplified scale.

The lameness scale was the same as used above however this was applied when the ewes were walking on the hill, not moved to areas of concrete ground as during the Assessments at Gather.

Indicator: Faecal soiling

Assessment: Five point scale based on the AgResearch, Wallaceville, New Zealand scale.

During the Assessment on the Hill the soiling of the rear was assessed using the same scale described above.

Data collection for indicators relating to Appropriate Behaviour

Indicator: Qualitative Behavioural Approach

Assessment: Undisturbed observation followed by qualitative scoring

In order for the observation to begin the ewe's body and face had to be fully visible. To avoid any disturbance and bias the QBA observations were conducted prior to any other measure. The QBA term development and data collection methods are explained in more detail in Chapter 5.

If ewes were disturbed by the presence of an observer (as determined by interruption of their behaviour or posture to focus attention on the observer) they were given time to resume the behaviour they were performing prior to disruption. Once normal behaviour had resumed observers waited for a further 30 seconds before the observation started. After each sixty second observation of an individual sheep each term on the QBA fixed list was scored using visual analogue scales.

Indicator: Nearest neighbour

Assessment: Visual estimation of distance between focal ewe and nearest adult sheep

The ewe's nearest neighbour was calculated by a visual estimation of the distance (m) between the focal ewe and the nearest adult ewe (not including any lambs). The ewes were considered to be part of a social group (temporary associations of adult sheep) if they were located within 30 meters from each other (as in Lawrence and Wood-Gush 1988). Ewes without nearest neighbours within a 30 metres radius were considered to be alone.

Indicator: Behavioural Synchrony and vigilance

Assessment: Scan sample

The social group containing the focal ewe was then observed during a scan sample. The total number of ewes and lambs in the social group were calculated by performing a head count. The number of ewes in each group performing the behaviours and postures described in Table 3.14 were counted.

Table 3.14 Ethogram describing behaviours and postures scored during scan samples performed during Assessments in the Field

Behaviour	Definition
Vigilance	Attention is focused on another ewe or stimulus. Head is raised above the level of the back, head and ears pointed forward.
Rumination	Regurgitating feed in to mouth and chews.
Feeding/drinking	Chewing or obtaining grass or foliage, or water from trough.
Locomotion	Walking or running.
Resting /Sleeping	Lying on ground, absence of other behaviour
Urination/defecation	Excretes urine or faeces
Scratching	Rubs body or head against fencing, wall or water trough.
Investigate	Looks at, sniffs or chews item in order to investigate an object, other sheep, or environment.
Attention is on lambs	Attention is focussed on the lambs.
Unclear	The ewe's behaviour is concealed by a visual barrier e.g. bush or another ewe.
Posture	**Definition**
Stand Head Up	Stands stationary on four limbs, head above, or level with shoulders
Stand head down	Ewe stands on four limbs with head lower than shoulders
Lie	Ewe lies with flank or chest on the ground, legs either tucked or extended
Walk	Ewe is upright on four limbs moving around environment

Synchrony of the social group was then calculated using Rook and Penning's (1991) kappa coefficient calculations:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

Where P(A) represents the proportion of synchrony across all observations, and P(E) is the synchronisation for each individual activity or posture. This equation calculates whether the synchrony observed was higher than could be expected by chance based on the number of behaviours or postures possible and previously observed. If K=1 complete true group synchrony was observed but if K=0 the synchrony occurred by chance.

Whether an individual ewe was synchronised with their social group was determined using Rukstuhl's group mean (Ruckstuhl 1999). This resulted in a ewe receiving a binary code of 0 or 1 depending upon whether their posture was the same or different to the highest proportion of the group.

Indicator: Ewe and Lamb vocalisations

Assessment: Number of vocalisations within a two minute period

The sub flock was then observed a further two minutes (timed using a stop watch) during which the total number of lamb and ewe vocalisations from within the group were counted.

Finally to assess the ewes Absence of General Fear during the Assessment in the Field, their reaction to a human approach, and response to and recovery from a surprise test were the last indicators to be assessed as it required the most disturbance of the ewe. If more than one focal animal was in the same group this was performed after all animals in the group had other indicator scores recorded. If the other focal ewes were disturbed by another ewe's testing they were left to resume previous behaviour before their assessment.

Indicator: Human approach test

Assessment: Distance to which a human can approach before sheep retreats

The assessment of the human approach indicator began when the ewe was orientated towards the assessor. The assessor approached the ewe head on at a slow and gentle pace at one step per two seconds. The person approaching the sheep held one arm out at a 45* angle with their palm facing the ground. The distance between the human, and the sheep at the point the ewe retreated from the human was estimated visually in meters and recorded. Once the ewe started to retreat the assessor stopped immediately.

Indicator: Response to surprise

Assessment: Distance fled during surprise test

When the ewe stopped retreating from the assessor and was stationary and looking towards the assessor, an umbrella which had previously been concealed behind their back was held out in front of the assessor at a 45° angle (Union Jack umbrella 115cm diameter when opened. Umbrella World, Lancashire, UK. Figure 3.11). The umbrella was opened by pushing a button on the handle. The ewe's flight from this surprise test was recorded in metres.

Figure 3.11 Example of umbrella used during surprise test. Image obtained from www.amazon.com

Indicator: Recovery from surprise test

Assessment: Time taken to recover following surprise test

When the umbrella was opened in the surprise test a stop watch was started. The time taken for the ewe to return to the behaviour she was performing prior to the approach test (e.g. grazing) was timed. At the end of this observation the Assessment on the Hill was complete.

Data preparation analysis

The data obtained during the assessments were recorded on preformatted paper sheets. Following the assessments information was transcribed in to Mircosoft Excel. The data collected during the Qualitative Behavioural Assessment was not collected using paper scoring sheets but using a specially developed app for use on Android tablets (described in detail in Chapter 5). This data was exported from the tablet app directly to Excel. As the data obtained differed considerably in type (continuous, categorical or binary) the subsequent analyses were approached in different ways. The data preparation and analysis techniques are explained in the relevant chapters.

3.4.4 Additional data collection during Assessments at Gather

To complement and further the interpretation of the indicators applied to sheep in this thesis physiological data were collected. Information collected from blood and faeces can be used to identify infections or parasite infestation. Although the collection of these is not feasible during an on-farm welfare assessment obtaining this information may be used to further validate indicators before accepting or rejecting them for use with extensively managed sheep.

As discussed in the previous chapter body condition scoring provides an indication of chronic nutritional imbalances in the body. The concentration of metabolites in the blood, non-esterified fatty acids (NEFA) and the ketone B-hydroxybutyrate (BHB), were selected to provide information on the current energy status of the animal. Ketones such as BHB are produced from the metabolism of NEFAs and other fatty acids in the liver (González et al., 2011).High concentrations of NEFA and BHB can be used to indicate metabolic disorders (Gonzales et al 2011) and that the glucose available in food is not sufficient in meeting the demands of the animals and so it is necessary to mobilise adipose tissue (fat) (Gonzales et al 2011). This typically occurs during times when energy demand is much increased, such as pregnancy or lactation (Regnault et al., 2004). As the levels of the metabolites and subsequently the levels of acids in the blood increase (and pH decreases) ruminants are at a greater risk of disease such as ketosis and pregnancy toxaemia (González et al., 2011; Ospina et al., 2010). Over time this will lead to the animals body condition score decreasing. These measures can be used ahead of live weight or body condition scoring to understand nutritional balance and inadequacies in nutrition can be addressed prior to potentially irreversible physiological shifts (Russel and Wright, 1983).

To further validate the FAMACHA scale and its use in recognising anaemia in sheep in northern European sheep breeds, additional blood parameters were selected: red blood cell count, haemoglobin concentrations, cell volumes and platelet counts. The function of red blood cells is to transport haemoglobin, the iron containing protein to which oxygen binds through respiration (Johnson-Wimbley and Graham, 2011). Low red blood cell counts and haemoglobin concentration have consequences on the transport of oxygen from the lungs to other body parts for aerobic respiration and are commonly used when diagnosing anaemia in mammals (Johnson-Wimbley and Graham 2011). Another way to assess the volume of red blood cells in a sample is the packed cell volume which provides a volume percentage of blood made up by red blood cells. These are important as although the count may be normal the cells may be microcytic (too small to function effectively) (Massey, 1992). Platelet counts can also be informative in the diagnosis of anaemia. Low numbers of platelets could lead to uncontrolled bleeding as the blood is unable to clot to stop blood flow outside the body, however too many platelets is also a problem as it can put the animal at risk of developing blood clots internally, and it can also indicate anaemia due to the composition of the blood being unbalanced with low numbers of red blood cells (Kulnigg-Dabsch et al., 2012).

Infections due to parasites may be discovered by analysing faecal material. As discussed in Chapter 1, the number of eggs found in a known volume of faeces can potentially provide an accurate assessment of the parasite burden faced by an individual, or group of animals (Oyewus et al., 2015). Such egg counts were chosen to further investigate the validation of the animal-based indicator, faecal soiling, previously selected by the expert panel.

Physiological sample collection

Blood samples were collected from the jugular vein while the animals restrained. The blood samples were collected in 4ml Vacutainer Lilac K3EDTA tubes using Vacutainer Precision Glide 20G 1" needles and Vacutainer needle holders (all BD (Becton, Dickinson and Company), Oxford, England).

Faecal samples were collected manually from the rectum of the ewes using a gloved finger. The levels of faeces provided by each ewe varied. The estimate average volume of faecal matter collected from each animal was approximately 20g which was transferred in to a 30ml sample container pot (Steralin, Newport, Wales).

Sample analysis

The blood samples were placed in a cool box and also taken to the SAC Veterinary Investigation centre (Bush Estate, Midlothian), and the Royal Dick School of Veterinary Studies Easter Bush Pathology lab (on one occasion). These labs performed biochemistry and haematology testing on the individual samples. Data were collected on various blood parameters: NEFA, BHB, RBCC, haemoglobin, packed cell volume, mean corpuscular volume, mean corpuscular haemoglobin concentration and platelets. These results were later provided to me in an excel spreadsheet for data analysis.

The faecal samples which were for analysis of egg counts were taken to the VI centre and analysed by SAC consulting laboratory staff. The numbers and species of GI parasites found in the faeces were counted and the results were returned provided in an Excel spreadsheet.

Chapter 4 Evaluation of indicators applied during the Assessments at Gather

Abstract

When developing an on-farm welfare assessment protocol decisions are faced such as how frequently the assessments should be performed. In order for an assessment to be successful and meaningful the results obtained should offer a representation of a long term situation, rather than being over sensitive to minor changes based on normal fluctuations. Welfare assessments repeated on the same farm at different times may not always yield identical results, perhaps due to normal fluctuations or it may be indicative of a genuine welfare problem. The purpose of the work reported in this chapter was to assess the long term repeatability of the indicators selected for inclusion in the Assessment at Gather protocol and to assess variation in welfare indicator scores when applied at different time of the calendar year, which coincided with changes in reproductive status. The relationship between farm management decisions during stock draw and the selected welfare indicators was also of interest. In total, data were collected on 135 ewes during ten Assessments at Gather which took place over two years on a Scottish hill farm. These five assessments took place at key stages of the animals' reproductive cycles: pre-mating, mid-pregnancy, late-pregnancy, lactation and weaning. The welfare indicators applied during the Assessments at Gather related to Good Feeding, Good Environment and Good Health. Time was shown to have a great impact upon each of the indicators applied during the assessments as variation occurred both within and between years. The fluctuations in scores may be attributed to normal shifts in the animal's physiology due to alterations in reproductive state or calendar season, or due to external environmental influences such as weather and parasite abundance. Ewes which were selected for removal from the breeding flock by the shepherd at the annual stock draws were found to fare significantly worse in terms of BCS, tooth loss and faecal soiling. The distinction between the tooth loss scores of removed and retained ewes

was seen almost a year in advance of the stock draw. Ewe age was found to have significant impact upon BCS and tooth loss suggesting that older ewes are more vulnerable to nutritional deficiencies than younger animals. Regardless of the cause of these fluctuations they represent meaningful changes to the animal's welfare. These findings have real implications for the interpretation of results from assessments carried out during professional audits of commercial farms especially when comparing between farms, or within farms at different time points.

4.1 Introduction

When developing an on-farm welfare assessment some difficult decisions are faced. One challenge is the decision regarding the frequency of assessments. Although there is no standard time interval for repeating on-farm welfare assessments for any livestock species, including sheep, some authors have proposed time scales e.g. the Welfare Quality should be applied at intervals of approximately six months (Knierim and Winckler, 2009). When assessments occur weeks or months apart, the long term repeatability of the indicators is an important consideration (Temple et al., 2013). For a welfare assessment to be successful and meaningful, the results obtained should offer a representation of a long term situation, rather than being over sensitive to minor changes based on temporary fluctuations (Plesch et al., 2010). Thus, the long term consistency of the welfare indicators in this project (as listed in Table 3.7), requires investigation before the indicators can be recommended for inclusion in a final protocol.

Another challenge is deciding when the assessments should be performed. Welfare assessments repeated on the same farm at different times of year may not always yield identical results. If discrepancies are found between the first visit and subsequent assessments, it may be difficult to identify whether the differences are due to a genuine welfare problem, poor intra- or inter-observer reliability (poor repeatability), or if it is merely linked to normal fluctuations in the animal's state (Phythian et al., 2015; Temple et al., 2013). When making comparisons between the scores received by the same farm over time, or comparing between farms, there are potential factors to consider; e.g. the current calendar season, the animals' reproductive history and current place in reproductive cycle and the age of the animals (Rushen et al., 2011b). As animal-based welfare assessments focus directly on the animals' experience of a situation, it is critical to remember the real impact these factors may have upon a welfare assessment.

The season in which a welfare assessment is performed may have significant effect on the result. Environmental factors such as seasonality and weather patterns may impact upon sheep welfare assessment outcomes as temperatures fluctuate widely in sheep farming regions and extremes of temperature are known to have a great impact upon sheep physiology (Marai et al., 2007; Nardon et al., 1991; Silanikove, 2000). Seasonal environmental patterns also typically coincide with alterations in sheep reproductive state i.e. summer typically coincides with lactation. During pregnancy and lactation there are specific alterations in ewe physiology (Lingis et al., 2012; Nasar and Rahman, 2006) and behaviour (Poindron et al., 1997, 1994; Viérin and Bouissou, 2001). Thus, a longitudinal approach spanning the entire reproductive cycle and calendar year is necessary when investigating the application of measures assessing ewe physiology and welfare.

The age of the animals assessed can also drastically influence the outcome of a welfare assessment. In dairy cattle, older animals have been found to have lower BCS and higher instances of lameness (de Vries et al., 2011). Although different welfare concerns have been identified for lambs and ewes (Phythian et al 2011) specific welfare issues experienced by adult sheep of different ages have not been reported in the literature. It is reasonable to expect that as with cattle, older ewes may not experience the same problems as younger individuals. Therefore the demographic of animals on a farm is an important consideration when interpreting the outcome of a welfare assessment. Extreme variation between individuals of the same flock is particularly likely to occur around an annual stock draw. Prior to a stock draw there may be a high proportion of older ewes which have reached the end of their productive life, whereas less than a month later these older animals may have been replaced by younger sheep. Welfare assessments on the same farm prior to, and following the stock draw would potentially yield substantially different results. Thus, the replacement of older animals with new individuals could have a large impact on the stability of the welfare assessment results.

The removal of animals at a stock draw may be due to their age or a decision made by the shepherd regarding their ability to survive another breeding season. The farmer makes these decisions based upon fundamental aspects of the animal's health, for example tooth loss and body condition (Ridler and West, 2010). Specific links between the decisions made by a shepherd during a stock draw and the welfare of animals either removed or retained in the breeding flock has not been previously reported. However, studies have been performed with dairy cattle. De Vris et al (2011) suggested that data on individual animals collected by dairy farmers during annual stock draw assessments could be used to estimate the welfare of animals on the farms. They performed a review and looked for associations between routine measurements (voluntary cull rates and the number of times a cow was mated before a successful conception) and welfare indicators include in the Welfare Quality protocol. BCS were found to be associated with management decisions i.e. to remove animals from the herd, and also related to reproductive success. Cows with lower BCS were more likely to be removed from the flock at stock draw, and overall had a shorter time in the herd than those with higher condition scores (de Vries et al., 2011; Hoedemaker et al., 2009; Machado et al., 2010). When examining the relationships between BCS and reproductive success, BCS was significantly positively related with pregnancy rates per service (Buckley et al., 2003) and low BCS at the end of lactation, and 10 weeks postpartumm was associated with low pregnancy rates (Hoedemaker et al., 2009; Machado et al., 2010). At a farm level the percentage of cows which were "very lean" was positively associated with calf mortality rate (Sandgren et al., 2009). Relationships between welfare and productivity at a farm level has also been found in sheep. Stott et al (2012) found that productivity, as measured by lambs weaned per ewe, was significantly positively correlated with the overall farm welfare score. They suggested that production indicators may be useful when predicting welfare and highlighting areas of concern. Thus, there is potential for production traits such as lamb weights and survival to be used as Iceberg Indictors. However the relationship between these traits and animal based welfare indicators applied to individual ewes is currently unknown and needs to be researched.

Aims of the work reported in this chapter:

The purpose of the work reported in this chapter was to assess the long term repeatability of the indicators selected for inclusion in the Assessments at Gather (see table 3.7) and to assess fluctuation in welfare indicator scores when applied at different time of the calendar year, which coincided with changes in reproductive status.

I was also interested in the effect of age on the repeatability of the indicators and whether welfare problems may be specific to older or younger ewes. The relationship between farm management decisions during stock draw and the selected welfare indicators was also of interest.

I also wanted to assess the potential of lamb birth and weaning weights for use an iceberg indicator and its relationship with other measures of ewe welfare.

The convergent validity of some indicators used with Scottish hill sheep also required further investigation, i.e. the FAMACHA anaemia scoring system and the novel simplified body condition scoring.

It is also important to assess the value of the information provided by the welfare indicators. The indicators selected should be complementary to each other. It is possible that some of the indicators selected for use during the Assessments at Gather may provide very similar information on the animal's welfare and therefore may be redundant.

4.2 Materials and Methods

The work presented in this chapter consists of the data collected as described in Chapter 3. As mentioned in Chapter 3, these Assessments at Gather were scheduled to include the five main periods of the commercial sheep reproductive cycle. Although not all ewes conceived and maintained their pregnancy through to lambing, and subsequently reared lambs until weaning, the times when the Assessments at Gather occurred are referred to as pre-mating, mid-pregnancy, late-pregnancy, mid-lactation and weaning.

As described in Chapter 3 the indicators included in the Assessment at Gather covered the principles of Good Feeding, Good Environment and Good Health. The indicators applied to the sheep during these assessments were: BCS, anaemia as assessed by the FAMACHA scale, tooth loss, coat cleanliness and faecal soiling, lameness, coughing and nasal discharge. During lactation udder assessment was also carried out for: discolouration, symmetry, fibrosis and evidence of orf and other lesions. Blood and faecal samples were collected on all gathers with the exception of 1 and 3 due to logistical constraints. On assessment 7 blood samples were collected and submitted to the lab however due to technical constraints only NEFA and BHB results were obtained, and no values for platelets were obtained for Assessment 9.

The short term repeatability of welfare indicators is acknowledged as important and therefore during the development of the AWIN sheep protocol a team in Spain carried out this work. However as this work was not performed by this student during this PhD project, it is not reported upon in this thesis.

4.2.1 Data preparation and analysis

The data obtained for some indicators did not offer sufficient variation or prevalence to be included in statistical analyses as only a very small proportion of the animals received a score above 0. These were: coughing, nasal discharge, lameness, udder discolouration, symmetry, fibrosis or presence of lesions on the udder.

The statistical analysis presented in this chapter was performed using Genstat 16 (Genstat for Windows International, UK) and Minitab 16 (Minitab Ltd. UK).

4.2.2 Repeatability of indicators over time and between ages

To investigate the consistency of ewe BCS using the Russel et al (1969) scale and physiological parameters, Linear Mixed Models were performed using Genstat. For these analyses BCS was included individually as the response variate while assessment number (1-10) and ewe birth year were included as fixed effects. Ewe identity was included in the models as a random effect to account for the repeated measures made on each ewe. A post-hoc comparison was performed on the data to further assess the variation between specific assessments. This was achieved using a pairwise differences test using Genstat.

To assess the consistency over time of the categorical indicators, and the effect of ewe age, Ordinal Logistic Regressions were performed using Minitab. In these models the categorical indicator results were included as the response, while time code and ewe birth year were included as categorical predictors. Ewe identity was again included as random effect in the model. Due to the low number of ewes born in 2005 (3 individuals) the birth years of 2005 and 2006 were combined for analysis.

4.2.3 Convergent validation assessment of selected indicators

The convergent validity of indicators was investigated using Linear Mixed Models using Genstat. Anaemia score, faecal soiling and the simplified BCS were selected for further convergent validation by assessing their relationship with the relevant blood and faecal parameters.

These analyses were initially performed on the data inclusive of all Assessments at Gather. Following the analyses of all assessments subsequent analyses were conducted in which the data were filtered to individual assessments to further assess the relationship between the indicators and physiological parameters.

In the Linear Mixed Model analyses the categorical indicator (anaemia, faecal soiling simplified BCS) was entered and ewe identity was included as a random effect. When analysing data from all ten Assessments at Gather the assessment number was also included as a random factor.

To analyse the convergent validity of the anaemia indicator the blood anaemia parameters were included as the response variate. The blood parameters included were haemoglobin, mean corpuscular volume (MCV), mean corpuscular haemoglobin concentration (MCHC), packed cell volume (PCV), red blood cell count (RBCC) and platelets. These measures were initially included individually and those which met a pre-requisite of $p<0.20$ were then included together to in a multivariate analysis.

Linear Mixed models were also used to analyse the relationship between faecal soiling and parasite egg counts present in faeces. The faecal soiling score was entered as a fixed effect and ewe identity as a random effect. As above, when analysing the data from all Assessments at Gather, assessment number was included as a random factor. The egg counts for species identified by the Veterinary Investigation lab (*Coccidia* and *Strongyle*) were included as the response variate.

4.2.4 Indicator relationship with management decisions

Generalised Linear Models (Genstat) were performed to assess the variation in indicator scores between animals which were either removed from the flock at stock draw or were retained for future production. Individual ewes received a binary score reflecting the shepherd's decision at stock draw at the end of each year: removed (0), or retained (1). This response was fitted in the model as the response variate with indicators entered individually as the fixed effects. The model assumed a binomial distribution based on the binary score and Logit link function was selected. Indicators which provided a statistically significant result when assessed individually were then included together in a final model for each assessment.

This analysis was initially performed on the fifth and tenth Assessments at Gather as they occurred in the closest temporal proximity to the stock draws. Subsequent analyses were then performed by working through the Assessments at Gather in reverse chronological order to assess the predicative ability of these indicators. The data collected during Year 1 was used in conjunction with the Year 1 stock draw, and the indicator results for Year 2 were used with the Year 2 stock draw.

4.2.5 Lamb weight

To investigate the potential of lamb birth and weaning weight to act as iceberg indicators for ewe welfare a Linear Mixed Model was performed. The total live lamb birth and weaning weights per ewe were calculated for each year. These were entered as response variates, and ewe identity as a random factor. Individual indicator results were included as fixed effects. For the indicators which provided categorical data (numerical scales, e.g. faecal soiling) this model was able to calculate predicted means and standard errors of differences, however for BCS (as assessed by the Russel scale), correlations were subsequently performed to further investigate the direction of the relationships.

4.2.6 Lamb survival between two years

To investigate variation in lamb survival between the two years a Chi square test was performed. The number of lambs born alive in Years 1 and 2 were entered along with the totals of those surviving to weaning or dying prior to this time.

4.2.7 Indicator redundancy

Cluster analysis was also performed in order to assess the similarity between the results obtained from the application of the indicators. Although no statistical significance is assigned, by assessing the association between the variables it gives information on which of the indicators may be providing similar information and can therefore be grouped together. This allocation of indicators into structured groups provides an understanding of the way in which the indicators relate to one another and individually contribute to the overall understanding of the animals' welfare. If multiple indicators are seen as offering identical information relating to the animal's welfare further this may imply some indicators are redundant, and the simplest to measure of the indicator cluster would then be the preferred measure for subsequent welfare assessments. The Average Linkage Method was selected to reduce the impact of any outliers and similarity level set to 75%.

4.3 Results

4.3.1 Good Feeding

4.3.1.1 Changes over time

Significant variation was found both within and between years for the indicators relating to Good Feeding: BCS and tooth loss. BCS altered significantly both within and between years (Wald=699.84, df=9, F=77.76, $p<0.001$). In Year 2, BCS was significantly higher pre-mating than in the previous year at this time, but for the other four assessments BCS wassignificantly lower in Year 2 in comparison to the same points in the previous years, shown in Figure 4.1 (all $p<0.001$). Significant differences were also discovered within years although these are not indicated in the figure. In Year 1 BCS was significantly lower in mid-lactation and weaning than the rest of the reproductive cycle ($p<0.002$). At pre-mating in Year 2 the ewes had a significantly higher BCS than the other assessment periods ($p<0.001$). Their BCS steadily decreased as by late-pregnancy their BCS were lower than mid-pregnancy ($p<0.001$), and mid-lactation was significantly lower than all other times ($p<0.001$). Between mid-lactation and weaning their condition score improved significantly ($p<0.001$).

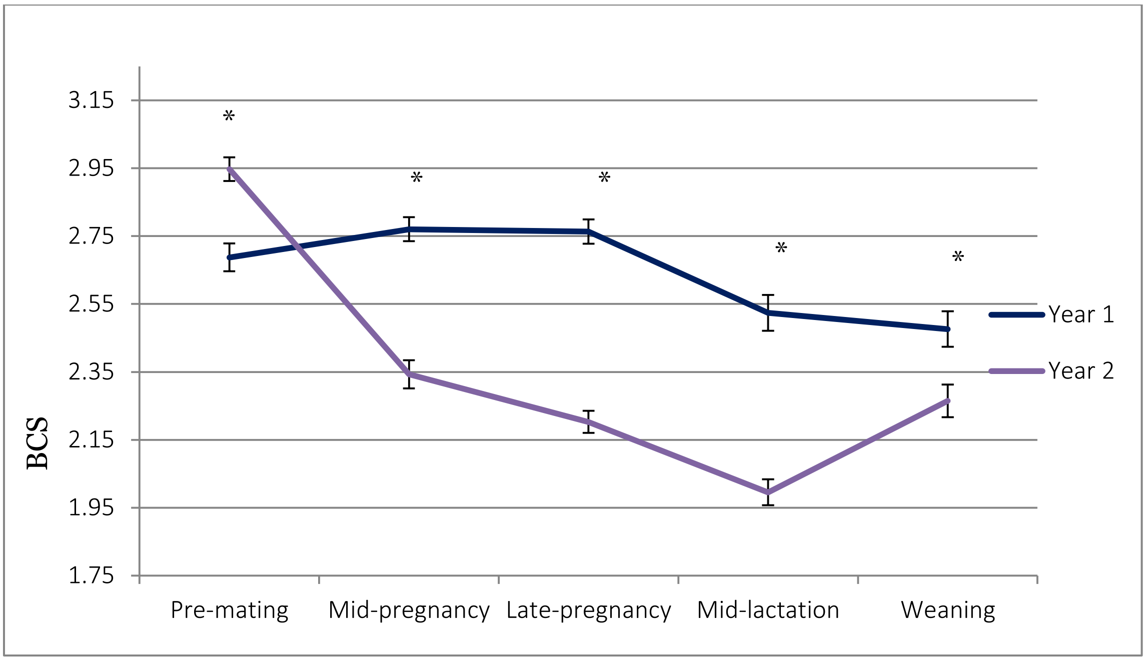


Figure 4.1 Variation in mean Russel et al BCS (±SEM) of ewes over Year 1 and Year 2. Asterisks indicate significant differences between years reproductive stage P<0.001

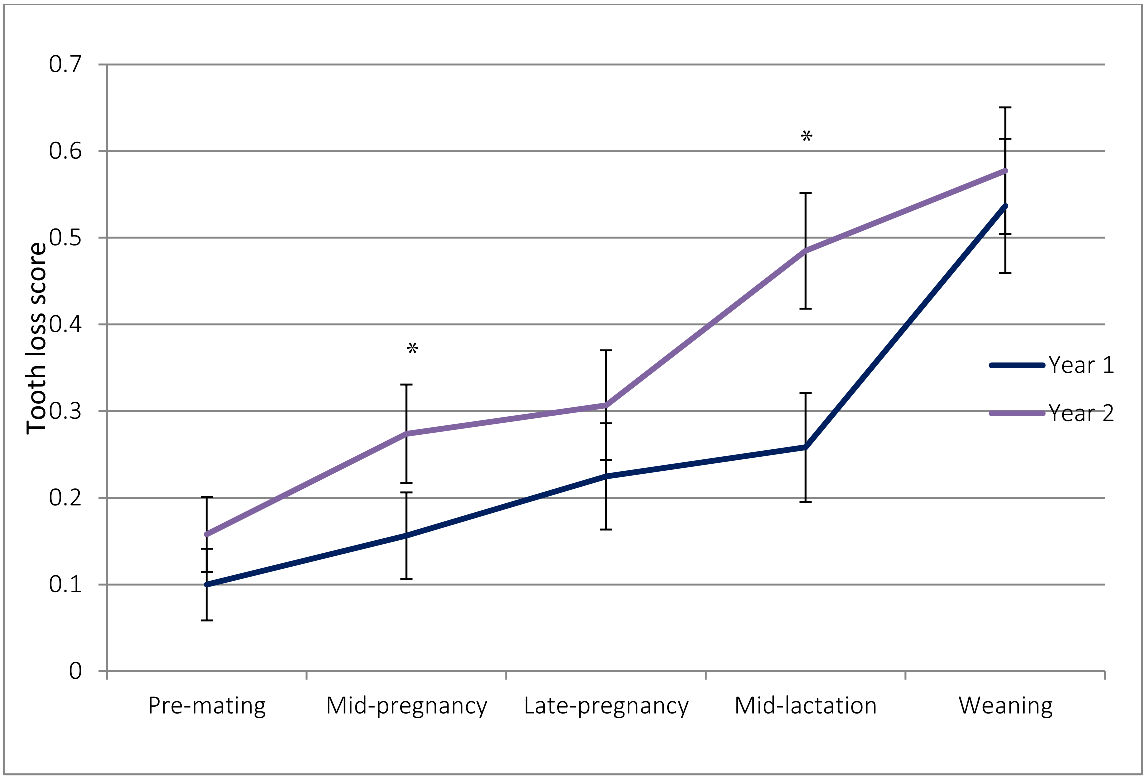


Figure 4.2 Variation in mean tooth loss scores (±SEM) for ewes over Years 1 and 2. Asterisks indicate significant differences between years P<0.001

Tooth loss also significantly differed between years at the same points in reproductive cycle, and between assessments conducted within the same year ($p<0.001$, $X^2=100.81$, df=9, Fig 4.2). The tooth loss scores were consistently higher (indicating greater tooth loss) in Year 2 compared to Year 1 and were significantly different at mid-pregnancy and mid-lactation ($p<0.02$). Both years followed a similar pattern with the lowest tooth loss scores occurring pre-mating, and the highest at weaning. In Year 1, the pre-mating tooth loss scores were significantly lower than those in late pregnancy, mid lactation and weaning (all $p<0.04$). The tooth loss values observed at weaning were significantly higher than all other times of that year ($p<0.001$). In Year 2, the pre-mating tooth loss scores were significantly lower than mid-pregnancy, mid-lactation and weaning ($p<0.048$). The scores given to the ewes during mid-lactation were significantly higher than those in mid- and late-pregnancy (all $p<0.003$).

4.3.1.2 Removal from flock

Ewes which were removed from the breeding flock at the end of Year 1, had significantly lower BCS than those retained in the flock as shown in Figure 4.3. This separation was apparent from late-pregnancy (Wald=0.35, df=1, $p<0.001$), through to mid-lactation (Wald=3.73, df=1, $p=0.023$), and weaning (Wald=0.612, df=1, $p=0.013$). Prior to late-pregnancy there were no significant differences between these groups of ewes (pre-mating and mid-pregnancy ($p>0.17$).

In Year 2 the ewes which were removed from the flock at stock draw did not significantly differ from those which were retained at any time period ($p>0.40$) other than pre-mating approximately 10 months prior to the stock draw (Wald=4.26, df=1, $p=0.039$, Figure 4.4).

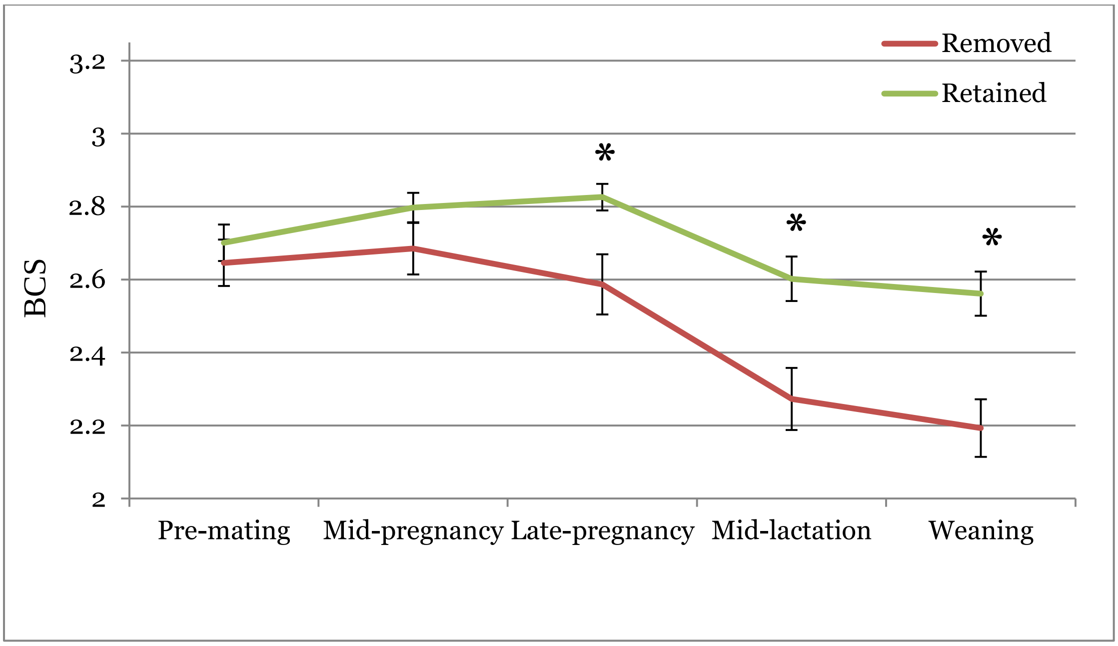


Figure 4.3 Variation in ewe BCS (±SEM) throughout Year 1 for ewes which were removed or retained in the breeding flock at the 2012 stock draw. Asterisks indicate significant differences between years (P<0.001)

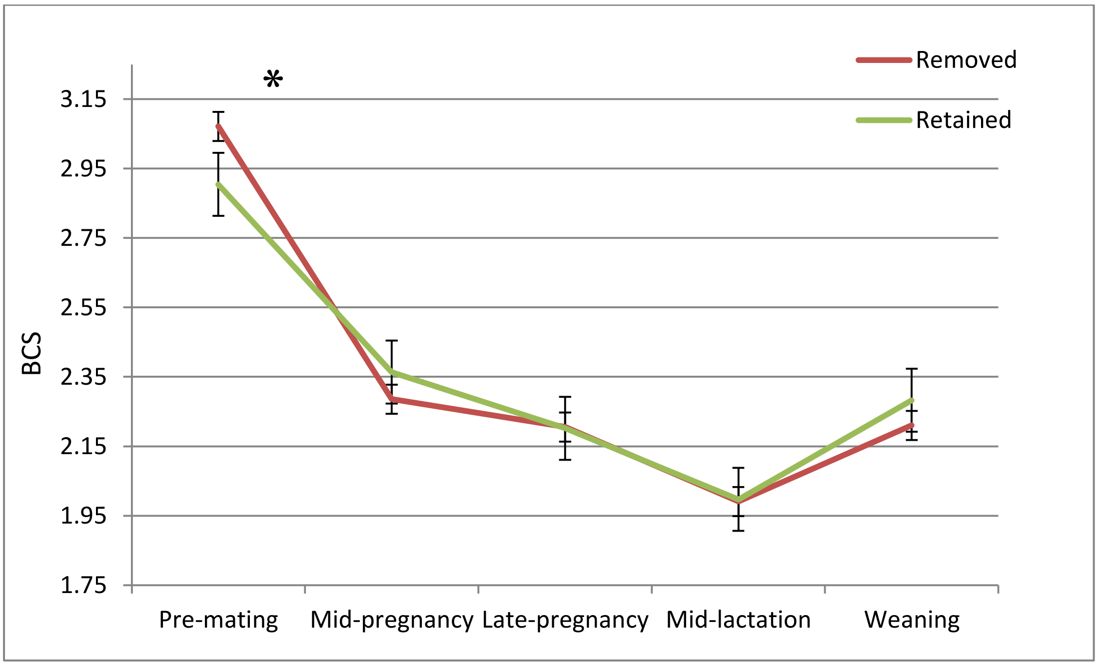


Figure 4.4 Variation in ewe BCS (±SEM) throughout Year 2 for ewes which were removed or retained in the breeding flock at the 2013 stock draw. Asterisks indicate significant differences between years (P<0.001)

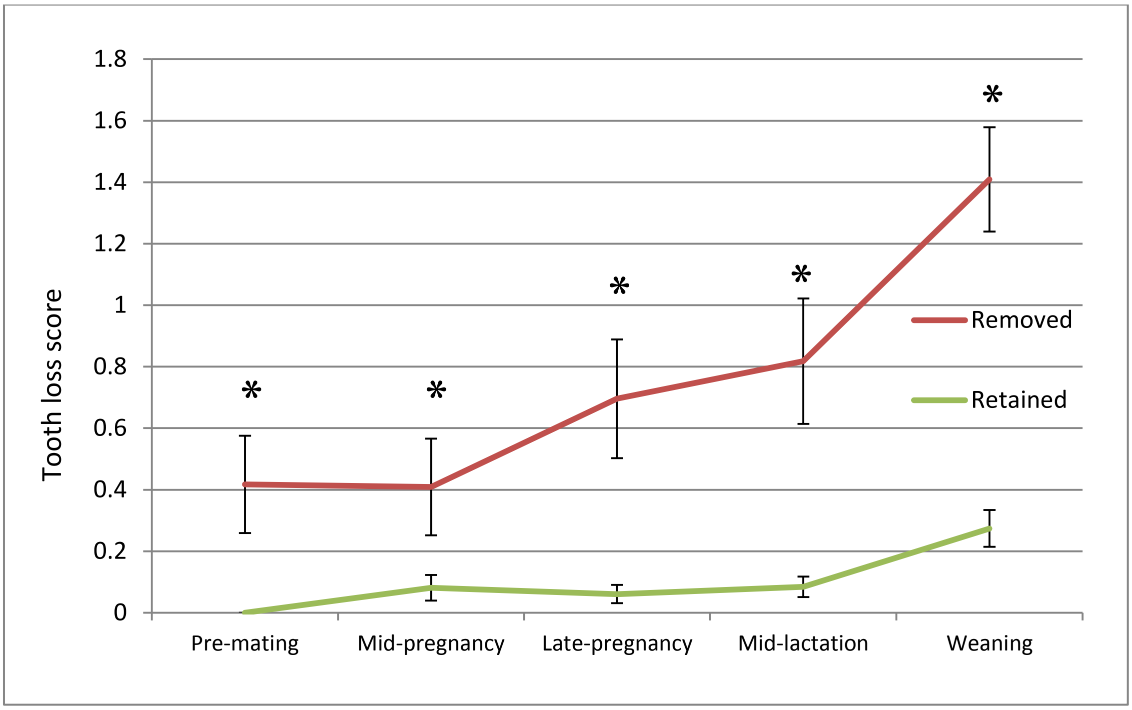


Figure 4.5 Variation in tooth loss scores (±SEM) throughout Year 1 for ewes which were removed or retained in the breeding flock at the 2012 stock draw. Asterisks indicate significant differences between years (P<0.001)

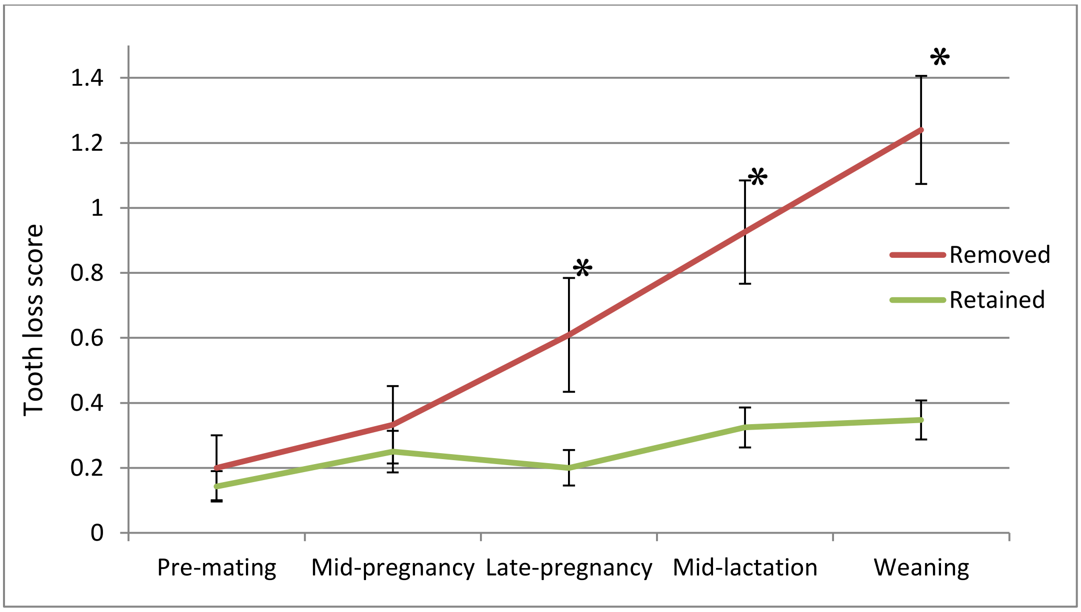


Figure 4.6 Variation in tooth loss scores (±SEM) throughout Year 2 for ewes which were removed or retained in the breeding flock at the 2013 stock draw. Asterisks indicate significant differences between years (P<0.001)

In Year 1 the ewes which were removed from the flock had consistently higher tooth loss scores than those which were retained (Fig 4.5). This occurred in the pre-mating assessment approximately 10 months prior to the stock draw (Wald=52.66, df=2, p<0.001). This discrepancy between the ewes which would later be removed or retained was maintained through mid-pregnancy (Wald=0.31, df=2, p<0.001), late-pregnancy (Wald=7.81, df=2, p=0.028), mid-lactation (Wald=0.847, df=2, <0.001) and weaning (Wald=16.60, df=2, p<0.001).

In Year 2 the ewes which were removed from the breeding flock at the end of the reproductive cycle had significantly higher tooth loss scores in late-pregnancy (Wald=6.74, df=2, p=0.034), mid-lactation (Wald=13.97, df=2, p<0.001) through to weaning (Wald=13.97, df=2, p<0.001) (see Figure 4.8).

4.3.1.3 Lamb birth and weaning weights

In Year 1, lamb birth and weaning weights were not related to ewe BCS during pregnancy, although a significant positive relationship was found between ewe BCS at Assessment 5 (weaning) and weight of lambs born in Year 1 (Wald-4.76, df=1, p=0.32. Table 4.1). A positive relationship was also found between ewe BCS and lamb weaning weights in Year 1 (Table 4.2). Significant correlations were found between ewe BCS and weaning weights at Assessment 4 (mid-lactation, p=0.01, Wald 6.84, df=1, F=6.84), and a trend towards significance was found at Assessment 2 (mid-pregnancy p=0.093, Wald 2.92, df=1, F=2.92).

Table 4.1 Correlation between lamb birth weight in 2012 and ewe BCS throughout Year 1

Assessment at Gather	R value	P-value
1 Pre-mating	0.08	0.51
2 Mid-pregnancy	0.14	0.22
3 late-pregnancy	0.01	0.99
4 mid-lactation	0.19	0.11
5 weaning	**0.25**	**0.03**

Table 4.2 Correlation between lamb weaning weight in 2012 and ewe BCS throughout Year 1

Assessment at Gather	R value	P-value
1 Pre-mating	0.09	0.47
2 Mid-pregnancy	**0.22**	**0.09**
3 late-pregnancy	0.01	0.94
4 mid-lactation	**0.32**	**0.01**
5 weaning	0.14	0.26

In Year 2, the birth weight of lambs born was significantly positively related to ewe BCS at Assessment 6 (pre-mating p=0.024, Wald 5.25, df=1, F=5.25), Assessment 7 (mid-pregnancy p=0.034, Wald=4.65, df=1, F=4.65), and Assessment 8 (late-pregnancy p=0.001, Wald=11.48, df=1, F=11.48). Birth weight also tended to be positively correlated with Assessments 9 and 10 (mid-lactation and weaning: p=0.053, Wald=3.83, df=1; p=0.094, Wald=2.87, df=1, respectively). Ewes with higher BCS therefore produced more lamb weight (Table 4.3). No significant relationships were found between ewe BCS at any time and lamb weaning weight in 2013 (p>0.14).

Table 4.3 Correlation between lamb birth weight in 2013 and ewe BCS throughout Year 2

Assessment at Gather	R value	P-value
1 Pre-mating	**0.24**	**0.02**
2 Mid-pregnancy	**0.23**	**0.03**
3 late-pregnancy	**0.34**	**0.01**
4 mid-lactation	**0.21**	**0.05**
5 weaning	0.18	0.09

No relationships were found between ewe tooth loss and lamb birth or weaning weights in Year 1 ($p>0.38$). Similarly, no relationships were found between ewe tooth loss and lamb birth weight in 2013. However, tooth loss at Assessment 6 (pre-mating, $p=0.002$, Wald=14.12, df=2), Assessment 7 (mid-pregnancy, $p=0.049$, Wald=6.32, df=2) and Assessment 8 (late pregnancy $p=0.009$, Wald=7.22, df=1) were significantly related to 2013 weaning weights, as greater lamb weaning weight was associated with higher tooth loss scores as shown in Table 4.4.

Table 4-4 Relationship between lamb weaning weights in 2013 and mean ewe tooth loss scores in 2013 (SEM in parentheses). Significant difference indicated by different letters.

	Tooth loss score			
Assessment at Gather	0	1	2	P-value
6 (Pre-mating)	**32.74 (1.33) [a]**	**43.84 (4.21) b**	**51.85 (7.23) b**	**0.002**
7 (Mid-pregnancy)	**32.46 (1.25) a**	**36.66 (3.58) b**	**44 (5.94) c**	**0.049**
8 (late-pregnancy)	**32.67 (1.36) a**	**38.45 (3.27) b**	**44.43 (7.17) b**	**0.009**
9 (mid-lactation)	33.66 (1.49)	34.82 (2.69)	38.46 (5.11)	0.421
10 (weaning)	33.62 (1.59)	34.95 (2.59)	37.41 (4.53)	0.365

4.3.1.4 Validity of simplified BCS

When analysed across all Assessments at Gather, significant relationships were found between the Russel et al (1969) BCS and the simplified scale (Wald 771.07, df=3, F=257, $p<0.001$, Table 4.5). The relationship between the simplified BCS and NEFA concentration tended towards significance (Wald= 7.40, df=3, F=2.47,$p=0.061$, Table 4.5) but no relationship was found between the simplified condition score and BHB (Table 4.5).

Table 4-5 Mean BCS (Russel et al scale), NEFA and BHB concentrations and relation to the simplified BCS when data from all Assessments at Gather were analysed together. (SEM in parentheses). Significant difference indicated by different letters.

	Simplified BCS score				
Indicator	Emaciated	Thin	Fit	Fat	P-value
Russel (1969) BCS	1.713 (0.08) a	1.958 (0.071) b	2.501 (0.065) c	3.217 (0.073) d	<0.001
NEFA	550.3 (126.5)	495 (100.3)	576.5 (91.1)	697.2 (105.7)	0.061
BHB	0.451 (0.085)	0.40 (0.069)	0.442 (0.065)	0.433 (0.072)	0.54

The significant relationship between the Russel et al 1969 scale and the simplified BCS was maintained at all individual assessments (Table 4.6).

Table 4-6 Mean Russel et al BCS and relationship with simplified scale score at individual Assessments at Gather. (SEM in parentheses).

Assessment at Gather	Simplified BCS score				
	Emaciated	Thin	Fit	Fat	Model output
1 Pre-mating		1.75 (0.32)	2.68 (0.229)	3.45 (0.321)	Wald=89.91, df=2, F=44, p<0.0001
2 Mid-pregnancy			**2.73 (0.29)**	3.5 (0.319)	Wald=38.91, df=1, p<0.001
3 Late-pregnancy			2.70 (0.032)	3.27 (0.097)	Wald=31.06, df=1, p<0.001
4 Mid-lactation	1.5 (0.34)	1.75 (0.20)	2.44 (0.091)	3.5 (0.10)	Wald=106.78, df=3, p<0.001
5 Weaning	1.5 (0.33)	1.75 (0.11)	2.46 (0.03)	3.44 (0.112)	Wald=124.41, df=3, F=41.7, p<0.001
6 Pre-mating			2.79 (0.02)	3.40 (0.046)	Wald=131.18, df=1, p<0.001
7 Mid-pregnancy	1.5 (0.21)	1.75 (0.09)	2.32 (0.032)	3.37 (0.15)	Wald=99.77, df=3, p<0.001
8 Late-pregnancy	1.5 (0.146)	1.75 (0.146)	2.25 (0.029)		Wald=34.98, df=2, F=17.49, p<0.001
9 Mid-lactation	1.5 (0.065)	1.75 (0.39)	2.19 (0.029)	3.75 (0.16)	Wald=219, df=3, p<0.001
10 Pre-mating	1.5 (0.16)	1.75 (0.59)	2.385 (0.033)	3.45 (0.125)	Wald=202.25, df=3, F=67.42 p<0.001

Table 4-7 Mean NEFA concentration and relationship with simplified scale score for individual Assessments at Gather. (SEM in parentheses). Significant difference indicated by different letters.

	Simplified body condition score				
Assessment at Gather	Emaciated	Thin	Fit	Fat	P-value
2 Mid-pregnancy			825.90 (488.3)	571.70 (557.7)	0.38
4 Mid-lactation	196.00 (355.0)	218.50 (251.1)	418.60 (45.5)	522.80 (125.5)	0.636
5 Weaning	51 (191.06)	252 (63.69)	303.80 (22.52)	393.00 (63.69)	0.229
6 Pre-mating			796.80 (48.31)	970.90 (84.27)	0.077
7 Mid-pregnancy	1012.00 (463.9)	742.80 (146.7)	879.5 (53.2)	1207.30 (267.8)	0.489
8 Late-pregnancy		544.00 (515.6)	914.0 (55.6)		0.478
9 Mid-lactation	456.90 (88.3)	380 (53.3)	421.10 (42.2)	245.00 (305.9)	0.811
10 Pre-mating	**157.30 (755.55)**	**183.00 (27.90) a**	**220.20 (16.62) b**	**441.20 (65.43) c**	**0.005**

When looking at individual Assessments at Gather the only significant relationship between NEFA concentration and the simplified BCS is at Assessment 10 (Wald=13.88, df=3, F=4.63, p=0.005). A trend was also observed at Assessment 6 as fat ewes tended to have higher circulating concentrations of NEFA than fit individuals as shown in Table 4.7 (Wald=3.21, df=1, P0.077).

No significant relationships were found at any individual point between the simplified body condition scale and BHB concentration (p>0.30).

4.3.1.5 Effect of ewe age

The age of ewes had a significant effect on BCS (Wald=13.34, df=5, p=0.044) as shown in Figure 4.7. Older ewes (those born in 2005 and 2006) had significantly lower BCS than ewes born between 2007 and 2011.

The ewes which were born in 2005, 2006 and 2007, had significantly higher tooth loss scores than ewes born from 2008 onwards (Figure 4.8, χ^2 10.74, df=5, p=0.05). There was a tendency for ewes born in 2008 to have higher tooth loss scores than those born in 2011 however this did not reach significance (p=0.056). The youngest ewes (born in 2011) had significantly lower tooth loss than all older ewes.

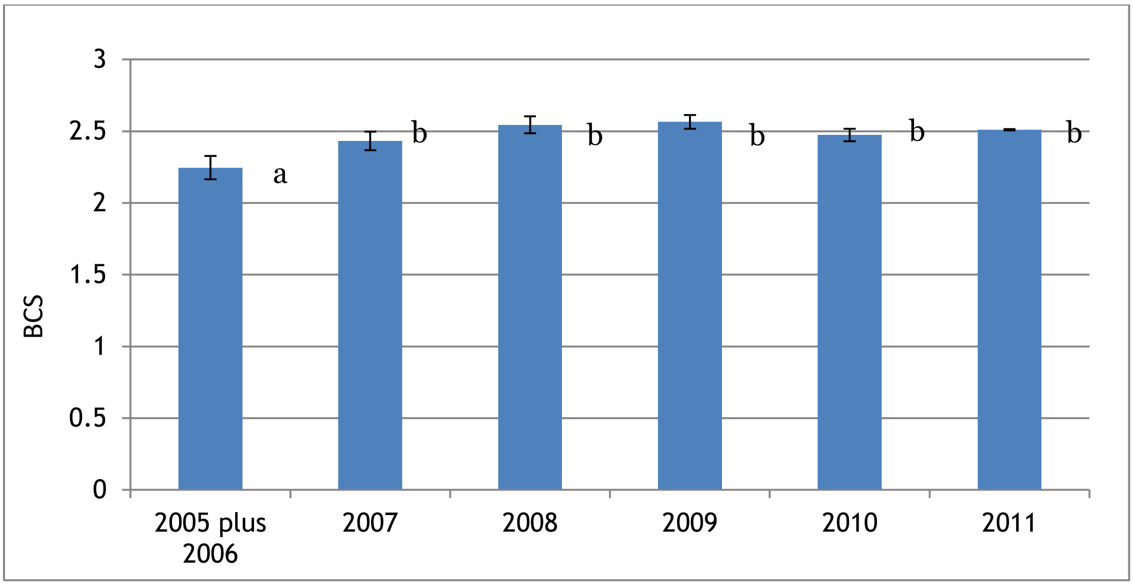


Figure 4.7 Mean Russel et al BCS (±SEM) of ewes born between 2005 and 2011. Different letters indicate statistical differences between years (P<0.04)

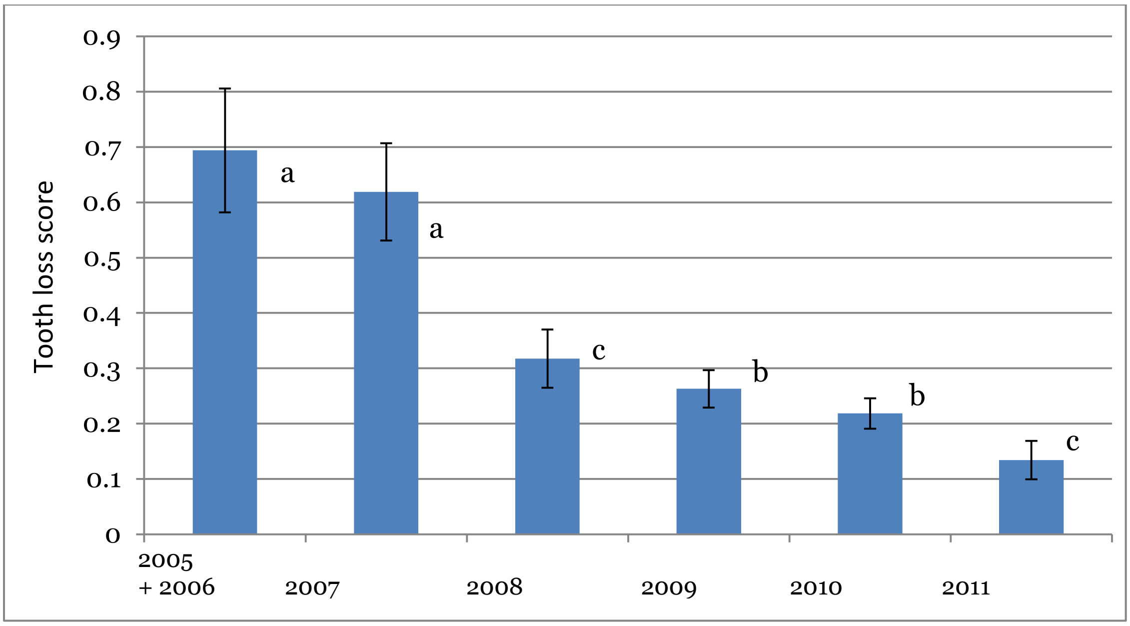


Figure 4.8 Mean tooth loss scores (±SEM) of ewes born between 2005 and 2011. Different letters indicate statistical differences between years (P<0.04)

4.3.2 Good Environment

4.3.2.1 Changes over time

As shown in Figure 4.9 significant changes in coat cleanliness were observed both within years and between years at the same point in the year (χ^2=251.20, df=9, p<0.001).

In both years the fleeces of the ewes were dirtiest in the pre-mating assessments but became significantly cleaner by mid-pregnancy, and improved significantly again by late pregnancy before getting significantly more soiled by mid lactation and again at weaning (all p<0.001). At pre-mating, the ewes in Year 1 had significantly dirtier coats than at the same time in Year 2 (p<0.001). However, by mid lactation the ewes in Year 2 were significantly more dirty than those in Year 1 (p=0.03).

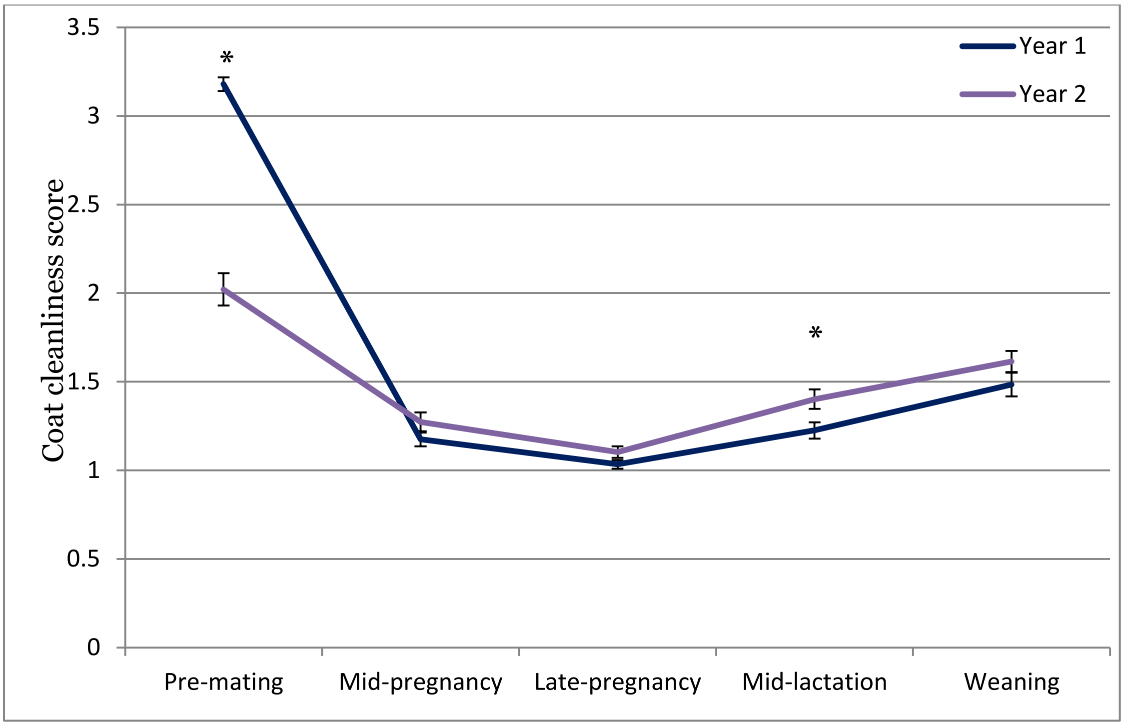


Figure 4.9 Variation in mean coat cleanliness (±SEM) during the Assessments at Gather in Years 1 and 2. Asterisks indicate significant differences between years (P<0.001)

4.3.2.2 Removal from flock

At no point did the ewes which were removed from the flock differ in coat cleanliness from those which were retained for breeding (all $p>0.2$).

4.3.2.3 Variation due to age

The age of ewes did not have a significant effect on their coat cleanliness scores (all $p>0.245$).

4.3.3 Good health

4.3.3.1 Changes over time

Significant variation in anaemia scores was observed both within years and at the same time between Year 1 and Year 2, with ewes in Year 2 consistently scoring higher than in Year 1 ($\chi2=261.94$, df=9, $p<0.001$). This was significant at each time stage ($p<0.001$), except mid-pregnancy ($p=0.32$) as shown in Figure 4.10.

In both years, the ewes received the lowest scores (indicating low anaemia) during the pre-mating Assessment at Gather. The scores received at this time were significantly lower than at any other time of the year (all P=0.001) with the exception of mid-pregnancy (Year 1 p=0.055, Year 2 p=0.90). In both years anaemia scores increased as the reproductive cycle progressed. In Year 1 the FAMACHA scores received by the ewes peaked in mid-lactation and remained constant through to weaning. These scores were significantly higher than any other assessment in Year 1 (p<0.001). In Year 2 the anaemia scores received by the ewes increased more rapidly and peaked in late-pregnancy. The anaemia scores were significantly higher in mid-pregnancy and weaning than any other assessment in Year 2 (p<0.001).

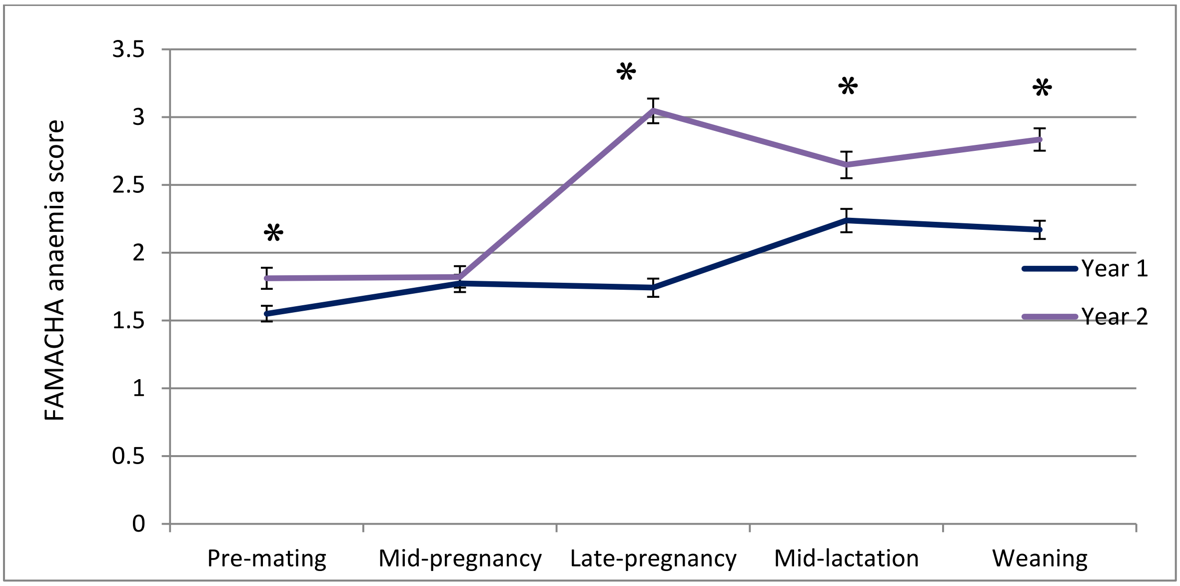


Figure 4.10 Variation in mean anaemia scores (±SEM) recorded during the Assessments at Gather in Year 1 and Year 2. Asterisks indicate significant differences between years (P<0.001)

4.3.3.2 Removal from flock

In Year 1 there were no significant differences in anaemia scores between ewes that were removed at stock draw and those that were retained. However, in Year 2 the ewes removed from the flock had significantly lower FAMACHA scores at weaning than those retained; (Wald=5.94, df=4, p<0.023, Fig 4.11).

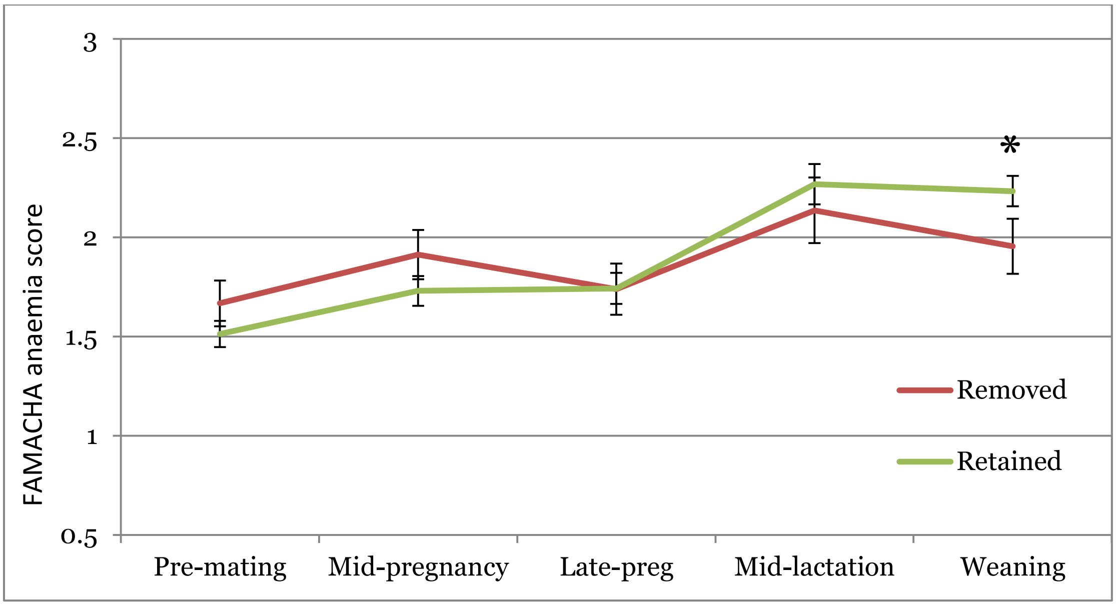


Figure 4.11 Variation in FAMACHA anaemia scores (±SEM) for ewes retained or removed from the breeding flock in Year 2. Asterisks indicate significant differences between years (P<0.001)

4.3.3.3 Lamb weights

No relationships were found between lamb birth or weaning weights and ewe FAMACHA score (all $p>0.35$).

4.3.3.4 Validation

When assessing overall Assessments at Gather the FAMACHA anaemia indicator was not significantly related to any of the Blood Anaemia Parameters.

4.3.3.5 Individual assessments

When looking at data collected at specific assessments significant relationships were found between the FAMACHA anaemia scale and relevant physiology. Anaemia score was found to be significantly related to red blood cell count at Assessment 2 (mid-pregnancy $p=0.04$, Wald=6.92, df=2, $F=3.46$), Assessment 9 (mid-lactation $p=0.039$, Wald=10.54, df=4, $F=2.64$) and a non-significant tendency was found between these measures at Assessment 4 (mid-lactation $p=0.084$, W=8.61, $F=2.15$, df=4) as shown in Table 4.8. At Assessment 2 low red blood cell counts are associated with higher anaemia scores but at Assessment 9 the direction of the relationship is less clear as ewes with the highest red blood cell counts tended to receive a 2 or 5 on the FAMACHA scale.

Table 4.4 Mean red blood cell counts and relationship with FAMACHA anaemia score at individual Assessments at Gather (SEM in parenthesis)

	Anaemia score					P-value
Assessments at Gather	1	2	3	4	5	
2 Mid-pregnancy	**9.23 (0.22)**	**9.50 (0.16)a**	**8.44 (0.38)b**			**0.04**
4 Mid-lactation	*9.02(0.37)*	*9.70 (0.16)*	*9.257 (0.18)*	*10.2 (0.52)*	*8.30 (0.91)*	*0.084*
5 Weaning	9.29 (0.30)	9.30 (0.15)	8.80 (0.21)	9.8 (1.07)		0.288
6 Pre-mating	9.33 (0.23)	9.51 (0.21)	10.14 (0.37)	9.75 (0.94)		0.321
8 Late-pregnancy		8.58 (0.18)	8.48 (0.14)	8.01 (0.19)	8.13 (0.50)	0.161
9 Mid-lactation	**7.74 (0.25)**	**8.05 (0.16)a**	**7.72 (0.15)a**	**7.25 (0.19)b**	**8.14 (0.85)**	**0.039**
10 Weaning	9.60 (0.78)	10.01 (0.20)	9.67 (0.19)	9.37 (0.26)	10.4 (1.11)	0.34

FAMACHA Anaemia scores were significantly related to packed cell volume at Assessments 2 (mid-pregnancy, p=0.013, W=9.5, df=2, F=4.75) and 9 (mid-lactation p=0.026, W=11.64, f=2.91, df=4. Table 4.9). The ewes with higher anaemia scores typically had lower PCV at Assessment 2, however assessment 9 the ewes with the highest PCV also scored the highest for anaemia.

Table 4.5 Mean packed cell volume (PCV) and relationship with FAMACHA anaemia score at individual Assessments at Gather (SEM in parenthesis)

	Anaemia score					
	1	2	3	4	5	P-value
2 Mid-pregnancy	**0.30 (0.01)**	**0.31 (0.01)a**	**0.27 (0.01)b**			**0.013**
4 Mid-lactation	0.2883 (0.01)	0.31 (0.01)	0.30 (0.01)	0.33 (0.02)	0.25 (0.03)	0.102
5 Weaning	0.31 (0.01)	0.31 (0.01)	0.30 (0.01)	0.32 (0.03)		0.319
6 Pre-mating	0.32 (0.01)	0.32 (0.01)	0.33 (0.01)	0.34 (0.02)		0.516
8 Late-pregnancy		0.29 (0.02)	0.30 (0.01)	0.29 (0.01)	0.3233 (0.03)	0.745
9 Mid-lactation	**0.33 (0.01)ac**	**0.33 (0.01)ac**	**0.32 (0.01)a**	**0.305 (0.08)bc**	**0.37 (0.03)**	**0.026**
10 Weaning	0.31 (0.03)	0.323 (0.01)	0.31 (0.01)	0.30 (0.01)	0.32 (0.04)	0.29

No other significant relationships were found between FAMACHA anaemia score and blood parameters $p>0.09$.

4.3.3.6 Age

Age did not significantly affect FAMACHA anaemia score or any of the anaemia related blood parameters (all $p>0.12$).

4.3.3.7 Changes over time

The faecal soiling scores received by the ewes significantly varied between and within years (Figure 4.10, $\chi2=249.23$, df=9, $p<0.001$). Both Years 1 and 2 showed the same pattern of faecal soiling scores: peak faecal soiling occurred in mid-lactation and the scores received at this time were significantly higher than all other time ($p<0.05$). In year 1 there was also a significant increase in faecal soiling at mid pregnancy in comparison to pre-mating and late pregnancy ($p<0.01$), which was not seen in Year 2.

Significant differences between years were found at mid-pregnancy at which the ewes in Year 1 received higher soiling scores, yet by mid-lactation and weaning the ewes in Year 2 had heavier soiling ($p<0.001$. Figure 4.10).

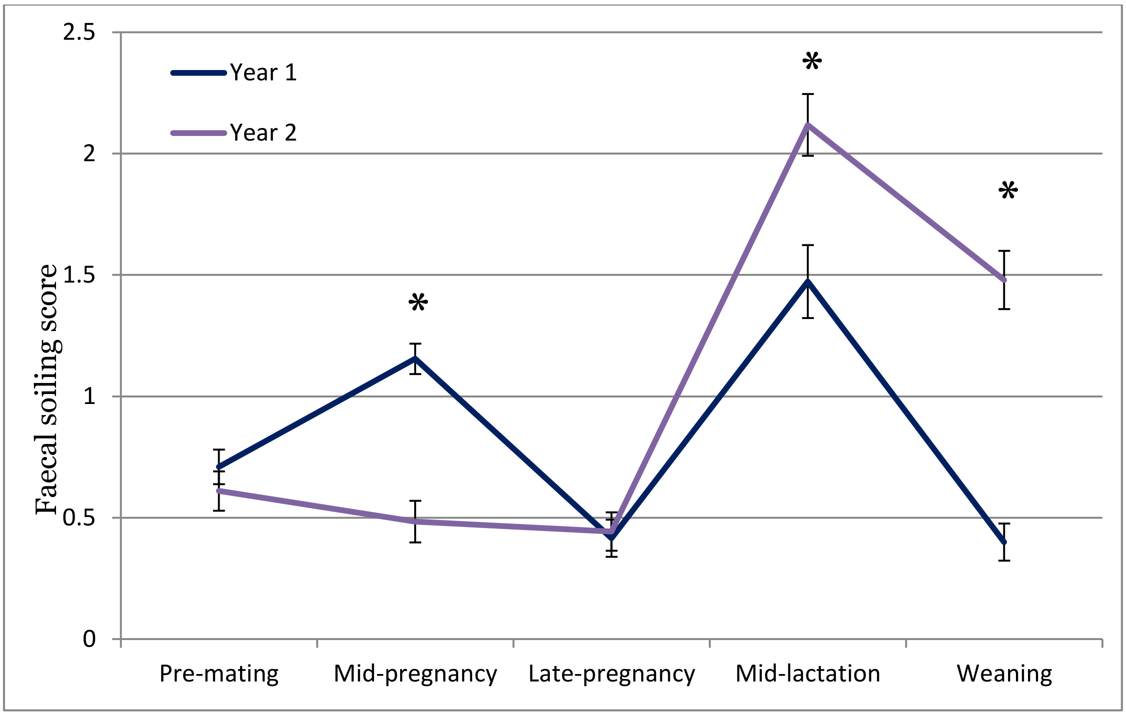


Figure 4.12 Variation in faecal soiling (±SEM) recorded during the Assessments at Gather in Year 1 and Year 2. Asterisks indicate significant differences between years ($P<0.001$)

4.3.3.8 Removal from flock

Ewes which were removed from the breeding flock at the end of Year 1 had significantly lower faecal soiling scores at the weaning assessment compared to those which were retained (Wald=8.88, df=4, p=0.024) although there were no significant differences at other points in time ($p>0.34$). Throughout Year 1 the ewes which were retained in the flock had consistently higher faecal soiling scores than those which remained in the flock as shown in Figure 4.11. In Year 2 no significant differences were identified in the faecal soiling scores for animals which were removed from the flock or retained for future breeding (p>0.18).

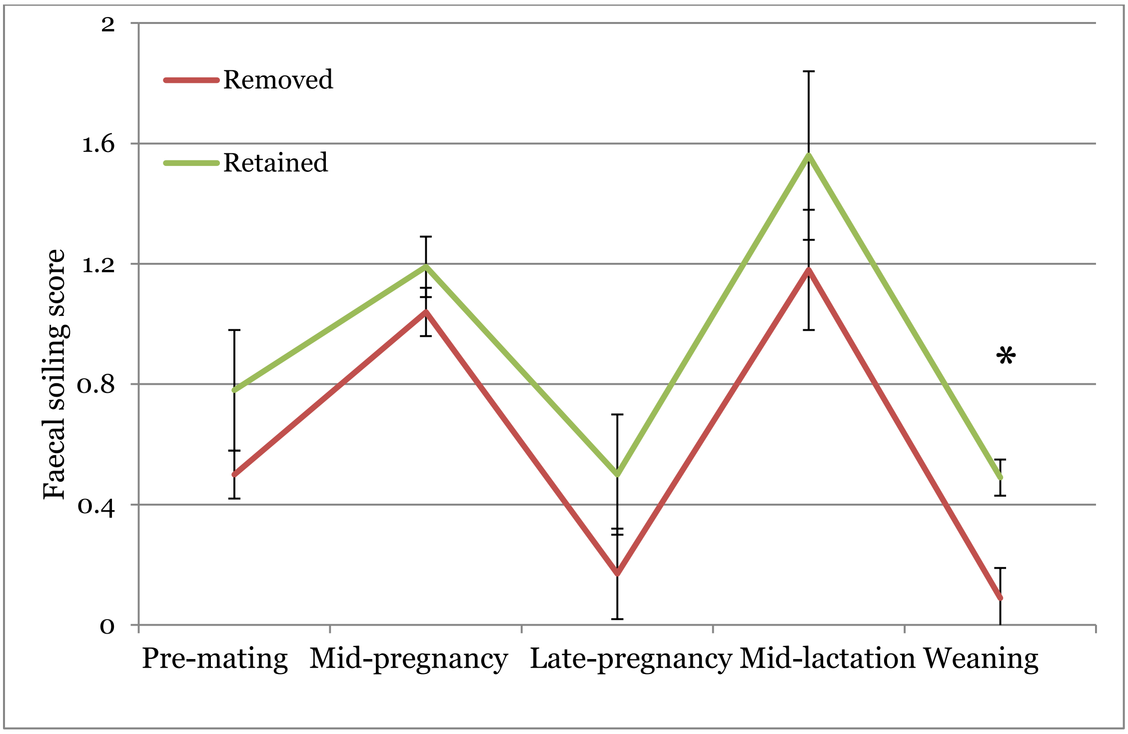


Figure 4.13 Variation in ewe faecal soiling scores (±SEM) throughout Year 1 for ewes which were removed or retained in the breeding flock at the end of the year in the 2012 stock draw

4.3.3.9 Lamb weights

Faecal soiling at Assessments 1 and 3 was significantly related to live lamb birth weight in Year 1 as shown in Table 4.11. At Assessment 1 (pre-mating) ewes with low faecal soiling were more likely to give birth to heavier lambs (p=0.049, Wald=6.30, df=2, F=3.15) than ewes with higher scores. At Assessment 3 (mid-pregnancy) the relationship between faecal soiling and lamb birth weight does not appear to be so straightforward as the ewes producing the heaviest lambs tended to receive the lowest and highest faecal soiling scores (Wald=12.75, df=3, F=4.25, p<0.008). The weaning weights of lambs in Year 1 were also significantly related to ewe faecal soiling at Assessment 3 (Wald=10.35, df=2, F=5.18, p<0.008). As for birth weight, the heaviest lambs were produced by ewes scoring a 0 or 3 for faecal soiling as shown in Table 4.12.

Table 4.6 Mean live lamb birth weight (total produced per ewe in 2012) and faecal soiling score. (SE Means are shown in parentheses).

	Faecal soiling score					
Assessment at Gather	0	1	2	3	4	P-value
1 (Pre-mating)	**6.25 (0.31) a**	**5.29 (0.28)**	**4.92 (0.72) b**			**0.049**
2 (Mid-pregnancy)	4.65 (0.63)	5.83 (0.25)	5.45 (0.54)	6.65 (1.26)		0.295
3 (late-pregnancy)	**6.05 (0.24) a**	**4.51 (0.4) b**	**4.5 (1.73)**	**7.5 (1.73)**		**0.008**
4 (mid-lactation)	6.16 (0.35)	6 (0.55)	5.13 (0.55)	5.57 (0.58)	4.83 (0.53)	0.225
5 (weaning)	5.51 (0.25)	5.98 (0.4)	8.15 (1.28)		5.1 (1.8)	0.194

Table 4.7 Mean live lamb weaning weight (total produced per ewe in 2012) and faecal soling score. (SE Means are shown in parentheses).

	Faecal soiling score					
Assessment at Gather	0	1	2	3	4	P-value
1 (Pre-mating)	39.89 (2.42)	34.93 (2.23)	24.83 (6.41)			0.063
2 (Mid-pregnancy)	32.19 (4.95)	37.36 (1.95)	33.14 (4.37)	57.6 (13.11)		0.258
3 (late-pregnancy)	**39.04 (1.83) a**	**28.48 (2.98) b**		**50 (12.29) c**		**0.008**
4 (mid-lactation)	41.67 (3.07)	39.48 (4.12)	32.95 (3.76)	36.7 (4.34)	31.45 (3.93)	0.228
5 (weaning)	35.58 (1.91)	37.17 (3.05)	63.1 (12.94)	3 (36.47)	4 (36.31)	0.114

No significant relationships were found between faecal soiling of the ewes and live lamb birth weight in Year 2 (all p>0.16). Trends were found between lamb weaning weight in Year 2 and faecal soling at Assessment 7 (mid-pregnancy p=0.093, Wald=6.7, df=3) and 9 (mid-lactation p=0.056, df=4, Wald=9.70). As reported in Table 4.13 the ewes with the second highest faecal soiling scores (3) in mid-pregnancy weaned the heaviest lambs, and by mid-lactation the ewes due to wean the heaviest lamb weights also scored 3 for faecal soiling.

Table 4.8 Mean lamb wean weight (total kg produced per ewe in 2013) and faecal soiling score. (SE Means are shown in parentheses).

	Faecal soiling score					
Assessment at Gather	0	1	2	3	4	P-value
1 (Pre-mating)	32.42 (1.76)	35.97 (2.56)	35.91 (4.21)	22 (11.15)		0.447
2 (Mid-pregnancy)	33.46 (1.46)	29.49 (2.53)	44.63 (5.65)	30.02 (4.38)		0.093
3 (late-pregnancy)	32.72 (1.56)	34 (2.99)	39.77 (4.4)	25 (10.78)		0.409
4 (mid-lactation)	28.23 (4.27)	33.34 (2.34)	37.31 (2.61)	36.53 (2.4)	27.39 (2.9)	0.056
5 (weaning)	28.75 (2.77)	31.71 (2.34)	36.93 (2.28)	33.39 (3.79)	39.52 (4.79)	0.132

4.3.3.10 Effects of age

No relationships were found between the age of the animal and faecal soiling (p=0.89), or parasitic egg counts (p>0.37).

4.3.3.11 Validation

No relationships were found between faecal soiling and parasite egg count when assessing the data across all Assessments at Gather (all p>0.05). However, *Coccidia* egg counts were found to be significantly related to faecal soiling at Assessment 8 (late-pregnancy) ($p=0.003$, Wald=16.25, df=3, F=5.42. Table 4.14): ewes with the highest worm burdens had the highest faecal soiling scores. No significant relationships were found between *Strongyle* egg counts and soiling ($p>0.148$).

Table 4.9 Mean Coccidia egg count in faeces and faecal soiling scores at individual Assessments at Gather (standard error of the means in parentheses).

	Faecal soiling					P-value
Assessment at Gather	0	1	2	3	4	
2 Mid-pregnancy	64.29 (68.87)	173.53 (31.25)	31.25 (64.43)	75 (128.85)		0.162
4 Mid-lactation	106.25 (42.03)	100 (48.53)	85.71 (44.93)	66.67 (68.63)	100 (44.93)	0.989
5 Weaning	303.6 (114.7)	104.2 (214.6)	1500 (743.4)			0.192
6 Pre-mating	221.9 (66.5)	346.2 (104.3)	50 (188)			0.353
7 Mid-pregnancy	205.4 (70.5)	145.8 (123.8)	100 (214.4)	175 (303.2)		0.952
8 Late-pregnancy	**106.4 (30.93) a**	**31.3 (68.29) a**	**150 (78.86) a**	**850 (193.16) b**		**0.003**
9 Mid-lactation	525 (299.9)	468.2 (221.5)	80 (232.3)	100 (203.7)	62.5 (367.3)	0.541
10 Weaning	81.58 (40.81)	58.7 (37.09)	111.9 (38.81)	145 (56.25)	37.5 (88.94)	0.668

4.3.3.12 Variation in Lamb survival between Years 1 and 2

Significantly more lambs died between birth and weaning in Year 2 in comparison with Year 1 ($\chi2=6.074$, df=2, $p=0.038$, Figure 4.12).

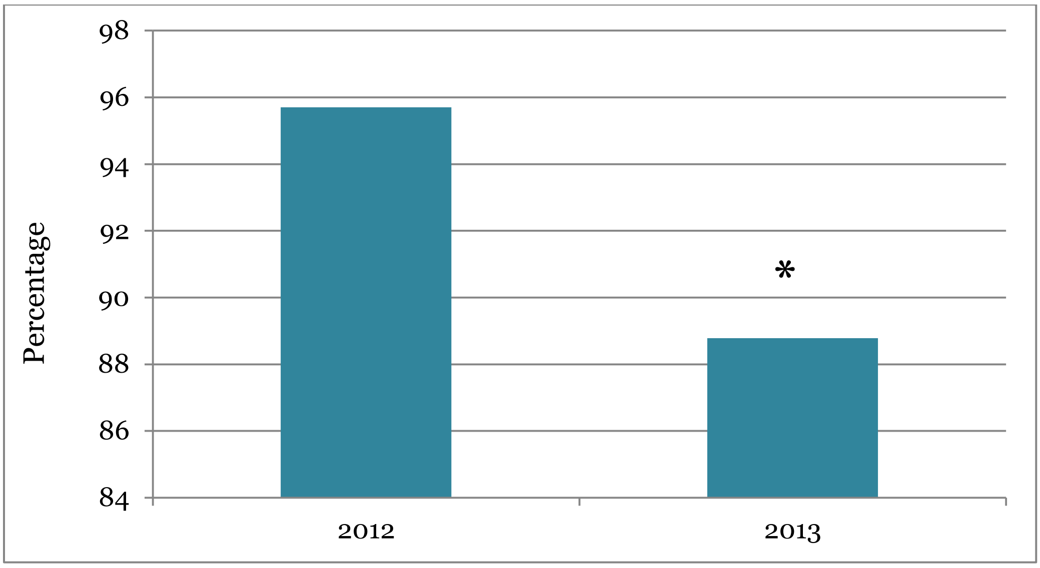


Figure 4.14 Percentage of lambs produced by the sub-flocks in Years 1 and 2 surviving until weaning

4.3.4 Cluster analysis

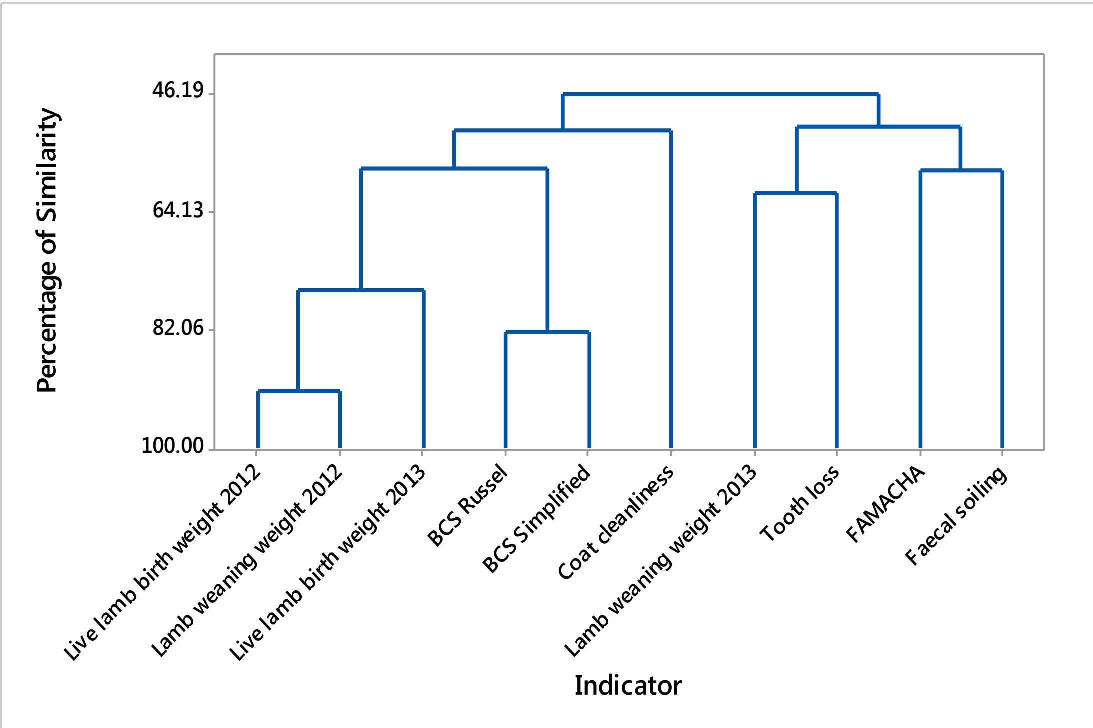


Figure 4.15 Dendogram showing similarity of information obtained using indicators applied during the Assessments at Gather

The results from the multivariate cluster analysis show the similarity in the information provided by the indicators applied during the Assessments at Gather. From the dendogram (Figure 4. 13) it can be seen that the two indicators which offer the most similar information are body condition when assessed using the Russel et al (1969) scale, and body condition using the simplified scale. The other indicators do no share similarity above 60% indicating that they offer information on discrete areas of the animal's welfare. This implies that none of the indicators applied here may be considered redundant as they all provide information on different aspects of welfare.

4.4 Discussion

This study appears to be the first to perform such a comprehensive assessment of sheep physical health on the same individual animals 10 times throughout a two year longitudinal time period. Time was found to have a significant impact upon each of the welfare indicators applied during the Assessments at Gather. The ewes in Year 2 consistently fared significantly worse in terms of BCS, tooth loss, anaemia (as measured by FAMACHA) and faecal soiling than at the equivalent points in reproductive cycle in Year 1. Repeating the assessments at the same time period in consecutive years was also generated significantly different results. These findings have real implications for the interpretation of results from assessments carried out during professional audits of commercial farms especially when comparing between farms, or within farms at different time points. At the annual stock draw, ewes which were selected for removal from the breeding flock were found to fare significantly worse in terms of BCS, tooth loss and anaemia. The distinction between the tooth loss scores of removed and retained ewes was seen almost a year in advance of the stock draw. Age of ewe was found to have a significant impact upon tooth loss and BCS suggesting that older ewes are more vulnerable to nutritional deficiencies than younger animals. The results from the cluster analysis indicate that the measures selected do indeed assess distinctive areas of welfare, and as such no measures are considered to be redundant. The indicators applied during the Assessments at Gather have also proven to be feasible when assessing gathered sheep.

4.4.1 Consistency over time

Effect of time on indicators assessing Good Feeding

The increase in tooth loss and decrease in BCS seen in Year 2 is believed to be reflective of a genuine alteration in welfare state. A possible reason for these results may be the substantial variation in environmental conditions which occurred during the ewes' pregnancy in Year 2 (2013). From late February, until the end of March 2013, heavy snow fell across Scotland and it was the coldest March since 1962 (The Met Office, 2015). This coincided with the ewes entering the mid-point of their pregnancies and as these animals are unhoused the environmental conditions undoubtedly had an effect on their welfare. The snowfall was approximately 60cm deep which impacted upon the ability of the ewes to forage and obtain sufficient food intake. They were provisioned with supplementary feed when possible by the shepherd however due to practical constraints this was not always feasible. There was also no guarantee that the supplementary feed would be consumed by all ewes. The effects of this adverse weather lasted for months after the snow had melted by mid-April when the ewes started to lamb, although the grass quality and quantity remained poor. Ewes struggled to maintain their body condition throughout lactation and the additional strain of providing for their lambs most likely also impacted upon their dentition. By weaning, body condition started to improve although it is not known the length of time it would have taken for it to return to levels typical of a normal year. It would have been interesting to follow the ewes for longer to investigate how long this bad weather impacted upon the welfare of the sheep.

The alterations seen within years correspond with findings in the literature and it is proposed they are related to the ewe's stage in her reproductive cycle. In the current study, the ewes' BCS was lowest during lactation, and tooth loss highest at this time in both years. Pregnancy and the following lactation are costly in terms of energy. It has been suggested that the full cycle of conception, pregnancy and lactation can be considered the most energetically expensive activities that a mammal can undertake (Wade and Schneider, 1992). In order for a ewe to produce a sufficient volume of milk for her lambs, and maintain a healthy body condition, she needs to be able to maximise her nutrient intake. The protein and energy requirements of lactating ewes are double those of non-lactating ewes (Ruckstuhl and Festa-Bianchet, 1998) and an increase in voluntary food intake helps to meet these requirements (Forbes, 1970). The stages of lactation have been shown to influence food intake in sheep (Molina et al., 2001) as food intake increases up to the fourth week post-lambing then declines by the seventh week of lactation along with milk yield (Hadjipieris and Holmes, 1966; Molina et al., 2001). Purely increasing the quantity of food consumed at this time may not be sufficient for the ewe to remain healthy if she is not consuming appropriate quantities of minerals. It is known that insufficient calcium intake during lactation can lead to premature tooth loss and weakness (Gunn, 1969) which in turn affects the ewes' feeding ability. The increased nutritional demands placed on the ewe during lactation is most likely the reason behind the poorer tooth loss and BCS recorded at this time.

Effect of time on Good Environment indicators

The ewes' coat cleanliness scores were also consistently worse in Year 2 compared to Year 1 with the exception of the pre-mating time point. It is proposed that this finding is also due to the poorer weather occurring in Year 2. In both Years 1 and 2 the worst coat cleanliness scores were recorded in the pre-mating time point which occurred in the winter. This was to be expected as the ewes are not housed and in the winter there is no guarantee of dry ground on which to lie. The increase in coat cleanliness as the spring and summer progressed indicates that during these seasons the environment provided to the ewes offered sufficient clean and dry lying space.

Effect of time on Good Health

As with the indicators described previously, anaemia (as measured by the FAMACHA scale) and faecal soiling were significantly worse in Year 2 compared to Year 1, and when comparing within years they too deteriorated during pregnancy through to mid-lactation and weaning. The increase in faecal soiling during the summer months may be a reflection of the alterations in environmental conditions and parasite abundance. The increase in demand for energy and grazing in the summer to support greater milk production corresponds with an increase in the number of gastrointestinal parasites in the environment during warm weather (Uriarte et al., 2003). This alteration in grazing behaviour can result in herbivores consuming high numbers of the parasites which cause faecal matter to become loose and adhere to the fleece. Alongside the increase in parasite numbers at this time of year, there are also alterations in the ewes' immune system during pregnancy and lactation. During this time ewes which would normally have good acquired immunity are experiencing the periparturient reduction of immunity which can cause a temporary increase in gastrointestinal parasite burdens (Barger, 1993;

Houdijk et al., 2000). This increase in parasite burden may explain the increase in faecal soiling, and could have also lead to the animals becoming more anaemic.

4.4.2 Convergent validity of indicators relating to Good Feeding

The simplified BCS consisting of four categories: emaciated, thin, fit and fat was found to be well related to the previously well validated Russel et al (1969) scale. This association helps to support the use of the simplified scale as a valid indicator for sheep allowing an assessor to identify animals which are either under-, or overweight and are therefore more likely to have impaired health and welfare. The connections between the simplified BCS and the blood metabolites, BHB and NEFA proved to be less conclusive. These blood measures may reflect long and short term fluctuations in the energy balance of the ewe, yet it may take weeks or months for the body condition of the sheep to alter sufficiently for them to become emaciated or fat. The exact time for BCS to increase or decrease to dangerous levels depends upon the breed of sheep, reproductive state, feed intake and exercise (Caldeira et al., 2007b; Verbeek et al., 2012). In a study involving dairy sheep it took 74 weeks for the BCS of a ewe to decrease from 4 to 1.25 (Caldiera et al 2007). Animals in the same body condition category may be in different metabolic states as a ewe with a BCS of 2, which would considered to be healthy, may still be chronically under-nourished if this score of 2 occurred during a downward trajectory (Caldeira et al., 2007a). It may be reasonable to assume that during a welfare assessment at a farm level that these examples may be rare and a high proportion of animals would receive condition scores relevant to their current welfare state. Also, in this study a relatively small proportion of the body condition observations identified animals which were “emaciated” (1%), “thin” (8%), or “fat” (6%), as in the majority of assessments animals were “fit” (85%) and therefore statistically the assessment of the relationships may not have been sensitive enough. Further investigation may be beneficial with more contrast between the BCS of the ewes.

4.4.3 Convergent validity of indicators relating to Good Health

This appears to have been the first time the FAMACHA chart has been used to identify anaemia in Scottish Blackface sheep. Relationships between the FAMACHA anaemia chart and blood parameters were found which suggested it may be assessing relevant aspects of the ewe's physiology. Over all assessments, a good relationship was found between the FAMACHA chart as animals with high red blood cell counts were typically recognised as being less anaemic. These relationships were not always strong, or in the predicted direction raising questions regarding the charts validity with these animals. More dubious relationships were found with packed cell volume and platelet counts. From the data presented here it appears that the FAMACHA scale may not be particularly well suited to Blackface hill sheep living in Northern Europe. It is possible that the skin pigmentation of the sheep caused the scale to be less sensitive than it may have been with paler animals. The scale was initially developed for use with white faced sheep in South Africa (Bath and Wyk 2009). When assessing the variation in FAMACHA score for white and blackface sheep in Germany, Moors and Gauly (2009) found significant differences between the FAMACHA scale of these groups even when their physiology results were similar. It is also possible that fluctuations in the light intensity may have led to the reduction in validity of the measure. The Assessments at Gather were performed outside and in a shed, at different times of day and across the course of the year. This may have led to subtle differences in light quality causing some variation in FAMACHA scores given to ewes. There are not specific guidelines regarding the lighting when performing an assessment using this chart, but for future studies investigating the validity of the chart maintaining consistency in ambient lighting may prove to be important. Further investigation in more controlled lighting conditions and comparison between breeds, including the Scottish Blackface, would aid in the further investigation of the suitability of this as a welfare indicator for hill sheep in Northern Europe.

Using faecal egg counts to help validate the faecal soiling score did not yield promising results. Only one occasion (Assessment 8; late-pregnancy) was heavier faecal soiling found to be related to increased *Coccidia* oocytes counts in the faeces. No relationships were found between faecal soiling and *Strongyle* worm egg counts. The validity and use of faecal soiling as a non-invasive measure of faecal egg counts has been established in other studies (e.g. Broughan and Wall 2007). The poor connection between the faecal egg counts and fleece soiling may have been due to management practises. If ewes are found to have heavy soiling the shepherd will frequently clip the wool from the tail, haunches and breech when the sheep are gathered in the summer months to prevent fly strike. Therefore ewes with the highest parasite burdens may have had the fleece removed and rear cleaned prior to the scoring during the assessments, and therefore received low faecal soiling scores which were unrepresentative of their natural state. If this work were to be repeated in future it is advised that the animals' rears would not be cleaned by the farmer during the observation period. Faecal soiling may also be related to other environmental conditions such as grass quality and protein content and thus high soiling scores may not be representative of high worm burdens, but rather to eating rich vegetation (Broughan and Wall 2007). However, as soiled fleece can lead to fly strike, a high faecal soiling score (indicating a dirty coat) is a welfare concern regardless of the cause.

4.4.4 Use of lamb survival and growth as an iceberg indicator

The percentage of lambs which survived until weaning was lower in Year 2 in comparison with Year 1, which corresponds with the overall decrease in ewe welfare seen at this time. The relationship between lamb weight and ewe welfare is more complex. Although ewes with higher BCS scores (indicative of better nutritional welfare) produced more lamb weight, higher lamb weight was also associated with an increase in tooth loss. The increased demands placed on the ewe at this time is the most likely the reason behind the relationship observed between high lamb weaning weights and poorer ewe dentition. In a study on hill sheep, Annett et al (2011) found high incidences of ewe tooth loss did not have a detrimental effect on the average growth rate of lambs, and they suggested that milk production was maintained at the expense of the ewes' health i.e. teeth and bone structure. Based on the results from the study reported in this thesis, and Annet et al (2011), the weight of lambs produced per ewe is not an accurate indication of ewe welfare state. Lamb survival on the other hand may prove to be a better indication of the overall physical health and welfare of the ewes. These findings are in line with Stott et al (2012) and add support to the use of lamb survival as an Iceberg Indicator of ewe welfare.

4.4.5 Effect of age of ewes on Good Feeding

Only indicators assessing the principle of Good Feeding were found to be influenced by ewe age. The body condition and tooth loss of the sheep was also found to be affected by age as older ewes were more likely to have lower BCS and higher levels of tooth loss. It is likely that the tooth loss contributed to malnutrition and the successive decrease in body condition. These findings are in accordance with studies in dairy cattle (De Vries et al 2011). The results from the work reported in this chapter suggest that the ages of the sheep assessed could have a significant impact upon the results obtained.

4.4.6 Removal from flock

BCS was found to be a predictor of removal at the Year 1 stock draw, as fatter ewes tended to be retained for future breeding. In Year 2, BCS did not differ between these groups, possibly due to fact that the flock as a whole were much thinner than the previous year. At the end of Year 2 the average BCS of the entire sub-flock was similar to the removed ewes at the Year 1 stock draw. This may have altered the shepherd's decisions when selecting individuals for removal in Year 2. Older ewes (approximately 6 years of age) are routinely removed from the flock as they are not expected to be able to sustain another pregnancy even if they have no existing welfare issues. Older ewes tend to lose more weight during pregnancy and lactation than younger sheep (Annett et al 2011) which could potentially cause serious welfare problems in the future.

Tooth loss was found to be the most consistent indicator of removal from flock with the retained and removed ewes differing almost a year in advance of the stock draw. In both Years 1 and 2 the ewes removed from the flock had significantly higher tooth loss than those remaining within the breeding flock. This is a particularly remarkable finding as these differences can be traced back as far as the pre-mating time point, thus these animals were significantly different from each other almost a year in advance of the stock draw. Ewes with more dental problems and lower BCS were less likely to survive to the next breeding season in a study conducted by Annett et al (2011). However, Annett et al (2011) did not report on the cause of death and whether the animals were removed by the shepherd from the flocks or whether the animals died due to lack of foraging ability. The work reported in this thesis appears to be the first to show that ewes removed from a flock during an annual stock draw could be predicted almost a year in advance based on their dentition. It is therefore suggested that the farmer could provide additional supplementation to these animals during pregnancy and lactation to minimise subsequent welfare problems.

4.5 Conclusion

In conclusion, the time at which welfare indicators are applied to animals has a considerable impact upon the outcome. The fluctuations in scores may be attributed to normal shifts in the animal's physiology due to alterations in reproductive state or calendar season, or due to external environmental influences such as weather patterns. Regardless of whether these shifts occur due to internal or external factors they represent meaningful changes to the animal's welfare and need to be taken in to account during an on-farm assessment. The welfare indicators applied in this chapter proved to be feasible for use during an Assessment at Gather on a commercially run farm. Tooth loss and body condition were also found to be particularly relevant to, and predictive of, management decisions at the annual stock draw. Further validation work is required before the FAMACHA anaemia chart can be accepted as a welfare indicator for use with Northern European sheep breeds. The use of lamb survival as an iceberg indicator is promising, but lamb birth and weaning weight did not accurately reflect the welfare of the ewes and should therefore not be considered a suitable welfare indicator. These findings have real implications for the interpretation of results from assessments carried out during professional audits of commercial farms especially when comparing between farms, or within farms at different time points.

Chapter 5 Qualitative Behavioural Assessment of sheep: inter-observer reliability and seasonal variation

Abstract

Assessing the welfare of extensively managed animals, such as sheep, requires welfare indicators which do not necessitate close contact. A whole-animal assessment method (Qualitative Behavioural Assessment (QBA) is therefore particularly useful as it can be performed with minimal disturbance. When implementing QBA, observers focus on how animals interact with their environment and perform behaviours, before translating this information into qualitative descriptors such as "calm" or "agitated". Two studies were performed to: 1) test the inter-observer reliability of QBA when applied to 48 individual ewes in an extensive environment and 2) track longitudinal changes in individual ewe expression over six months, spanning late pregnancy to post-weaning. Both studies used the same fixed list of 21 qualitative terms, which were analysed by Principal Components Analysis (PCA) to determine the main dimensions of ewe behavioural expression. In the first study, three trained observers achieved good agreement on two PCs, and moderate agreement on a third during two farm visits (PC1 $W=0.77$, PC2 $W=0.70$, PC3 $W=0.54$; all significant at $P<0.001$). In the second study, the behavioural expression of up to 49 individual ewes was tracked by one of those three observers on 13 occasions over six months (spanning late gestation to post-weaning). Four meaningful PCs were identified; the first two accounting for more than half of the explained variation between sheep, classifying the main components of 'General Mood' (PC1) and 'Arousal' (PC2). Significant systematic changes in affective state occurred across the observational period with increases in General Mood and Arousal. These may have been related to external (environmental) and internal (physiological alterations in reproductive state) factors occurring during this time. We conclude that QBA is capable of recognising meaningful variation in the

expression of individual sheep and should be further investigated for use as a welfare indicator for on-farm welfare assessments in extensive systems.

5.1 Introduction

For animals living in extensive systems, logistical challenges can arise when performing a welfare assessment. Extensively managed animals may be difficult to identify individually and a large flight distance may prevent assessors approaching and handling the animals in the field. Animals may range across a large area or number of locations making gathering animals for inspection expensive in terms of both time and labour. Gathering may also alter the welfare of the animals being assessed and might be unsuitable at particular times of year (e.g. during mating and when lambs are at foot). Thus, for these animals, welfare indicators which do not demand close contact are needed (Fitzpatrick et al. 2006; Turner & Dwyer 2007; Clare J Phythian et al., 2012). A measure which may be particularly well suited to the assessment of extensively managed sheep is Qualitative Behavioural Assessment (QBA). QBA requires relatively few resources, allows animals to be observed in their normal environment and can be performed from a reasonable distance with minimal disturbance. These aspects of the method are advantageous to the assessors, producers and to the animals themselves (Phythian et al., 2015). During conventional quantitative behaviour scoring, the observer records which behaviours are present, but the QBA approach asks the observer how behaviours are being executed. When performing QBA the observer assimilates many pieces of information about the animal's body language and the way in which it interacts with the environment and translates this in to qualitative descriptors such as "calm" or "agitated" (Wemelsfelder et al., 2000). Unlike other animal-based welfare indicators, which focus on specific individual aspects of physical health and behaviour of an animal, QBA does not isolate these facets, since in doing so the concept of the "whole-animal" is lost. This "whole-animal" information cannot be regained at a later stage and potentially leads to important information being omitted when using quantitative scoring (Wemelsfelder & Lawrence 2001; Walker et al. 2010; Napolitano et al. 2012). Another advantage of the QBA approach is its ability to capture both positive and negative aspects of welfare (Wemelsfelder, 2007). The information gained when using QBA enhances and complements quantitative scoring measures used for assessing animal welfare (Wemelsfelder et al., 2000; Wemelsfelder et al. 2001; Andreasen et al, 2013). The descriptive terms scored by assessors may be either generated individually by the assessors, through a procedure known as free-choice profiling (Wemelsfelder, 2007), or by using a fixed list in which terms are predefined and agreed upon by a panel of people with experience of the species (Forkman and Keeling, 2009c). In on-farm welfare assessments for which the location, time and observers may differ, the fixed list approach offers greater consistency of terms and definitions than the free choice profiling method (Forkman and Keeling, 2009a, 2009b, 2009c). As all observers use the same fixed list of terms and definitions, direct comparisons can be made between observers, farms and animals (Forkman and Keeling, 2009a, 2009b, 2009c).

As shown in the previous chapter, seasonality and weather may have a significant impact upon sheep welfare assessment outcomes. Temperatures fluctuate widely in sheep farming regions and extremes of temperature are known to have a great impact upon sheep physiology (Marai et al., 2007; Nardon et al., 1991; Silanikove, 2000). Pregnancy also places unique demands upon the ewe and is known to affect physiology (Lingis et al., 2012; Nasar and Rahman, 2006) and behaviour (Poindron et al. 1994; Poindron et al. 1997; Viérin & Bouissou 2001).

Considering the impact that such factors may have on animal-based welfare assessment scores, the current studies reported in this chapter had two main aims 1) to test the inter-observer reliability of QBA when applied to extensively managed ewes (maintained outdoors throughout the year on a Scottish hill farm), and 2) to track longitudinal changes in ewe behavioural expression through over six months, spanning late gestation to post-weaning.

5.2 Methodology

5.2.1 Term development

The two main aims of the project were addressed in two separate studies. A fixed list of qualitative terms was first developed and then used for both studies. The terms used in the fixed list were established in a two stage process. In stage one a focus group was held which consisted of 11 participants (all female, aged between 24 – 54 years of age) with experience of sheep in a variety of roles: veterinary, research, shepherding/husbandry and research technicians. The focus group reported here viewed the same 12 one-minute long video clips as used in the study reported by Phythian et al (2012). Six of the videos focussed on individual animals, while the others allowed views of the animals at a group level. The videos featured adult sheep (ewes and rams) with and without lambs, and a variety of production systems consisting of indoor, lowland, upland and hill with the aim of covering as many aspects of behaviour and the expressive repertoire of sheep as possible. Further information on the individual videos can be found in Phythian et al (2012). The focus group discussed the 12 terms previously developed by Phythian et al (2012), however the group consensus was that this list should be adjusted to meet the needs of welfare assessment in extensive sheep systems and additional terms were suggested and discussed. The focus group reached agreement on 25 terms and meanings to be used when observing ewes under extensive conditions.

Stage two of the process consisted of refining the number of terms by sending the 25 terms and meanings to seven sheep experts, both local and international. Following their feedback, the list of QBA terms was reduced from 25 to 21. The finalised terms ranged across negative to positive welfare states, and low to high energy levels. The terms were then arranged as to avoid any bias during scoring, and the terms remained in this order throughout the data collection period. The terms, order and their meanings are shown in Table 5.1.

Table 5. 1 Qualitative terms and descriptions for ewe expression used in both studies

Qualitative Term	Description
Alert	Observant and vigilant.
Active	Animal is physically active and engaged in task e.g. grazing, or walking.
Relaxed	At ease, free from anxiety, agitation or tension. The animal appears to be unthreatened.
Fearful	Attention is focussed on one specific object/being which is either a real or perceived threat. Animal may, or may not be fleeing.
Content	Satisfied and at peace. Their needs are met, or the animal is successfully working towards their completion.
Agitated	Excessive cognitive and/or motor activity due to tension or anxiety. The animal is uneasy and restless. If moving, their actions are erratic.
Sociable	Seeking and interacting with other sheep. The sheep appears to be enjoying/taking comfort from their contact. The sheep is actively choosing to be part of a flock and not fully isolate themselves.
Aggressive	Hostile and tense. Either attacking or ready to attack. Usually unprovoked or to compete for resource.
Vigorous	The animal is carrying out task in an energetic or forceful way. If stationary or moving slowly the animal expresses an inner strength and energy. May imply good physical health.
Subdued	Submissive and docile. Often removed from social group and appear to be self-absorbed.
Physically Uncomfortable	Giving impression of pain or other physical discomfort through posture/movement.
Defensive	Ready to potentially defend herself or lamb from harm/perceived threat.
Calm	Placid and sedate. If physically active the animal's movements are smooth and unhurried.
Frustrated	Dissatisfied. Unable to fulfil satisfaction and achieve goal.
Apathetic	Unresponsive and dull.
Wary	Shy, cautious, apprehensive and possibly distrustful.
Tense	Uneasy and/or on-edge. Posture may show physical tension.
Bright	Alert, lively and aware of environment.
Inquisitive	Curious, interested and intrigued by the environment or other animals.
Assertive	Displaying confidence or determination.
Listless	Lack of vigour and energy. Animal appears lacklustre.

The 21 terms were each scored on a visual analogue scale with two anchors, a left end "minimum" point meaning the expression denoted by the term was considered absent, and a right-end "maximum" point indicating that the animal could not show the expression more strongly. Scores were measured as the distance (in mm) from the minimum point, to where the observer placed their mark. The visual analogue scales were displayed on electronic tablets (Ainol Novo7 flame, China) and observers' scores were exported automatically into excel files, using a QBA assessment app recently developed at SRUC. The development of the QBA app is ongoing at SRUC, with grateful acknowledgement of start-up-funding by the AssureWel project as supported by the Tubney Charitable Trust.

5.2.1.1 Location and animal identification

The studies used 57 Scottish Blackface ewes, selected from a flock of 700 ewes and were representative of age (aged between 2 and 7 years), weight (between 35.5 to 62.4 kg) and body condition (ranging from 1.75 to 3.5). The sample group were selected following ultrasound scanning for pregnancy determination and litter size in February 2013. All selected ewes were carrying at least one lamb when scanned, however at lambing four ewes did not produce any offspring. The ewes were managed outdoors on natural pasture all year round. Due to management decisions during late gestation, and while lambs were at foot, the ewes were moved by the shepherd to different grazing areas. The sheep were relocated between free grazing on the hill (unfenced, unimproved pastures of low nutritional quality), fenced areas of unimproved hill grazing (250 hectares unimproved/less favoured grazing), or improved fields near the farmstead (22 hectares of fenced improved grazing). They had unlimited access to grazing and were provided with nutritional supplementation approximately six weeks prior to lambing (16% Protein Ewe concentrate, East Coast Viners, UK –approx. 250-500g/ewe per day).

The ewes were individually identified with plastic ear tags. However, to increase the distance from which the animals could be identified a green stripe was applied with marker spray across their shoulders (Ritchey super sprayline stockmarker, England), and uniquely coloured and numbered tags (Alflex maxi female cattle ear tag, Alflex Europe) attached around their neck with string.

The sheep were managed on the farm to achieve commercial growth rates, using commercially relevant management, and decisions influencing sheep husbandry were made independently from this study. In August, the male lambs were weaned and relocated to an area in which they no longer had visual or auditory contact with their dams. The female lambs remained with the ewes until they were weaned at the annual stock draw in mid-September (after the end of this study). The period following the removal of the male lambs is termed “post-weaning”, even though female lambs remained with their dams.

5.2.2 Assessment procedure

For both studies each individual ewe was observed from a distance of between 5-20 metres for 60 seconds, which was timed using a stopwatch. Ewes were observed in the order in which they were found and for the observation to begin the ewe’s body and face had to be fully visible. In poor weather, or at distances greater than 10 metres, binoculars were used to ensure a good view of the animal. If ewes were disturbed by the presence of an observer (as determined by interruption of their behaviour or posture to focus attention on the observer) they were given time to resume the behaviour they were performing prior to disruption. Once normal behaviour had resumed, observers waited for a further 30 seconds before the observation started. After each observation of an individual sheep, each term on the fixed list was scored on the visual analogue scales on the tablet. The time of day was recorded by the tablet app, whereas weather (categorical: rain/sun/snow etc), temperature (°C using a thermometer) and location (field number, hill area) were recorded into the tablet manually along with notes of any disturbances e.g. the presence of hill walkers or dogs.

5.2.3 Study 1 - Inter-observer reliability

5.2.3.1 Training session

The assessors attended a training session lasting approximately three hours. The 21 terms and meanings were discussed at the beginning of the session, followed by a viewing and scoring of the video footage described above in the "term development" section. The observers practiced scoring using the electronic tablets, and although the scoring occurred in silence, after each video clip (12 clips) a dialogue took place regarding any discrepancies in how observers had used each term to score that clip. The training session ended when all observers were confident in using the tablets, and had achieved good understanding and agreement on the terms and their definitions.

5.2.3.2 Assessor information

The three assessors in the first study (all female) consisted of one PhD student (SER), and two research assistants (JD and MF) all of whom were familiar with sheep and had previously performed QBA training, but had varying levels of experience of using the approach and applying it to sheep.

5.2.3.3 Assessment procedure

The three observers performed QBA on the marked individual ewes on two separate visits 26 days apart in the summer of 2013. The first visit occurred while the ewes were lactating with lambs present, and the second visit after weaning. The assessors always observed the same ewe at the same time for 1 minute before turning away and scoring all 21 terms individually on the tablets without conferring. Observations were conducted between 0900 and 1700. From the 57 ewes selected for this work as reported in Chapter 3, complete data sets were collected for 47 ewes at visit one, and 48 at visit two. The discrepancy in these numbers was due to the observers not being able to locate all individual sheep.

5.2.4 Study 2 – Longitudinal variation

The data for the second study were collected by one observer (observer SER) who was one of the three observers involved in the first study.

5.2.4.1 Assessment procedure

The QBA assessments were performed on the marked individual ewes (from the same subset of 57) on 13 visits ranging from the 5th of March to the 31st of August 2013. Visits 1 to 4 took place pre-lambing, visits 5 to 11 took place post-lambing (lambs present), and visits 12 and 13 post-weaning. The observations were not evenly spaced, with the shortest time between visits being 3 days, and the longest 65 days, as no QBA observations were conducted during the lambing period. Observations were made between 0800 and 1800. Complete data sets were collected for 43 ewes in the pre-lambing time period, 49 post-lambing and 46 post-weaning. As in Study 1 the discrepancy in these numbers was due to the observers not being able to locate all individual sheep.

5.2.5 Data Preparation and Analysis

The scores for the first study (two visits by three observers) were analysed together using a Principal Component Analysis (PCA) with correlation matrix (no rotation) in Minitab 16 (Minitab Inc). A PCA generates a number of Principal Components (PCs) which describe the variation between sheep expression across all individuals and visits. This culminates in individual sheep receiving a score for each visit on the generated PCs. Agreement between the three observers in how they ranked these scores was calculated using Kendall's coefficient of concordance (Martin and

Bateson, 1993). In addition, agreement was investigated by testing whether there was an observer effect on the mean values of the scores for PCs one, two and three, using a random coefficient REML mixed model. Visit number was fitted as a fixed effect and observers and ewe identity as random effects. These analyses were all performed in Genstat 15 (Genstat for Windows International, UK).

The data collected in the second study (13 visits over six months by one observer), were analysed together using PCA with correlation matrix (no rotation). The effects of visit number, lamb sex, time of day (a.m./p.m.) and ambient temperature on PCs scores were investigated using REML (Genstat 15). Ewe identity was fitted as a random effect. The temperature data were categorised for analysis in three groups: Cold: between -1 and 8°C; Cool: between 9-18°C; and Warm: from 19-25°C. The effect of reproductive state (pregnant pre-lambing, lactating post-lambing and post-weaning following the removal of the male lambs) was investigated using a one-way repeated measures ANOVA (with reproductive state as a factor), followed by a post-hoc Tukey test in Minitab 16. By including lamb sex in the analyses we were able to further investigate the potential impact of weaning on the ewes with male lambs, in comparison to ewes remaining with their female lambs.

5.3 Results

5.3.1 Study 1 - Inter-observer reliability

Three meaningful components were identified from the PCA with Eigenvalues greater than 1.5, which explained a total 56.3% of the variation between sheep (Table 5.2). The first component (PC1) explained 27.4% of the variation and had terms describing high energy (e.g. tense, wary, alert, active) at one end and low energy (calm, relaxed, content, subdued) at the other. This PC was labelled 'Arousal'. The second component (PC2), explained 21.2% of the variation and was described by positive emotional states at one end (bright, content, vigorous, relaxed) and negative emotional states at the other (subdued, apathetic, listless, physically uncomfortable). This PC was labelled 'General Mood'. The third component (PC3) explained 7.7% of the variation between sheep, and had the terms inquisitive and vigorous loading highly at one end, and defensive and fearful at the other. Though comparatively not a strong component, this PC appeared to indicate the sheep's tendency to express either defensiveness or curiosity and is labelled as 'Defensiveness'.

Table 5.2 Study 1: PCA loadings from combined analysis of two assessments by three observers. Bold type indicates terms are highest or lowest loading on the principle component

PCA Loadings					
PC1 "Arousal" (27.4% variation explained)		**PC2 "General Mood"** (21.2% variation explained)		**PC3 "Defensiveness"** (7.7% variation explained)	
Calm	**0.31**	**Bright**	**0.31**	**Inquisitive**	**0.34**
Relaxed	**0.30**	**Content**	**0.26**	**Vigorous**	**0.27**
Content	0.24	Vigorous	0.23	Active	0.24
Subdued	0.15	Relaxed	0.21	Apathetic	0.24
Apathetic	0.12	Active	0.15	Assertive	0.24
Listless	0.12	Alert	0.14	Listless	0.23
Sociable	-0.05	Assertive	0.13	Subdued	0.16
Phys. Uncomfortable	-0.06	Calm	0.13	Alert	0.10
Aggressive	-0.10	Inquisitive	0.13	Bright	0.06
Defensive	-0.17	Sociable	0.10	Phys. Uncomfortable	0.05
Inquisitive	-0.17	Aggressive	-0.03	Calm	-0.09
Vigorous	-0.18	Defensive	-0.06	Sociable	-0.11
Bright	-0.19	Fearful	-0.08	Frustrated	-0.11
Frustrated	-0.19	Wary	-0.08	Tense	-0.12
Assertive	-0.25	Tense	-0.18	Wary	-0.12
Agitated	-0.25	Agitated	-0.20	Aggressive	-0.17
Fearful	-0.26	Frustrated	-0.21	Relaxed	-0.18
Active	-0.27	Phys. Uncomfortable	-0.32	Content	-0.24
Alert	-0.29	Listless	-0.35	Agitated	-0.24
Wary	**-0.29**	**Apathetic**	**-0.36**	**Fearful**	**-0.31**
Tense	**-0.31**	**Subdued**	**-0.38**	**Defensive**	**-0.46**

The distribution of ewe scores for each observer over PC1 and PC2 is shown in Figure 5.1. The observers showed good agreement in ranking sheep on the main components: PC1 Arousal: W=0.77, PC2 General Mood: W=0.71, and moderate agreement PC3 Defensiveness: W=0.54. All correlations were significant at $P<0.001$ (n=95 for all). Thresholds for categorisation as 'good' and 'moderate' are in accordance with Martin and Bateson (1993).

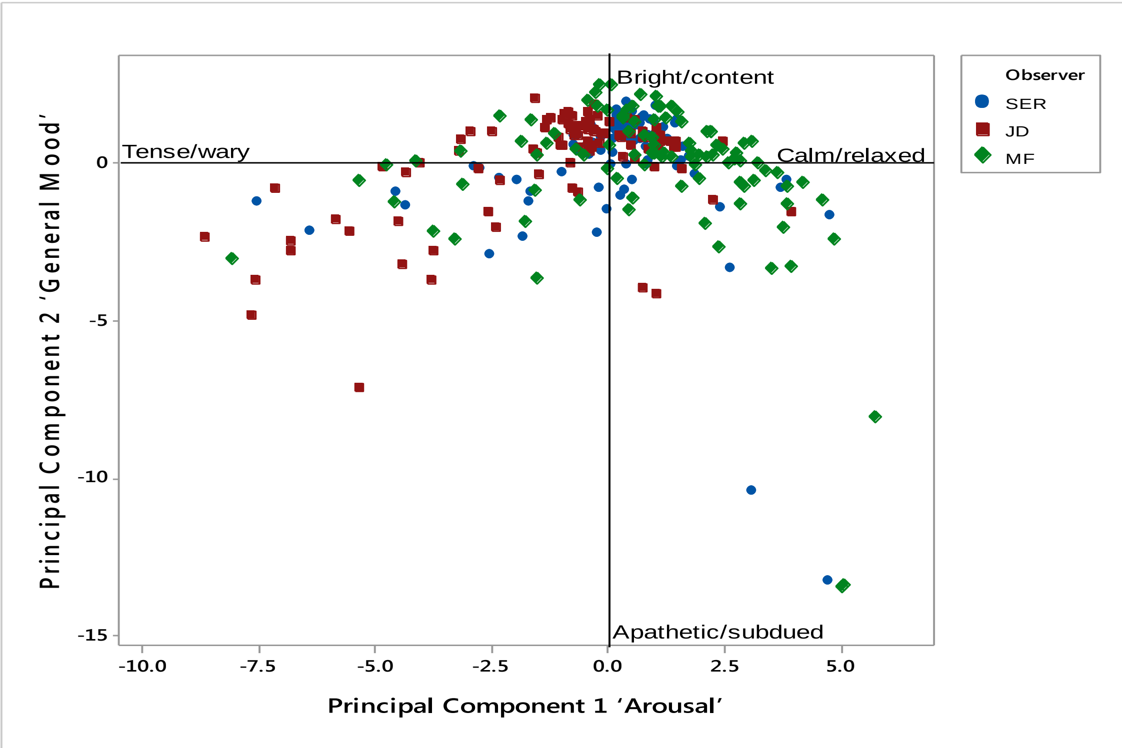


Figure 5.1 Study 1: Score plot showing the distribution of the three observers' QBA scores along PC1 and PC2, for 48 individual sheep at two on-farm visits

Significant effects of observer were found on PC1 and PC3 scores, but not PC2 scores. On the first component, observer JD had lower mean value of scores than observers SER and MF (JD mean=-1.115, SER mean=0.26, MF mean=0.914, SE=0.18, d.f.=2, P=0.03), implying that she consistently judged sheep to be in a higher Arousal state than the other two observers. On PC3, observer MF had a higher mean score than JD and SER (MF mean= 0.85, SER mean = -0.45, JD mean

= -0.45. d.f.=2, P=0.04), indicating she consistently judged sheep to be more inquisitive and vigorous than the other two observers.

A significant effect of visit was found on PC2 (W=5.71, D.F.=1, P=0.02) as the General Mood state of ewes was higher on the second visit, which occurred post-weaning, compared to the first visit when lambs were present (visit 1 mean=-0.33, visit 2 mean=0.3, S.E.=0.27). No significant effect of visit was found for either PC1 or PC3 (P=0.16 and P=0.14).

5.3.2 Study 2 - Longitudinal variation

Four PCs with Eigenvalues higher than 1.5 were identified which together explained 60.6% of the variation between sheep (Table 5.3). PC1 explained 22.6% of the variation and ranged from positive emotional states (content, bright, vigorous, relaxed) to negative states (apathetic, subdued, listless, tense) and can be summarised as ‘General Mood’. PC2, explaining 17.3% of the variation, ranged from low energy states (relaxed, calm, content, listless) to high energy states (tense, agitated, frustrated, defensive) and relates to ‘Arousal’. PC3 explained 11.8% of the variation and had the terms listless and apathetic loading strongly at one end, and wary and fearful at the other. Again, PC3 explains a relatively small percentage of variation and therefore could be considered as a sub-component. PC3 appears to indicate the sheep’s tendency to express listlessness and apathy, and as such can be termed `Listlessness´. PC4 explained 8.9% of the variation and had the terms aggressive and defensive at one end, and wary and fearful at the other. PC4 could thus be considered a sub-component representing the sheep’s “fight or flight” response to a stressor and will be referred to by as `Aggressiveness´.

Table 5. 3 PCA loadings from longitudinal data collected by one observer over 13 visits. Bold type indicates terms are highest, or lowest loading on the principle component

PCA Loadings							
PC1 "General Mood" (22.6% variation explained)		**PC2 "Arousal"** (17.3% variation explained)		**PC3 "Listlessness"** (11.8% variation explained)		**PC4 "Aggressiveness"** (8.9% variation explained)	
Content	**0.37**	**Relaxed**	**0.17**	**Wary**	**0.20**	**Aggressive**	**0.48**
Bright	**0.36**	**Calm**	**0.12**	**Fearful**	**0.11**	**Defensive**	**0.44**
Vigorous	0.35	Content	0.12	Tense	0.11	Frustrated	0.17
Relaxed	0.33	Listless	0.05	Alert	0.05	Assertive	0.13
Active	0.29	Apathetic	0.04	Agitated	0.03	Relaxed	0.03
Calm	0.26	Subdued	0.02	Bright	-0.03	Agitated	0.02
Assertive	0.20	Vigorous	-0.13	Inquisitive	-0.04	Content	0.02
Alert	0.18	Phys. Uncomfortable.	-0.15	Sociable	-0.06	Active	-0.02
Inquisitive	0.16	Bright	-0.16	Vigorous	-0.11	Phys. Uncomfortable.	-0.04
Sociable	0.09	Active	-0.17	Frustrated	-0.14	Vigorous	-0.07
Aggressive	-0.04	Sociable	-0.19	Defensive	-0.14	Bright	-0.11
Defensive	-0.05	Inquisitive	-0.23	Aggressive	-0.14	Sociable	-0.13
Agitated	-0.05	Fearful	-0.25	Active	-0.15	Inquisitive	-0.15
Fearful	-0.06	Assertive	-0.25	Assertive	-0.16	Listless	-0.17
Wary	-0.07	Aggressive	-0.27	Content	-0.20	Apathetic	-0.19
Frustrated	-0.14	Wary	-0.27	Calm	-0.22	Tense	-0.19
Phys. Uncomfortable.	-0.15	Alert	-0.28	Relaxed	-0.25	Alert	-0.19
Tense	-0.17	Defensive	-0.30	Phys. Uncomfortable.	-0.36	Calm	-0.21
Listless	-0.21	Frustrated	-0.31	Subdued	-0.40	Subdued	-0.21
Subdued	**-0.25**	**Agitated**	**-0.32**	**Apathetic**	**-0.43**	**Fearful**	**-0.29**
Apathetic	**-0.25**	**Tense**	**-0.35**	**Listless**	**-0.43**	**Wary**	**-0.39**

There were significant effects of visit number on the main components of General Mood and Arousal, and on the sub-component of Listlessness, but not Aggressiveness (P=0.862). Over time ewes showed an increase in General Mood scores (PC1: F=150.6, d.f.=1, P<0.001,) and increased Arousal scores (PC2: F=4.15, d.f.=1, P=0.042). However, on PC3 ewe the ewes became significantly more Listless over time (PC3: F=19.75, d.f=1, P<0.001). The change in ewe scores for all four PCs over the 13 visits is shown in Fig 5.2.

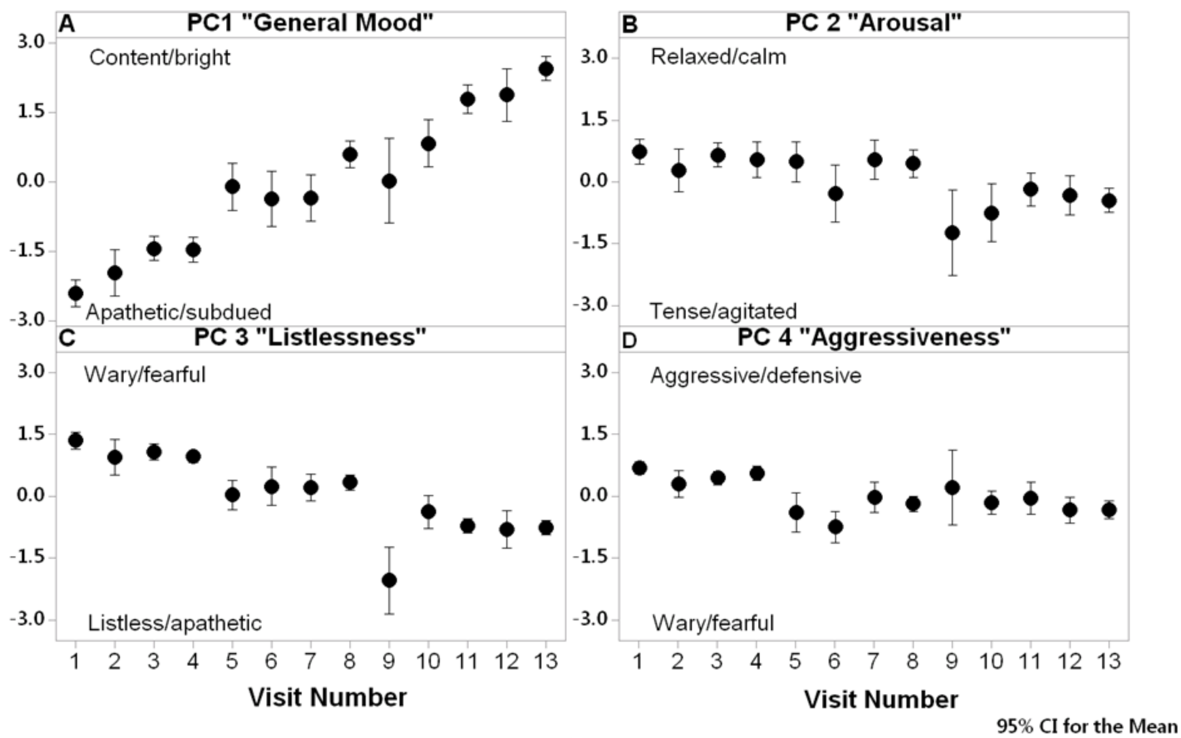


Figure 5.2 Study 2: Interval plots displaying PCA scores for 48 ewes observed on 13 occasions over a six month time period

Significant effects of ambient temperature were found on PCs 2, 3, and 4. The ewes expressed significantly higher Arousal (PC2) levels when temperatures were categorised as Warm in comparison with Cold and Cool weather conditions (mean score for Cold=0.22, Cool=0.15, Warm =-0.54, W=12.09, d.f.=2, P=0.003). On PC3, a rise in temperature was associated with sheep becoming more Listless (mean score for Cold =0.60, Cool=0.02, Warm=-0.74, W=35.01, d.f.=2, P=<0.001,), while on PC4, Cold and Warm temperatures were associated with increased levels of Aggressiveness in comparison to when the weather was Cool (mean score for Cold=0.52, Cool=-0.34, Warm =-0.01, W=22.01, d.f=2, P<0.001). Although there was no significant impact of temperature on General Mood (PC1), as temperatures rose above 19°C there was a tendency for ewe General Mood scores to decrease (P=0.084). The warmer weather occurred during visits 9, 10 and 11, and the marked effect on ewe scores at visit nine, and the increased variation in individual animal scores at this visit, can be seen in Figure 5.2.

There were significant differences in the ewe's PC scores on all four components across the three reproductive states of pre-lambing, post-lambing and post-weaning. The ewes' General Mood scores significantly increased in the post-lambing period compared to pre-lambing observations, and significantly increased again post-weaning (F=177.15, P<0.001, Figure 5.3a). Ewes experienced significantly higher Arousal in post-lambing and post-weaning conditions compared to pre-lambing, but there was no difference between post-lambing and post-weaning (F=10.68, P<0.001, Figure 5.3b). Ewes were significantly more Listless post-lambing than pre-lambing (PC3: F=75.88, P<0.001), and this increased somewhat further post-weaning. Ewes also showed a reduction in Aggressiveness (PC4) between pre-lambing and post-lambing/post-weaning, but there were no significant differences between post-lambing and post-weaning on this PC (PC4 F=17.92, P<0.001, Figure 5.3d).

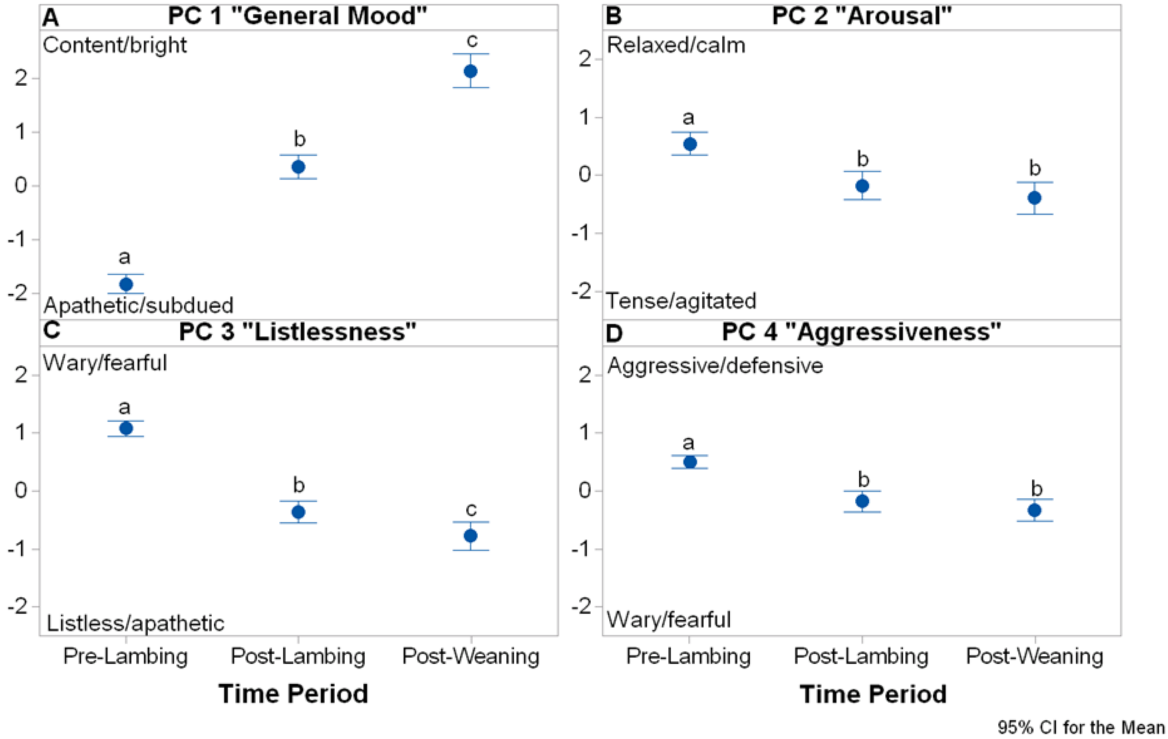


Figure 5.3 Study 2: Mean Principle Component Scores for 48 individual ewes in the three time periods of pre- and post-lambing and post-weaning. Different lower-case letters indicate a significant difference between time periods

The time of the observations (a.m or p.m) had no significant effect on the ewe's scores (P>0.21 on all PCs), and there were no significant effects of lamb sex on the ewe's QBA scores at any time point (P>0.38 on all PCs).

5.4 Discussion

In the two studies reported in this paper, we found ewe demeanour to be predominantly expressed along two main components of General Mood (spanning negative to positive emotional states), and Arousal (low to high energy states), with sub-components of Defensiveness, Listlessness and Aggressiveness. In study one, three trained observers showed a good level of agreement in scoring individual sheep on these expressive dimensions. In study two there was a gradual and systematic change in both General Mood and Arousal over a six month period, which was inferred to reflect changes in season (winter to summer) and ewe reproductive state (pregnant to lactating to weaning).

5.4.1 Inter-observer reliability

The high levels of correlation between sheep scores on PCs 1, 2 and 3 for the three observers found in the first study indicate that the qualitative ranking of sheep on expressive dimensions was robust, a result which corresponds with other studies using video footage of sheep (Clare Phythian et al., 2013) or reporting on-farm welfare assessments of other species (Wemelsfelder, et al., 2009; Wemelsfelder & Millard, 2009; Andreasen et al., 2013). However, the observer effects we found related to PCs 1 and 3 indicate that the three observers differed in how they quantified their terms, causing their mean sheep values on PC1 and 3 to differ significantly. Thus the relative patterns of expression identified by QBA seem robust across observers, but these patterns' exact level of quantification appears more variable across observers.

However such observer effects are not unusual in the use of visual analogue scales (Torrance, et al 2001). The visual analogue scale used for QBA only has two given anchors at either end of the scale, and so opens up the rest of the scale to be interpreted and used by individuals in their own way. Even slight discrepancies in the way in which observers quantify their assessments, if they persistently occur, could lead to significant observer effects. For example, the phenomenon of end-aversion bias is well known to affect scores and measurements scored on continuous scales where there are few anchors (Torrance et al., 2001; Torrance, Furlong, & Feeny, 2002; Parkin & Devlin, 2006). Some people consistently score at either ends of the scale while others appear to prefer using a smaller portion of the scale confining themselves to the middle (Torrance et al., 2001; Torrance et al., 2002; Hasson & Arnetz, 2005; Parkin & Devlin, 2006). Although attaching multiple well-defined anchors on to a scale may help to diminish the effects of end-aversion bias, they may lessen the sensitivity and subtle distinctions which are a main advantage of using visual analogue scales and in fact ill-defined anchors may make the problem worse (Torrance et al., 2001; Torrance et al., 2002). A better solution is to provide observers with detailed instructions and training in the use of visual analogue scales prior to data collection (Fleming et al., 2015). This may be achieved by dedicating time to comparing and adjusting how individual observers use the visual analogue scale to create scoring patterns for the different QBA descriptors. In combination with discussion of the meaning of QBA terms, this will help to better align observers' understanding of, and agreement in, how to score QBA terms. In any case, the effects of environmental and internal physiological factors on QBA scores reported in Study 2 of this paper were all based on assessments made by one observer, and were therefore not vulnerable to discrepancies with other observers' styles of quantification.

5.4.2 Terms and applicability to extensively managed sheep

Recent studies have shown that sheep have the potential to experience a wide range of emotions including fear and boredom (Boissy et al 2007; Boissy et al 2011; Veissier et al 2009). The qualitative terms developed for the QBA fixed term list in this paper were similar to the key descriptors of sheep expression developed in previous studies (Wemelsfelder & Farish, 2004 ; Phythian et al., 2013), but some novel terms were added by the expert focus groups. The resulting list of sheep QBA terms was found to effectively discriminate between expressive demeanours observed in this study, and so it appears they are applicable to, but not limited to, extensively managed sheep.

The two studies identified similar expressive dimensions of Mood and Arousal, which corresponds with QBA outcomes of previous longitudinal studies in sheep (Phythian et al., 2013; 2015). Mendl et al (2010) suggest that animal emotion, conceived as 'core affect', can be described in terms of valence and arousal dimensions. This concept provides a framework for integrating discrete individual emotions (e.g. fear, anger, contentment, excitement) into an overall dimensional picture of general mood. Our findings suggest that in sheep, these individual emotions can be observed and discriminated by trained observers, and integrated by multivariate analysis to provide a core affect describing sheep experience. At first glance, it appears that there are contradictions in the results relating to PC 1 and 3 in the longitudinal study: as sheep are assessed to become less apathetic on the main PC (General Mood), they are simultaneously assessed as becoming more apathetic on sub component PC3 (Listlessness). However, these results are not necessarily incompatible, but reflect discrete segments of the sheep's demeanour identified through multivariate analysis. The third and fourth PCs are sub-components explaining smaller proportions of variation in the ewes' demeanour, within the larger expressive patterns of general mood and arousal being described by the two main components. As a general trend, the ewes can show a gradual increase in mood, but within that trend they can exhibit minor fluctuations in mood caused by more incidental factors.

The gradual and systematic increase in General Mood and Arousal states of ewes across the six month observation period suggests that ewe affective state is being influenced by both internal (reproductive state) and external (environmental) factors. The time period from late gestation to post-weaning involves marked shifts in ewe physiology including endocrine activity, metabolism and behaviour, due to alterations in their reproductive state. These internal physiological shifts coincide with external environmental changes that occur with the progression of winter to summer, which affect the quantity and quality of pasture available to sheep, climatic variables, and disease challenges, such as the changing gastrointestinal parasite populations to which the sheep are exposed. It was not possible in the present studies to separate the differential effects that reproductive state and environmental variables may have on ewe affective states, and therefore the potential role of each of these factors are discussed below.

5.4.3 Hormonal and social changes

During the course of these studies ewes were observed in mid gestation, during lactation, and after the male lambs were weaned. These reproductive changes are associated with marked alterations in physiology, social and emotional responses. Sheep are social animals and isolation and separation from their social group is known to cause distress (Pendu et al. 1995; Degrabriele & Fell 2001). However, prior to parturition ewes become more individualistic and separate themselves from the social group (Poindron et al., 1994). Many studies have shown that pregnant ewes show decreased anxiety-related behaviours compared to non-pregnant sheep and are less responsive in stress tests (Bell et al. 1991; Poindron et al. 1997; Viérin & Bouissou 2001; Vierin et al. 2002). An increase in circulating progesterone, which occurs in late gestation, may contribute to the observed reduction in stress reactivity as it is known to mediate emotional activity in several mammalian species, including sheep (Bell et al., 1991; Viérin and Bouissou, 2001). The ewes in the current studies had lower Arousal states pre-lambing than in any other reproductive stages, which may reflect these changes.

Ewes develop a selective ewe-lamb bond at lambing, which is established quickly after birth (Orgeur et al., 1998), and it has been suggested that this social bond with the lamb takes the place of the ewe's peers during the lactation period (Hinch et al., 1987). The close bond between lamb and ewe is particularly strong in the Scottish Blackface breed and may be indicative of a weaker attachment to the flock (Dwyer and Lawrence, 2000), at least during the summer months, whereas other breeds (e.g. Merinos) tend to be more attached to the rest of the flock than their own (McBride et al., 1967). Oxytocin release in the brain plays an important role in the onset of maternal behaviour (Dwyer et al., 2004), and is also released in the ewe upon recognition of the sight and sound of her lamb (Fuchs et al., 1987), and during suckling (Keverne and Kendrick, 1991). Many studies have shown that oxytocin is involved in the regulation of social behaviours and anti-stress effects in mammals including sheep, as well as in the mechanisms underlying the development and maintenance of social attachments (Coulon et al., 2013; Gordon et al., 2011). Phythian et al (2015) also reported that ewes were more relaxed post-lambing compared to any other time of year, and suggested that the presence and suckling behaviours of the ewes' young lambs (less than 12 weeks of age) had a 'relaxing' effect on the ewes due to the physiological changes associated with maternal bonding behaviours.

As the ewe's social network and bonds transform with the weakening relationships with peers and becoming focused on the lamb, there are alterations in her anti-predator behaviour. Studies investigating ewe behaviour in the post-lambing time period have found increased levels of vigilance in lactating ewes compare to non-lactating ewes at the same time point, and this increase in vigilance leads to an increased protection of her lamb (Pickup and Dwyer, 2011). Livestock with good maternal care have been shown to be more vigilant and defend their young by displaying threatening and defensive behaviour towards perceived predators and threats (Pickup & Dwyer 2011; Turner & Lawrence 2007). These changes may explain the increased Arousal and higher energy states that the present study reports in lactating ewes compared to the same ewes during late pregnancy.

Vigilance levels vary throughout the lactation period (Pickup and Dwyer, 2011), typically peaking at one month post-lambing and then steadily decreasing coinciding with the reduction in suckling behaviour in lambs. The frequency and duration of lambs suckling naturally decreases after four to five weeks as the ewe begins to limit the sucking activities of the lamb (Gordon and Siegmann, 1991). Progressive natural weaning has very little apparent negative consequences on social groups of ewes and lambs (Orgeur et al., 1998) in comparison to abrupt artificial weaning. In the current study, the male lambs were abruptly weaned at approximately four months of age, which is earlier than natural weaning would occur (approximately 6-9 months Arnold & Pahl, 1974). At this age, however, the ewe-lamb bond is beginning to weaken and lambs are no longer relying on the ewe for their nutrition (Arnold and Pahl, 1974; C. Dwyer, 2008). Ewe-lamb bonds however do not cease completely at the end of suckling as both partners can remain associated for several weeks, months or years if allowed (Arnold and Pahl, 1974; Hinch et al., 1990; Hunter and Milner, 1963). Weaning at three months has been shown to lead to distress in both the lambs and ewes even if performed gradually (Cockram et al., 1993; Orgeur et al., 1998). Ewes react to abrupt separation at three months by showing behavioural and postural alterations indicative of stress: increased vocalisation, raised head, erect ears and a decrease in lying and sleeping behaviours (P Orgeur et al. 1998; Cockram et al. 1993), although these behavioural responses disappeared after two days. The negative consequences of weaning on the ewes thus appear to be a short term effect. In the present study no observations were made within the first two days post-weaning, and it is therefore likely that by the time our observations were made, this effect had subsided and the changes observed in ewe affective state were not directly related to the removal of the lambs. This proposal is supported by the fact that there were no statistical differences in the QBA scores of the ewes which underwent abrupt weaning of their male lambs, and those remaining with their female lambs. Following the removal of the lamb, the ewe will need to re-establish her previous social relationship with her flock members, which may be associated with the improvement in General Mood following weaning observed in both parts of this study.

5.4.4 Nutrition and environment

Lawrence & Wood-Gush (1988) proposed an alternative reason for the shift in ewe social groupings during the summer months. Grouping behaviour of sheep is also influenced by season (Lawrence & Wood-Gush, 1988; Dwyer & Lawrence, 1999) and in summer the marked decrease in gregariousness and considerable reduction in group size coincides with decreased behavioural and grazing synchrony, potentially due to greater grazing dispersal with greater choice of plant species on which to forage.

The reproductive season in sheep is timed so that lamb birth and ewe lactation coincide with the most favourable environmental conditions in the spring. Gestation and the following lactation are energetically expensive, with the full cycle of conception, gestation and lactation considered to be one of the most energetically expensive activities that a female mammal can undertake (Wade and Schneider, 1992). In order for the ewe to produce a sufficient volume of milk for her lambs, while maintaining a healthy body condition, she needs to be able to maximise her nutrient intake. The protein and energy requirements of lactating ewes are double those of non-lactating ewes (Ruckstuhl and Festa-Bianchet, 1998). There is an increase in voluntary food intake during lactation, which goes some way to meeting the requirements (Forbes, 1970). The stages of lactation influence food intake in sheep (Molina et al., 2001), increasing up to approximately the fourth week then declining by the seventh week of lactation along with milk yield (Hadjipieris and Holmes, 1966; Molina et al., 2001). The extent of this increased metabolic demand depends partly upon parity, litter size and milking or suckling frequency (González-García et al., 2015). A need to spend more time grazing to meet this increased requirement for nutrients may also explain the increased Arousal of ewes during lactation.

The increase in demand for energy and grazing in the summer corresponds with an increase in the number of gastrointestinal parasites in the environment (Uriarte et al., 2003). During lactation, ewes that would normally have good acquired immunity to parasitic challenge, are still experiencing a periparturient relaxation of immunity which can cause a temporary increase in gastrointestinal parasite burdens (Barger, 1993; Houdijk et al., 2000). Higher worm burdens can cause altered behavioural responses in choices of grazing area or even anorexia (Kyriazakis et al. 1998; Hutchings et al. 1999; Smith et al. 2006). In the current study, the increase in grass quality may have also led to the improvement in General Mood as the summer progressed, however the competition between ewes for resources may have also led to the animals becoming more tense and agitated on the Arousal dimension. In addition, the tendency for ewes to become more Listless (on PC3) as summer progressed, could perhaps be seen as indicative of a failure to meet their nutritional needs.

5.4.5 Temperature effects

In extensive systems animals will spend the vast majority, if not all, of their lives outside and have to cope with extremes of weather and temperature and fluctuations in ambient temperature. These fluctuations can have direct and indirect effects on the animals' wellbeing (Silanikove, 2000). In this study, the Cold weather condition coincided with the ewes' lowest mood, and despite their fleeces providing insulation and allowing the maintenance of homeostasis (Piccione et al., 2002) the ewes' lowered General Mood and increased Aggressiveness during the Cold conditions suggests they still found the hard winter unpleasant and challenging. As the summer progressed the ambient temperature increased from -2°C with snow (Cold condition) in the pre-lambing period to more clement weather in the post-lambing time period. This rise in temperature from Cold to Cool in the spring coincided with re-growth of grass and continued to the optimal grass growth conditions (20-25°C) in the Warm temperature range (Parsons and Williams, 2010) . The increased availability of herbage quantity and quality may have prompted the gradual improvements observed in the ewes' General Mood throughout the summer.

The Warm weather condition also provided the sheep with a challenge. Ruminants can experience heat stress at elevated temperatures, particularly if this is associated with high humidity (Marai et al., 2007; Silanikove, 2000). The effects of exposure to elevated ambient temperature in sheep has been well documented with alterations to the animals' biological functions: reduction in feed intake (Nardon et al., 1991) and decrease in feed efficiency and utilisation (Marai et al., 2007). Numerous studies have reported the wide variety of negative effects high temperatures have on humans including a dislike of other people (Griffit and Veitch, 1971), increased aggression (Anderson, 1989; Vrij et al., 1994) and negative affective state (Griffit & Veitch 1971; Vrij et al. 1994). Although relatively little has been reported regarding the effect of heat stress on behaviour of animals in extensive systems, Sherwin & Johnson (1987) investigated shade seeking and dominance and aggression levels in Merino sheep exposed to high temperatures. They found shade use increased in hot weather, and a significant positive correlation between the initiation of head butting and the time the animals spent in the shade – a relationship which became even stronger as temperatures increased further. Similarly we suggest that the short periods of hot weather seen in the current study during visits 9, 10 and 11, caused the remarkable shifts in the ewes' scores on PC2 and PC3. On PC2 it is apparent that the ewes became more tense/agitated during these visits and this is similar to the behavioural changes reported by Sherwin and Johnson, and in the human literature (Griffith & Veitch 1971; Vrij et al 1994). On PC3 the ewes were shown to be more Listless in particular on visit nine, in which temperatures peaked at 25°C. In addition it can be seen the variation between individual ewes increases on all four PCs during the Warm weather, in particular visit nine, not seen in the adjacent Cool or Cold temperature ranges. This implies that when the sheep were faced with a physiological challenge they responded in different ways. For some animals the increase in temperatures may have led to some animals becoming more tense/agitated (PC2) and show heightened Aggressiveness (PC4), while others, perhaps those which were recipients of aggression, became more Listless (PC3) as they struggled in the heat. In our study, the three visits during which temperatures were Warm occurred post-lambing, and during this time period the sheep showed an overall improvement in General Mood (PC1). However, it has been suggested that some weather conditions can have a negative impact on an animal's welfare state, overshadowing an otherwise positive situation (Mellor, 2015), and our findings support this hypothesis.

5.5 Conclusion

The evidence provided by this study supports that QBA can be applied reliably in field conditions, and is capable of recognising meaningful variation in the expression of extensively managed individual sheep over a six month time period. Sheep emotional expression was found to vary across two main dimensions of General Mood and Arousal, showing significant systematic changes throughout the late-gestation to post-weaning phase of the reproductive cycle. In addition, there were shifts on smaller sub-components of 'Defensiveness' and 'Listlessness'. These fluctuations may relate to external (environmental) and internal (physiological) changes, and indicate alterations in the ewes' welfare state. The good observer agreement, and the ability of QBA to pick up on subtle and meaningful changes in sheep expression, adds further support for its use in on-farm welfare assessments. Such precision, however, could potentially mean that QBA is too sensitive to temporary fluctuations in the animals' welfare, which may not be representative of a farm's overall welfare level in the longer term. Thus, further research is required to investigate at which point of the reproductive cycle, and how frequently, QBA observations can best be applied in order for it to serve as a robust and useful on-farm welfare indicator.

Chapter 6 Evaluating welfare indicators applied during the Assessment in the Field

Abstract

In the previous chapter the affective state of ewes, as measured by QBA, was shown to fluctuate throughout a six month time period spanning late pregnancy to post-weaning. The relationships between QBA and the other indicators selected for use during the Assessments in the Field was unknown. The indicators applied during the Assessment in the Field were intended to measure aspects of social behaviour (distance to nearest neighbour, group size, composition, synchronicity and vocalisation), general fear (distance to which a human could approach, flight distance and recovery following a surprise test) and physical health and comfort (lameness, faecal soiling and coat cleanliness). The aim of the work presented in this chapter was to assess the validity and feasibility of these welfare indicators when applied to ewes during the Assessment in the Field. One of the main aims of this work was to understand the relationships between the selected welfare indicators and the four QBA principle components which were identified during the longitudinal study reported in Chapter 5. Using scan sampling techniques, data were collected on social groupings, behaviour and physical indicators of health and comfort. Five assessments were conducted in total. Three when the lambs were at foot (post-lambing) and two post-weaning. Data sets were collected on 44 individual ewes. Fluctuations in the ewes' affective state was found to be related to the size and composition of their social group, and the perceived levels of predation threat. Ewes with a lower General Mood had larger distances between themselves and their nearest neighbour than ewes in higher mood states. Higher numbers of vocalisations during the two minute observation was observed in groups in which the focal ewes had low General Mood, increased Arousal state, higher Listlessness and heightened Aggressiveness. Ewes in a higher Arousal state tended to flee earlier from an approaching human than those experiencing low levels of Arousal. The physical

health and comfort indicators did not appear to have any significant association with the affective state of the animals, perhaps due to the low variability in the physical health data. In conclusion, these findings support the validity of the QBA approach. Group size and composition, and flight distance following a surprise event, were also affected by the time period in which the Assessments in the Field were conducted implying that the removal of the male lambs has a significant impact upon ewe social behaviour. The indicators applied during the Assessment in the Field proved to be feasible supporting their use as welfare indicators for use with extensively managed sheep.

6.1 Introduction

Sheep living in extensive systems are directly affected by alterations in their physical environment such as forage availability and weather. Seasonal environmental patterns typically coincide with alterations in their reproductive state i.e. summer typically coincides with their period of lactation. As a result, the outcomes of welfare assessments performed on these animals are likely to fluctuate across time. The indicators applied during the Assessment in the Field were intended to measure aspects of social behaviour (distance to nearest neighbour, group size and composition and vocalisation), general fear (distance to which a human could approach, flight distance and recovery following a surprise test) and physical health and comfort (lameness, faecal soiling and coat cleanliness). The first aim of this chapter was to investigate whether the results of these welfare indicators applied to ewes observed in a typical extensive environment fluctuated over post-lambing and post-weaning time periods.

As stated in previous chapters, a major concern in the development of on-farm welfare assessments is the challenge of interpreting fluctuations in welfare indicator results over time (Phythian et al 2015). In order to fully comprehend these fluctuations and be certain that they are an accurate reflection of welfare, one must ensure that the indicators used during the assessments are valid. The main aim of this work was to assess the convergent validity and feasibility of the welfare indicators applied during the Assessment in the Field.

The relationships between the above indicators and the four QBA principle components of General Mood, Arousal, Listlessness and Aggressiveness identified during the longitudinal study reported in **Chapter 5** were of particular interest. In **Chapter 5**, it was shown that there were systematic meaningful changes in ewe affective state across a six month time period spanning pre-lambing to post-weaning. For example, as time progressed, gradual and systematic increases in ewe General Mood and Arousal were observed. The impact of potential alterations in ewe social and anti-predator behaviours during these time points were discussed, although without data on these aspects further conclusions could not be drawn. In the current chapter, the data collected during the Assessments in the Field were used to further explore these speculations. The third aim of this work was to understand the relationships between the selected welfare indicators and the four QBA principle components which were identified during the longitudinal study reported in **Chapter 5**.

6.2 Methodology

6.2.1 Animal Selection and Data collection

The data presented in this chapter includes five QBA observations, visits 9 – 13, described in Chapter 5. The data were collected between the 13th of July and the 31st of August 2013 at intervals of approximately two weeks (see Figure 3.2 and Table 3.13 for further information). The shortest period between assessments was 11 days and the longest 15 days. Assessments occurred between 0800 and 1800. As in Chapter 5, removal of the male lambs occurred between visits 11 and 12 and therefore visits 9 to 11 are "post-lambing" and visits 12 and 13 are referred to as "post-weaning". Complete data sets were collected for 44 ewes during post-lambing, and 46 post-weaning. As in the previous chapter the discrepancy in these numbers was due to the ewes' ability to freely range across the hills and avoid observation. The ewes produced 44 lambs in total, 28 of them were male and therefore removed from their dams in August. The remaining 16 lambs were females who stayed with the flock until the end of this experiment. All data reported in this chapter were collected by one observer (SER).

Following the QBA observations, the animals were assessed using the Assessment in the Field protocol. The methodology of this protocol is described in detail in Chapter 3. The Assessment in the Field measures consisted of measurements designed to allow the assessment of the ewes' social behaviour, response to potentially surprising and fear inducing stimuli and their physical health and comfort. The indicators applied during the Assessment in the Field are listed in Table 3.12 in Chapter 3. Although group size is not a specific welfare indicator included in the evaluation, gathering this information during the Assessment in the Field allowed further interpretation of the data set. While the ewes remained undisturbed, data were collected on: nearest neighbour distance (m), group size and composition, behavioural and postural synchrony (see Table 3.14 for ethogram), and number of vocalisations by ewes and lambs during a two minute observation. The assessor then performed the human approach test and visually estimated the distance (m) to which the human could approach before the ewe retreated. The ewes' response to, and recovery from, the surprise test was performed and recorded as described in Chapter 3. The movement of the ewe during these tests allowed the assessor to score lameness, coat cleanliness and faecal soiling score using the scales also defined in Chapter 3. With the exception of the group size, composition and vocalisation, data was collected at an individual sheep level.

6.2.2 Data preparation

Following the analysis reported in Chapter 5, the QBA Principal Components of General Mood and Arousal, and the sub-effects of Listlessness and Aggressiveness were included in the analysis reported here. In order to improve readability, the directions of two of the four PCs, Arousal and Listlessness, have been reversed by performing a simple calculation in which scores were multiplied by -1. This has not affected the absolute value of the results, but rather the directions of the components are more intuitive. This aids in the description of the relationship between these PCs and the other indicators. The direction of PC loadings can be seen in Table 6.1.

Table 6.1 Highest and lowest loading terms for each of the QBA PCs following reversal calculations

Principal Component	General Mood	Arousal	Listlessness	Aggressiveness
Highest loading terms	Content/bright	Tense/agitated	Listless/apathetic	Aggressive/defensive
Lowest loading terms	Apathetic/subdued	Relaxed/calm	Wary/fearful	Wary/fearful

One ewe was removed from the post-weaning time period. This ewe contracted acute mastitis which resulted in her death within 24 hours of her last QBA and Assessment in the Field observation (assessment 12). She was therefore considered as an outlier and not representative of other animals in the data set for this time period, but her data are included during other assessments prior to illness.

The behaviour and posture of the ewes were recorded during scan samples and synchrony calculated, however in the analysis presented in this chapter only data on the postural synchrony is reported. Postural synchrony, whether ewes were lying or standing, was calculated for groups (as per Rook and Penning, 1991, described in Chapter 3). Whether or not the individual ewe was synchronous with the majority of the group was determined using the criteria described by (Ruckstuhl and Neuhaus (2001) (Chapter 3). The data obtained during the scan samples were also used to assess the percentage of animals in each social group performing vigilance behaviours.

6.2.2.1 Data analysis

Repeated measures REML was used, in GenStat, to assess the variation in indicator measures between the post-lambing and post-weaning time points for continuous data (Qualitative Behavioural Assessment scores, nearest neighbour distances, approach distance, flight distance, recovery time, number of ewes and lambs in social group and number of vocalisations). Indicators were included as the response variates, and individual focal ewe identity was selected as the "subject" and time points of assessments included. For the physical health and comfort indicators, for which categorical data were obtained, Ordinal Logistic Regressions were performed in Minitab. The indicator data were entered as the Response, ewe identity was included as a random effect in the model, and the time periods of post-lambing and post-weaning were include as categorical predictors. A Generalised Linear Model was performed in Genstat to assess the variation in individual focal ewe synchrony (binary trait) between post-lambing and post-weaning. Binomial distribution was selected with a Logit Link Function.

The relationships between the indicators were assessed. The results from the QBA, social behaviour of the ewe and their response to a potentially fear inducing stimulus (all continuous data) were included in a Pearson's correlation using Minitab. Firstly, correlations were performed for all five visits together and then subsequently the data set was filtered into post-lambing and post-weaning only. Spearman's correlations were used to analyse the relationships between physical health and comfort indicators (categorical data).

Mann-Whitney tests were performed in Minitab to assess whether there were significant differences in the QBA scores of ewes which were posturally synchronised with the flock compared to those which were asynchronous.

In order to further assess the interactions between the behaviour of the ewe and physical health on the QBA scores, a Linear Mixed Model was performed in Genstat. The General Mood, Arousal, Listlessness and Aggression scores were fitted individually as response variates. Initially univariate analysis was carried out where one variable was included in the model with animal ID and time period (post-lambing or post-weaning) as a random effect. Variables which provided a P value of below 0.2 were retained and included together in the final multivariate models. Four final models were built, General Mood, Arousal, Listlessness and Aggressiveness, in which the random effects remained the same as reported above.

6.3 Results

6.3.1 Effect of time period on indicator scores

As shown in Table 6.2 the size of social groups were found to significantly decrease following weaning with fewer ewes, and lambs congregating together. The number of vocalisations by ewes and lambs during the two minute observations also significantly declined between periods. No significant variation in nearest neighbour distance was observed during these two time periods. Following weaning the flight distance of the ewe following the surprising event significantly decreased compared to before weaning. However, no significant differences were found for the distance to which a human could approach or length of time taken for the ewe to recover from surprise test between periods. Statistically, no significant differences were observed in the proportion of ewes performing vigilance behaviour between the post-lambing and post-weaning time periods.

The ewes received significantly higher faecal soiling scores following weaning than during the post-lambing observations. The coat cleanliness and lameness of the ewes did not significantly differ between post-lambing and post-weaning time periods. Further descriptive information on the variation in indicator scores before and after weaning is shown in Table 6.2.

There were no statistical differences in the synchrony of the focal ewe with the rest of the flock before or after weaning (P=0.10).

Table 6.2 Variation in indicator means between the post-lambing and post-weaning time periods. Indicators for which there was a significant difference between time periods are in bold (SEM in parentheses)

Principle	Indicator	Post-lambing Mean (SEM)	Post-weaning Mean (SEM)	Test statistics
Good Environment	Coat cleanliness	0.51 (0.04)	0.68 (0.05)	P=0.10
Good Health	Faecal soiling	**0.24 (0.048)**	**0.59 (0.078)**	**Chi=2.44, DF=1, P<0.001**
	Lameness	0.29 (0.05)	0.24 (0.07)	P=0.369
Appropriate Behaviour	Percentage of group performing vigilance	7.36 (1.38)	5.55 (1.68)	P=0.40
	Nearest neighbour	10.73 (0.814)	11.10 (0.98)	P=0.769
	Ewe vocalisations	**0.96 (0.27)**	**0.09 (0.32)**	**Wald=4.12, DF=1, P<0.044**
	Lamb vocalisations	**1.20 (0.226)**	**0.12 (0.27)**	**Wald=9.23, DF=1, P<0.003**
	Human approach distance (m)	5.63 (0.28)	5.33 (0.31)	P=0.311
	Flight distance response (m)	**8.23 (0.36)**	**6.755 (0.4)**	**Wald=11.56, DF=1, P<0.001**
	Recovery from surprise test (s)	10.43 (0.51)	10.01 (0.6)	P=0.586
	Number of lambs in group	**6.16 (0.45)**	**2.13 (0.54)**	**Wald=32.02, DF=1, P<0.001**
	Number of ewes in group	**6.53 (0.33)**	**4.60 (0.48)**	**Wald=9.77, DF=1, P<0.001**

6.3.2 Relationships between indicators: all 5 assessments

When the results from all five Assessments in the Field were analysed together significant correlations were found between indicators assessing social behaviour and general fear as shown in Table 6.3.

Social behaviour

A positive correlation was found between the number of ewes in a group and the number of lambs ($P<0.001$). In smaller social groups made up of fewer ewes and lambs, there were significantly greater distances between the focal ewe and her nearest neighbour (both $P<0.001$). Numbers of ewes and lambs were significantly positively correlated with vocalisations (both $P<0.001$) indicating that in larger groups there were higher numbers of bleats by ewes and lambs.

General fear

Ewes which fled while the human was at a greater distance also tended to flee farther and took longer to recover from the surprising event (both $P<0.001$). Recovery time and flight distance were also significantly positively correlated with each other ($P<0.001$). These animals were also more likely to be found in groups with higher lamb vocalisations ($P=0.02$).

Ewes in groups with a high percentage of individuals engaged in vigilance behaviour during the scan samples were found to be closer to their nearest neighbours than those in groups with low vigilance rates ($P=0.011$). Significant positive relationships were seen between the percentage of ewes in a group performing vigilance behaviour and the total number of ewes in social group ($P<0.001$) and vocalisation by ewes ($P=0.005$). A higher percentage of the social group performing vigilance behaviour tended to be associated with a higher number of lambs in group and lamb vocalisations (both $P=0.097$).

Table 6.3 Correlation strength and significance for indicators applied during all five Assessments in the Field

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Ewe vocalisations	-0.087							
P-value	*0.197*							
Lamb vocalisations	-0.08	**0.55**						
P-value	*0.237*	***<0.001***						
Percentage of ewes in group performing vigilance	**-0.171**	**0.188**	0.112					
P-value	***0.011***	***0.005***	*0.097*					
Human approach distance (m)	-0.01	0.111	**0.157**	0.022				
P-value	*0.88*	*0.1*	***0.02***	*0.744*				
Flight distance response (m)	0.078	0.089	0.082	-0.052	**0.492**			
P-value	*0.246*	*0.188*	*0.228*	*0.442*	***<0.001***			

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Recovery from surprise (s)	-0.05	-0.086	0.109	-0.084	**0.346**	**0.434**		
P-value	*0.463*	*0.204*	*0.108*	*0.212*	***<0.001***	***<0.001***		
Number of ewes in group	**-0.516**	**0.264**	**0.487**	**0.285**	0.036	0.008	0.083	
P-value	***<0.001***	***<0.001***	***<0.001***	***<0.001***	*0.6*	*0.905*	*0.222*	
Number of lambs in group	**-0.251**	**0.418**	**0.37**	0.112	0.068	0.108	0.059	**0.796**
P-value	***<0.001***	***<0.001***	***<0.001***	*0.097*	*0.313*	*0.111*	*0.385*	***<0.001***

Physical comfort and health

No significant correlations were found between the physical health and comfort measures across all five assessments (table 6.4).

Table 6.4 Correlation strength and significant between the physical health and comfort measures across all five Assessments in the Field

Indicator	Coat cleanliness	Faecal soiling
Faecal soiling *P-value*	-0.006 *0.930*	
Lameness *P-value*	-0.047 *0.491*	-0.074 *0.273*

Individual ewe postural synchrony

The time taken for ewes to recover from the surprise test was found to be longer in ewes which were previously synchronised with the flock (Synchronised mean recovery time=10.66, Non-synchronised mean recovery= 8.78, Wald=4.07, DF=1, P=0.038). No other indicators were found to be significantly related to individual synchrony (P>0.05).

6.3.3 Relationships between indicators: post-lambing time period

When the data set was considered for post-lambing only, many of the significant correlations remained and are shown in Table 6.6.

Social behaviours

As before, when analysing the post-lambing data separately, ewes with larger nearest neighbour distances were found in smaller groups (P<0.001). Groups with more ewes typically contained a higher number of lambs (P<0.001) and made higher numbers of vocalisations during the observation period (P<0.001).

General fear

In groups with higher vigilance there were more ewes and lambs (both $P<0.001$), and the ewes were more vocal ($P=0.004$). The ewes which fled from the human at a greater distance also fled farther and took longer to recover from the surprising event (all $P<0.001$).

Table 6.5 Correlations between Assessments in the Field behavioural indicators during the post-lambing period (assessments 9-11)

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Ewe vocalisations	-0.098							
P-value	0.268							
Lamb vocalisations	-0.115	**0.543**						
P-value	0.191	**<0.001**						
Percent of ewes in group performing vigilance	-0.154	**0.247**	**0.144**					
P-value	0.079	**0.004**	**<0.001**					
Human approach distance (m)	-0.061	0.131	**0.189**	0.106				
P-value	0.49	0.137	**0.031**	0.229				
Flight distance response (m)	0.062	0.075	0.06	0.001	**0.586**			
P-value	0.483	0.393	0.497	0.991	**<0.001**			
Recovery from surprise (s)	-0.049	-0.113	0.126	-0.046	**0.353**	**0.447**		
P-value	0.577	0.198	0.153	0.602	**<0.001**	**<0.001**		

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Number of ewes in group	**-0.48**	**0.267**	**0.383**	**0.289**	0.095	-0.019	0.115	
P-value	**<0.001**	**0.002**	**<0.001**	**0.001**	0.28	0.83	0.189	
Number of lambs in group	**-0.278**	**0.416**	**0.478**	**0.37**	0.089	0.021	0.069	**0.847**
P-value	**0.001**	**<0.001**	**<0.001**	**<0.001**	0.313	0.811	0.436	**<0.001**

Physical comfort and health

No significant correlations were found between physical health/comfort measures at this time as shown in table 6.7.

Table 6.6 Correlations between the Assessment in the Field physical health and comfort measures during post-lambing assessments (9-11)

Indicator	Coat cleanliness	Faecal soiling
Faecal soiling	0.005	
P-value	*0.951*	
Lameness	-0.048	0.058
P-value	*0.588*	*0.508*

6.3.4 Relationships between indicators: post-weaning time period

Social behaviours

Following weaning, ewes which were further away from neighbours were more likely to be in smaller groups with fewer ewes and lambs (P<0.001 and P=0.004 respectively) and tended to have fewer ewes engaged in vigilance (P=0.071).

General Fear

Ewes which ran from humans from a greater distance had a larger flight distance (P=0.007), and took longer to recover (P=0.002). Flight distance and recovery time were also significantly correlated with each other (P<0.001). Ewes in larger groups tended to be more likely to allow the human to approach closer before they fled (P=0.058). Groups with higher numbers of ewes were more likely to contain more lambs (P<0.001) and have a higher percentage of ewes performing vigilance behaviour (P=0.003).

Table 6.7 Correlations between behavioural measures during the Assessments in the Field in the post-weaning period (assessments 12 and 13)

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Ewe vocalisations	-0.055							
P-value	0.608							
Lamb vocalisations	0.039	0.133						
P-value	0.717	0.216						
Percentage of ewes in group performing vigilance	**-0.192**	-0.026	-0.031					
P-value	**0.071**	0.808	0.77					
Human approach distance (m)	0.082	-0.05	-0.041	-0.117				
P-value	0.446	0.639	0.7	0.274				
Flight distance response (m)	0.13	0.007	-0.083	-0.183	**0.283**			
P-value	0.226	0.948	0.441	0.087	**0.007**			

Indicator	Nearest Neighbour distance (m)	Ewe vocalisations	Lamb vocalisations	Percentage of ewes in group performing vigilance	Human approach distance (m)	Flight distance response (m)	Recovery from surprise (s)	Number of ewes in group
Recovery from surprise (s)	-0.049	0.064	-0.011	-0.159	**0.326**	**0.404**		
P-value	0.646	0.549	0.921	0.137	**0.002**	**<0.001**		
Number of ewes in group	**-0.718**	-0.053	-0.077	**0.31**	**-0.202**	-0.091	-0.055	
P-value	**<0.001**	0.623	0.475	**0.003**	**0.058**	0.394	0.607	
Number of lambs in group	**-0.303**	-0.054	-0.038	-0.13	-0.114	0.16	-0.047	**0.404**
P-value	**0.004**	0.614	0.725	0.223	0.287	0.135	0.665	**<0.001**

Physical health and comfort

Post-weaning there were no significant relationships between coat cleanliness and Faecal soiling score, or lameness. However, lameness was significantly negatively correlated with faecal soiling scores (P=0.048) as shown in table 6.10.

Table 6.8 Correlations between the Assessment in the Field physical health and comfort indicators during post-weaning assessments (12 and 13)

Indicator	Coat	Faecal soiling
Faecal soiling	-0.078 0.466	
Lameness	-0.029 0.791	-0.211 0.048

6.3.5 Relationships between QBA and behavioural indicators during all five assessments

Significant correlations were found between the QBA PCs and other Assessment in the Field indicators when the full data set consisting of all five visits was analysed. The correlations are shown in Table 6.5.

Compared to ewes with high General Mood, those with lower General Mood were found in larger groups with higher numbers of ewes (P=0.039) and lambs (P<0.001), and these groups were more vocal (P<0.001). A non-significant negative trend (P=0.089) was found between General Mood and nearest neighbour suggesting that ewes with lower General Mood scores had greater distances between themselves and other individuals. A non-significant negative trend was also found between General Mood and flight distance (P=0.074) meaning that ewes in lower General Mood states fled farther following the surprise test than those in a higher mood state.

Ewes in higher Arousal state were more likely to flee when the human was at a greater distance ($P=0.034$). These ewes were in more vocal groups with higher numbers of bleats by ewes and lambs (both $P<0.001$). A non-significant positive trend was found between Arousal and number of lambs present in group ($P=0.078$).

Animals with larger approach distances ($P=0.032$) were found to be more Listless, and those with longer recovery times also tended to be more Listless ($P=0.095$). Listless animals were more likely to be in larger groups with more ewes and lambs (both $P<0.001$), by whom there were higher vocalisation rates ($P=0.039$ and $P<0.001$). A higher percentage of these group members were engaged in vigilance ($P=0.026$) than those with lower Listlessness scores.

Ewes which were more Aggressive were more likely to be in more vocal groups (both $P=0.001$), with higher numbers of lambs ($P=0.015$). Higher numbers of ewes in a social group tended to be associated with increased Aggressiveness ($P=0.094$).

Table 6.9 Correlations between QBA PCs and behavioural measures for all five Assessments in the Field

Indicator	General Mood	Arousal	Listlessness	Aggressiveness
Nearest neighbour distance (m)	-0.115	0.071	-0.068	-0.109
P-value	*0.089*	*0.296*	*0.319*	*0.108*
Ewe vocalisations	**-0.241**	**0.496**	0.228	**0.413**
P-value	***<0.001***	***<0.001***	*0.001*	***<0.001***
Lamb vocalisations	-0.103	**0.283**	0.14	**0.396**
P-value	*0.129*	***<0.001***	*0.039*	***<0.001***
Percentage of group performing vigilance	-0.078	0.008	0.15	-0.041
P-value	*0.253*	*0.909*	*0.026*	*0.544*
Human approach distance (m)	-0.098	**0.143**	**0.144**	0.024
P-value	*0.147*	***0.034***	***0.032***	*0.727*
Recovery from surprise test (s)	-0.008	-0.11	0.113	-0.107
P-value	*0.907*	*0.103*	*0.095*	*0.115*
Flight distance response (m)	-0.121	0.079	0.054	0
P-value	*0.074*	*0.243*	*0.428*	*0.999*
Number of ewes in group	**-0.139**	0.038	**0.27**	0.113
P-value	***0.039***	*0.58*	***<0.001***	*0.094*
Number of lambs in group	**-0.256**	0.119	0.292	**0.163**
P-value	***<0.001***	*0.078*	*<0.001*	***0.015***

Individual ewe postural synchrony

Ewes which were posturally synchronised with the rest of the group were significantly more Listless than those which were not synchronised (Synchronised mean=-0.892, Non-synchronised mean=-0.911; Wald=0.40, DF=1, P=0.028). No other QBA PCs were significantly related to ewe synchrony P>0.34.

6.3.6 Relationships between QBA and behavioural indicators in the post-lambing time period

The correlations between the post-lambing QBA results and other welfare indicators are shown in Table 6.8. In the post-lambing time point ewes with higher General Mood were typically in groups with lower rates of ewe vocalisation (P=0.013), and these groups tended to have lower numbers of lambs (P=0.077).

Ewes in higher Arousal states were more likely to flee when the assessor was at a greater distance (P=0.044). The ewes in higher Arousal tended to be in groups with more lambs (P=0.073) and higher numbers of vocalisations (P<0.001).

Ewes scoring higher on Listlessness were found in larger groups with more lambs (P=0.012) and ewes (P=0.002), which were more vocal (P=0.012) and were more likely to flee when the human was at a greater distance (P=0.041). These ewes also tended to be in groups with high levels of vigilance behaviour (P=0.041).

Ewes which received higher Aggressiveness scores recovered from the surprising event more quickly than those with lower Aggressiveness scores (P=0.036). Those with high Aggressiveness scores were found in groups with high vocalisation rates (P<0.001), and tended to be found in groups with more lambs (P=0.087).

Table 6.10 Correlations between QBA PCs and behavioural indicators applied during the Assessment in the Field during the post-lambing time period (assessments 9-11)

Indicator	General Mood	Arousal	Listlessness	Aggressiveness
Nearest neighbour	-0.136	0.02	-0.045	-0.084
P-value	0.122	0.823	0.612	0.341
Vocalisation by ewes	**-0.218**	**0.543**	**0.22**	**0.432**
P-value	**0.013**	**0**	**0.012**	**0**
Vocalisation by lambs	-0.035	**0.31**	0.124	**0.418**
P-value	0.691	**0**	0.157	**0**
Percentage of group performing vigilance	-0.08	0.025	**0.188**	-0.06
P-value	0.364	0.773	**0.031**	0.493
Approach distance	-0.115	**0.177**	**0.179**	0.036
P-value	0.192	**0.044**	**0.041**	0.685
Response flight	-0.052	0.121	0.036	-0.052
P-value	0.556	0.168	0.681	0.554
Recovery time	0.004	-0.119	0.115	**-0.183**
P-value	0.966	0.177	0.19	**0.036**
Number of ewes	-0.071	0.077	**0.271**	0.093
P-value	0.422	0.385	**0.002**	0.292
Number of lambs	-0.155	*0.157*	**0.29**	*0.15*
P-value	0.077	*0.073*	**0.001**	*0.087*

6.3.7 Relationships between QBA and behavioural indicators in the post-weaning time period

Significant correlations were found between the QBA results and other Assessment in the Field indicators in the post-weaning time point and are shown in Table 6.11.

Ewes which were experiencing higher levels of Arousal had smaller distances between themselves and their nearest neighbours (P=0.034), and were found in larger groups with more ewes (P=0.043) and lambs (P=0.021).

Ewes which were more Listless had larger distances between themselves and their nearest neighbours (P=0.036). Whereas those ewes which in higher states of Aggressiveness had larger distances between themselves and others (P=0.038).

Table 6.11 Correlations between QBA PCs and Assessment in the Field indicators during the post-weaning time period (assessments 11 and 12)

Indicator	General Mood	Arousal	Listlessness	Aggressiveness
Nearest neighbour (m)	-0.172	**-0.226**	**0.223**	**-0.22**
P-value	0.107	**0.034**	**0.036**	**0.038**
Vocalisation by ewes	0.137	-0.033	-0.114	0.07
P-value	0.202	0.758	0.286	0.513
Vocalisation by lambs	0.04	0.005	0	0.014
P-value	0.708	0.963	0.997	0.897
Percentage of group performing vigilance	-0.019	0.039	-0.061	-0.015
P-value	0.86	0.72	0.568	0.887
Approach distance (m)	0.042	-0.04	0.024	-0.041
P-value	0.695	0.713	0.824	0.703
Response flight (m)	-0.075	0.1	-0.018	0.155
P-value	0.487	0.349	0.865	0.148
Recovery time (s)	0.025	0.097	-0.123	*0.191*
P-value	0.817	0.368	0.252	*0.073*
Number of ewes	-0.068	**0.215**	-0.129	0.156
P-value	0.525	**0.043**	0.23	0.145
Number of lambs	-0.015	**0.245**	0.009	0.129
P-value	0.89	**0.021**	0.936	0.228

6.3.8 Further investigation of the relationship between QBA and assessment indicators

General Mood

From the univariate analysis the factors nearest neighbour distance and ewe vocalisation were found to explain some of the variance in General Mood at P=0.2 or lower and were retained in the multivariate model.

Following the multivariate model analysis, significance was retained for nearest neighbour distance (Wald=4.61, DF=1, P=0.02) and ewe vocalisation (W=5.51, DF=1, P=0.02), Ewes with higher General Mood were nearer to their neighbours and were in groups with higher numbers of ewe vocalisations.

Arousal

From the univariate analysis the factors approach distance and ewe vocalisation were found to explain the variance in Arousal at P=0.2 or lower and were retained in the multivariate model.

When all were included in the multivariate model, approach distance (Wald=4.28, DF=1, P=0.04) and ewe vocalisation (Wald=3.51, DF=1, P<0.001) were found to be significantly positively related to Arousal.

Listlessness

From the univariate analysis the factors approach distance, number of lambs in group, recovery from surprise, and ewe vocalisation were found to explain the variance in Listlessness at P=0.2 or lower and were retained in the multivariate model.

In the multivariate model only ewe vocalisation tended increase with Listlessness (Wald=0.49, DF=1, P= 0.059).

Aggressiveness

When included in the univariate models number of lambs, recovery time and vocalisations by lambs and ewes explained the variance in Aggressiveness at P=0.2 or lower and were therefore retained for inclusion in the multivariate model.

In the multivariate model only the ewe and lamb vocalisations had significant effects on Aggressiveness (ewe vocalisations: Wald=44.1, DF=1 P<0.001; Lamb vocalisations: Wald=37.23, DF=1, P<0.001).

6.4 Discussion

6.4.1 Summary

Alterations in ewe behaviour, social grouping and affective state were observed across the assessment period throughout July and August 2013. Group size and composition, and flight distance following a surprise event differed between post-lambing and post-weaning time periods. These alterations in behaviour indicate that the removal of male lambs and progression of calendar season may have significant effects on the outcome of a welfare assessment. It is proposed that the alterations in ewe behaviour was due to environmental factors, such as forage availability, and changes in reproductive state i.e. the removal of the male lambs at weaning. A decrease in group size and vocalisation were seen in the post-weaning time period compared to the observations during which all lambs were at foot. This indicates that the removal of the male lambs had consequences on the flock dynamics. The ewes also showed a reduction in their vigilance behaviours, postural synchronisation and flight distance from an approaching human in the post-weaning period compared to the post-lambing time point. The relationships found between ewe behaviour and affective state also altered during this time. Alteration in the ewes' affective state, measured by QBA as in the previous chapter, was found to be related to the size and composition of their social group, and the perceived levels of predation threat. Ewes with a lower General Mood had larger distances between themselves and their nearest neighbour than ewes in higher mood states. Higher numbers of vocalisations during the two minute observation were observed in groups in which the focal ewes had low General Mood, increased Arousal state, higher Listlessness and heightened Aggressiveness. Ewes in a higher Arousal state tended to flee earlier from an approaching human than those experiencing low levels of Arousal. The physical health and comfort indicators did not appear to have any significant association with the affective state of the animals, perhaps due to the low variability in the physical health data. In this discussion the convergent validity and feasibility of the indicators applied during the Assessment in the Field are discussed, and conclusions formed on the suitability of each of the measures as a welfare indicator for extensively managed sheep.

6.4.2 Nearest neighbour

Grazing as a group, and flocking together are among the most prominent characteristics of the social behaviour of sheep (Sibbald and Hooper, 2003). The animals' natural instinct to flock has been enhanced during domestication and selective breeding to improve ease of moving and handling. In group situations the decisions made by individuals are also influenced by conspecifics and the desire to be near to each other (Michelena et al., 2008). The breed used in this research, the Scottish Blackface, are highly gregarious and each individual within a group will tend to maintain a consistent distance from their nearest neighbours when grazing (Shackleton and Shank, 1984; Sibbald and Hooper, 2004) although it can vary with reproductive status (Poindron et al., 1994), environmental factors (Dwyer et al., 1999), and with activity type (Lynch et al., 1985).

The effect of reproductive state on social behaviours such as separation from social group has been discussed in Chapter 5. In this study no significant differences in nearest neighbour distance were found between the post-lambing and post weaning time periods. Previous studies have found that the reproductive status of an animal affects their grazing behaviour and lactating ewes are more likely to graze farther away from conspecifics providing their lamb remains in close proximity (Poindron et al., 1994). It is possible that the findings in this work were due to the fact that not all lambs were removed in the post-weaning time points, and the ewes with female lambs were therefore largely unaffected by the event of "weaning". It is also possible that the presence of the female lambs also affected the behaviour of the ewes which had their male lambs removed. If this study were to be repeated all lambs should be removed at the same time in order to fully understand the impact of lamb presence on the grazing behaviour of ewes.

In this study, nearest neighbour distance was found to be related to the General Mood state of the ewes as those with lower General Mood were more likely to have greater nearest neighbour distances than those with in high mood states. These findings are in accordance with Phythian et al (2015). In their study the overall mood of sheep was negatively correlated with a measure called Dull Physical Demeanour. Dull Physical Demeanour described sheep which have removed themselves from the social group and may have altered posture e.g. head down with a reduction in responsivity. This description is similar to that described by others as "sickness behaviour". Although these symptoms are well recognised by vets and farmers when assessing animals (Eksebo, 2011), the distance between a ewe and her nearest neighbour does not appear to have been used in an animal based welfare assessment for extensively managed animals before.

Distances between grazing animals are also affected by the perceived risk of predation. Sheep flock together in response to fear-inducing stimuli, a behaviour thought to aid survival. Animals on the edge of a social group are more likely to be predated upon, whereas those in the centre of the flock are less conspicuous (Sibbald and Hooper 2003). Although information regarding the position of the ewe in that social group was not collected during this study it would have been interesting to assess the effect of this on her social behaviours and QBA assessments. This would be a potential additional measure to include in future work assessing the impact of environmental and social effects on the variation of sheep expression.

In some situations a trade-off has to be made by a grazing animal such as sheep as the motivation to stay close to conspecifics may conflict with the animal's internal motivation to graze on patches of preferred vegetation. Thus, the social nature of the sheep is in direct conflict with their nutritional requirements (Lynch et al., 1985). The potential effects of season and forage availability on ewe grazing behaviour and dispersal from social group was discussed in Chapter 5. The data collected in this chapter did not include the type of forage that was currently available to animals during their Assessments, only the most abundant plant species were known as shown in Chapter 3. In future work the forage type and quality may be examined in order to further understand the effects of these environmental conditions on the outcomes of the welfare indicators.

Nearest neighbour as a welfare indicator

From the work presented in this chapter it can be concluded that estimating nearest neighbour distance by sight is a feasible measure for use with freely ranging extensively managed sheep. In Chapter 2 of this thesis it was stated that due to expert opinion this measure is considered to have face-validity. Experts advocate that distance between a ewe and her nearest neighbour is potentially an informative welfare indicator (Phythian et al 2011) although convergent validity between this measure and other welfare indicators were previously lacking in the scientific literature. The results of this study, more specifically the relationships found between nearest neighbour distance and affective state, strengthen the validity of this measure as a welfare indicator for sheep. Inter-observer agreement was not investigated during this study and would therefore be required before a decision can be made regarding the measure's reliability.

6.4.3 Postural synchrony

Sheep display instinctive flocking behaviour. They maintain appropriate nearest neighbour distances, and follow the flock's movements, postures and behaviours (Dwyer 2004). Sheep typically congregate with animals showing similar behavioural patterns or requirements to themselves. They are most likely to be synchronised with the behaviour of their nearest neighbours (Le Pendu et al., 1996) and other group members of the same age (K. Ruckstuhl and Neuhaus, 2001)). The postural synchrony of ewes in this chapter was calculated at both the group and individual levels. The group synchrony was found to be higher during the post-lambing time period in comparison with the post-weaning assessments. However, the overall synchrony levels of the groups were relatively low, indicating that the sheep tended not to be particularly synchronised. It is possible that the reason the ewes in this study displayed low levels of group synchrony was due to the fact lambs were at foot throughout all observations. As described in the previous Chapter, social networks between the dam and other ewes are affected by the presence of her lamb. The ewe tends to behave in a more individualistic way during this time as the lamb becomes the focus of the ewe's social attention (Hinch et al., 1987). The removal of the male lambs may have had an impact on the social behaviour of their dams, although as the female lambs remained with their mothers their attention may have continued to be directed primarily to their lambs. The fact that some ewes remained with their lambs while others were weaned caused a decrease in the uniformity of the animals observed.

At an individual level ewes which were asynchronous with their group were more Listless. This implies that the ewes which were Listless were less motivated or able to maintain synchronicity with the rest of the flock. It has been proposed that the synchronous behaviour of sheep may function as an anti-predator strategy by ensuring that individual animals are not noticed (Dwyer, 2004). The degree of synchrony can vary with group composition and size (Ruckstuhl and Neuhaus 2001). Synchrony with other individuals may be achieved through two main processes (Conradt and List, 2009; Gautrais et al., 2007). The first is allelomimcry or social facilitation. The individual adopts the postures or behaviours of other individuals in the group which are near them. Through this process the entire group may achieve synchronicity (Sumpter, 2010). Synchrony may also occur through combined responses. Animals make their own decisions as to their posture and behaviour and does so at the same time of other animals due to external factors, such as the arrival of feed, or internal factors, such as similar circadian rhythms. A group of animals of similar ages are likely to have similar biological needs and will all therefore alternate between periods of activity (grazing and foraging), and inactivity (resting) (Sumpter 2010). In order to stay in close proximity with other group members, individuals need to coordinate their movement from one grazing area to another. The animal must then compromise between its own needs and remaining in synchrony with the flock (Meldrum and Ruckstuhl, 2009; Ruckstuhl, 1999). The activity patterns of a group are influenced by, and change, in response to many factors e.g. the presence of predators, and group size (Tadesse and Kotler, 2013). In the study reported in this chapter Listless ewes may have been less engaged with their surroundings and therefore less likely to actively attempt synchronisation with other flock members.

The use of lying synchrony as an indicator of welfare is increasing, particularly for large ruminants such as cattle (Færevik et al., 2005; Fregonesi and Leaver, 2001; Napolitano et al., 2009; O'Driscoll et al., 2008). The ability to calculate behavioural or postural synchrony during an instantaneous scan sample makes it an attractive welfare indicator due to the speed and ease at which it can be performed (Napolitano et al., 2009). Synchrony can indicate whether there are sufficient resources available for all individuals (Færevik et al ,2005). Resources such as space at a feeder and lying space may be more important for housed animals or those in intensive systems (Napolitano et al 2009), but for extensively managed sheep, living outside, the provision of adequate clean lying space may be of concern. The low levels of synchrony found in these groups of sheep, and low mean coat cleanliness score, indicate that although the sheep were not highly synchronised there was not a problem with dry lying space for the animals observed during this time.

Postural synchrony as a welfare indicator

Assessing postural synchrony of a group of extensively managed sheep using scan sampling techniques proved to be feasible during this study. The inter-observer reliability of the measure has not been tested and therefore a conclusion on this aspect of the measure cannot be drawn. The potential influence of offspring on the results of this measure imply that if applying this measure to extensively managed sheep some seasonal variability may be expected but this may not necessarily mirror shifts in the animals' welfare states. The validity of synchronicity of states such as feeding and resting has been shown in other studies particularly with cattle (Færevik et al., 2005; Fregonesi and Leaver, 2001; Napolitano et al., 2009; O'Driscoll et al., 2008) in which it is generally agreed that high levels of synchronisation indicate high welfare states. These studies tend to be focussed on animals which are housed or live in environments with limited resources and thus, synchronicity implies all animals have access to basic resources. In an extensive environment this competition may not be so great and therefore synchronicity may not be a suitable welfare indicator. The low levels of synchrony observed in the study reported in this chapter mean it is not possible to make a conclusion on the validity of this measure from the animals observed here.

6.4.4 Group size

The number of sheep in a social group depends upon many factors: openness of terrain, the availability and distribution of resources, and predator pressures (Alexander, 1974). Large group numbers may not always prove to be beneficial as it can result in individuals feeding on less preferred food types or on lower quality feeds (Sibbald and Hooper 2004). Although group size is not a specific welfare indicator included in the evaluation, gathering this information during the Assessment in the Field allowed further interpretation of the data set.

In the work reported in this chapter, group size was significantly affected by the event of weaning. Following weaning, the groups in which the animals congregated were smaller than when all lambs were at foot during the summer. The increase in group size found during the post-lambing time point compared to post-weaning may have been due to the increased perceived predation risks occurring while lambs are present in the group (Pickup and Dwyer, 2011). This finding is, however, in contrast to previous work on free ranging sheep as ewe group sizes have been reported to decrease when lambs were at foot (Alistair B. Lawrence and Wood-Gush, 1988). The group size of wild sheep has been found to vary seasonally (Oli and Rogers, 1996). These authors found that groups of Blue sheep were larger during spring and autumn, than during the winter. They suggested that these differences may have been due to forage abundance and when good quality forage became scarce the sheep were forced to prioritise their nutritional needs over their desire to remain part of a close social group. In the work by Oli and Rogers (1996), they did not follow the sheep between the months of May and September, when lambs remained in close proximity to their dams. It is possible that the ewes observed during the study reported in this chapter were also affected by forage availability. Following weaning, in August, there may have been lower forage abundance and quality available forcing the ewes to behave more individualistically.

During the Assessments in Field, higher Aggressiveness was found in larger groups. Research looking at aggression levels and group size in sheep has proposed that aggression in housed sheep is more susceptible to changes in space allowance rather than the number of group members (Averós et al., 2014b). Although space allowance is not a concern for the ewes located on Castlelaw farm, the increase in Aggressiveness may have been caused by competition of resources as animals may want to defend a food patch (Taillon and Côté, 2006) or due to alterations in the social hierarchy (Cassinello, 1995). The larger groups in which Aggressiveness levels were higher typically occurred in the post-lambing time period. It is possible that the presence of the lambs influenced the ewe's Aggressiveness as they may have been protecting their own lambs, preventing other lambs from suckling from them, or defending their own grazing space. In a study with free-ranging Barbary sheep, Cassinello (1995) found that mating and parturition increased a female's social rank and these rankings continued throughout the period at which the young were at foot. During this time the number of aggressive interactions observed by Casellino (1995) increased. Following weaning, and the cessation of lactation, the female's rank reduced. This increased social rank and aggression levels during lactation may assist the dam in finding and gaining access to higher quality forage (Cassinello, 1995;

Taillon and Côté, 2006). Although Barbary sheep are not “true” sheep, but rather a predecessor of modern sheep breeds, these findings may aid in the understanding of the fluctuations in levels of Aggression and variation in social behaviour observed in this study. As discussed previously the removal of male lambs but retention of female lambs may have caused further discrepancy in the social structure of the group. If the ewes were followed for an entire reproductive cycle (pre-mating to post-weaning) the differences in social behaviour described by Lawrence and Wood-Gush (1988), Oli and Rogers, (1996) and Casellino (1995) may have been found.

Aggression levels can be influenced by social rank, but also by the animal’s body condition and the quality and value of a resource such as forage (Taillon and Cote 2006). The body condition score of the ewes was assessed at the “weaning” Assessment at Gather and if more time were available it may have been interesting to assess the relationship between these scores and the Aggressiveness displayed by the ewes.

Ewes which were more Listless were more likely to be found in larger groups, which may be indicative of their feelings of insecurity and desire to surround themselves with conspecifics to avoid predation. Group size was also found to be related to General Mood as higher General Mood was associated with a smaller social group. This may be because when threat of predation is high ewes naturally flock together, however when the predation risk is decreased the ewes maintain their smaller social groups (Hopewell et al 2005) and it could be this reduction in general fearfulness and interruptions to grazing that led to their improved General Mood.

6.4.5 Vigilance

Group level

In gregarious animals, such as sheep, the relationship between group size and function of vigilance levels are complex. The composition of groups, inter-individual distances and relative positions of individuals have all been found to influence vigilance by individuals because they are affected by the judgments of group size and predation risk (Hopewell et al., 2005; Treves, 2000). In the data collected for this chapter, vigilance was measured at a group level. Higher rates of vigilance occurred in larger groups, in which there were increased call rates by ewes and lambs. The distance between the focal ewe and her nearest neighbour was found to be shorter in groups in which a larger proportion of the animals were engaged in vigilance. No associations were found between the percentage of individuals in a group performing vigilance behaviours and the focal ewe's reaction to a human approach, or response to and recovery from the surprise test.

Vigilance has several functions besides predator detection as some information may be obtained about behaviour of conspecifics, signals between ewes and lambs, and the availability of food (Berger, 1978; Hopewell et al., 2005). Free ranging sheep must trade off the benefits of spending time grazing with the potential costs of compromising their vigilance whilst they feed with their heads down (Hopewell 2005). The costs include a reduced likelihood of detecting a predator (Quenette, 1990), loss of social contact (Beauchamp 2003), and missing out on food in other locations (Barbosa, 2002). Ewes in larger groups may increase their vigilance levels in order to synchronise their behaviour and movement patterns. Individuals which are unable to synchronise may be separated from the flock, and therefore risk increased predation (Aivaz and Ruckstuhl, 2011). The posture associated with vigilance, in which the ewe stands with her head raised above her shoulders, is also known to be employed during communication between the ewe and her lamb (Lawrence and Woodgush, 1987; Pickup and Dwyer, 2011). The ewe uses this "head up" posture to signal to the lamb that they may approach to suckle (Lawrence 1985, Pickup and Dwyer 2011). Whether sheep seek larger groups so they may reduce time spent visually scanning the environment for signs of predators, or because conspecifics may then buffer them from predators is unclear but these categories are not mutually exclusive and both factors may interplay in decision making (Berger, 1978). In large groups there is more opportunity to find food by watching others and therefore individuals may spend more time on vigilance, even if there is little threat of predation (Beauchamp, 2001).

The percentage of ewes engaged in vigilance behaviours during the scan sampling was significantly higher pre-weaning when all lambs were at foot. The findings reported in this chapter are in accordance with those reported in the other literature. The reproductive status of a ewe is known to affect vigilance behaviours as lactating ewes are more vigilant than non-lactating ewes (Berger, 1991; Bleich et al., 1997; Bon, 2012; Rieucau and G. A. Martin, 2008). The increase in vigilance during this may be related to the "head-up" posture during which she invites the lamb to approach her (Pickup and Dwyer 2011). The reduction in head-up posture found in the study reported in this chapter may have been caused by the scheduled weaning of the male lambs, or due to the lambs naturally weaning from the ewe's milk on to grass (Arnold and Pahl, 1974). Vigilance levels aids early detection of predators and therefore lamb survival (Rieucau and Martin 2008). Female vigilance is also increased when their offspring are active (White and Berger, 2001) and when predator density is high (Laundre et al., 2001). Young and inexperienced ewes are known to show poorer vigilance and suffer higher predation of their lambs (Warren et al., 2001).

Previous studies have shown that gregarious animals adjust their vigilance levels according to their physical position within the group and due to the behaviours performed by their nearest neighbours and other members of the group (Fernández-Juricic and Kacelnik, 2004). This information was not obtained during this study but if similar work were to be performed in the future the ewe's position in the group may provide valuable information assisting in the interpretation of vigilance behaviour.

6.4.5.1 Vigilance of individual sheep

A decrease in vigilance by individuals as group size increases is a widely reported effect that has been explained in terms of an increase in the likelihood of predator detection (many eyes hypothesis, (Roberts, 1996) or a decrease in the perceived threat of predation (dilution effect: (Elgar, 1989; Quenette, 1990). Michelena and Deneubourg (2011) found vigilance and cortisol levels decreased in larger groups, however ewes in smaller groups had more variability in their cortisol than those in larger groups. Larger group sizes in locations and time points with higher predation risks have been found for a range of prey species, including sheep (Hopewell et al., 2005; Jørgensen et al., 2009; Sibbald et al., 2009). For Soay sheep group size depends mainly upon the need for anti-predatory vigilance (Hopewell 2005). An increase in group size also allows animals to reduce individual vigilance behaviours leaving the animal free to devote more time to other behaviours such as feeding and social interactions (Roberts, 1996). Whether the reduction in vigilance behaviour per animal is due to the increased group vigilance (Pulliam, 1973), or because an increase in group size reduces predation risk and therefore the need for vigilance is reduced, or a combination of the two, is unknown (Roberts, 1996).

Vigilance as a welfare indicator

Vigilance behaviour has been assessed and used as a welfare indicator for other livestock species. The use of vigilance as a measure of fearfulness in dairy cattle was assessed by (Welp et al., 2004). They concluded that vigilance could be used to measure fearfulness in cattle and that the vigilance level of individual cows was determined by the threat posed by the stimuli presented. Vigilance behaviour during feeding and during a novel object test was included in the Welfare Quality on-farm welfare assessment protocol for use with cattle, and during a human-approach test with pigs (Forkman and Keeling, 2009a, 2009c). Assessing vigilance behaviour in undisturbed groups, as in this study, does not appear to have been reported in the literature or used in other on-farm welfare assessments. Assessing undisturbed animals may provide a more accurate indication of overall vigilance levels in a group/flock compared to artificial situations discussed above. The validity of using vigilance behaviour as a welfare indicator may be questioned, thus care must be taken when interpreting these results as discussed above vigilance behaviour may not solely indicate the predation risk. Humans cannot easily distinguish between the intention of the vigilance and studies have shown that individual vigilance does not necessarily decrease with group size (Treves 2000). Another potential problem with using vigilance behaviour as a welfare indicator is the definition and the human's ability to recognise such behaviours. The vigilance behaviours recorded in this study were scored when the ewes had their heads raised and were notably scanning their environment however, it is possible that ewes may also be vigilant with their heads down. In a recent study (Banks, Sprague, Schmoll, Parnell, & Love, (2015) found that sheep were able to rotate their eyes when their heads are lowered for grazing allowing them to maintain their ability to scan the environment. These eye movements maintain the alignment of the pupil with the horizon. It is possible then that humans performing scan samples could misattribute vigilance which is directed at conspecifics. A ewe synchronising her behaviour with another flock member, or to communicating with her lamb could be mistaken for one scanning the environment for predators. Or if an animal has their head down and is scanning for predators the observer might be unaware of, and not score, genuine vigilance. The intention of vigilance may, however, not be the most important aspect of this behaviour when assessing the welfare of animals on farm. When raising her head the ewe interrupts her grazing patterns. High levels of this behaviour may also indicate that the normal behavioural repertoire and activity budgets are constrained.

In summary, the use of vigilance as a welfare indicator is a complex issue. The intent of the animal performing vigilance behaviour may not be readily understood by an observer, but abnormally high levels of this behaviour can have negative consequences on the ewe's activity budget and impact upon her ability to graze or rest for a sufficient time. In this instance her welfare may be compromised. For this reason vigilance behaviour may be considered to be a valid welfare indicator for use with extensively managed sheep. The measure has proven to be feasible when performed as part of a visual scan. As with other measures in this chapter, inter-observer reliably requires further work prior to its confirmation as a welfare indicator for extensively managed sheep.

6.4.6 General Fear

The assessment of general fear included the indicators of response to human and response to, and recovery from, a surprise test. The results for these indicators were found to be very well correlated with one another. Due to their relatively small size sheep are too small to successfully defend themselves against large predator and therefore their main anti-predator defensive is to flee to cover and refuge (Bleich et al., 1997; Lima, 1998). In the study reported in this chapter it was found that the ewes which were in a higher General Mood had a lower response flight, meaning that the human could get closer before the ewe fled. Whereas ewes in a higher Arousal state, and those who were more Listless, fled while the human was still farther away indicating a higher fear response. Phythian et al. (2015) found that ewes with lowered mood also had a decrease in “responsivity”. “Responsivity” appeared to be whether the sheep “noticed” the presence of observers or stimulus but the “response” wasn’t explicitly stated.

The common behavioural response both of wild and domestic sheep to the presence of a predator is to flock together and run (Dwyer 2004). However, if attacked by small predators mountain sheep will stand over their lamb to protect it (Geist 1971, Dwyer 2008). Most prey animals will permit a predator to approach to a certain point before responding (the approach distance). In Himalayan sheep, perceived danger is tolerated at a “safe” distance based on the previous experience of sheep (Schaller, 1977). Mouflon are also known to modify their flight distance in the presence of different types of intruders (e.g. hikers or dog walkers. (Martinetto and Cugnasse, 2001)). The environment of the sheep also influences flight behaviour. Bighorn ewes are more likely to run when threatened with predation in the open than when near slopes or vegetation cover (Berger 1991, Bleich 1999). A human on foot approaching a prey animal, such as sheep may represent a predator (Frid and Dill, 2002) and as such the sheep’s reaction to the human in this study may be used to indicate their generalised response to predators and their general fear levels. Although the sheep on the SRUC hill farm have relatively few predators, and those that are present are more likely to attack lambs rather than adult ewes, the innate behaviours of the sheep as a prey species are likely to have been retained. As stated in Chapter 3 the land on which the farm is situated is owned by the Ministry of Defence and therefore there are often large numbers of soldiers present performing various manoeuvres during the day and night. The presence of the people, loud noises, ammunition and flares likely disturbs the sheep and may lead to increased general anxiety.

The optimal anti-predator response for an individual animal is a trade-off between interruption to grazing and increased energy expenditure and the risk of getting caught (Cherry et al., 2015). The predator may be noticed ahead of the possible attack either by the animal itself or by signals from other group members (Sirot and Touzalin, 2009).In this situation the "economics of fleeing" are assessed (Ydenberg and Dill, 1986). Initially the costs of remaining in the current position are small but as the predator approaches it increases the probability of attack. The animal should flee from the approaching predator when the cost of staying overcomes the costs of fleeing (energy requirements and time taken away from current behaviour e.g. grazing). These costs vary between animals and therefore the ideal departure time for one animal may be too late for another (Ydenberg and Dill 1986). In this study the variability of the approach distance of the ewes remained relatively constant with low variation seen before or after weaning. This indicates that the ewes on the farm had similar motivations and trade-offs.

Although the distance to which a human could approach the ewes did not vary over time, flight distances following the surprise test decreased following weaning. The effect of the presence of young on the flight distance of sheep does not appear to have been reported in the literature. It is therefore suggested that the ewes may have fled further when lambs were at foot in order to ensure the survival of her offspring. The increase in energy expenditure and time taken from other behaviours may be worth the investment at this time in order to protect her lamb from immediate danger. The distance between the ewe and cover such as bushes and trees may also have played a part in her reactivity to the surprise test, especially when the lambs were at foot. No data were collected on the ewe's proximity to such cover but it is possible the presence of the lambs influenced this behaviour.

With the exception of approach and flight distances, the length of time taken for the ewe to recover from the surprising event was not related to any of the other welfare indicators. The exact stimuli an animal uses to assess whether it is safe to resume normal activity following escape from a predator has received little attention, but it has been documented that in situations which are perceived to be more dangerous animals tend to remain inactive for longer (Dwyer 2008, Lima 1998). It has been found however that animals in poorer body condition may be less risk averse and resume feeding activities earlier than animals in better nutritional states (Lima 1998). There was little variation in the recovery time of ewes observed in the study reported in this chapter potentially indicating that all sheep used similar cues and performed behavioural patterns when recovering from the surprising event.

General Fear measurements as welfare indicators

In summary, the assessment of general fear during on farm welfare assessments of extensively managed sheep still requires some further evaluation. In the study reported in this chapter the human approach and surprise tests proved to be feasible for use with extensively managed animals. The results were consistent and good relationships' with other indicators such as QBA support the measures' validity. Information on inter-observer reliability would be required before the methods could be accepted, especially for the human approach and flight distance as this relied upon a subjective estimation of distance. The use of electronic devices such as laser range finders may offer a more reliable estimate of distance between the assessor and the sheep although this does not appear to have been previously reported in the literature and may pose additional challenges.

6.4.7 Vocalisation

In this chapter ewe vocalisation was found to be related to all four QBA PCs: General Mood, Arousal, Listlessness and Aggression and the number of vocalisations by both lambs and ewes fell in the post-weaning time period.

Ewe-lamb vocalisations are important when forming and maintaining the ewe-lamb bond (Pickup and Dwyer 2011). Vocalisations by the ewe and lamb are used to identify each other and sustain contact over distance (Pickup and Dwyer 2011, Shillito and Walser et al 1981). The vocalisations of young lambs may be used a signal of immediate distress, such as during separation (Dwyer et al 1998), or be related to the degree of need (Weary and Fraser 1995). The use of high pitched vocalisations are known to decrease as lambs age and rely less on the dam for guidance and protection (Pickup and Dwyer 2011).

The use of vocalisation as part of a welfare assessment has many advantages as it is an objective and non-invasive indicator which at first glance appears to be easy to measure (Cordeiro et al., 2013). Distress related vocalisation is of particular interest as welfare indicator, particularly of impaired welfare (Marx et al., 2003; Weary and Fraser, 1995). Using vocalisation rates has been used as a simple indicator to identify handling and equipment problems in slaughterhouses (Grandin, 2001, 1998). The work by Grandin has typically focussed on cattle and as a prey species sheep may use vocal communication less than other livestock (Dwyer 1998). The inability or antipathy of sheep to use vocalisations during the presence of perceived predators (humans or a dog) is most likely to be a strategic attempt to prevent unwanted attention (Gougoulis et al., 2010; Romeyer and Bouissou, 1992).

Even if the animals do vocalise, the vocalisation by itself is not a welfare indicator suitable for use during farm assessments as without knowing the intent of the caller one cannot deuce whether the call is a signal warning of danger, cry for help, or indicative of positive mother-young bonding (Guilford and Dawkins, 1995). Using the pitch of the call may offer an insight into the messages and their relation to well-being (Manteuffel 2004). In the data collected in this study we do not have pitch or volume records however it is possible that with some additional technology vocalisation type and rate may provide useful information when assessing welfare of extensively managed sheep.

Vocalisation as a welfare indicator

From the results in this study it is not possible to confirm that vocalisation is a valid or reliable welfare indicator for use with sheep in extensive environments. The feasibility of this measure may also be questioned as it is possible that in windy weather the observer may not be able to hear all vocalisations or attribute them to the correct group of animals. For these reasons assessing vocalisation without specialised equipment is not considered to be an appropriate welfare indicator for extensively managed sheep at this time.

6.5 Conclusion

In conclusion, the indicators applied during the Assessment in the Field proved to be feasible, supporting their use as welfare indicators for use with extensively managed sheep. Distance to nearest neighbour, vigilance, QBA and behavioural measures assessing general fear were found to be valid and promising welfare indicators for use with extensively managed sheep. The physical health and comfort measures provided consistent results between the time periods of post-lambing and post-weaning but changes in behaviour and affective state were found when time periods were compared. Group size and composition, and flight distance following a surprise event were affected by the time period in which the Assessments in the Field were conducted. These alterations in behaviour indicates that the removal of male lambs, and progression of calendar season may have significant effects on the outcome of a welfare assessment. Interesting and strong relationships between indicators were found in across both time periods. Affective state, as measured by QBA, was affected by the size and composition of the social group, and the perceived levels of predation threat. For example ewes with a lower General Mood were found to be further from their nearest neighbours than ewes in higher mood states. Higher numbers of vocalisations during the two minute observation was observed in groups in which the focal ewes had low General Mood, increased Arousal state, higher Listlessness and heightened Aggressiveness. When approached by a human the ewes which tended to flee earlier were those which expressed a higher arousal state. These findings provide additional support to the use of QBA as a welfare indicator for extensively managed sheep. Further, inter-observer reliability for some of these measures is required (e.g. nearest neighbour, group size, vigilance, response to human), or revision of data collection methods with additional technology (e.g. laser range finders) before they could be used during on-farm welfare assessments. The results presented here are promising and support the use of some novel welfare indicators for use with extensively managed sheep.

Chapter 7 General Discussion

7.1 Introduction

The focus of this thesis was the evaluation of potential animal-based indicators which could be applied during an on-farm welfare assessment of extensively managed sheep. For indicators to be included in a welfare assessment protocol they should be valid, feasible and reliable. In this PhD project potential welfare indicators were assessed to determine whether they met these criteria and could therefore be accepted as a foundation of a novel welfare protocol. To achieve this, data from 135 sheep were obtained over a two year period and analysed. The results from this work have contributed to a published welfare assessment protocol. In the last chapter of this thesis the main findings and conclusions are summarised. Limitations and the potential impact of this work are discussed.

7.2 Summary of principle findings

7.2.1 Identification of potential welfare indicators

Initially the welfare issues faced by extensively managed sheep and the demand a specialised welfare assessment protocol were identified in **Chapter 1**. The majority of current welfare indicators and assessment schemes have been developed for other species and/or have been devised for use with housed animals, and are therefore not readily transferrable to sheep in extensive systems. For example the Welfare Quality project developed welfare assessment protocols for the assessment of cattle, chickens and pigs which are generally housed or live in confined paddocks. The difficulties faced when trying to apply these assessment methods in the extensive environment were described in **Chapter 1**.

Potential welfare indictors for use with extensively managed sheep were first identified during a literature review as described in **Chapter 2**. In common with work developed in the Welfare Quality project (Forkman 2009), potential welfare indicators were initially categorised according to 4 principles: Good Feeding, Good Environment, Good Health and Appropriate Behaviour. Within these principles 12 criteria encompass the main aspects of animal welfare. During an expert panel meeting previous support of these potential indicator's validity, reliability and feasibility were examined and discussed (**Chapter 2**). Following this discussion, 22 indicators were found to either meet all three requirements, or were considered to be promising but required further work before they could be accepted for use with extensively managed sheep (reported in **Chapter 2**). The prospect of Iceberg Indicators was suggested during this meeting and two lamb production traits were identified as possible candidates: lamb growth and survival.

7.2.2 Development of a protocol

Once the potential indicators for use with extensively managed sheep were selected a preliminary protocol was devised (**Chapter 3**). The indicators were assigned to either/both Assessments at Gather which could be applied to animals when gathered and comprised of a physical health inspection, and Assessments in the Field in which the animals could be assessed on their grazing land with minimal disturbance. These preliminary protocols were then be applied to animals on an SRUC research farm throughout a 2 year time frame covering key stages in reproductive cycles and across calendar year/seasonal variation. The protocols and timings of these assessments were described in more detail in **Chapter 3**.

7.2.3 Application of protocol and evaluation of indicators

The application of the Assessments at Gathers was performed across two years (2011-2013) and is described in detail in **Chapter 4**. The welfare indicators applied during these assessments covered three main aspects of welfare: Good Feeding, Good Environment and Good Health. These assessments, which comprised of a physical health check, were performed at five key points of the reproductive cycle on both years: pre-mating, mid-pregnancy, late-pregnancy, mid-lactation and weaning. The long term repeatability of the selected indicators was investigated and the fluctuation in scores were explored in **Chapter 4**. The potential use of lamb production traits as Iceberg Indicators was investigated by looking at the association between these measures and other ewe indicators. The validity of some physical health measures i.e. the novel simplified BCS and FAMACHA scoring system required further testing before they could be accepted as meaningful indicators of welfare. To the author's knowledge this was the first time the FAMACHA scoring system for anaemia had been applied to Scottish Blackface Hill sheep. Time was shown to have a great impact upon each of the indicators applied during the assessments as variation occurred both within and between years. The fluctuations in scores may be attributed to normal shifts in the animal's physiology due to alterations in reproductive state or calendar season. The predictive value of the indicators and their relationship with management decisions was also investigated in **Chapter 4**. Ewes which were selected for removal from the breeding flock by the shepherd at the annual stock draws were found to fare significantly worse in terms of BCS, tooth loss and faecal soiling. The distinction between the tooth loss scores of removed and retained ewes was seen almost a year in advance of the stock draw. It was also found that the indicators applied during the Assessments at Gather were complementary to each other which confirmed that no measures were redundant.

In **Chapter 5** one of the indicators selected for use during the Assessment in the Field, QBA was applied to sheep over a six month time period. A novel fixed list of 21 descriptive terms was developed and applied to 48 sheep over this time period. On two of these occasions three observers assessed the sheep in order to test the inter-observer reliability of the measure. Longitudinal changes in ewe behavioural expression over the 6 month time period was assessed. Four meaningful PCs were identified; the first two accounting for more than half of the explained variation between sheep, classifying the main components of ‘General Mood’ (PC1) and ‘Arousal’ (PC2). Good inter-observer agreement was found between three observers on these main components. Significant systematic changes in affective state occurred across the observational period with increases in General Mood and Arousal. These may have been related to external (environmental) and internal (physiological alterations in reproductive state) factors occurring during this time. The potential influence of both external and internal factors were discussed in **Chapter 5**.

In **Chapter 6** the entire protocol developed for the Assessment in the Field was applied to 44 ewes on 5 occasions. These indicators covered Appropriate Behaviour, Good Environment and Good Health. The main aim of the work in this chapter was to assess the convergent validity and feasibility of the welfare indicators applied during these assessments including the four QBA principle components previously reported in **Chapter 5**. An additional aim of the work described in **Chapter 6** was to investigate whether the results of the welfare indicators applied to ewes observed in a typical extensive environment fluctuated over post-lambing and post-weaning time periods. In **Chapter 6,** ewes’ affective state, as measured by QBA, was found to be related to the size and composition of their social group, and the perceived levels of predation threat. Ewes with a lower General Mood had larger distances between themselves and their nearest neighbour than ewes in higher mood states. Higher numbers of vocalisations during the two minute observation was observed in groups in which the focal ewes had low General Mood, increased Arousal state, higher Listlessness and heightened Aggressiveness. Ewes in a higher Arousal state tended to flee earlier from an approaching human than those experiencing low levels of Arousal. The physical health and comfort indicators did not appear to have any significant association with the affective state of the animals, perhaps due to the low variability in the physical health data. These findings support the validity of the QBA approach. Group size and composition, and flight distance following a surprise event, were also affected by the time period in which the Assessments in the Field were conducted implying that the removal of the male lambs has a significant impact upon ewe social behaviour. The indicators applied during the Assessment in the Field proved to be feasible supporting their use as welfare indicators for use with extensively managed sheep.

7.3 Acceptance or rejection of indicators

Analyses of the data collected during these assessments led to the potential indicators being either accepted and therefore recommended for use with extensively managed sheep, rejected, or deemed to require further investigation before such a judgement could be made.

In **Chapter 4,** six animal-based welfare indicators were accepted after being applied to ewes. These were BCS (Russel et al 1969 scale), tooth loss, coat cleanliness, lameness (Kaler and Green 2006 scale) and faecal soiling. Lamb survival until weaning was the only Iceberg Indicator which was considered to be a valid indicator of ewe welfare. However, lamb birth and weaning weights were rejected as welfare indicators. In some instances, the data obtained in **Chapter 4** did not consist of sufficient variation for analyses to be performed. Before a decision can be made regarding the applicability of nasal discharge, coughing, the FAMACHA anaemia chart, BCS and lameness (both 4 point scales) and udder condition as measured by colour, symmetry and presence of fibroids, further investigating is be required.

In **Chapter 5**, QBA was found to be feasible when performed on an extensively managed farm. Good inter-observer reliability was also found between three observers on two occasions. This appears to be the first study to report these findings which supports the use to QBA as an indicator for use with extensively managed sheep.

The results obtained from the work reported in **Chapter 6** led to the acceptance of six indicators: QBA, vigilance behaviour, nearest neighbour distance, human-approach test, response to surprise and recovery from surprise (these findings are summarised in Table 7.1).

Table 7.1 Measures applied during Assessments at Gather or Assessments in the Field and the outcome of their evaluation

Accepted	Rejected	Requiring further investigation
Body condition score (Russel et al 1969) Tooth loss Lamb survival until weaning Coat cleanliness Lameness (Kaler and Green 2006) Faecal soiling Nearest neighbour distance Qualitative Behavioural Assessment Synchronised vigilance behaviour Human approach Response to surprise Recovery from surprise	Lamb birth and weaning weights	Body condition score - 4 point Nasal discharge Coughing FAMACHA anaemia chart Lameness - 4 point Udder condition: colour, symmetry and presence of fibroids Vocalisation

As mentioned previously, the indicators selected for evaluation in this project were categorised using the framework previously developed by Welfare Quality of Good Feeding, Good Environment, Good Health and Appropriate Behaviour. The indicators and specific rationale behind the decision regarding their acceptance are discussed below.

7.3.1 Indicators relating to Good Feeding

The indicators selected in this study for the assessment of Good Feeding were body condition score (using two assessment methods: the scale devised by Russel et al 1969, and a novel simplified scale), tooth loss, and lamb production traits: growth and survival. These indicators were assessed and reported on in **Chapter 4**. Prior to this project the Russel et al 1969 body condition score was already proven to be valid, feasible and reliable when applied to sheep on-farm (Caldeira et al 2007, Morgan-Davis et al 2008). Work by Phythian et al (2012) suggested that during an on-farm welfare assessment this 6 point scale may be too intricate and a simplified scale may prove to be beneficial. A simplified scale was thought to allow assessors to quickly identify animals which were likely to be too fat or too thin, and therefore likely to be in poorer welfare states. The validity of a novel four point (emaciated, thin, fit, fat) scale was therefore assessed. A similar scale was developed by Phythian (2012) in which they divided animals in to three categories, thin, fit or fat and the inter- and intra-observer reliability of their scale was found to be good. In the project reported in this thesis the additional category of emaciated was included following suggestions made during a consultation with industry. For extensively managed animals inadequate nutrition is a major welfare problem (Humann-Zeikhank et al 2008, Phythian et al 2011) and therefore an additional category was deemed necessary in order to assess the severity of this problem and identify animals in a severely undernourished state. This four point scale was found to correspond well with the Russel scale but not with physiological measures collected at the same time. The possible reasons for this were discussed in **Chapter 4** and although promising, the simplified body condition score requires further investigation before it can replace the Russel et al scale and be accepted as a welfare indicator for use with extensively managed sheep.

The animals' tooth loss scores were found to provide valuable information regarding their welfare state. The validity of tooth loss as a welfare indicator was already established (Tatara et al 2014, Erjaved and Crossely 2010), and it proved to be a feasible measure to perform during the Assessments at Gather. In this study, the measure was found to be a predictor of removal from the flock up to a year in advance. This appears to have been an original finding not previously reported in the literature and provided additional validity for this measure. The inter- and intra-observer reliability of the measure has not been tested formally although based on the validity and feasibility of the measure it is considered to be a suitable measure for inclusion in a welfare assessment of sheep.

The possibility of using lamb growth and mortality as iceberg indicators was considered. Based on the relationships between ewe welfare and data collected on the lambs in **Chapter 4**, lamb mortality, but not lamb growth, was accepted as a potential iceberg indicator. This finding was in accordance with previous literature (Stott et al 2012).

7.3.2 Indicators relating to Good Environment

The only potential welfare indicator deemed suitable for unhoused sheep and assessed in this project was coat cleanliness. The inter- and intra-observer reliability of this measure using different scales has been established in previous studies (Phythian et al 2012, Napolitano et al 2009) and further detail of this can be found in **Chapter 2**. In the work presented in this thesis, the coat cleanliness of the ewes was assessed during the Assessments at Gather and Assessments in the Field. Coat cleanliness was found to vary considerably when assessed across the year. In both years the ewes were the most dirty prior to mating which coincided with the beginning of winter. From this result it appears that coat cleanliness is a valid indicator of the environmental cleanliness and conditions in which the animal is living. This measure was found to be feasible when the animals were handled during the Assessments at Gather and from a reasonable distance during the Assessment in the Field. For these reasons coat cleanliness was accepted as a welfare indicator for unhoused sheep.

7.3.3 Indicators relating to Appropriate Behaviour

The indicators selected to assess the Appropriate Behaviour of the ewes, were applied during the Assessments in the Field, reported in **Chapters 5** and **6**. These were: QBA, nearest neighbour distance, synchrony and vigilance, vocalisations during a two minute period, human approach test, and reaction to, and recovery from, a surprise test. All indicators listed were found to be feasible for use with sheep in an open environment. The only indicator in this project which was assessed for inter-observer reliability was QBA. This was the first time the inter-observer reliability of a fixed list of QBA terms has been assessed using individually identified sheep in extensive conditions across a six month time period. Previous studies used the free choice profiling method, or observed sheep on video (Phythian 2013, Wickham et al 2012). In **Chapter 5**, good inter-observer reliability was found for three observers on two occasions. Systematic and meaningful changes in sheep expression were identified over the six month time period when observed by one person (SER). In **Chapter 5**, four PCs of sheep expression were described. The two primary components identified were General Mood and Arousal, with a further two subcomponents of Listlessness and Aggressiveness. The components were in line with results reported in the literature (Phythian et al 2012, Wemelsfelder and Farish 2004). In **Chapter 6**, these components were found to relate to other welfare indicators (including established social, and behaviours indicative of fear) maintaining the established validity of the measure (Muri et al 2013, Phythian et al 2011, Sant'Anna & Paranhos da Costa, 2013).

The social behaviours of the ewes: nearest neighbour distance, group synchrony and vigilance, and vocalisations were found to be informative, feasible and repeatable measures. The findings from the Assessments in the Field are discussed in detail in **Chapter 6**. The behaviours indicative of general fear, i.e. response to human during an approach test, and the response to and recovery from a surprise test were found to be well correlated with each other, indicating they are providing similar information regarding the animal's general fear levels. The feasibility and validity of the measures applied during the Assessment in the Field support their acceptance as welfare indicators for use with sheep in extensive environments. However, they do still require inter- and intra-observer reliability testing, particularly those requiring the observer to visually estimate a distance (as discussed in **Chapter 6**). As indicators such as the response to, and recovery from a surprise test appear to be measuring very similar aspects of sheep welfare, once they have been tested for inter-observer reliability it may be worthwhile to assess whether it is possible to obtain the same information regarding the animals' welfare using only one of the measures. The measure which is deemed to be the most informative and reliable may be selected. This may improve the feasibility for use during a welfare assessment as it would reduce the time taken to perform the assessment.

These behavioural indicators were not found to be affected by physical health and comfort traits also observed during the Assessment in the Field: coat cleanliness, faecal soiling and lameness. This indicates they were assessing a different and complementary aspect of the animals' welfare and therefore should be retained as welfare indicators.

7.3.4 Indicators relating to Good Health

Three indicators assessing Good Health (lameness, coat cleanliness and faecal soiling) were applied during the Assessment in the Field as described above and in

Chapter 6. These indicators were also applied during the Assessments at Gather (**Chapter 4**) in addition to measures assessing respiratory disease, udder condition, and anaemia, and were considered to address the most prevalent and important health conditions affecting sheep welfare.

Previous work reported in the literature reported that the FAMACHA chart was a valid assessment of anaemia in sheep and goats (Mederos et al 2014, Bath and Van Wyk 2009, Papadopolous et al 2013). The results from the work presented in this thesis (**Chapter 4**) did not support these findings. When evaluated using the data obtained from the Assessments at Gather, although feasible, it was not found to be statistically related to physiological parameters indicative of anaemia e.g. red blood cell count. Based on these results it cannot be recommended as a welfare indicator with extensively managed sheep breeds such as the Scottish Blackface without further evaluation, and possible modification. As discussed in **Chapter 4** it is possible that it was the dark pigmentation of the sheep used in this project may have affected the results, but also standardisation of the environment, such as light intensity also merits further investigation.

Data on faecal soiling and lameness were collected during the Assessments at Gather and Assessments in the Field (**Chapters 4** and **6**). Faecal soiling scores peaked in the summer months during lactation and weaning. The validity of faecal soiling as an indicator of gastro-intestinal parasite burden has been validated in other studies however in this work it was only found to be significantly related to the *Coccidia* species egg counts during late pregnancy in Year 2. The potential reasons for these results are discussed in **Chapter 4**. The inter- and intra-reliability of this measure does not appear to have been explicitly tested in the literature and further research regarding the use of the pictorial scale may prove to be valuable. The measure was found to be feasible for use in both assessment approaches and despite the low agreement with other indicators in this study, it has been accepted as a welfare indicator for extensively managed sheep.

As for the novel body condition score, the simplified lameness score was developed and proposed to increase feasibility while retaining the integrity of the indicator and maintaining the ability to identify animals experiencing poor welfare. When scoring the lameness of ewes during the Assessments in the Field, Kaler and Green's scale offered sufficient variation in scores to allow statistical analyses, however the simplified scale did not. This was problematic as one of the reasons for selecting the simplified scale was the prospect that it would be less sensitive to fluctuations in the animal's gate caused by irregular terrain leading to erroneous recording. Further work is required before a decision can be made regarding the simplified scale and whether it is appropriate for use with animals walking on uneven ground. In the meantime as the validity and reliability of the Kaler and Green (2009) scale is already established and it proved to be feasible to assess in both assessments it is considered suitable for use when assessing the welfare of these sheep.

Although feasible to assess during the Assessments at Gather, coughing and nasal discharge were not sufficiently prevalent to be included in the analysis. The measures selected for use to assess udder condition during lactation also did not show enough variability between ewes or assessments.

7.4 Limitations of study

7.4.1 Lack of variation/prevalence for some health indicators

Without being able to perform analyses on indicators due to low prevalence e.g. coughing, an informed decision regarding their use as a welfare indicator could not be made. This problem is not unique to this study and has been faced by other authors. For example in their development of a welfare assessment Phythian et al (2012) found insufficient numbers of animals showing coughing, skin irritation, panting and abdominal soiling. Stubsjoen et al (2011) did not observe between variation in lesion scores, disease, lameness, udder problems and coughing among others. The opposite problem in which diarrhoea, problems associated with heat stress, missing teeth and ectoparasites were too prevalent in a sample group of working horses also resulted in Burn et al (Burn et al., 2009) being unable to satisfactorily assess these measures statistically. In order to fully examine the validity and reliability of these measures, data would have to be collected on a population of sheep in which the prevalence of respiratory and udder problems is higher.

It has been suggested that the optimum prevalence of conditions for these kind of assessments and evaluations is approximately 50% (Hoehler 2000). As the work reported in this thesis was not conducted under controlled experimental conditions this was not achievable or realistic. Although respiratory and udder conditions are accepted as valid welfare indicators for sheep (Blagitz et al 2014, Bell et al 2008), and the assessments proved to be feasible it is not possible to make a statement regarding their use as a welfare indicator for extensively managed sheep. In future work, controlled experiments may be performed using flocks with higher incidences of these problems in order to confirm their used as welfare indicators.

7.4.2 One farm

A major limitation of this work was the fact the data were collected on one farm on which there was only one breed of sheep. This low variability in genetics, environment and management procedures was most likely the cause of the low variability in results of some welfare indicators (e.g. udder problems). These limitations make it difficult to extrapolate the data further and care must be taken when making generalised statements regarding other breeds, management systems, or animals in other countries. Another limitation is the lack of control I had regarding the management of the animals. The management decisions were made independently of this work and these may have had an impact upon the strength of the findings, for example only male lambs being weaned in August 2013. As the ewes were permanently unhoused it was not possible for me to control their environment in the same way as may be possible in an intensive or more typical experimental environment. Also the land at Castlelaw farm is owned not by the college but by the Ministry of Defense, and the presence of soldiers and cadets was also an additional factor over which I had no control. It is likely the behaviour of the sheep may have been affected by the soldiers' regular presence and made them either over or under sensitive to human presence. When considering some of these results and extrapolating to other farms care must be taken before suggesting that these ewes are representative of all sheep in extensive systems.

7.4.3 Implications of reproductive cycle and calendar season on the outcomes of a welfare assessment

Reproductive cycle stage and season of calendar year were both found to have significant effects on the results of indicators applied during the Assessments at Gather and the Assessments in the Field. The specific details and proposed reasons for these effects have been discussed in **Chapters 4**, **5** and **6**. These findings are pertinent in terms of the interpretation of results collected during on-farm welfare assessments. Assessments comparing between farms, or the same farm at different times of year, may yield different results. This potential problem is not specific to this study and has been faced by others (Phythian et al 2015, Temple et al 2013). In this thesis it is not possible to give a definite answer to this challenge which is part of a larger quandary faced by this area of science, but adds to the debate and highlights the importance of understanding these potentially confounding factors. A potential way in which to account for this is to build up a database of a reference population and compare the scores of commercial farms to others in this group. Providing the data collected for such a reference population consisted of sufficient variation in terms of animals in different stages of reproduction and lactation, and the time at which the assessments were performed to account for disparity due to environmental factors, it would further our understanding of the normal predicted shifts in indicator scores. The comparison between an individual farm and the reference population would allow a judgement to be made regarding the relative welfare state of animals on that particular farm. The farms assessed could be ranked according to their relationship with the reference population and the stability of the ranking may be an indicator of the severity of the welfare fluctuations seen on individual farms. Tracking the ranking of farms would allow for genuine improvements or reductions in welfare states to be recognised regardless of reproductive or calendar season.

7.4.4 Implications of environmental conditions

The bad wintery weather in the spring of 2013 undoubtedly had an effect on the welfare of the ewes in Year 2. The difference in results obtained during the Assessments at Gather in Year 2 compared to Year 1 was apparent in Chapter **4**. It was expected that animals which are unhoused all year round, such as extensively managed sheep, are susceptible to welfare problems caused by extremes in weather and environmental conditions (Dwyer 2008). However, the length of time taken for such animals to return to a welfare level prior to the challenge is unknown. As the ewes were pregnant during this time it would also be particularly interesting to know the potential long term effects of on their offspring. The effects of maternal stress during gestation is of interest both from a welfare and production perspective.

Studies in livestock species have shown that acute stressors faced by a female while she is pregnant can result in the offspring experiencing alterations in their physiology, such as increased pain reactivity and poorer growth rates, and heightened emotional reactivity and behaviour responses to fearful stimuli (Averos et al 2015, Chojnacki et al 2014, Rutherford et al 2009). Prolonged periods of hunger during gestation have been shown to have effects on the lamb's health in with a decrease in birth weight, glucose tolerance, cardio-vascular dysfunction and blood pressure (Todd et al 2009, Ford et al 2007). Under nutrition during pregnancy has also been shown to affect the maternal behaviour of the ewe towards the lamb, as ewes which received a 30% reduction in feed intake during pregnancy spent less time licking the lamb and displayed more aggressive behaviours towards their lambs post-birth (Dwyer et al 2003). The impacts of these events on the foetal and neonatal lamb may significantly affect its welfare both as a juvenile and adult. It would have been interesting to have followed these ewes and lambs for an extended time in order to investigate the longitudinal consequences of this weather and subsequent nutritional limitations on both parties.

7.5 Subsequent work

The measures which were accepted as welfare indicators based on the results from this project were included in a preliminary protocol for use in a further development study, which took place on another SRUC owned sheep farm, where experimental treatments (worm control) were applied to induce differences in welfare state. Although I provided advice regarding the practicalities of collecting this data the further development study was conducted by research assistants from the Animal Behaviour and Welfare team. This further development project aimed to assess the inter-observer reliability of the physical health indicators. Additional data were also collected by AWIN members from Neiker Tecnalia in Spain. They applied the preliminary protocol when performing welfare assessments on 30 Spanish dairy sheep farms. Following these assessments the indicators which were which were found to be reliable were included in a preliminary protocol for use on commercial farms. Additional management-based measures such as the length of docked tails, were also incorporated for the commercial farm visits. Following the commercial farm visits by SRUC research assistants an expert meeting was held during which AWIN stakeholders (including members of Quality Meat Scotland and National Farmers' Union Scotland) were able to voice their scientific opinions on the indicators and preliminary protocol. The participants at this meeting, along with additional 30 international stakeholders, were asked to complete a pair wise ranking of the indicators and specify which indicators they believed to be the most informative and valid regarding the welfare issues faced by extensively managed sheep. Following this meeting and results obtained from the pair wise comparison a final protocol for use with extensively managed sheep for the AWIN project was devised (AWIN, 2015).

7.6 Final AWIN protocol

The work presented in this thesis was the first step in the development of an on-farm welfare assessment which could be applied to extensively managed sheep. The resulting protocol has been delivered to the EU and has been published. Although the intent of the EU in terms of who will perform the application of the final welfare protocol remains unclear the project has ultimately delivered a practical tool. This tool can be used to assess the welfare of sheep in a feasible manner in extensive conditions extending from remote hill farms in Scotland to freely ranging sheep flocks in southern Europe.

The protocol is easy to use and therefore sheep farmers could use it routinely assess their own animals for health problems, such as lameness and tooth loss. They could apply the protocol themselves and monitor their flock, potentially identifying welfare problems earlier and preventing avoidable suffering on their farms.

There are some indicators, such as QBA, which although offer a quick and reliable assessment of welfare require specific training. For these indicators training could be given to allow farmers to perform these measures themselves or the EU may wish to train specific inspectors who could regularly visit farms as part of an official assessment. If the EU were to use this protocol and standardise assessor training it would help to maintain the consistency in application between farms and ensure the results obtained are meaningful. The final published AWIN protocol could be used for an accreditation scheme which are becoming increasingly popular (Mellor et al 2015) to provide consumers with the option of making informed decisions when buying products derived from sheep living in extensive systems.

7.7 Conclusion

The main aim of this PhD project was to evaluate potential welfare indicators for use with extensively managed sheep. This was achieved by applying welfare indicators to Scottish Blackface ewes on a Scottish hill farm over a two year period. Out of 19 indicators evaluated in this project 13 met the pre-determined criteria of feasibility, reliability and validity. These indicators were therefore considered to be suitable for use when assessing the welfare of extensively managed sheep and were accepted for use with adult breeding ewes. These indicators covered aspects of welfare relating to

Good Feeding, Good Health, Good Environment and Appropriate Behaviour. Some of these indicators had been previously developed for use with other species or for sheep living in intensive housing systems and required modification before use in extensive environments. In addition, entirely novel approaches, such as the "surprise test", were created and developed in this PhD project to meet the specific needs of assessors performing welfare assessments open environments. The work presented in this thesis was the first step in the development of an on-farm welfare assessment protocol which could be applied to extensively managed sheep as part of the AWIN project. The resulting final protocol has been published and delivered to the EU. Applying the final protocol to sheep in meat and dairy production systems has the potential to have a real impact and improve the welfare of these animals across the European Union.

References

Aivaz, a. N., Ruckstuhl, K.E., 2011. Costs of behavioral synchrony as a potential driver behind size-assorted grouping. Behav. Ecol. 22, 1353–1363.

Albenzio, M., Taibi, L., Musico, A., Sevi, A., 2002. Prevalence and etiology of subclinical mastitis in intensively managed flocks and related changes in the yield and quality of ewe milk. Small Rumin. Res. 43, 219–226.

Alexander, R.D., 1974. The evolution of social behaviour. Annu. Rev. Ecol. Evol. Syst. 5, 325–383.

Allerton, G.R., Gogoleweski, R.P., Rugg, D., Plue, R.E., Barrick, R.A., Eagleson, J.S., 1998. Field trials evaluating ivermectin controlled release capsules for weaner sheep and breeding ewes. Aust. Vet. J. 76, 39–43.

Anderson, C.A., 1989. Temperature and aggression: ubiquitous effects of heat on occurance of human violence. Psychol. Bull. 46, 91–97.

Andreasen, S.N., Wemelsfelder, F., Sandøe, P., Forkman, B., 2013. The correlation of Qualitative Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. Appl. Anim. Behav. Sci. 143, 9–17.

Annett, R.W., Carson, A.F., Dawson, L.E.R., Irwin, D., Gordon, a W., Kilpatrick, D.J., 2011. Comparison of the longevity and lifetime performance of Scottish Blackface ewes and their crosses within hill sheep flocks. Animal 5, 347–55. doi:10.1017/S1751731110002107

Arnold, G.W., Pahl, P.J., 1974. Some aspects of social behaviour in domestic sheep. Anim. Behav. 22, 592–600.

Asher, L., Collins, L.M., 2012. Assessing synchrony in groups: Are you measuring what you think you are measuring? Appl. Anim. Behav. Sci. 138, 162–169.

Averós, X., Lorea, A., Beltrán de Heredia, I., Arranz, J., Ruiz, R., Estevez, I., 2014a. Space availability in confined sheep during pregnancy, effects in movement patterns and use of space. PLoS One 9, 947 – 967.

Averós, X., Lorea, A., Heredia, I. de, 2014b. The behaviour of gestating dairy ewes under different space allowances. Appl. Anim. Behav. Sci. 150, 17–26.

AWIN, 2015. AWIN Welfare assessment protocol for sheep 1.1. doi:10.13130/AWIN

Azizi, S., Tehrani, A., Dalir-Naghadeh, B., Hemmati, M., 2011. The effects of farming system and season on the prevalence of lameness in sheep in northwest Iran. N. Z. Vet. J. 59, 311–316.

Banks, M.S., Sprague, W.W., Schmoll, J., Parnell, J.A.Q., Love, G.D., 2015. Why do animal eyes have pupils of different shapes ? Sci. Adv. 1, 1–9.

Barbosa, A., 2002. Does vigilance always co-vary negatively with group size? Effects of foraging strategy. Acta Ethol. 5, 51–55.

Barger, I.A., 1993. Influence of sex and reproductive status on suceptibility of ruminants to nematode parasitism. Int. J. Parasitol. 23, 463–469.

Barnard, S., Matthews, L.R., Messori, S., Podaliri Vulpiani, M., Ferri, N., 2015. Behavioural reactivity of ewes and lambs during partial and total social isolation. Appl. Anim. Behav. Sci. 163, 89–97.

Barnett, J.L., Hemsworth, P.H., 2009. Welfare monitoring schemes: using research to safeguard welfare of animals on the farm. J. Appl. Anim. Welf. Sci. 12, 114–31.

Bartussek, H., Leeb, C., Held, S., 2000. Animal Needs Index for cattle. Bal gumpenstein, Irdning.

Bateson, P., 1991. Assessment of pain in animals. Anim. Behav. 42, 827–839.

Bath, G., Malan, F., van Wyk, J.A., 1996. "FAMACHA" ovine anaemia guide to assist with the control of haemonchosis, in: 7 Th Annual Congress of Livestock Health and Production Group of the South African Veterinary Association. Port Elizabeth, pp. 5 – 7.

Bath, G.F., van Wyk, J.., 2009. The Five Point Check© for targeted selective treatment of internal parasites in small ruminants. Small Rumin. Res. 86, 6–13.

Battini, M., Andreoli, E., Barbieri, S., Mattiello, S., 2011. Long-term stability of Avoidance Distance tests for on-farm assessment of dairy cow relationship to humans in alpine traditional husbandry systems. Appl. Anim. Behav. Sci. 135, 267–270.

Beauchamp, G., 2001. Should vigilance always decrease with group size? Behav. Ecol. Sociobiol. 51, 47–52.

Bell, M.E., Wood, C.E., Keller-Wood, M., 1991. Influence of reproductive state on pituitary-adrenal activity in the ewe. Domest. Anim. Endorinology 8, 245–254.

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. OIE Rev. Sci. Tech. 33, 189–196.

Berger, J., 1991. Pregnancy incentives , predation constraints and habitat shifts : experimental and field evidence for wild bighorn sheep. Anim. Behav. 41, 61–77.

Berger, J., 1978. Group size, foraging, and anti-predator ploys: an analysis of bighorn sheep decisions. Behav. Ecol. Sociobiol. 4, 91–99.

Bergonier, D., Berthelot, X., 2003. New advances in epizootiology and control of ewe mastitis. Livest. Prod. Sci. 79, 1–16.

Binns, S., Cox, I., Rizvi, S., Green, L., 2002. Risk factors for lamb mortality on UK sheep farms. Prev. Vet. Med. 52, 287–303.

Black, H., Matthews, L.R., Bremner, K.J., 1994. The behaviour of male lambs transported by sea from New Zealand to Saudi Arabia. N. Z. Vet. J. 42, 16–23.

Blagitz, M.G., Souza, F.N., Batista, C.F., Diniz, S.A., Haddad, J.P.A., Benites, N.R., Melville, P.A., Della Libera, A.M.M.P., 2014. Clinical findings related to intramammary infections in meat-producing ewes. Trop. Anim. Health Prod. 46, 127–32.

Bleich, V.C., Bower, R.T., Wehausen, J.D., 1997. Sexual segregation in mountain sheep: resources or predation? Wildl. Monogr. 134, 3–50.

Blockhuis, H.J., Miele, M., Veissier, I., Jones, B., 2013. Improving farm animal welfare, 1st ed. Wageningen Academic Publishers, Springer, The Netherlands.

Blokhuis, H.J., Jones, R.B., Geers, R., Miele, M., Veissier, I., 2003. Measuring and monitoring animal welfare: Tranparency in the food product quality chain. Anim. Welf. 12, 445–455.

Bloxham, G.P., Purton, D.G., 1991. Demineralisation and incisor wear: An in vitro study. New Zeal. J. Agric. Res. 34, 277–279.

Bøe, K., 2012. Feed intake and competition in ewes on ad libitum or semi-restricted feeding of grass silage and hay. J. Anim. Sci. 90, 4014–4019.

Bøe, K.E., Berg, S., Andersen, I.L., 2006. Resting behaviour and displacements in ewes—effects of reduced lying space and pen shape. Appl. Anim. Behav. Sci. 98, 249–259.

Boissy, A., Aubert, A., Desire, L., Greiveldinger, L., Delval, E., Veissier, I., 2011. Cognitive sciences to relate ear postures to emotions in sheep. Anim. Welf. 20, 47–56.

Boissy, A., Bouissou, M., 1995. Assessment of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations. Appl. Anim. Behav. Sci. 46, 17–31.

Boissy, A., Dumont, B., 2002. Interactions between social and feeding motivations on the grazing behaviour of herbivores: sheep more easily split into subgroups with familiar peers. Appl. Anim. Behav. Sci. 79, 233–245.

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. Physiol. Behav. 92, 375–97.

Boivin, X., Lensink, J., Tallet, C., Veissier, I., 2003. Stockmanship and farm animal welfare. Anim. Welf. 12, 479–492.

Boivin, X., Neindre, P. Le, Garel, J.P., Chupin, J.M., 1994. Influence of breed and rearing management on cattle reactions during human handling. Appl. Anim. Behav. Sci. 39, 115–122.

Bokkers, E., de Vries, M., Antonissen, I., de Boer, I., 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. Anim. Welf. 21, 307–318.

Bon, R., 2012. Group size elicits specific physiological response in herbivores. Biol. Lett. 537–539.

Botreau, R., Veissier, I., Perny, P., 2009. Overall assessment of animal welfare: strategy adopted in Welfare Quality ®. Anim. Welf. 18, 363–370.

Boyazoglu, J., Morand-fehr, P., 2001. Mediterranean dairy sheep and goat products and their quality: A critical review. Small Rumin. Res. 40, 1–11.

Bracke, M., Edwards, S., Engel, D., Buist, G., Algers, B., 2008. Expert opinion as "validation" of risk assessment applied to calf welfare. Acta Vet. Scand. 50, 29–40.

Brambell, F.W.R., 1969. Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems. London: Her Majesty's Stationary Office., London.

Breuer, K., Hemsworth, P.., Coleman, G.., 2003. The effect of positive or negative handling on the behavioural and physiological responses of nonlactating heifers. Appl. Anim. Behav. Sci. 84, 3–22.

Breuer, K., Hemsworth, P.H., Barnett, J.L., 2000. Behavioural response to humans and the productivity of commercial dairy cows. Appl. Anim. Behav. Sci. 66, 273–288.

Broom, D., Corke, M., 2002. Effects of disease on farm animal welfare. Acta Vet. Brno 71, 133–136.

Broughan, J.M., Wall, R., 2007. Faecal soiling and gastrointestinal helminth infection in lambs. Int. J. Parasitol. 37, 1255–68.

Buckley, F., O'Sullivan, K., Mee, J.F., Evans, R.D., Dillon, P., 2003. Relationship among milk yeild, body condition, cow weight, and reproduction in spring-calved Holstein Freisians. J. Dairy Sci. 2308–2319.

Burkholder WJ, 2000. Assessment of the provision of optimal nutrition. J. Am. Vet. Med. Assoc. 217, 650–654.

Burn, C., Pritchard, J., Whay, H., 2009. Observer reliability for working equine welfare assessment: problems with high prevalences of certain results. Anim. Welf. 18, 1–27.

Caldeira, R.M., Belo, A.T., Santos, C.C., Vazques, M.I., Portugal, A.V., 2007a. The effect of body condition score on blood metabolites and hormonal profiles in ewes. Small Rumin. Res. 68, 233–241.

Caldeira, R.M., Belo, A.T., Santos, C.C., Vazques, M.I., Portugal, A.V., 2007b. The effect of long-term feed restriction and over-nutrition on body condition score, blood metabolites and hormonal profiles in ewes. Small Rumin. Res. 68, 242–255.

Canali, E., Keeling, L., 2009. Welfare Quality ® project: from scientific research to on farm assessment of animal welfare. Ital. J. Anim. Sci. 8, 900–903.

Capdeville, J., Veissier, I., 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Agric. Scand. 30, 62–68.

Carneiro, H., Louvandini, H., Paiva, S.R., Macedo, F., Mernies, B., Mcmanus, C., 2010. Morphological characterization of sheep breeds in Brazil, Uruguay and Colombia. Small Rumin. Res. 94, 58–65.

Carol, C., Huntingdon, P., 1988. Body condition scoring and weight estimation in horses. Equine Vet. J. 20, 42–45.

Caroprese, M., Casamassima, D., 2009. Monitoring the on-farm welfare of sheep and goats. Ital. J. Anim. Sci. 8, 343–354.

Cassinello, J., 1995. Factors modifying female social ranks in Ammotragus. Appl. Anim. Behav. Sci. 45, 175–180.

Caulfield, M.P., Cambridge, H., Foster, S.F., McGreevy, P.D., 2014. Heat stress: a major contributor to poor animal welfare associated with long-haul live export voyages. Vet. J. 199, 223–8.

Cherry, M.J., Conner, L.M., Warren, R.J., 2015. Effects of predation risk and group dynamics on white-tailed deer foraging behavior in a longleaf pine savanna. Behav. Ecol. 26, 1091–1099.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R.R., Pemberton, J.M., Beraldi, D., Stear, M.J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M.H., Kazwala, R.R., Zhang, Y., Arranz, J.J., 2009. Revealing the history of sheep domestication using retrovirus integrations. Science. 324, 532–536.

Chronbach, L., Meehl, P., 1955. Construct validity in phsychological tests. Psychol. Bull. 524, 281–302.

Clark, C., Mendl, M., Jamieson, J., Arnone, A., Waterman-Pearson, A., Murrell, J., 2011. Do psychological and physiological stressors alter the acute pain response to castration and tail docking in lambs? Vet. Anaesth. Analg. 38, 134–45.

Clark, C., Murrell, J., Fernyhough, M., Rourke, T.O., Mendl, M., 2014. effects of neonatal experience on sheep behaviour. Biol. Lett. 10, 1–4.

Clements, A.C., Taylor, D.J., Fitzpatrick, J.L., 2003. Evaluation of diagnostic procedures for subclinical mastitis in meat-producing sheep. J. Dairy Res. 70, 139–148.

Cockram, M.S., Hughes, D., 2011. Health and disease, in: Appleby, M., Hughes, B., Mench, J. (Eds.), Animal Welfare. CABI Publishing, Wallingford, U.K., p. Chapter 8.

Cockram, M.S., Imlah, P., Goddard, P.J., Harkiss, G.D., Waran, N.K., 1993. The behavioural, endocrine and leucocyte response of ewes to repeated removal of lambs before the age of natural weaning. Appl. Anim. Behav. Sci. 38, 127–142.

Colebrook, E., Wall, R., 2004. Ectoparasites of livestock in Europe and the Mediterranean region. Vet. Parasitol. 120, 251–74.

Conington, J., Bishop, S.C., Grundy, B., Waterhouse, A., Simm, G., 2001. Multi-trait selection indexes for sustainable UK hill sheep production. Anim. Sci. 73, 413–423.

Conradt, L., List, C., 2009. Group decisions in humans and animals: a survey. Philos. Trans. R. Soc. B 1518, 719–742.

Cordeiro, A.F., Naas, I.A., Oliviera, S.R.M., Violaro, D., de Almeida, A.C., Neves, D.P., 2013. Understanding vocalisation might help to assist stressful conditions in piglets. Animals 3, 923–934.

Coulon, M., Nowak, R., Andanson, S., Ravel, C., Marnet, P., Boissy, A., Boivin, X., 2013. Human-lamb bonding: oxytocin, cortisol and behavioural responses of lambs to human contacts and social separation. Psychoneuroendocrinology 38, 499–508.

Cronin, G., Barnett, J., Edge, J., Hemsworth, P., 2002. Identifying animal welfare issues for sheep in Australia. Int. J. Sheep Wool Sci. 50, 534–550.

D'Souza, D.N., Leury, B.J., Dunshea, F.R., Warner, R.D., 1998. Effect of on-farm and pre-slaughter handling of pigs on meat quality. Aust. J. Agric. Res. 49, 1021 – 1025.

Das, N., 2001. Factors influencing the inactive behaviours of stall-fed sheep under experimental conditions. Small Rumin. Res. 42, 39–47.

Dawkins, M.., 1990. From an aimals point of view motivation fitness and animal welfare. Behav. Brain Sci. 13, 1–10.

Dawkins, M.S., 2006a. Behaviour as a tool in the assessment of animal welfare. Zoology 106, 383–387.

Dawkins, M.S., 2006b. Through animal eyes: what behaviour tells us. Appl. Anim. Behav. Sci. 100, 4–10.

de Vries, M., Bokkers, E. a M., Dijkstra, T., van Schaik, G., de Boer, I.J.M., 2011. Invited review: associations between variables of routine herd data and dairy cattle welfare indicators. J. Dairy Sci. 94, 3213–28.

Deag, J.M., 1996. Behavioural ecology and the welfare of extensively farrned animals. Appl. Anim. Behav. Sci. 95, 9–22.

DEFRA, 2014. Identifying and recording sheep and goats: Requirements for English slaughterhouses. London, UK.

DEFRA, 2003. Code of recommendations for the welfare of livestock: Sheep, Defra Publications.

Degrabriele, R., Fell, L, R., 2001. Changes in behaviour, cortisol and lymphocyte types during isolation and group confinement of sheep. Immunol. Cell Biol. 79, 583–589.

Donkelaar, H.J., Nemcova, V., Lammens, M., Overeem, S., Keyser, A., 2011. The autonomic nervous system, in: Donkelaar, H.J. (Ed.), Clinical Neuroanatomy. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 565–601.

Dove, H., Milne, J., 1991. An evaluation of the effects of incisor dentition and of age on the performance of lactating ewes and their lambs. Anim. Prod. 53, 183–190.

Duddy, G., Bell, A., Shands, C., Hegarty, R., 2007. Feedlotting Lambs, Primefact. New South Wales, Australia.

Dukas, R., Clark, C.W., 1995. Sustained vigilance and animal performance. Anim. Behav 49, 1259–1267.

Dwyer, C., 2008. The welfare of the neonatal lamb. Small Rumin. Res. 76, 31–41.

Dwyer, C.M., 2009. Welfare of sheep: Providing for welfare in an extensive environment. Small Rumin. Res. 86, 14–21.

Dwyer, C.M., 2008. Environment and the Sheep. Breed Adaptations and Welfare Implications, in: Dwyer, C.. (Ed.), The Welfare of Sheep. Springer Science and Buiness Media, pp. 41 – 79.

Dwyer, C.M., 2004. How has the risk of predation shaped the behavioural responses of sheep to fear and distress? Anim. Welf. 13, 269–281.

Dwyer, C.M., Bornett, H.L.I., 2004. Chronic stress in sheep: assessment tools and their use in different management conditions. Anim. Welf. 13, 293–304.

Dwyer, C.M., Dingwall, W.S., Lawrence, A.B., 1999. Physiological correlates of maternal-offspring behaviour in sheep: a factor analysis. Physiol. Behav. 67, 443–54.

Dwyer, C.M., Gilbert, C.L., Lawrence, A.B., 2004. Prepartum plasma estradiol and postpartum cortisol, but not oxytocin, are associated with interindividual and breed differences in the expression of maternal behaviour in sheep. Horm. Behav. 46, 529–43.

Dwyer, C.M., Lawrence, A.B., 2008. The Welfare of Sheep, in: Dwyer, C.M. (Ed.), The Welfare of Sheep. Springer Science and Buiness Media B.V, pp. 1–40.

Dwyer, C.M., Lawrence, A.B., 2005. A review of the behavioural and physiological adaptations of hill and lowland breeds of sheep that favour lamb survival. Appl. Anim. Behav. Sci. 92, 235–260.

Dwyer, C.M., Lawrence, A.B., 2000. Effects of maternal genotype and behaviour on the behavioural development of their offspring in sheep. Behaviour 137, 1629–1654.

Dwyer, C.M., Lawrence, A.B., 1999. Ewe – ewe and ewe – lamb behaviour in a hill and a lowland breed of sheep: a study using embryo transfer. Appl. Anim. Behav. Sci. 61, 319–334.

Dwyer, C.M., Lawrence, A.B., Bishop, S.C., Lewis, M., 2003. Ewe-lamb bonding behaviours at birth are affected by maternal undernutrition in pregnancy. Br. J. Nutr. 89, 123–36.

Edgar, J., Mullan, S., Pritchard, J., McFarlane, U., Main, D., 2013. Towards a “Good Life” for Farm Animals: Development of a Resource Tier Framework to Achieve Positive Welfare for Laying Hens. Animals 3, 584–605.

Edwards, D.S., Johnston, A.M., 1999. Welfare implications of sheep ear tags. Vet. Rec. 144, 603–606.

Edwards, D.S., Johnston, A.M., Pfeiffer, D.U., 2001. Acomparison of commonly used ear tags on the ear damage of sheep. Anim. Welf. 10, 141–151.

EFSA Panel on Animal Health and Welfare, 2014. Scientific Opinion on the welfare risks related to the farming of sheep for wool, meat and milk production. EFSA J. 12, 1–128.

EFSA Panel on Animal Health and Welfare, 2012a. Statement on the use of animal-based measures to assess the welfare of animals. EFSA J. 10, 1–29.

EFSA Panel on Animal Health and Welfare, 2012b. Scientific Opinion on the use of animal-based measures to assess welfare of dairy cows. EFSA J. 10, 81.

Eksebo, I., 2011. Farm animal behaviour: Characteristics for assessment of health and welfare, 1st ed. CABI Publishing, Wallingford, U.K.

Elgar, M.A., 1989. Predator vigilance and group size in mammals and birds. Biol. Rev. 64, 13–33.

Elsheikha, H.M., 2009. Coccidiosis incidence, control and epidemiology put in focus. Vet. Times 6–9.

Erjavec, V., Crossley, D., 2010. Initial observations of cheek tooth abnormalities in sheep in Slovenia. Vet. Rec. 167, 134–7.

Escobar, M.P., Demeritt, D., 2015. Records, regulation and the practice of animal husbandry. HowSAFE Work. Pap. 4 1–23.

Estevez, I., Andersen, I.-L., Nævdal, E., 2007. Group size, density and social dynamics in farm animals. Appl. Anim. Behav. Sci. 103, 185–204.

Eurostat, E.U., 2015. Meat Production Statistics [WWW Document]. Eurostat Dissem. Database. URL http://ec.europa.eu/eurostat/web/products-datasets/-/apro_mt_lssheep

Everett-Hincks, J.M., Dodds, K.G., 2008. Management of maternal-offspring behavior to improve lamb survival in easy care sheep systems. J. Anim. Sci. 86, E259–70.

Everitt, G., 1962. On the assessment of body composition of live sheep and cattle. Anim. Prod. 4, 79–89.

Færevik, G., Andersen, I.L., Bøe, K.E., 2005. Preferences of sheep for different types of pen flooring. Appl. Anim. Behav. Sci. 90, 265–276.

Farm Animal Welfare Council, 1992. FAWC upates the five freedoms. Vet. Rec. 131, 537–544.

Fernández-Juricic, E., Kacelnik, A., 2004. Information transfer and gain in flocks: the effects of quality and quantity of social information at different neighbour distances. Behav. Ecol. Sociobiol. 55, 502–511.

Fisher, A.D., Morris, C.A., Matthews, L.R., 2000. Cattle behaviour: comparison of measures of temperament in beef cattle. Proc. New Zeal. Soc. Anim. Prod. 60, 214–217.

Fisher, M., 2004. A review of the welfare implications of out-of-season extensive lamb produc tion systems in New Zealand. Livest. Prod. Sci. 85, 165–172.

Fisher, M.W., Gregory, N.G., Kent, J.E., Scobie, D.R., Mellor, D.J., Pollard, J.C., 2004. Justifying the appropriate length for docking lambs ' tails - a review of the literature 64, 293–296.

Fitzpatrick, J., Scott, M., Nolan, A., 2006. Assessment of pain and welfare in sheep. Small Rumin. Res. 62, 55–61.

Fleming, P.A., Wickham, S.L., Stockman, C.A., Verbeek, E., Matthews, L., Wemelsfelder, F., 2015. The sensitivity of QBA assessments of sheep behavioural expression to variations in visual or verbal information provided to observers. Animal 9, 878–87.

Forbes, J.M., 2006. Feeding in domestic vertebrates: from structure to behaviour, in: Bels, V. (Ed.), Feeding in Domestic Vertebrates from Structure to Behaviour. CABI Pub, Wallingford, U.K., pp. 349–358.

Forbes, J.M., 1970. Voluntary food intake of pregnant ewes. J. Agric. Sci. 31, 1222–1227.

Forkman, B., Boissy, A., Meunier-salaun, M., Canali, E., Jones, R., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiol. Behav. 92, 340–374.

Forkman, B., Keeling, L.J., 2009a. Assessment of animal welfare neasures for dairy cattle, beef bulls and veal calves. Welfare Quality Report No. 11. Cardiff, UK.

Forkman, B., Keeling, L.J., 2009b. Assessment of Animal Welfare Measures for Layers and Broilers. Welfare Quality Report No.9.

Forkman, B., Keeling, L.J., 2009c. Assessment of Animal Welfare Measures for Sows, Piglets and Fattening Pigs. Welfare Quality Report No. 10. Cardiff, UK.

Fragkou, I.., Boscos, C.M., Fthenakis, G.C., 2014. Diagnosis of clinical or subclinical mastitis in ewes. Small Rumin. Res. 118, 86–92.

Fraser, A, F., 1983. The behaviour of maintenance and the intensive husbandry of cattle, sheep and pigs. Ecosyst. Env. 9, 1–13.

Fraser, D., 2006. Animal welfare assurance programs in food production: a framework for assessing the options. Anim. Welf. 15, 93–104.

Fraser, D., 1993. Assessing animal well-being: common sense, uncommon science., in: Proceedings of the Food Animal Well-Being Conference Proceedings and Deliberations. West Lafayette, Indiana, pp. 37–54.

Fregonesi, J., Leaver, J., 2002. Influence of space allowance and milk yield level on behaviour, performance and health of dairy cows housed in strawyard and cubicle systems. Livest. Prod. Sci. 78, 245–257.

Fregonesi, J.A., Leaver, J.D., 2001. Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems. Livest. Prod. Sci. 68, 205–216.

Frid, A., Dill, L., 2002. Human-caused disturbance stimuli as a form of predation risk. Conserv. Ecol. 6.

Fthenakis, G., Jones, J., 1990. The effect of inoculation of coagulase-negative staphylococci into the ovine mammary gland. J. Comp. Pathol. 102.

Fuchs, A.R., Ayromlooi, J., Rasmussen, A.B., 1987. Oxytocin reponse to conditioned and nonconditioned stimuli in lactating ewes. Biol. Reprod. 37, 301–305.

Gautrais, J., Michelena, P., Sibbald, A., Bon, R., Deneubourg, J.-L., 2007. Allelomimetic synchronization in Merino sheep. Anim. Behav. 74, 1443–1454.

Goddard, P., Waterhouse, T., Dwyer, C., Stott, A., 2006. The perception of the welfare of sheep in extensive systems. Small Rumin. Res. 62, 215–225.

Goddard, P.J., 2008. The Management of Sheep, in: Dwyer, C.M. (Ed.), The Welfare of Sheep,. Springer Science and Buiness Media B.V, pp. 291–323.

Gonyou, H.W., Stookey, J.M., Mcneal, L.G., 1985. Effect of double decking and space allowance on the performance and behaviour of feeder lambs. J. Anim. Ecol. 60, 1110–1116.

González, F., Muiño, R., Pereira, V., Campos, R., Benedito, J., 2011. Relationship among blood indicators of lipomobilization and hepatic function during early lactation in high-yielding dairy cows. J. Vet. Sci. 12, 251–255.

González-García, E., Tesniere, A., Camous, S., Bocquier, F., Barillet, F., Hassoun, P., 2015. The effects of parity, litter size, physiological state, and milking frequency on the metabolic profile of Lacaune dairy ewes. Domest. Anim. Endocrinol. 50, 32–44.

Gordon, I., Martin, C., Feldman, R., Leckman, J.F., 2011. Oxytocin and Social Motivation. Dev. Cogn. Neurosci. 1, 471–493.

Gordon, K., Siegmann, M., 1991. Suckling behaviour of ewes in early lactation. Physiol. Behav. 50, 1079–1081.

Gougoulis, D.A., Kyriazakis, I., Fthenakis, G.C., 2010. Diagnostic significance of behaviour changes of sheep: A selected review. Small Rumin. Res. 92, 52–56.

Grace, D., Himstedt, H., Sidibe, I., Randolph, T., Clausen, P., 2007. Comparing FAMACHA eye colour chart and haemoglobin colour scale tests for detecting anaemia and improving of bovine panosomosis in West Africa. Vet. Parasitol. 147, 26–39.

Grandin, T., 2001. Cattle vocalisations are associated with handling and equipment problems at beef slaughter plants. Appl. Anim. Behav. Sci. 71, 191–201.

Grandin, T., 1998. Objective scoring of animal handling and stunning practices at slaughter plants. J. Am. Vet. Med. Assoc. 212, 36–39.

Greiveldinger, L., Veissier, I., Boissy, A., 2007. Emotional experience in sheep: predictability of a sudden event lowers subsequent emotional responses. Physiol. Behav. 92, 675–83.

Griffit, W., Veitch, R., 1971. Hot and crowded: Influences of population density and temperature on interpersonal affective behaviour. J. Pers. Soc. Psychol. 17, 92–98.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., 2010. Genetic diversity in farm animals - A review. Anim. Genet. 41, 6–31.

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. Anim. Behav. Sci. 135, 51–56.

Guilford, T., Dawkins, M.S., 1995. What are conventional signals? Anim. Behav. 49, 1689–1695.

Gunn, R., 1969. The effects of calcium and phosphorus supplementation on the performance of Scottish Blackface hill ewes, with particular reference to the premature loss of permanent incisor teeth. J. Agric. Sci. 72, 371–378.

Hadjipieris, G., Holmes, W., 1966. Studies on feed intake and feed utilisation by sheep. J. Agric. Sci. 106, 351–367.

Hasson, D., Arnetz, B.B., 2005. Validation and findings comparing VAS vs . likert scales for psychosocial measurements. Int. Electron. J. Health Educ. 8, 1–14.

Healy, W.B., Cutress, T.W., Michie, C., 1967. Wear of sheep's teeth. New Zeal. J. Agric. Res. 10, 201–209.

Heath, C.A.E., Browne, W.J., Mullan, S., Main, D.C.J., 2014. Navigating the iceberg: reducing the number of parameters within the Welfare Quality® assessment protocol for dairy cows. Animal 8, 1978–1986.

Hemsworth, P., Mellor, D., Cronin, G., Tilbrook, A., 2015. Scientific assessment of animal welfare. N. Z. Vet. J. 63, 24–30.

Hemsworth, P.., 2003. Human–animal interactions in livestock production. Appl. Anim. Behav. Sci. 81, 185–198.

Hemsworth, P.H., Coleman, G.J., 1994. Improving the attitude and behaviour of stockpersons towards pigs and the consequences on the behaviour and reproductive performance of commercial pigs. Appl. Anim. Behav. Sci. 39, 349–362.

Henderson, D.C., 2010. The veterinary book for sheep farmers. Ipswich, UK.

Hinch, G.N., Lecrivain, E., Lynch, J.J., Elwin, R.L., 1987. Changes in maternal-young associations with increasing age of lambs. Appl. Anim. Behav. Sci. 17, 305–318.

Hinch, G.N., Lynch, J.J., Elwin, R.L., Green, G.C., 1990. Long-term associations between Merino ewes and their offspring. Appl. Anim. Behav. Sci. 27, 93–103.

Hoedemaker, M., Prange, D., Gundelach, Y., 2009. Body condition change and ante- and post-partum healthand reproductive performance in German Holstein cows. Reprod. Domest. Anim. 44, 167–173.

Hopewell, L., Rossiter, R., Blower, E., Leaver, L., Goto, K., 2005. Grazing and vigilance by Soay sheep on Lundy island: influence of group size, terrain and the distribution of vegetation. Behav. Processes 70, 186–93.

Horton, G.M.J., Malinowski, K., Burgher, C.C., Palatini, D.D., 1991. The effect of space allowance and sex on blood catecholamines and cortisol , feed consumption and average daily gain in growing lambs. Appl. Anim. Behav. Sci. 32, 197–204.

Houdijk, J.G.M., Kyriazakis, I., Jackson, F., Huntley, J.F., Coop, R.L., 2000. Can an increased intake of metabolizable protein affect the periparturient relaxation in immunity against Teladorsagia circumcincta in sheep? Vet. Parasitol. 91, 43–62.

Humann-Ziehank, E., Ganter, M., Hennig-Pauka, I., Binder, A., 2008. Trace mineral status and liver and blood parameters in sheep without mineral supply compared to local roe deer (Capreolus capreolus) populations. Small Rumin. Res. 75, 185–191.

Hunter, R.F., Milner, C., 1963. The behaviour of individual, related and groups of south country cheviot hill sheep. Anim. Behav. 6, 507–513.

Hutchings, M.R., Kyriazakis, I., Gordon, I.J., Jackson, F., 1999. Trade-offs between nutrient intake and faecal avoidance in herbivore foraging decisions: the effect of animal parasitic status, level of feeding motivation and sward nitrogen content. J. Anim. Ecol. 68, 310–323.

Institute of Grassland and Environmental Research, n.d. How grass grows.

Jackson, F., Varady, M., Bartley, D.J., 2012. Managing anthelmintic resistance in goats — Can we learn lessons from sheep? Small Rumin. Res. 103, 3–9.

Johnson-Wimbley, T.D., Graham, D.Y., 2011. Diagnosis and management of iron deficiency anemia in the 21st century. Therap. Adv. Gastroenterol. 4, 177–84.

Jones, R., Waddington, D., 1992. Modification of fear in domestic chicks, Gallus gallus domesticus, via regular handling and early environmental enrichment. Anim. Behav. 43, 1021–1033.

Jongman, E.C., Morris, J.P., Barnett, J.L., Hemsworth, P.H., 2000. EEG changes in 4-week-old lambs in response to castration, tail docking and mulesing. Aust. Vet. J. 78, 339–343.

Jørgensen, G.H.M., Andersen, I.L., Bøe, K.E., 2009. The effect of different pen partition configurations on the behaviour of sheep. Appl. Anim. Behav. Sci. 119, 66–70.

Kaler, J., Green, L.E., 2008. Naming and recognition of six foot lesions of sheep using written and pictorial information: a study of 809 English sheep farmers. Prev. Vet. Med. 83, 52–64.

Kaler, J., Medley, G.F., Grogono-Thomas, R., Wellington, E.M.H., Calvo-Bado, L. a, Wassink, G.J., King, E.M., Moore, L.J., Russell, C., Green, L.E., 2010. Factors associated with changes of state of foot conformation and lameness in a flock of sheep. Prev. Vet. Med. 97, 237–44.

Kaler, J., Wassink, G.J., Green, L.E., 2009. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. Vet. J. 180, 189–194.

Kaler, J., Wassink, G.J., Green, L.E., 2008. Observer reliability of a locomotion scoring scale in sheep. Ecology and Epidemiology Group, University of Warwick, Coventrary, England.

Keverne, E.B., Kendrick, K.M., 1991. Oxytocin facilitation of maternal behavior in sheep. Ann. New York Acad. Sci. 83–101.

Kilgour, R.J., Waterhouse, T., Dwyer, C.M., Ivanov, I.D., 2008. Farming systems for sheep production and their effect on welfare, in: Dwyer, C.M. (Ed.), The Welfare of Sheep. Springer Science and Buiness Media B.V, pp. 213–265.

Kılıç, İ., Bozkurt, Z., 2013. The Relationship between F armers ' Perceptions and Animal Welfare Standards in Sheep Farms. Asian Australas. J. Anim. Sci. 26, 1329–1338.

Kirkwood, J., 2007. Quality of life: the heart of the matter. Anim. Welf. 16, 3–7.

Knierim, U., Winckler, C., 2009. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality Approach. Anim. Welf. 18, 451–458.

Krebs, S., Danuser, J., Regula, G., 2001. Using a herd monitoring system in the assessment of welfare. Acta Agric. Scand. 30, 78–81.

Kulnigg-Dabsch, S., Evstatiev, R., Dejaco, C., Gasche, C., 2012. Effect of Iron Therapy on Platelet Counts in Patients with Inflammatory Bowel Disease-Associated Anemia. PLoS One 7, 1-6.

Kyriazakis, I., Tolkamp, B.J., Hutchings, M.R., 1998. Towards a functional explanation for the occurrence of anorexia during parasitic infections. Anim. Behav. 56, 265–274.

Lauber, M., Nash, J. a., Gatt, A., Hemsworth, P.H., 2012. Prevalence and Incidence of Abnormal Behaviours in Individually Housed Sheep. Animals 2, 27–37.

Laundre, J.W., Hernandez, L., Altendorf, K.B., 2001. Wolves, elk, and bison: reestablishing the “landscape of fear” in Yellowstone National Park, USA. Can. J. Zool. 79, 1401–1409.

Lawrence, A.B., Conington, J., 2008. Sheep Welfare: A Future Perspective, in: Dwyer, C.M. (Ed.), The Welfare of Sheep. Springer Science Business Media B.V, pp. 343–360.

Lawrence, A.B., Woodgush, D., 1987. Social behaviour of hill sheep; more to it than meets the eye. Appl. Anim. Behav. Sci. 17, 382.

Lawrence, A.B., Wood-Gush, D.G.M., 1988. Home range behaviour and social organisation of Scottish Blackface sheep. J. Appl. Ecol. 25, 25–40.

Le Pendu, Y., Maulblanc, M.L., Briedermann, L., Dubois, M., 1996. Spatial structure and activity in groups of Mediterranean Mouflon (Ovis gmelini): a comparative study. Appl. Anim. Behav. Sci. 46, 201–216.

Leask, R., van Wyk, J.A., Thompson, P.N., Bath, G.F., 2013. The effect of application of the FAMACHA© system on selected production parameters in sheep. Small Rumin. Res. 110, 1–8.

Legarra, A., Ramón, M., Ugarte, E., Pérez-Guzmán, M.D., Arranz, J., 2007. Economic weights of somatic cell score in dairy sheep. Animal 1, 205–12.

Leone, E.H., Estevez, I., 2008. Use of space in the domestic fowl: separating the effects of enclosure size, group size and density. Anim. Behav. 76, 1673–1682.

Lima, S.L., 1998. Stress and decision making under the risk of predation: recent developments from behavioural, reproductive and ecological perspectives. Adv. Study Behav. 27, 215–290.

Lingis, M., Richards, E., Perrone, D., Keller-Wood, M., 2012. Serotonergic effects on feeding, but not hypothalamus-pituitary-adrenal secretion, are altered in ovine pregnancy. Am. J. Physiol. Endocrinol. Metab. 302, 1231–8.

Livesey, C., Marsh, C., Metcalf, J., Laven, R., 2002. Hock injuries in cattle kept in staw yards or cubicles with rubber mats or mattresses. Vet. Rec. 150, 677–679.

Lovatt, F.M., 2010. Clinical examination of sheep. Small Rumin. Res. 92, 72–77.

Lui, S., Masters, D., Ferguson, M., Thompson, A., 2014. Vitamin E status and reproduction in sheep: potential implications for Australian sheep production. Anim. Prod. Sci. 54, 694–714.

Lynch, J.J., Wood-gish, D.G.M., Davies, H.J., 1985. Aggression and nearest neighbours in a flock of Scottish Blackface ewes. Biol. Behav. 10, 215–225.

Machado, V.S., Caixeta, J.A., McArt, A.A., Bicalho, R.C., 2010. The effect of claw horn disruption lesions and body condition score at dry off on survivability, reproductive performance and milk production in the subsequent lactation. J. Dairy Sci. 93, 4071–4078.

Maia, D., Rosalinski-Moraes, F., van Wyk, J.A., Weber, S., Sotomaior, C.S., 2014. Assessment of a hands-on method for FAMACHA© system training. Vet. Parasitol. 200, 165–71.

Main, D.C.J, Leeb, C., Whay, H.R., Hovi, M. and Webster, J., 2004. Bristol Welfare Assurance Programme: Animal based assessment tool for farm animal welfare certification. Bristol, 1–4.

Manteuffel, G., Puppe, B., Schön, P.C., 2004. Vocalization of farm animals as a measure of welfare. Appl. Anim. Behav. Sci. 88, 163–182.

Marai, I.F., El-Darawany, A., Fadiel, A., Abdel-Hafez, M.A., 2007. Physiological traits as affected by heat stress in sheep—A review. Small Rumin. Res. 71, 1–12.

Marchewka, J., Watanabe, T., 2013. Welfare assessment in broiler farms: Transect walks versus individual scoring. Poult. Sci. 92, 2588–2599.

Markwick, G., 2007. Water requirements for sheep and cattle. Prime Facts, NSW Dep. Prim. Ind. 326, 1–4.

Marsden, M., Wood-gush, D., 1986. The use of space by group-housed sheep. Appl. Anim. Behav. Sci. 15, 178–182.

Martelli, G., 2007. Consumers' perception of farm animal welfare: in Italian and European perspective. Itallian J. Anim. Sci. 8, 31–41.

Martin, P.R., Bateson, P.P.G., 1993. Measuring Behaviour: An Introductory Guide, Second. ed. Cambridge University Press.

Martinetto, K., Cugnasse, J.M., 2001. Reaction distance in Mediterranean Mouflon (Ovis gmelini musimon × Ovis sp) in the presence of hikers with a dog on the Caroux plateau. Rev. Ecol. 56, 231–242.

Marx, G., Horn, T., Thielbein, J., Knubel, B., von Borrel, E., 2003. Analysis of pain-related vocalisation in pigs. J. Sound Vib. 266, 687–698.

Massey, A.C., 1992. Microcutic anaemia. Differential diagnosis and management of iron deficiency anaemia. Med. Clin. North Am. 76, 549–566.

Matthews, L.R., 1996. Animal welfare and sustainability of production under extensive conditions: a non-EU perspective. Appl. Anim. Behav. Sci. 49, 41–46.

McBride, G., Arnold, G.W., Alexander, G., Lynch, J.J., 1967. Ecological aspects of behaviour of domestic animals, in: Proceedings of the Ecological Society of

Australia. pp. 133–165.

McCarthy, M., 2005. Pilot monitoring of shipboard environmental conditions and animal performance. Meat Livest. Aust. 223, 1–31.

McGregor, B.A., 2011. Incisor development, wear and loss in sheep and their impact on ewe production, longevity and economics: A review. Small Rumin. Res. 95, 79–87.

McKiernan, F., Houchins, J.A., Mattes, R.D., 2008. Relationships between human thirst, hunger, drinking, and feeding. Physiol. Behav. 94, 700–8.

Mcmanus, C., Paiva, S.R., Araújo, R.O. De, 2010. Genetics and breeding of sheep in Brazil. Rev. Bras. Zootec. 39, 236–246.

Mederos, A., Kelton, D., Peregrine, A.S., VanLeeuwen, J., Fernández, S., LeBoeuf, A., Menzies, P., Martin, R., 2014. Evaluation of the utility of subjective clinical parameters for estimating fecal egg counts and packed cell volume in Canadian sheep flocks. Vet. Parasitol. 205, 568–74.

Meldrum, G.E., Ruckstuhl, K.E., 2009. Mixed-sex group formation by bighorn sheep in winter: trading costs of synchrony for benefits of group living. Anim. Behav. 77, 919–929.

Mellor, D., 2015. Positive animal welfare states and reference standards for welfare assessment. N. Z. Vet. J. 63, 17–23.

Mellor, D.J., Stafford, K.J., 2000a. Acute castration and/or tailing distress and its alleviation in lambs. N. Z. Vet. J. 48, 33–43.

Mellor, D.J., Stafford, K.J., 2000b. Review Article Acute castration and / or tailing distress and its alleviation in lambs Pain-induced distress and its assessment.

N. Z. Vet. J. 48, 33–43.

Mench, J.A., 2003. Assessing Animal Welfare at the Farm and Group Level: A United States Perspective. Anim. Welf. 12, 493–503.

Mendl, M., Burman, O.P., Paul, E.S., 2010. An integrative and functional framework for the study of animal emotion and mood. Proc. R. Soc. Biol. 277, 2895–904.

Michelena, P., Deneubourg, J.-L., 2011. How group size affects vigilance dynamics and time allocation patterns: the key role of imitation and tempo. PLoS One 6, 1-8.

Michelena, P., Gautrais, J., Gérard, J.-F., Bon, R., Deneubourg, J.-L., 2008. Social cohesion in groups of sheep: Effect of activity level, sex composition and group size. Appl. Anim. Behav. Sci. 112, 81–93.

Molina, E., Ferret, A., Caja, G., Calsamiglia, S., Such, X., Gasa, J., 2001. Comparison of voluntary food intake, apparent digestibility, digesta kinetics and digestive tract content in Manchega and Lacaune dairy sheep in late pregnancy and early and mid lactation. Anim. Sci. 72, 209–221.

Mollenhorst, H., Rodenburg, T., Bokkers, E., Koene, P., Deboer, I., 2005. On-farm assessment of laying hen welfare: a comparison of one environment-based and two animal-based methods. Appl. Anim. Behav. Sci. 90, 277–291.

Molony, V., Kent, J.E., 1997. Assessment of Acute Pain in Farm Animals Using Behavioral and Physiological Measurements. J. Anim. Sci. 75, 266–272.

Moors, E., Gauly, M., 2009. Is the FAMACHA chart suitable for every breed? Correlations between FAMACHA scores and different traits of mucosa colour in naturally parasite infected sheep breeds. Vet. Parasitol. 166, 108–11.

Morgan, K.L., Brown, P.J., Wright, A.I., Steele, F.C., Baker, A.S., 1986. An investigation into the aetiology of “wool slip”: alopecia in ewes which are housed and shorn in winter. Vet. Rec. 119, 621–625.

Morgan-Davis, C., Waterhouse, A., Pollock, M.L., Milner, J.M., 2008. Body condition score as an indicator of ewe survival under extensive conditions. Anim. Welf. 17, 71–77.

Möstl, E., Palme, R., 2002. Hormones as indicators of stress. Domest. Anim. Endocrinol. 23, 67–74.

Moura, D., Silva, W., Naas, I., Tolon, Y., Lima, K., Vale, M., 2008. Real time computer stress monitoring of piglets using vocalisation analysis. Comput. Electron. Agric. 64, 11–18.

Munksgaard, L., Passille, A.M. De, Rushen, J., Ladewig, J., 1999. Dairy cows ’ use of colour cues to discriminate between people. Appl. Anim. Behav. Sci. 65, 1–11.

Muri, K., Stubsjøen, S.M., Valle, P.S., 2013. Development and testing of an on-farm welfare assessment protocol for dairy goats. Anim. Welf. 22, 385–400.

Napolitano, F., De Rosa, G., Ferrante, V., Grasso, F., Braghieri, A., 2009. Monitoring the welfare of sheep in organic and conventional farms using an ANI 35 L derived method. Small Rumin. Res. 83, 49–57.

Napolitano, F., De Rosa, G., Grasso, F., Wemelsfelder, F., 2012. Qualitative behaviour assessment of dairy buffaloes (Bubalus bubalis). Appl. Anim. Behav. Sci. 141, 91–100.

Napolitano, F., Derosa, G., Braghieri, A., Grasso, F., Bordi, A., Wemelsfelder, F., 2008. The qualitative assessment of responsiveness to environmental challenge in horses and ponies. Appl. Anim. Behav. Sci. 109, 342–354.

Napolitano, F., Grasso, F., Bordi, A., Tripaldi, C., Saltamacchia, F., Pacelli, C., De Rosa, G., 2005. On-farm welfare assessment in cattle and buffolos: evalution of some animal based parameters. Anim. Welf. 18, 451–458.

Nardon, A., Ronchi, B., Valentini, A., 1991. Effects of solar radiation on water and food intake and weight gain in Sarda and Comisana female lambs. Anim. Husb. Warm Clim. EAAP Publ. 55, 149–150.

Nasar, A., Rahman, A., 2006. Hormonal changes in the uterus during pregnancy – Lessons from the ewe: A review. J. Agric. Rural Dev. 4, 1–7.

Nieuwhof, G.J., Bishop, S.C., 2005. Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. Anim. Sci. 81, 23–29.

Norring, M., Manninen, E., de Passillé, A., Rushen, J., Munksgaard, L., Saloniemi, H., 2008. Effects of sand and straw bedding on the lying behavior, cleanliness, and hoof and hock injuries of dairy cows. J. Dairy Sci. 91, 570–6.

Nowak, R., Porter, R.H., Blache, D., Dwyer, C.M., 2008. Behaviour and the welfare of the sheep, in: Dwyer, C.M. (Ed.), The Welfare of Sheep. Springer Science and Buiness Media B.V, pp. 81–134.

O'Connor, L.J., Walkden-Brown, S.W., Kahn, L.P., 2006. Ecology of the free-living stages of major trichostrongylid parasites of sheep. Vet. Parasitol. 142, 1–15.

O'Driscoll, K., Hanlon, A., Boyle, L., 2008. The effect of out-wintering pad design on the synchrony of dairy cow behavior. J. Dairy Sci. 91, 4651–60.

Oldham, C., Thompson, A., 2011. The birthweight and survival of Merino lambs can be predicted from the profile of liveweight change of their mothers during pregnancy. Anim. Prod. Sci. 51, 776–783.

Olechnowicz, J.A.N., Jaśkowski, J.M., 2014. Mastitis in small ruminants Mastitis in small ruminants. Med Weter 70, 67–70.

Oli, M.K., Rogers, M.E., 1996. Seasonal pattern in group size and population composition of Blue Sheep in Manang, Nepal. J. Wildl. Manage. 60, 797–880.

Orgeur, P., N, M., Yvore, P., Bernard, S., Nowak, R., Schaal, B., Levy, F., 1998. Artificial weaning in sheep: consequences on behavioural, hormonal and immuno-pathological indicators of welfare. Appl. Anim. Behav. Sci. 58, 87–103.

Ospina, P.A.D., Nydam, T., Stokol, T., Overton, T.R., 2010. Association between the proportion of sampled transition cows with increased nonesterified fatty acids and beta-hydroxybutyrate and disease incidence, pregnancy rate, and milk production at the herd level. J. Dairy Sci. 93, 3595–3601.

Oyewus, A., Oyewusi, K., Takeet, A., Mustapha, A., Awoyomi, A., 2015. Ovine Coccidiosis: A Case Report. J. Vet. Adv. 5, 1023.

Papadopoulos, E., Gallidis, E., Ptochos, S., Fthenakis, G.C., 2013. Evaluation of the FAMACHA© system for targeted selective anthelmintic treatments for potential use in small ruminants in Greece. Small Rumin. Res. 110, 124–127.

Parkin, D., Devlin, N., 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Econ. 15, 653–64.

Parsons, A., Williams, T., 2010. Grass. Its production and utilisation, Third Edit. ed. Blackwell Publishing Ltd, London, UK.

Pendu, Y.L., Briedermann, L., Gerard, J.F., Maublanc, M.L., 1995. Inter-individual associations and social structure of a mouflon population (Ovis orientalis musimon). Behav. Processes 34, 67–80.

Petherick, J.C., 2007. Spatial requirements of animals: Allometry and beyond. J. Vet. Behav. Clin. Appl. Res. 2, 197–204.

Petherick, J.C., Phillips, C.J.C., 2009. Space allowances for confined livestock and their determination from allometric principles. Appl. Anim. Behav. Sci. 117, 1–12.

Phillips, C.J.C., 2009. A review of mulesing and other methods to control flystrike (cutaneous myiasis) in sheep. Anim. Welf. 18, 113–121.

Phythian, C., Cripps, P.C., Grove-White, D., Jones, P.H., Michalopoulou, E., Duncan, J.S., 2013. Observing lame sheep: evaluating test agreement between group-level and individual animal methods of assessment. Anim. Welf. 22, 417–422.

Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F., 2013. Inter-observer reliability of Qualitative Behavioural Assessments of sheep. Appl. Anim. Behav. Sci. 144, 73–79.

Phythian, C.., Michalopoulou, E., Cripps, P.., Duncan, J.., Wemelsfelder, F., 2015. On-farm qualitative behavioural assessment in sheep: repeated measurements accross time and association with physical indicators of flock health and welfare. Appl. Anim. Behav. Sci. In press.

Phythian, C.J., Cripps, P.J., Michalopoulou, E., Jones, P.H., Grove-White, D., Clarkson, M.J., Winter, A.C., Stubbings, L. a, Duncan, J.S., 2012. Reliability of indicators of sheep welfare assessed by a group observation method. Vet. J. 193, 257–63.

Phythian, C.J., Michalopoulou, E., Jones, P.H., Winter, A.C., Clarkson, M.J., Stubbings, L.A., Grove-White, D., Cripps, P.J., Duncan, J.S., 2011. Validating indicators of sheep welfare through a consensus of expert opinion. Animal 5, 943–52.

Phythian, C.J., Toft, N., Cripps, P.J., Michalopoulou, E., Winter, A.C., Jones, P.H., Grove-White, D., Duncan, J.S., 2013. Inter-observer agreement, diagnostic sensitivity and specificity of animal-based indicators of young lamb welfare. Animal 7, 1182–90.

Phythian, C.J.J., Hughes, D., Michalopoulou, E., Cripps, P.J., Duncan, J.S., 2011. Reliability of body condition scoring of sheep for cross-farm assessments. Small Rumin. Res. 104, 1–7.

Piccione, G., Caola, G., Refinetti, R., 2002. Effect of shearing on the core body temperature of three breeds of Mediterranean sheep. Small Rumin. Res. 46, 211–215.

Pickup, H.E., Dwyer, C.M., 2011. Breed differences in the expression of maternal care at parturition persist throughout the lactation period in sheep. Appl. Anim. Behav. Sci. 132, 33–41.

Pines, M.K., Petherick, J.C., Gaughan, J.B., Phillips, C.J.C., 2007. Stakeholders ' assessment of welfare indicators for sheep and cattle exported by sea from Australia. Anim. Welf. 16, 489–498.

Plant, J.W., 2006. Sheep ectoparasite control and animal welfare. Small Rumin. Res. 62, 109–112.

Platzer, B., Prosl, H., Cieslicki, M., Joachim, A., 2005. Epidemiology of Eimeria infections in an Austrian milking sheep flock and control with diclazuril. Vet.

Parasitol. 129, 1–9.

Plesch, G., Broerkens, N., Laister, S., Winckler, C., Knierim, U., 2010. Reliability and feasibility of selected measures concerning resting behaviour for the on-farm welfare assessment in dairy cows. Appl. Anim. Behav. Sci. 126, 19–26.

Poindron, P., Caba, M., Gomora Arrati, P., Krehbiel, D., Beyer, C., 1994. Responses of maternal and non-maternal ewes to social and mother-young separation. Behav. Processes 31, 97–110.

Poindron, P., Soto, R., Romeyer, A., 1997. Decrease of response to social separation in preparturient ewes. Behav. Processes 40, 45–51.

Pritchard, J.C., Lindberg, A.C., Main, D.C.J., Whay, H.R., 2005. Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. Prev. Vet. Med. 69, 265–83.

Pulliam, H.R., 1973. On the advantages of flocking. J. Theor. Biol. 38, 419–425.

Quenette, P.Y., 1990. Functions of vigilance in mammals: a review. Acta Oecologica 11, 801–818.

Raussi, S., Jauhiainen, L., Saastamoinen, S., Siivonen, J., Hepola, H., Veissier, I., 2011. A note on overdispersion as an index of behavioural synchrony: a pilot study in dairy cows. Animal 5, 428–32.

Regnault, T.R.H., Oddy, H. V, Nancarrow, C., Sriskandarajah, N., Scaramuzzi, R.J., 2004. Glucose-stimulated insulin response in pregnant sheep following acute suppression of plasma non-esterified fatty acid concentrations. Reprod. Biol. Endocrinol. 2, 64.

Reid, T.C., Cottle, D.J., 1999. New Zealand Society of Animal Production online archive. Proc. New Zeal. Soc. Anim. Prod. 59, 52–54.

Ridler, A.L., West, D.M., 2010. Examination of teeth in sheep health management. Small Rumin. Res. 92, 92–95.

Rieucau, G., G. A. Martin, J., 2008. Many eyes or many ewes: vigilance tactics in female bighorn sheep Ovis canadensis vary according to reproductive status. Oikos 11, 501–506.

Rinaldi, L., Catelan, D., Musella, V., Cecconi, L., Torgerson, P.R., Mavrot, F., Waal, T. De, Selemetas, N., Bosco, A., Biggeri, A., Cringoli, G., 2015. Haemonchus contortus: spatial risk distribution for infection in sheep in Europe. Geospat. Health 9, 325–331.

Roberts, G., 1996. Why individual vigilance decline as group size increases. Anim. Behav. 51, 1077–1086.

Roche, J.R., Friggens, N.C., Kay, J.K., Fisher, M.W., Stafford, K.J., Berry, D.P., 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92, 5769–801.

Roeber, F., Jex, A.R., Gasser, R.B., 2013. Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance - an Australian perspective. Parasit. Vectors 6, 1.

Roger, P. a., 2013. Why do sheep need umbrellas? Small Rumin. Res. 110, 75–77.

Romeyer, A., Bouissou, M., 1992. Assessment of fear reactions in domestic sheep, and influence of breed and rearing conditions. Appl. Anim. Behav. Sci. 34, 93–119.

Rook, A., Penning, P., 1991. Synchronisation of eating, ruminating and idling activity by grazing sheep. Appl. Anim. Behav. Sci. 32, 157–166.

Rousing, T., Ibsen, B., Sørensen, J.T., 2005. A note on: On-farm testing of the behavioural response of group-housed calves towards humans; test–retest and inter-observer reliability and effect of familiarity of test person. Appl. Anim. Behav. Sci. 94, 237–243.

Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Appl. Anim. Behav. Sci. 101, 40–53.

Ruckstuhl, K., Neuhaus, P., 2001. Behavioral Synchrony in Ibex Groups : Effects of Age , Sex and Habitat. Behaviour 138, 1033–1046.

Ruckstuhl, K.E., 1999. To synchonise or not to synchronise: a dilema for young bighorn males? Behaviour 136, 805–818.

Ruckstuhl, K.E., Festa-Bianchet, M., 1998. Do reproductive status and lamb gender affect the foraging behaviour of big horn ewes? Ethology 104, 941–954.

Ruckstuhl, K.E., Neuhaus, P., 2001. Behavioural synchrony in ibex groups: effects of age, sex and habitat. Behaviour 138, 1033–1046.

Rushen, J., Butterworth, A., Swanson, J.C., 2011a. Animal behavior and well-being symposium: Farm animal welfare assurance: science and application. J. Anim. Sci. 89, 1219–28.

Rushen, J., Butterworth, A., Swanson, J.C., 2011b. Animal behavior and well-being symposium: Farm animal welfare assurance: science and application. J. Anim. Sci. 89, 1219–28.

Rushen, J., Depassille, A.., 2005. Can we measure human–animal interactions in on-farm animal welfare assessment? Some unresolved issues. Appl. Anim. Behav. Sci. 92, 193–209.

Rushen, J., Taylor, A.A., Passille, A.M. De, 1999a. Domestic animals ' fear of humans and its effect on their welfare 285–303.

Rushen, J., Taylor, A.A., Passille, A.M. De, 1999b. Domestic animals' fear of humans and its effect on their welfare. Appl. Anim. Behav. Sci. 65, 285–303.

Russel, A.J.F., Doney, J.M., Gunn, R.G., 1969. Subjective assessment of body fat in live sheep. J. Agric. Sci. 72, 451–454.

Russel, A.J.F., Wright, A.I., 1983. The use of blood metabolites in the determination of energy status in beef cows. Anim. Prod. 37, 335–343.

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

Saether, N., 2002. Norway Country Report on Farm Animal Genetic Resources.

Sandgren, C.H., Lindberg, A., Keeling, L.J., 2009. Using a national dairy database to identify herds with poor welfare. Anim. Welf. 18, 523–532.

Sant'Anna, A.C., Paranhos da Costa, M.J.R., 2013. Validity and feasibility of qualitative behavior assessment for the evaluation of Nellore cattle temperament. Livest. Sci. 157, 254–262.

Sapolsky, R., 1992. Do glucocorticoid concentrations rise with age in the rat? Neurobiol. Aging 13, 171–147.

Sargison, N.D., 2013. Understanding the epidemiology of gastrointestinal parasitic infections in sheep: What does a faecal helminth egg count tell us? Small

Rumin. Res. 110, 78–81.

Schaller, G.B., 1977. Mountain Monarchs - wild sheep and goats of the Himalaya, 1st Editio. ed. University of Chicago Press, Chicago, USA.

Scholtz, A.J., Cloete, S.W.P., Cloete, J.J.E., Kruger, A.C.M., Wyk, J.B. Van, 2012. Divergent selection for reproduction affects dag score, breech wrinkle score and crutching time in Merinos. S. Afr. J. Anim. Sci. 42, 274–279.

Schreiner, D. a, Ruegg, P.L., 2003. Relationship between udder and leg hygiene scores and subclinical mastitis. J. Dairy Sci. 86, 3460–5.

Science, C.V., Kingdom, U., Sciences, B., Main, D.C.J., Webster, F., Green, L.E., 2001. Animal Welfare Assessment in Farm Assurance Schemes 108–113.

Scott, E., Fitzpatrick, A., Nolan, J., 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. Acta Agric. Scand. 51, 5–10.

Sevi, A., Casamassima, D., Pulina, G., Pazzona, A., 2007. Factors of welfare reduction in dairy sheep and goats. Ital. J. Anim. Sci. 8, 81–101.

Sevi, B.A., Massa, S., Annicchiarico, G., Dell, S., Antonio, A., Agraria, Z., 1999. Effect of stocking density on ewes' milk yield, udder health and microenvironment. J. Da 66, 489–499.

Shackleton, D.M., Shank, C.C., 1984. A review of the social behaviour of feral and wild sheep and goats. J. Anim. Sci. 58, 500–509.

Shelton, J., Usherwood, N.M., Wapenaar, W., Brennan, M.L., Green, L.E., 2011. Measurement and error of hoof horn growth rate in sheep. J. Agric. Sci. 150, 373–378.

Sherwin, C.M., Johnson, K.G., 1987. The influence of social factors on the use of shade by sheep. Appl. Anim. Behav. Sci. 18, 143–155.

Sibbald, A.M., Erhard, H.W., McLeod, J.E., Hooper, R.J., 2009. Individual personality and the spatial distribution of groups of grazing animals: an example with sheep. Behav. Processes 82, 319–26.

Sibbald, A.M., Hooper, R.J., 2004. Sociability and the willingness of individual sheep to move away from their companions in order to graze. Appl. Anim. Behav. Sci. 86, 51–62.

Sibbald, A.M., Hooper, R.J., 2003. Trade-offs between social behaviour and foraging by sheep in heterogeneous pastures. Behav. Processes 61, 1–12.

Sibbald, A.M., Oom, S.P., Hooper, R.J., Anderson, R.M., 2008. Effects of social behaviour on the spatial distribution of sheep grazing a complex vegetation mosaic. Appl. Anim. Behav. Sci. 115, 149–159.

Silanikove, N., 2000. Effects of heat stress on the welfare of extensively managed domestic ruminants. Livest. Prod. Sci. 67, 1–18.

Sirot, E., Touzalin, R., 2009. Coordination and synchronization of vigilance in groups of prey: the role of collective detection and predators' preference for stragglers. Am. Nat 173, 47–59.

Smith, E.M., Green, O.D.J., Calvo-Bado, L. a, Witcomb, L. a, Grogono-Thomas, R., Russell, C.L., Brown, J.C., Medley, G.F., KilBride, A.L., Wellington, E.M.H., Green, L.E., 2014. Dynamics and impact of footrot and climate on hoof horn length in 50 ewes from one farm over a period of 10 months. Vet. J. 201, 295–301.

Smith, L.A., White, P.C.L., Hutchings, M.R., 2006. Effect of the nutritional environment and reproductive investment on herbivore-parasite interactions in grazing environments. Behav. Ecol. 17, 591–596.

Sørensen, J.T., Fraser, D., 2010. On-farm welfare assessment for regulatory purposes: Issues and possible solutions. Livest. Sci. 131, 1–7.

Sorge, U.S., Cherry, C., Bender, J.B., 2014. Perception of the importance of human-animal interactions on cattle flow and worker safety on Minnesota dairy farms. J. Dairy Sci. 97, 4632–4638.

Spoolder, H.A.M., Geudeke, M.J., Van der Peet-Schwering, C.M.C., Soede, N.M., 2009. Group housing of sows in early pregnancy: A review of success and risk factors. Livest. Sci. 125, 1–14.

Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P. a., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naïve and habituated to road transport. Anim. Prod. Sci. 51, 240–249.

Stott, A., Vosough Ahmadi, B., Dwyer, C., Kupiec, B., Morgan-Davies, C., Milne, C., Ringrose, S., Goddard, P., Phillips, K., Waterhouse, A., 2012. Interactions between profit and welfare on extensive sheep farms. Anim. Welf. 21, 57–64.

Stubsjøen, S.M., Hektoen, L., Valle, P.S., Janczak, A.M., Zanella, A.J., 2011. Assessment of sheep welfare using on-farm registrations and performance. Anim. Welf. 20, 239–251.

Sumpter, D.J.T., 2010. Collective Animal Behaviour. Princeton University Press, Princeton, New Jersey.

Swan, R., Chapman, H., Hawkins, C., Howell, J., Spalding, V., 1984. The epidemiology of squamous cell carcinoma of the perineal region of sheep: abattoir and flock studies. Aust. Vet. J. 61, 146–151.

Tadesse, S. a., Kotler, B.P., 2013. Effects of habitat, group-size, sex-age class and seasonal variation on the behavioural responses of the mountain nyala (Tragelaphus buxtoni) in Munessa, Ethiopia. J. Trop. Ecol. 30, 33–43.

Taillon, J., Côté, S.D., 2006. The role of previous social encounters and body mass in determining social rank: an experiment with white-tailed deer. Anim. Behav. 72, 1103–1110.

Tallet, C., Veissier, I., Boivin, X., 2006. A note on the consistency and specificity of lambs' responses to a stockperson and to their photograph in an arena test. Appl. Anim. Behav. Sci. 98, 308–314.

Tatara, M.R., Szabelska, A., Krupski, W., Tymczyna, B., Luszczewska-Sierakowska, I., Bieniaś, J., Ostapiuk, M., 2014. Morphometric, densitometric and mechanical properties of mandibular deciduous teeth in 5-month-old Polish Merino sheep. BMC Vet. Res. 10, 45.

Taylor, M., 2012. Emerging parasitic diseases of sheep. Vet. Parasitol. 189, 2–7.

Taylor, M., Catchpole, J., 1994. Review article: coccidiousis of domestic ruminants. Appl. Parasitol. 35, 73–86.

Teixeira, D.L., Villarroel, M., María, G. a., 2014. Assessment of different organic beddings materials for fattening lamb. Small Rumin. Res. 119, 22–27.

Temple, D., Manteca, X., Dalmau, A., Velarde, A., 2013. Assessment of test–retest reliability of animal-based measures on growing pig farms. Livest. Sci. 151, 35–

45.

The Met Office, 2015. Fact sheet: UK Minimum Temperatures. Exeter: Met Office.

Thogerson, C.M., Hester, P.Y., Mench, J.A., Newberry, R.C., Okura, C.M., Pajor, E.A., Talaty, P.N., Garner, J.P., 2009. The effect of feeder space allocation on productivity and physiology of Hy-Line W-36 hens housed in conventional cages. Poult. Sci. 88, 1793–9.

Thomas, D.L., Waldron, D.F., Lowe, G.D., Morrical, D.G., Meyer, H.H., High, R.A., Berger, Y.M., Clevenger, D.D., Fogle, G.E., Gottfredson, R.G., Loerch, S.C., Mcclure, K.E., Willingham, T.D., Zartman, D.L., Zelinsky, R.D., 2003. Length of docked tail and the incidence of rectal prolapse in lambs. Am. Soc. Anim. Sci. 2725–2732.

Thompson, C.S., Connington, J., Williams, J., Rutherford, K.M., 2015. Farmers' perceptions of pain and their attitudes towards pain and analgesia, in: The Association for Veterinary Teaching and Research Work.

Torrance, G.W., Feeny, D., Furlong, W., 2001. Visual Analog Scales: Do they have a role in the measurement of preferences for health states? Med. Decis. Mak. 21, 329–334.

Torrance, G.W., Furlong, W., Feeny, D., 2002. Health utility estimation. Expert Rev. Pharmacoeconomics Outcomes Res 2, 99–108.

Treves, A., 2000. Theory and method in studies of vigilance and aggregation. Anim. Behav. 60, 711–722.

Tucker, C.B., Rogers, A.R., Verkerk, G. a., Kendall, P.E., Webster, J.R., Matthews, L.R., 2007. Effects of shelter and body condition on the behaviour and physiology of dairy cattle in winter. Appl. Anim. Behav. Sci. 105, 1–13.

Turner, S.P., Dwyer, C.M., 2007. Welfare assessment in extensive animal production systems: challenges and opportunities. Anim. Welf. 16, 189–192.

Turner, S.P., Lawrence, A.B., 2007. Relationship between maternal defensive aggression, fear of handling and other maternal care traits in beef cows. Livest. Sci. 106, 182–188.

Tuyttens, F.A.M., 2005. The importance of straw for pig and cattle welfare: A review. Appl. Anim. Behav. Sci. 92, 261–282.

Underwood, E., Suttle, N., 1999. The detection and correction of mineral imbalances in animals, in: Suttle, N. (Ed.), The Mineral Nutrition of Livestock. CABI Pub, Wallingford, U.K.

Uriarte, J., Llorente, M.M., Valderrábano, J., 2003. Seasonal changes of gastrointestinal nematode burden in sheep under an intensive grazing system. Vet. Parasitol. 118, 79–92.

Veissier, I., Boissy, A., Desire, L., Greiveldinger, L., 2009. Animals ' emotions: studies in sheep using appraisal theories. Anim. Welf. 18, 347–354.

Verbeek, E., Waas, J.R., McLeay, L., Matthews, L.R., 2011. Measurement of feeding motivation in sheep and the effects of food restriction. Appl. Anim. Behav. Sci. 132, 121–130.

Verbeek, E., Waas, J.R., Oliver, M.H., McLeay, L.M., Ferguson, D.M., Matthews, L.R., 2012. Motivation to obtain a food reward of pregnant ewes in negative energy balance: behavioural, metabolic and endocrine considerations. Horm. Behav. 62, 162–72.

Vierin, M., Bouissou, M., Vie, M., 2002. Influence of maternal experience on fear reactions in ewes. Appl. Anim. Behav. Sci. 75, 307–315.

Viérin, M., Bouissou, M.F., 2001. Pregnancy is associated with low fear reactions in ewes. Physiol. Behav. 72, 579–87.

Vokey, F.J., Guard, C.L., Erb, H.N., Galton, D.M., 2001. Effects of alley and stall surfaces on indices of claw and leg health in dairy cattle housed in a free-stall barn. J. Dairy Sci. 84, 2686–99.

Vrij, A., Van der Steen, J., Koppelaar, L., 1994. Aggression of police officers as a function of temperature. An experiment with the fire arms training system. J. Community Appl. Soc. Psychol. 4, 365–370.

Wade, G.N., Schneider, J.E., 1992. Metabolic fuels and reproduction in female mammals. Neurosci. Biobehav. Rev. 16, 235–272.

Waghorn, G.C., Gregory, N.G., Todd, S.E., Wesselink, R., 1999. Dags in sheep; a look at faeces and reasons for dag formation. Proc. New Zeal. Grassl. Assoc. 49, 43–49.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M., Janczak, A., Visser, E., Jones, R., 2006. Assessing the human–animal relationship in farmed species: A critical review. Appl. Anim. Behav. Sci. 101, 185–242.

Waiblinger, S., Knierim, U., Winckler, C., 2001. The Development of an Epidemiologically Based On-Farm Welfare Assessment System for use with Dairy Cows. Acta Agric. Scand. Sect. A - Anim. Sci. 51, 73–77.

Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M., Wemelsfelder, F., 2010. The assessment of emotional expression in dogs using a Free Choice Profiling methodology. Anim. Welf. 19, 75–84.

Wall, R., 2007. Ectoparasites: future challenges in a changing world. Vet. Parasitol. 148, 62–74.

Warren, J.T., Mysterud, I., Lynnebakken, T., 2001. Mortality of lambs in free-ranging domestic sheep (Ovis aries) in northern Norway. J. Zool. 254, 195–202.

Waterhouse, A., 1996. Animal welfare and sustainability of production under extensive conditions—A European perspective. Appl. Anim. Behav. Sci. 49, 29–40.

Weary, D., Fraser, D., 1995. Calling by domestic piglets: reliable signals of need? Anim. Behav. 50, 1047–1055.

Weary, D.M., Niel, L., Flower, F.C., Fraser, D., 2006. Identifying and preventing pain in animals. Appl. Anim. Behav. Sci. 100, 64–76.

Webster, J., Schütz, K., Sutherland, M., Stewart, M., Mellor, D., 2015. Different animal welfare orientations towards some key research areas of current relevance to pastoral dairy farming in New Zealand. N. Z. Vet. J. 63, 31–36.

Welp, T., Rushen, J., Kramer, D.L., Festa-Bianchet, M., de Passille, A.M.B., 2004. Vigilance as a measure of fear in dairy cattle. Appl. Anim. Behav. Sci. 87, 1–13.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25–31.

Wemelsfelder, F., Farish, M., 2004. Qualitative categories for the interpretation of sheep welfare: a review. Anim. Welf. 13, 261–268.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. Appl. Anim. Behav. Sci. 67, 193–215.

Wemelsfelder, F., Hunter, T.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the “whole animal”: a free choice profiling approach. Anim. Behav. 62, 209–220.

Wemelsfelder, F., Lawrence, A.B., 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Acta Agric. Scand. 30, 21–25.

Whay, H.R., Main, D.C.J., Greent, L.E., Webster, A.J.F., 2003. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. Anim. Welf. 12, 205–217.

White, K.S., Berger, J., 2001. Antipredator strategies of Alaskanmoose: arematernal trade-offs influenced by offspring activity? Can. J. Zool. 79, 2055–2062.

Wickham, S.L., Collins, T., Barnes, A.L., Miller, D.W., Beatty, D.T., Stockman, C., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012. Qualitative behavioral assessment of transport-naïve and transport-habituated sheep. Am. Soc. Anim. Sci. 90, 4523–4535.

Williams, D.L., 2010. Welfare issues in farm animal ophthalmology. Vet. Clin. North Am. Food Anim. Pract. 26, 427–35.

Winckler, C., Capdeville, J., Gebresenbet, G., Roiha, U., Tosi, M., Waiblinger, S., 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Anim. Welf. 12, 619–624.

Windschnurer, I., Boivin, X., Waiblinger, S., 2009. Reliability of an avoidance distance test for the assessment of animals’ responsiveness to humans and a preliminary investigation of its association with farmers' attitudes on bull fattening farms. Appl. Anim. Behav. Sci. 117, 117–127.

Windschnurer, I., Schmied, C., Boivin, X., Waiblinger, S., 2008. Reliability and inter-test relationship of tests for on-farm assessment of dairy cows' relationship to humans. Appl. Anim. Behav. Sci. 114, 37–53.

Wingfield, J., Romero, L., 2001. Adrenocortical responses to stress and their modulation in free-living vertebrates., in: McEwan, B., Goodman, H. (Eds.), Handbook of Physiology, The Endocrine System, Coping with the Environment: Neural and Endocrine Mechanisms. Oxford University Press, New York, pp. 211–234.

Winter, A.C., 2008. Lameness in sheep. Small Rumin. Res. 76, 149–153.

Winter, A.C., 1995. Wool loss in sheep. Vet. Annu. 35, 313–319.

Ydenberg, R.C., Dill, L.M., 1986. The economics of fleeing from predators. Adv. Study Behav. 16, 229–249.

Zimmerman, P., Koene, P., van Hooff, J., 2000. Thwarting behaviour in different contexts and the gakel-call in the laying hen. Appl. Anim. Behav. Sci. 69, 255–264.

Zubrigg, K., Kelton, D., Anderson, N., 2005. Tie-stall design and its relationship to lameness, injury, and cleanlinesson 317 Ontario dairy farms. J. Dairy Sci. 88, 3201–3210.