UNIVERSITY OF LIVERPOOL

DEVELOPMENT OF INDICATORS FOR THE ON-FARM ASSESSMENT OF SHEEP WELFARE

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Clare Joan Phythian

June 2011

Abstract

The objective of the work presented in this thesis was to develop valid, reliable and feasible indicators for the on-farm assessment of sheep welfare. In the absence of a reference test for animal welfare assessment, the welfare indicators in this thesis were developed within the Farm Animal Welfare Council (FAWC) Five Freedoms framework.

A scientific literature review and the consensus opinion of a panel of experts were used to judge the face and consensual validity of a selection of indicators of sheep welfare. Experts identified 193 current on-farm welfare issues for sheep and subsequently suggested a range of animal- (n = 26), resource- (n = 13) and management-based indicators (n = 22) in order to assess the on-farm welfare of adult sheep (> 1 year-old), growing lambs (> 6 weeks – < 1 year-old) and young lambs (< 6 weeks-old).

The diagnostic validity of 49 non-invasive, animal-based indicators was tested during a cross-sectional study in which 8 observers independently assessed the indicators on 4686 sheep and lambs from 50 farms in England and Wales. This study found that many indicators, including measures of lameness, body condition, and cleanliness, were reliable, sensitive and specific between observers of differing occupations and levels of training and experience. The measures were also feasible to apply and capable of detecting between-farm variation in conditions associated with sheep welfare. Studies in the use of qualitative behaviour assessment (QBA) also found good levels of reliability for observer assessments of video-clips of sheep behaviour.

The ability of animal-based indicators to detect seasonal variation in sheep welfare conditions was investigated on 5740 adult sheep and growing lambs from 12 sheep farms during a one-year longitudinal study. Animal-based indicators including measures of lameness, body condition and QBA, were found to be capable of detecting seasonal variation, suggesting that the tests were valid under different management conditions and across the different events of the annual sheep production cycle. Overall, a low proportion of the sample population was observed with conditions that affected sheep welfare, which may have been the result of non-random sampling of farms. However, for the purposes of this thesis the ability of the indicators to detect important welfare conditions at a low prevalence provided further evidence of their validity.

Resource-based assessments were feasible to perform but assessments were limited to certain periods of the production cycle. As management-based indicators relied on the accuracy of farmer interviews and access to farm records, the use of animal-based measures may be a more appropriate means of assessing some aspects of flock welfare.

A final set of valid, reliable and feasible indicators of sheep welfare, comprising 28 animal- and 11 resource- and management-based measures, was recommended on the basis of field validation results and expert opinion. Key animal-based indicators that were found to be reliable, responsive and robust under extensive and intensive farming systems and suitable for assessing both sheep and lambs were lameness, demeanour and body condition. It is suggested that these indicators should be applied in future on-farm protocols by trained assessors who are calibrated to the Standard Operating Procedures (SOP's). The interpretation of animal-based indicator assessments was guided by expert opinion in the form of preliminary cut-off points, which defined the level of acceptable and unacceptable flock welfare. As a result, the work presented in this thesis can inform the method of assessment and interpretation of a selection of valid, reliable and feasible on-farm indicators of sheep and lamb welfare.

Contents

List of tables

List of figures

Acknowledgments

Firstly I would like to thank my main supervisors, Dr Jennifer Duncan and Dr Eleni Michalopoulou for giving me the opportunity to develop my interest in sheep welfare and experience of clinical research and for their continued support, advice and encouragement throughout the course of my PhD. I'm very grateful to Dr Peter Cripps and Dr Dai Grove-White for taking on the supervisory role mid-way through the project and for providing statistical and epidemiological expertise and invaluable support with data analysis. Special thanks are also due to other members of the research team - Professor Agnes Winter for her initial role as supervisor, expert panel member and for encouraging me to follow a career in sheep health, welfare and production; Dr Phil Jones for particular support with the expert panel meetings and his pivotal role in the analysis of the expert panel cut-off data; Professor Michael Clarkson for general advice, participating as an observer and expert panel member; Lesley Stubbing for advice regarding the sheep industry and application of the indicators.

I'd like to gratefully acknowledge the support and advice of Dr Francoise Wemelsfelder, from the Scottish Agricultural College, who kindly and very generously shared her QBA methodology with me and spent considerable time and effort in guiding me in the understanding of QBA and Principal Component Analysis. Thanks also to Dr Nils Toft, University of Copenhagen, for introducing me to Latent Class Analysis, inviting me to attend his diagnostic test evaluation course and for checking all the model outputs.

I'm grateful to the undergraduate students - Dan Hughes, Steven Beer, Rachel Willey, Emily Capes, Matthew Turner and Rosie Higgs, who remained enthusiastic, optimistic and hardworking as they accompanied me on farm visits, handled animals, performed indicator assessments and cleaned the mobile handling unit and equipment. Thanks also to Lynda Jeff for help with maintenance of the clothing, Chris Broadbent for dealing with University administration and project accounts and Louisa Platt for assistance with the scanning of workbooks from the final expert panel meeting.

The welfare indicators could not have been developed without the voluntary contribution from the members of the expert panel who provided a wealth of knowledge and opinions. They played a large role in the development of indicators and the final guidance on the potential use of the measures presented in this thesis and I am very grateful to them for participating in the panel meetings and for their commitment and support of this project.

I would also like to acknowledge the role of Defra, who funded this PhD as part of the AW1025 project to develop objective indicators for the on-farm assessment of sheep welfare. Thanks also to Iain Rogerson of the Soil Association and the team of farm assurance inspectors who kindly tested the reliability of QBA during a video clip study.

I enjoyed immensely the time spent working on the 50 sheep farms and would like to thank all the farmers involved for agreeing to participate in the study and allowing me to spend considerable time assessing their sheep, learning from their experiences and expertise. I am also grateful to all the farmers, shepherds and families for welcoming me onto their farms and homes and providing me with the very best of on-farm hospitality.

Thank you to all my friends for their support, patience and understanding during the completion of this PhD. Finally, I'm indebted to my parents for the outstanding level of support provided throughout the course of this project. I really could not have completed the research or this thesis without their generosity, kindness and support. Therefore, I would like to take this opportunity to thank them whole heartedly for their continued and unfailing support and for their enthusiasm and genuine interest in this research.

Chapter 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

The objective of this thesis was to develop valid, reliable and feasible welfare indicators for sheep that were non-invasive and sensitive to the current on-farm welfare issues for the species. A key driver for the development of welfare indicators for use in on-farm animal welfare assessments has been the increased consumer awareness and concern for farm animal welfare, which has created a demand for products from high welfare systems, assurance of farm animal welfare standards and welfare labelling of food (Main *et al.*, 2001; Waiblinger *et al.*, 2001). Welfare assessments are also undertaken to assess the on-farm compliance with national welfare legislation (EFSA, 2007). Furthermore, stockpeople, agricultural advisors and veterinary surgeons also need to possess the necessary skills and tools to assess whether good standards of animal welfare are maintained or improved following health or management changes.

Sheep are farmed for a variety of reasons, including the production of meat, milk, wool, breeding replacements, research animals, and are increasingly being kept as companion animals. The general public may associate a very "green" image with extensively managed sheep that are seen to roam wide open spaces and perceived to exhibit natural behaviours (Matthews, 1996; Waterhouse 1996). However, the on-farm management of sheep is very diverse and varies according to the time of the year and period within the production cycle (FAWC, 1994). For example, sheep can be managed under relatively extensive conditions during summer periods but more intensive managements systems, with periods of housing, may occur at other times of the year and there are welfare issues for sheep at both extremes of management (Dwyer, 2008; Goddard, 2008).

An animal welfare indicator is a measure of an input or an outcome and is used to gauge the past, present or future welfare state of an animal (Sørensen *et al.*, 2001). However, few scientifically valid and robust indicators of sheep welfare are currently available. There is therefore a need to develop valid, reliable and feasible measures for use in on-farm welfare assessments. Welfare indicators for sheep were developed, as described in this thesis, by combining methodology of animal welfare science with epidemiological principles regarding diagnostic test evaluation. The objective of this chapter was to review the scientific literature to identify the current on-farm welfare issues for sheep in the United Kingdom (UK) and identify potential indicators of sheep welfare. The results from this review were then used to inform the subsequent development of indicators, which involved using the opinion of an expert panel and evaluation of indicator performance during field studies.

1.1.2 Defining animal welfare

In order to develop valid indicators of sheep welfare, it was necessary to define the meaning of animal welfare to provide a basis for the development of scientific assessment methods. However, there is currently no universally accepted definition of animal welfare and definitions of animal welfare provide much debate for scientists, producers, consumers and animal protectionists alike (Phillips *et al.*, 2009). Therefore, several theories and schools of thought regarding animal welfare are in existence, including the biological, naturalness and feelings concepts, animal needs, the quality of life approach, and the Five Freedoms. As a result, each of these theories was examined before a definition was selected to be used as the framework in the development of indicators of sheep welfare in this thesis.

Definitions of welfare that are concerned with physical health and well-being are part of the "biological functioning" school. This concept is concerned with the underlying physiology and ability of an animal to cope with its environment (Broom, 1986) and is central to animal welfare assessments performed by stockpeople and veterinary surgeons (von Keyserlingk *et al.*, 2009). Consequently, the functioning approach often focuses on the assessment of environmental resources and measurements of physical outcomes such as disease, injury, and productivity (Broom, 1986). However, the biological functioning approach to animal welfare may not address all aspects of an animal's life. Although good health is an essential component of good welfare, poor welfare does not always limit the level of productivity. Therefore, production outputs, such as reproductive performance, may not be altered even when animals are maintained under conditions associated with poor standards of welfare (von Keyserlingk *et al.*, 2009).

An alternative view of animal welfare focuses on the psychological health and emotional experiences of animals - the so-called "feelings-based approach" (Dwyer, 2008b). The feelings concept suggests that if an animal is feeling well, it has good welfare (Duncan, 1996). Therefore, unpleasant feelings such as pain, fear, suffering and distress, are the antithesis of good welfare (Dawkins, 1990) – aspects that are covered under the welfare legislation of the Protection of Animals Act (1911). As it can be difficult to measure the subjective concept of feelings, animal feelings are often measured by assessing behaviour, such as fear responses or the strength of the motivation to obtain food or another valuable resource (Greiveldinger *et al.*, 2007).

The third concept is the "naturalness" school of animal welfare. This is the 'teleos' concept of animal welfare (Rollin, 1991) which suggests that animals should live as nature intended (Hughes, 1976) and be capable of expressing 'natural' behaviours (Dawkins, 1990). The natural approach is thought to conflict with production systems where animals live indoors under intensive management conditions (Harrison, 1964), which has led to increased interest in organic farming methods. However, the naturalness approach does create some ethical concerns because the exposure of domesticated animals to natural living conditions means that they may experience extreme weather, chronic conditions, periods of chronic and untreated disease, predation, and deprivation of food and water which can all have negative implications for animal welfare (Dwyer, 2008b). As many farm animals have been selected for certain production traits, it may be difficult to define the natural states for all domestic animals (Smidt, 1982). This motivated research into animal behaviour to identify if the needs and motivations of a particular species are provided by certain farming systems (Bracke and Hopster, 2006).

The 'animal needs' approach is based on the idea that animals have a set of intrinsic physiological and behavioural motivations, such as thermoregulation or interactions with other members of the same species (Broom and Johnson, 1993). If the need for a particular resource, or motivation to perform a specific behaviour i.e. an ethological need, is not fulfilled then this can have deleterious consequences on the physiology and/or behaviour of the animal. Obviously there is a hierarchy of needs, as life-sustaining needs are thought to be more important than those for health and comfort (Hurnik, 1988).

Accordingly, the 'animal needs' concept has led to the development of a welfare assessment system known as the Animal Needs Index (ANI) (Bartussek, 1999). This system can be used to assess the quality of resources such as food, water and thermal comfort, although there can be difficulties in the interpretation and measurement of ethological needs such as satisfaction and frustration (Bracke and Hopster, 2006). The animal needs approach is also encompassed under the European welfare legislation of the Council of Europe (CoE) directives on the Protection of Animals kept for farming purposes and the national legislation (England and Wales) of the Animal Welfare Act

(2006). The Act dictates a duty of care on owners to promote the welfare of animals under their responsibility and contains the general laws relating to minimal standards of animal welfare, including the provision of a suitable environment and diet, the ability to exhibit normal behaviours, and protection from pain, injury, suffering and disease. Furthermore, the Treaty of Amsterdam (1997) recognises the ability of farm animals to consciously experience positive and negative emotions. Consequently, a quality of life view on animal welfare has developed (Fraser *et al.*, 1997), which has extended the view of welfare from the mere absence of suffering and physical fitness to ask whether an animal feels good and experiences positive emotions – i.e. the animal has 'a life worth living' (Wemelsfelder, 1997).

It is clear that the current understanding of animal welfare encompasses both subjective and objective features of life (Fitzpatrick *et al.*, 2006), and there is overlap in the concepts of physical, mental and natural welfare. For example, a lame animal may be in pain (feelings), with reduced growth rates and poor body condition (functioning) and have impaired movement (natural behaviour) (von Keyserlingk *et al.*, 2009). The quality of life approach is focused on the positive aspects of welfare, although there is currently no clear means of assessing the on-farm emotional experience of sheep (Roger, 2008). Rather than focusing on a single concept, a holistic definition that covers the physical, emotional and naturalness aspects of welfare was considered to be more appropriate for the development of objective, welfare indicators for sheep in this thesis. Therefore, the concepts defined in the Five Freedoms were examined.

1.1.3 Five freedoms framework

Following public scrutiny of the welfare of production animals (Harrison, 1964), the UK government launched a scientific committee to investigate the welfare of intensively managed farm animals. The technical scientific report by the Brambell Committee (1965) was concerned with the physical and mental components of an animal's life and defined ideal states of welfare. This report proposed that an animal should have sufficient freedom of movement to be able to move, turn around, groom itself, get up, lie down and stretch all limbs without difficulty - criteria which became known as the "Brambell Five Freedoms". These concepts were later refined by the Farm Animal Welfare Council (FAWC) into the "Five Freedoms" (FAWC, 1994), and are defined as:

1. **Freedom from hunger and thirst** – by ready access to fresh water and a diet to maintain full health and vigour
2. **Freedom from discomfort** – by providing an appropriate environment including shelter and a comfortable resting area
3. **Freedom from pain, injury and disease** – by prevention or rapid diagnosis and treatment
4. **Freedom from fear and distress** – by ensuring conditions and treatment which avoid mental suffering
5. **Freedom to express normal behaviour** – by providing sufficient space, proper facilities and company of the animal's own kind

The Five Freedoms can be considered as a list of the inputs needed to provide good animal welfare and cover the biological functioning, feelings and naturalness based definitions of animal welfare (Dwyer, 2009). For example, the freedom to express normal behaviours is associated with natural states of welfare; the biological functioning approach is covered by the freedom from pain, injury and disease, and animal feelings are considered under the freedom from fear and distress. Accordingly, the Freedoms criteria have been used as the principle framework for UK farm assurance schemes including the Bristol Welfare Assessment Program (BWAP) (Whay *et al.*, 2003c) and Royal Society for the Prevention of Cruelty to Animals (RSPCA) freedom foods scheme (www.rspca.org.uk). They also form the framework to the code of recommendations for the welfare of sheep (Defra, 2003), which apply under the Animal Welfare Act (2006) and Welfare of Farmed Animals (2007) Regulations. Given that the Five Freedoms are an accepted method for the scientific assessment of animal welfare (FAWC, 2009; Whay *et al.*, 2003b) the Freedoms framework was selected as the welfare criteria for the development of sheep welfare indicators in this thesis.

1.1.4 Measuring animal welfare

Welfare indicators are generally categorised into those that are based on observations and examinations of animals (animal-based), those that rely on assessment of housing and grazing areas (resource-based) or those that assess farmer policies and management practices (management-based) (Capdeville and Veissier, 2001; Smidt, 1982). Animal-based measures tend to focus on measures of production, behaviour, physiology, anatomy, pathology or health (Smidt, 1982). Assessments of resource inputs include the type of housing, space allowance, facilities for water and nutrition, and ventilation system (EFSA, 2007). Management-based indicators evaluate policies for disease management, mutilations such as castration and tail-docking, records of drug consumption and assessment of the human-animal bond (EFSA, 2007).

In the past, farm assurance schemes were often solely based on the assessment of resource- and management-based inputs. Consequently, there have been criticisms directed towards welfare measures that focus on aspects of housing and facilities rather than assessing welfare at the level of the animal. An advantage of animal-based indicators is that they are an integrated welfare measure that can reveal the outcome of resource inputs and management decisions and therefore can be viewed as more direct measures of animal welfare (Alban *et al.*, 2001). Additionally, animal-based outcomes allow benchmarking between different farms (Huxley *et al.*, 2004) and have been suggested to be more reliable than resource-based measures (Mullan *et al.*, 2009). Animal-based outcomes can use observations of animal behaviour and measures akin to those performed during clinical examinations, so they may require experienced and trained personnel. As the responsiveness of animal-based indicators can vary, some measures may be capable of assessing chronic welfare issues but these measures may not capture more acute compromises in animal welfare (Mason and Latham, 2004). Therefore, the use of resource- and management-based measures may be more appropriate for the assessment of certain animal welfare issues.

Assessments of resource and management inputs have been shown to have high validity (Alban *et al.*, 2001) and can be easier to quantify compared to animal-based measures (Mollenhorst *et al.*, 2005). Several assessments of farm management practices rely on examination of on-farm records and are reliant on the accuracy and quality of records maintained. As techniques for easy and reliable collection of farm health records are not currently available (Waiblinger *et al.*, 2001), the examination and retrieval of this type of material can also prove to be time and labour consuming (Sørensen *et al.*, 2007). The attitude, behaviour and skills of the stockperson are crucial to the standard of farm animal welfare (Hemsworth, 2007), but the human-animal bond can be challenging to assess and may not always be possible given the snapshot of time available during an on-farm visit (de Passillé and Rushen, 2005). Therefore, a mix of animal-, resource-, and management-based indicators may be the most useful means of assessing the on-farm welfare issues for production animals (Capdeville and Veissier, 2001).

1.1.5 Developing indicators of animal welfare

Following the identification of a suitable animal welfare definition, the next step in the development of animal welfare indicators is to establish the welfare concerns for the species of interest. In previous studies, the opinion of stakeholders and experts from animal health, welfare and production research and industry has been used to identify welfare issues for sheep managed under Australian farming conditions (Phillips *et al.*, 2009), including those transported long-distances by sea (Pines *et al.*, 2007). In addition, during the Economics and Welfare of Extensively-managed Sheep ('EWES') project, an expert panel was used to identify the key concerns for sheep managed under extensive farming conditions – nutrition, lameness, health, shepherding and the adequacy of farm facilities and equipment (Waterhouse *et al.*, 2003).

The methodology used to elicit the opinion of experts and stakeholders regarding animal welfare concerns has included the use of conferences (Pines *et al.*, 2007), web-based surveys (Fernie *et al.*, 2009) or individual face-to-face interviews of experts (Phillips and Phillips, 2010). Alternatively, a list of welfare issues can be initially identified using a literature review which is later refined by a panel of experts (Phillips *et al.*, 2009).

Although similarities in the welfare issues for sheep have been identified by different research projects, it is evident that welfare issues for sheep can be influenced by the farming system, period of production and geographical location. Compared to other countries in which extensive flocks are shepherded onto newer grazing pastures, British sheep managed under extensive condition are often left to fend for themselves for large parts of the year and are not frequently inspected (Waterhouse, 1996). In addition, specific management practices, such as mulesing (removal of skin over perineum and hindquarters) is routinely performed in Australia (Cronin *et al.*, 2002) but is not practised on UK sheep farms and so is not highly relevant to the development of welfare indicators in this thesis. As such, there is a need to ascertain the current knowledge regarding the welfare issues for sheep managed under particular farming conditions and within a specific location. As the welfare indicators developed in this thesis have the potential to be used on sheep farms across the UK, it is important to consider the management systems used in these regions. Therefore, a review of the scientific literature was needed in order to identify the current on-farm welfare issues for sheep in the UK. Welfare issues found by the literature review were categorised into the relevant area of the Five Freedoms and were used to inform the development of indicators of sheep welfare in this thesis.

1.2 On-farm welfare issues for sheep identified by the literature review

1.2.1 Freedom from hunger and thirst – *by ready access to fresh water and a diet to maintain full health and vigour*

Provision of an adequate diet is stipulated in most welfare legislation and codes of practice (Defra, 2003; FAWC, 1994). For sheep, the availability of a diet that meets behavioural needs for rumination is also important, plus an adequate quantity and quality of feed (Lynch *et al.,* 1992). Under-nutrition is a welfare issue because of the effects on rumen and physiological functioning and the adverse experience of hunger (Mellor and Stafford, 2004; Villalba *et al.,* 2010). Particular welfare concerns for sheep pertain to the provision of inadequate feed, particularly a shortage of energy intake in pregnant ewes, and the presence of thin, old ewes in the flock (Winter, 1997). In the UK, most sheep production is dependent on grassland management, so sheep may be at the greatest risk of under-nutrition when they are over-wintered on the hill and managed under extensive conditions (Waterhouse, 1996). As well as providing sufficient feed, sheep must be able to access troughs or feed containers. For housed sheep this depends on the stocking density, number of troughs and trough design (FAWC, 1994). Other factors that can affect food intake include dental condition (Gordon *et al.,* 1996), food preferences and general health (Villalba *et al.,* 2010).

Under-nutrition of pregnant ewes also has consequences for maternal bonding, as well as reduced colostrum and milk production which has deleterious effects on lamb health and welfare. Therefore, a common on-farm management practice is to supply supplementary feed to pre-partum ewes in order to improve lamb birth weight and survival (Goddard *et al.,* 2006). Other management practices, such as the early removal of lambs from the ewe and their artificial rearing on replacement milk, have received welfare attention (Prettejohn, 1990). Artificial rearing is commonly used in dairy flocks in which lambs are separated 18 – 36 days post-partum (Sevi *et al.,* 2001a). Early separation of lambs has been associated with poor standards of welfare because these animals lack the nutritional, immunological and physical protection of their dams (Napolitano *et al.,* 2002). However, other research has shown that the weight gain of artificially reared lambs is not significantly different if lambs are offered dam milk or a mixture of ewe milk and substitute post-weaning (Sevi *et al.,* 2001a).

Another on-farm management practice that is associated with lactating animals is the deprivation of water and restricted nutrition of newly weaned ewes. The method, known as 'drying off', is claimed to reduce the level of post-weaning mastitis, however, there are no findings to suggest that this is beneficial for ewe health or welfare. Indeed, the complete lack of water is a serious welfare concern (FAWC, 1994) as lactating animals have higher water intakes and should be provided with a continuous source of clean water (Lynch *et al.*, 1992). The Welfare of Farmed Animals (England) regulations 2007 (SI 2078) state that "all animals must either have access to a suitable water supply and be provided with an adequate supply of fresh drinking water each day, or be able to satisfy their fluid intake needs by other means". Although, the Defra welfare codes suggest that the water content of feedstuffs, such as root crops is not a sufficient source of water (Defra, 2003). It is clear that further research is needed in order to investigate the welfare and physiological requirements for water in housed and grazed sheep and this would also be useful for informing the on-farm provision of water resources and setting welfare standards and statutory regulations.

1.2.2 Freedom from discomfort – *by providing an appropriate environment including shelter and a comfortable resting area*

As adverse weather conditions can affect the on-farm welfare of grazing sheep (Macnab, 1998a, b), these animals should be provided with a source of shelter and shade. Whilst housing of sheep during specific periods of the production cycle, such as during the lambing season, can have beneficial effects on welfare, there is often a trade-off between economic viability, productivity and animal welfare (Webster, 1994). This is because housed sheep have restricted areas for exercise and cannot make choices regarding their environment and nutrition. Therefore it has been recommended that sheep are housed for short periods of time (FAWC, 1994) and are provided with dry clean, comfortable conditions, appropriate stocking densities, sufficient feeding space, clean water, sufficient lambing pen size, adequate lighting and comfortable flooring surfaces (Lynch *et al.*, 1992; McGreevy *et al.*, 2007a). Whilst for grazing sheep, the provision of a comfortable and well drained lie-back area meets the criteria for comfortable lying conditions.

Environmental hygiene is also important in order to minimise diseases such as mastitis (Sevi *et al.*, 2001b) and infectious diseases of neonatal lambs. Therefore, housing systems also need to provide effective ventilation systems to control air temperature and humidity. Space allowance for sheep depends on the age, size and stage of production and minimal requirements are detailed in the Defra codes of welfare (Defra, 2003). Although the stated requirements often rely on expert judgement and industry opinion and have not been scientifically validated (Lynch *et al.,* 1992). It is clear that adequate space is required in housing systems to prevent overcrowding of sheep, which can have deleterious effects to health and welfare due to the increased pathogen exposure, trauma and climatic stresses (Sevi *et al.,* 2001b). Housed sheep may also be subjected to psychosocial stressors, such as bullying and dominance behaviours including wool picking or wool-biting; these are behaviours that appear to specifically occur in crowded and intensive housing conditions (Done-Currie *et al.*, 1984).

1.2.3 Freedom from pain, injury and disease – *by prevention or rapid diagnosis and treatment*

Sources of poor welfare identified under this freedom were management practices that were associated with pain and injury and the management of specific sheep diseases. Sheep farm management practices that were identified to cause pain included procedures permitted under the Mutilations (Permitted Procedures) (England) regulations (2007), such as ear notching, ear tagging, castration, vasectomy, dehorning, and tail docking, as well as electro-ejaculation and obstetrical care, and are firstly discussed below.

Electro-ejaculation

The physical and behavioural reactions of pain displayed by some rams undergoing electro-ejaculation had led to concerns over this management practice (Stafford, 1995). Although research has found that the transit time to enter a handling pen for electro-ejaculation or part-shearing was not significantly different, this does not mean that the rams do not find the procedure to be aversive (Stafford, 1995; Stafford *et al.,* 1996). In humans, the procedure is known to be painful and is performed under anaesthesia, so it is suggested that electo-ejaculation of rams is performed under sedation and analgesia (Stafford, 1995). Since the stimulation of skeletal muscles as well as erectile, smooth muscle is thought to occur, it has been suggested that a probe with ventrally positioned electrodes may reduce the level of pain and discomfort experienced (Stafford, 1995).

Artificial insemination and embryo transfer

The use of embryo transplant and artificial insemination (AI) has been raised as an animal welfare issue, due to concerns over the pain endured, the lack of analgesia and sedation, the withdrawal of water and food, together with surgical complications and post-operative infections (Fisher and Scobie, 2003). Jeffam (1987) suggested that if laparoscopic procedures for artificial insemination (AI) were stressful, conception rates over 70% would be unobtainable. However, it is recognised that animals can maintain high standards of production and still continue to be able to reproduce under conditions of poor welfare (Webster, 1994). Given the technical difficulties encompassed with trans-cervical AI and embryo transfer (ET) (FAWC, 1994), these procedures must only be performed under the direction of a veterinary surgeon under the Mutilations (Permitted Procedures) (England or Wales) regulations 2007.

Obstetrical care

The highest incidence of ewe mortality occurs around lambing time when estimates of 5 – 25% have been reported (Binns *et al.*, 2002; Green and Morgan, 1994; Winter, 2007). Mortality may be due to inadequate supervision due to staffing shortages, lack of obstetrical management skills and economics of sheep production which prevent the acquisition of appropriate veterinary attention (Scott, 2005a). Although shepherding at lambing time can improve the health and welfare of lambs and ewes, human disturbance can affect the contractility of the uterus leading to delays in lambing and can increase the risk of dystocia (Fisher and Mellor, 2002). In addition, intervention with parturition conflicts with the freedom to express normal behaviour by allowing the ewe to lamb unaided and develop maternal bonding (Fisher and Mellor, 2002). At the same time, the freedom from pain and injury indicates that mal-presentations and dystocia should be corrected and shepherds need to take steps to minimise predation and disease (Fisher and Mellor, 2002). Protracted lambing also affects the quality of maternal behaviour, impinging on the development of the crucial ewe-lamb bond, with implications for lamb morbidity and mortality (Nowak, 1996). Studies have found that lambing assistance is being carried out using unhygienic practices (Scott, 2003) and inappropriate methods of replacing vaginal prolapses are being used, including the lack of analgesia and epidurals (Scott *et al.*, 1995). Following the correction of dystocia, less than two-thirds of study farmers provided antibiosis for the ewe and when this was supplied, over 75% of farmers provided ewes with an inadequate course of antibiotic (Scott, 2003). In cases of maternal rejection, or morbidity and mortality of the ewe, lambs may be fostered onto another ewe. To facilitate bonding, the head of the ewe may be restrained, which can result in physical and behavioural isolation as well as restricting the supply of food and water.

Tail-docking

Tail-docking - the removal of a portion of the tail from young lambs has been used to prevent the accumulation of faeces around tail and breach areas and the risk of myiasis (French *et al.,* 1994b; Wood and Molony, 1992). Results of research regarding the benefits of tail-docking for myiasis prevention have been conflicting. Whilst it has been suggested that, compared to docked lambs, undocked lambs are 5 times more likely to be affected by myiasis (blow fly strike) (French *et al.,* 1994b), other authors suggest that the incidence of fly strike was greatest in animals with a short docked tail (Fisher, 2004).

Tail-docking can be performed using four techniques: application of an 'elastrator' or rubber ring, use of a sharp knife, application of a bloodless instrument e.g. nipper or burdizzo instruments, or application of a hot iron (Wood and Molony, 1992). The application of a rubber ring to the tail severs the vascular supply so that the lower part of the tail sloughs several weeks later. Alternatively, the use of a sharp knife allows immediate removal of a proportion of tail, but there is a risk of severe blood loss and the open wound is susceptible to infection (FAWC, 1994). These methods of tail docking are permitted under the Mutilations (Permitted Procedures) (England) Regulations 2007. Methods of tail docking that sever vascular supply, such as elastrator rings, can only be performed without the use of anaesthetic within the first 7 days of life (Wood and Molony, 1992). For other methods, an anaesthetic must be used in lambs over 3 months of age. Regardless of the technique and instruments used, all methods of tail-docking are associated with pain and discomfort. The greatest level of acute pain is associated with docking by rubber ring although, subcutaneous application of a local anaesthetic can reduce this response (Graham *et al.,* 1997; Kent *et al.,* 2004).

As well as the acute pain caused, the short term welfare issues of tail-docking include the stress caused by handling of lambs, haemorrhage, infection, burns, and subsequent fly strike (Wood and Molony, 1992). Long term welfare issues include chronic pain, inflammation and infection (French and Morgan, 1992; Wood and Molony, 1992). Under welfare legislation, the amount of tail left on each animal must cover the vulva of a female and the anus of a male sheep (Defra, 2003). There are concerns regarding short tail-docking in ewes, which can be associated with vulval tumours (Scott *et al.,* 2007).

There are also inconsistencies in the practice of tail-docking on farms within the British Sheep Stratification (BSS) system (Pollott and Stone, 2006). Whilst tail-docking is routinely performed on lowland flocks; hill-bred lambs, which may be sent for fattening on lowland farms, are not routinely docked. So it is not clear why there are differences in the tail-docking policies of these farm types (Scott *et al.*, 2007). As such FAWC (2010) have called for farmers to consider whether tail-docking needs to be routinely applied and are encouraging the use of alternative management strategies to prevent and control outbreaks of myiasis.

Castration

Castration is considered to cause even greater pain and discomfort than tail-docking (Molony *et al.*, 2002). The method is practised on farms to maximise fattening of lambs of slow maturing breeds, prevent unplanned mating, prevent carcass downgrading from "ram effects", and reduce fighting and associated injuries (Wood and Molony, 1992). Although there have been claims regarding improved taste, growth rates and quality of castrated lamb, these claims may be unsubstantiated (Molony and Kent, 2007). The techniques used for castration are similar to those described for tail-docking. The application of a tight rubber ring to the base of the scrotum is the most commonly used castration technique in the UK (FAWC, 1994). The ring severs the testicular vascular supply immediately, although the nervous supply persists for a few hours (Kent *et al.*, 2004). Similar to tail-docking, the use of a rubber ring for castration is restricted to lambs less than 1 week old (Defra, 2003). Alternative methods include the use of bloodless castrators which crush the spermatic cords causing testicular atrophy, or surgical removal of the testes (Wood and Molony, 1992).

Research has clearly demonstrated that all methods of castration cause acute pain and distress (Molony and Kent, 2007; Thornton and Waterman-Pearson, 1999). Surgical castration has been suggested to cause the greatest level of acute pain (Thornton and Waterman-Pearson, 1999) and this is apparent even when analgesia is used (Melches *et al.*, 2007). Surgery also carries the risk of intestinal herniation into the open scrotal wound, haemorrhage and infection (FAWC, 1994; Wood and Molony, 1992). 'Bloodless' methods of castration lead to sloughing of the testes and the inflammatory responses within the scrotal skin and tissue may also lead to chronic pain (Thornton and Waterman-Pearson, 1999). There are also risks with the burdizzo technique as there is considerable variation in the maintenance, quality and rushing pressure applied by burdizzo instruments, which can mean that ram lambs are injured or remain uncastrated (Hosie *et al.,* 1996). There can therefore be welfare concerns if uncastrated ram lambs are unintentionally managed with ewe lambs as these lambs may be sent for slaughter in the late stages of pregnancy (Wood and Molony, 1992).

Due to concerns over the short- and long-term consequences of castration for lamb welfare, FAWC (2008) have suggested that there should be a reduction in the use of castration and alternative management practices, such as maintaining separated groups of ewe and rams lambs, are used. Where there is a particular need for castration, it is recommended that local anaesthesia and analgesia are provided (FAWC, 2008). However, in spite of the evidence that shows that improved welfare is afforded by using local anaesthesia in conjunction with rubber ring castration (Graham *et al.,* 1997; Kent *et al.,* 1998), the time and cost implications may influence the routine uptake of this method on farms (Kent *et al.,* 2004).

Ear tagging

In common with all other countries in the European Union (EU), the UK is required to operate a national scheme for the identification of sheep under EU directive 92/102. Ear tagging has been identified as a welfare issue due to the pain, inflammation and infection caused (Edwards *et al.,* 2001). All methods of ear tagging can result in long-term damage and tags placed nearer to the ear tip cause more damage (Edwards *et al.,* 2001). A study on the effect of the main tags available in the UK discovered that "all flex" style flexible plastic tags caused the fewest problems, followed by golf tee-shaped plastic tags, whilst metal loop and plastic loops tags were linked to the highest number and greatest severity of lesions (Edwards *et al.,* 2001). As tagging is a legal requirement, it is not possible to eliminate the welfare issues associated with the method but the research could be used to inform best on-farm practices in order to attain the highest standards of sheep welfare.

Diseases of sheep

Disease is considered to be a major cause of poor welfare in sheep. In particular, diseases that are found at a high prevalence and long duration are of concern for sheep welfare. On-farm management is key to the control and prevention of disease as management practices and environmental factors such as exposure to extreme climatic conditions or overcrowding of groups of animals may predispose to certain diseases of sheep (Roger, 2008). Management of disease may involve preventive management schemes such as vaccination and use of flock health and welfare planning, which may reduce the impact of disease and adverse welfare conditions for sheep (Scott *et al.*, 2007). In addition, sheep with signs of pain and disease need to be treated appropriately. Although it is recognised that non-steroidal anti-inflammatory drugs (NSAIDS) are effective in reducing pain, they are currently not licensed for use in sheep in the UK (Fitzpatrick *et al.*, 2006). In addition, concerns, including the cost of analgesia (Waterhouse *et al.*, 2003) and the need for veterinary diagnosis and advice (Scott, 2005b), may influence the uptake of methods of reducing the impact of disease and pain on sheep farms.

Sheep diseases of welfare relevance include Johnne's disease, Pasteurellosis, contagious conjunctivitis, entropion, tick-borne disease, myiasis, abortion, Clostridial diseases, urolithiasis, dental disease and orf (Edwards, 2005; Lovatt, 2005; Scott *et al.*, 2007; Winter, 2007). Overall, lameness due to footrot, endo- and ectoparasitic disease and mastitis were identified as the most important on-farm welfare issues for sheep because of the severity of pain and discomfort, potential high on-farm prevalence and the duration of their effects on both the short-term and long-term welfare (Fitzpatrick *et al.*, 2006; Harkins, 2005). Accordingly, lameness, parasitic diseases and mastitis will be discussed in further detail below.

Lameness

Lameness is a significant and serious welfare issue for sheep because of the pain, discomfort and debilitation caused (Fitzpatrick *et al.*, 2006; Ley *et al.*, 1995; Welsh *et al.*, 1993). There are a range of causes of sheep lameness - in young lambs the most common condition is suggested to be infectious polyarthritis ('joint ill'), whereas for older sheep, the main concern is footrot (Winter, 2004b). Footrot has received particular attention because of the degree and chronicity of pain, lameness and the debility caused, which persists several months after treatment (Fitzpatrick *et al.*, 2006; Ley *et al.*, 1995; Welsh *et al.*, 1993). Footrot is also a preventable disease but it is apparent that the optimal methods to prevent and control footrot are not being performed by farmers (Grogono-Thomas, 2001; Peddie *et al.*, 2003). Another foot condition, known as Contagious Ovine Digital Dermatitis (CODD) infection is associated with severe lameness, recumbency and pain due to loss of the hoof capsule and appears to be of increasing importance to flock welfare and may not respond to standard footrot treatments (Winter, 2004a; Winter, 2008). The necessity, type and quality of any routine foot care treatment or therapeutic intervention for lameness are important as sheep welfare can be compromised by inappropriate and ineffective treatments (Grogono-Thomas, 2001). The ability to observe early stages of lameness and determination of the correct diagnosis are therefore important for the effective treatment of foot lesions and flock lameness (Kaler and Green, 2008a, b, 2009). The pain and discomfort caused by lameness can result in reduced feed intakes and affected sheep may be of poorer body condition. Furthermore, the presence of an infectious foot lesion, such as footrot, can also predispose to other welfare conditions. Sheep affected with foot rot may be recumbent and infected material from the foot can rub onto the fleece attracting blowflies, predisposing sheep to myiasis (Winter, 2004b).

Mastitis

Although recognised as a welfare issue for sheep, the scientific knowledge regarding the risk factors, prevention and control of mastitis in ewes managed under UK farming systems is currently limited. At present, two forms of mastitis are clinically recognised in sheep, both of which appear to be associated with pain and discomfort (Dolan *et al.*, 2000). Acute mastitis, associated with peak lactation in the post-lambing period, can be fatal or result in recovery of the ewe but permanent damage to the mammary gland can ensue. By contrast, the identification of chronic mastitis cases may be delayed until ewes are fully inspected during the pre-tupping period (Winter, 2001). In addition to pain and discomfort, mastitis can affect ewe appetite and body condition, milk yield and quality (Sevi, 2007), with consequences for the lamb including starvation and death.

Endoparasitism

Parasitic gastroenteritis (PGE) is an important health and welfare issue for sheep due to the severe metabolic and physical effects of protein and electrolyte losses, which can lead to anorexia, anaemia, diarrhoea, and death (Athanasiadou *et al.*, 2008). Faecal soiling of the tail and perineal area also increases the risk of blow fly strike (French *et al.*, 1996). In sheep, high levels of parasitism are associated with chronic stress and this has an effect on productivity including reduced immunity, poorer meat quality, reduced reproductive performance and poor body and fleece growth (Coop *et al.*, 1982). Accordingly, gastrointestinal nematodes have been controlled in sheep using the application of anthelmintic drugs (Taylor *et al.*, 2007), but there is increasing concern regarding the excessive and inappropriate use of anthelmintics which have been associated with the development of parasite resistance (Jackson and Coop, 2000). In comparison, concerns have been raised regarding the limited number of anthelmintic treatments permitted for parasite control in organically-reared sheep (Benoit and Laignel, 2002; Hovi *et al.*, 2003). So, the welfare concerns regarding endoparasitism may be influenced by the system of sheep farming.

Ectoparasitism

The major ectoparasites of relevance to sheep welfare are sheep scab (*Psoroptes ovis*), lice (*Bovicola ovis*), ked (*Melophagus ovis*) and blowflies (*Lucilia* and *Calliphora* species) (Plant, 2006). Concerns regarding sheep scab are attributed to the intense pruritus caused, and in severe cases, epileptiform fits and death may occur. The severity of pruritus appears to be associated with the age and size of the lesion and the duration of infestation (Berriatua *et al.*, 2001). The time spent rubbing, scratching and biting of skin lesions can also interfere with grazing, ruminating and other normal behaviours, resulting in weight loss (Berriatua *et al.*, 2001; Corke and Broom, 1999). The introduction of sheep scab may be due to lack of knowledge, understanding of the welfare effects of scab and previous on-farm experience of the disease (Morgan-Davies *et al.*, 2006). Until 1992, sheep scab was classified as a notifiable disease in the UK and the low incidence of outbreaks was associated with compulsory dipping with organophosphates. The method of dipping remains an effective means of treatment and control but requires specific on-farm resources and skilled and competent labour. There can be problems with quality of handling and gathering of sheep and outbreaks of post-dipping lameness which can follow poorly managed dipping (Plant, 2006).

Myiasis – the invasion of living animals by blowfly larvae, was also identified as an on-farm concern for sheep (Farkas *et al.*, 1997; Macleod, 1992). The *Lucilia sericata* species of blowfly are a primary cause of myiasis and are capable of invading intact skin, resulting in severe tissue damage, pain, discomfort, debilitation and even death (Hall and Wall, 1995; Macleod, 1992). Myiasis generally affects the skin or body orifices and blowflies are attracted to areas of faecal soiling, urine contamination and infected wound. The risk of myiasis ('blowfly strike') is influenced by the size of the fly population, susceptibility of sheep, location and altitude of the farm (French *et al.*, 1994a). Diarrhoea is a major risk factor and associated with the development of flystrike in the breech region (French and Morgan, 1996). The feeding activity of *L. sericata* produces extensive tissue damage and sheep affected with myiasis may appear restless and anxious. The severe annoyance caused to the sheep reduces feeding and grazing leading to loss in body condition (Farkas *et al.*, 1997). Other species that are relevant to the welfare of sheep include *Oestrus ovis,* also known as the sheep nasal bot fly, which grows and develops in the sinuses and nasal passages (Hall and Wall, 1995).

Lamb mortality

Mortality rates of neonatal lambs (aged $\leq$ 3 days) managed under UK and New Zealand farming systems have been suggested as 5 – 25 % (Binns *et al.*, 2002; Fisher and Mellor, 2002; Mellor and Stafford, 2004). The majority of neonatal deaths are attributed to dystocia, starvation, hypothermia, and placental circulatory compromise (Binns *et al.*, 2002; Christley *et al.*, 2003; Nowak, 1996). Hypothermia of lambs is a major problem for both indoor and outdoor lambing systems as low temperatures compromise the ability of lambs to locate teats may lead to starvation (Mellor and Stafford, 2003; Nowak, 1996). On-farm lambing management skills have the potential to alleviate certain welfare issues and the level of husbandry is associated with lamb survival (Ducker and Fraser, 1973). For example, provision of shelter for outdoor lambs reduces neonatal mortality (Broster *et al.*, 2010), whilst providing supplementary feed to ewes improves their body condition, increases the quantity of maternal behaviour and lamb birth weight (Christley *et al.*, 2003; Nowak, 1996). Lamb mortality can be reduced by establishing good physical areas for lambing and allowing good maternal-offspring bonds to form. Greater levels of lamb survival and less separation occur when ewes are penned at the site of birth (Nowak 1996). Bonding behaviour encourages lambs to suck and promotes vocalisation between ewe and lamb. In addition to the quality of resources provided during lambing time, factors such as ewe genotype, breed (Dwyer and Lawrence, 2000; Dwyer *et al.*, 1996) and prolificacy (Mellor and Stafford, 2004) affect the quality of maternal behaviour and subsequent survival of the lamb.

Veterinary attention

A major welfare issue for sheep is the failure to take prompt action to treat and control disease and the concern is that veterinary attention is not sought early, if at all (Waterhouse *et al.*, 2003). A survey of farmers and their veterinary practices in the Scottish borders calculated that veterinary visits amounted to less than 2 visits per flock each year (Clements *et al.*, 2002). Farmers indicated that the main reason they would acquire veterinary assistance was the degree of pain and discomfort observed, followed by concerns for the health of other sheep in the flock (Clements *et al.*, 2002). Flock health planning can tailor specific prevention and control in management, husbandry and use of veterinary treatments to a specific flock (Scott, 2005b). However, reports suggest that despite a free-of-charge visit, farmers delayed obtaining veterinary advice during flock disease outbreaks. It appears that the cost of veterinary fees and treatment are clearly visible to farmers but the financial cost of production losses, management inputs and welfare implications are less obvious (Lovatt, 2005). The low level of veterinary input on some flocks may not only be due to financial factors but also the perceived competence of sheep farmers (Evans and Scott, 1999)

1.2.4 Freedom to express normal behaviour – *by providing sufficient space, proper facilities and company of the animal's own kind*

The freedom to express normal behaviour is strongly influenced by the management of the flock. As sheep are highly social animals, the maintenance of the social stability of groups of sheep is important to flock behaviour and welfare (Lynch *et al.,* 1992). Management practices such as the physical and visual isolation from other sheep, conditions of feed restriction or changing the composition of groups of sheep is perceived to be an aversive experience, as evidenced by the increased incidence of stereotypies, such as wool-biting, wool pulling and aggressive interaction, that have been observed (Done-currie *et al.,* 1984; Lynch *et al.,* 1992). In addition, the ability to express normal and natural behaviours, such as rumination and exploration, may not be fulfilled when sheep are managed under intensive, housing conditions. Rather than changing the environment to afford improvements in animal welfare, more recent research efforts have focused on producing breeding animals with particular behavioural traits (D'Eath *et al.,* 2010) that may be able to cope better with intensive farming and management practices.

1.2.5 Freedom from fear and distress – *by ensuring conditions and treatment which avoid mental suffering*

Appropriate management system and breed

Selecting a breed suited to the farm management system and environment is an important concept for good welfare. Highly prolific breeds and those with little fleece cover and lean carcasses are not considered to be appropriate for extensive sheep farming systems (Goddard *et al.*, 2006). There has also been debate as to whether hill ewes on extensive systems are required to lamb every year (Goddard *et al.*, 2006). As a result, there has been interest in the use of 'easy-care' breeds which produce a single lamb with minimal human assistance, with the aim of increasing lamb survival, preventing dystocia and reducing labour requirements at lambing (Sargison, 2000). However, extensive sheep production requires considerable labour input so the 'easy-care' system or management may not be suitable for many UK farms (Goddard *et al.*, 2006; Sargison, 2000).

As well as selecting the right breed of ram and ewe, it is important to select rams of high health and welfare status. This is because a single ram can play a major role in the prevention of certain diseases and traits, with the potential for positive effects on a large number of animals. Therefore, it is of concern that on-farm welfare inspections have identified deficiencies in the management of rams in the UK, including the failure to remove rams at the end of the tupping period, and a lack of year-round inspection and attention (Macnab, 1998b).

Predation

Predation of lambs is considered to be a greater welfare issue for extensively managed sheep, who are more at risk of attack by wild animals (Dwyer, 2009; May *et al.*, 2008). The main flock defence against predators involves 'refugeing', in which the flock groups together and run for cover (Dwyer, 2004). Although, differences in predation density and sheep breed may affect the behavioural response to the threat of a predator (Dwyer, 2009). As well as predation by wild animals, the presence of shepherding dogs and humans may also be perceived as predators by sheep. Sheep worrying, mauling and killing caused by untrained dogs plus the use of shepherding dogs may therefore be an aversive experience for sheep (Cockram, 2005).

Weaning procedure

The bond that develops between a ewe and her offspring can be very strong and selective (Nowak, 1996), and so removal of the lamb as part of weaning has been suggested to be an aversive experience as both ewes and lambs exhibit strong fearful behavioural reactions following separation (Cockram *et al.,* 1993; Napolitano *et al.,* 2002). In practice, abrupt weaning methods are routinely used on farms in England and Wales. The fear and distress associated with weaning might be reduced by modifying the method of weaning since research has demonstrated that ewes and lambs can habituate to repeated separation for short periods of time (Cockram *et al.,* 1993; Orgeur *et al.,* 1998). This method would probably be much more time demanding and labour intensive and so does not appear to be feasible for many sheep flocks managed under the BSS system.

The effect of the stockperson on the Five Freedoms

Ultimately the greatest influence for the on-farm welfare of the individual and flock of sheep is the skill, attitude and behaviour of the stockperson. This has the ability to affect each one of the Five Freedoms criteria. The welfare issues associated with the freedom from hunger and thirst are concerned with the provision of feed of an adequate quality and quantity, provision of a clean and accessible supply of water and the recognition that nutritional demands vary according to age, health and physiological stage of the sheep. Similarly, the freedom from discomfort implies that the stockperson recognises the need for thermal and physical comfort by providing a dry lying area, shade and shelter for sheep managed both indoors and outdoors. These are all concerns which rely on a range of resource inputs and management decisions. A high level on-farm labour input has been positively correlated with high welfare levels (Goddard *et al.,* 2006), although this is reliant on the quality of care afforded by the individual shepherd. The correct use of dosing guns, careful and proper foot trimming methods and sterile administration of treatments and vaccines are important sheep management practices that have the potential to maintain good animal welfare (FAWC, 1994).

Freedom from pain, injury and disease also implies that shepherds select the right animals as breeding stock – those free from conformation or inheritable defects and disease. The decision to practice mutilations, such as tail-docking and castration is likely to be influenced by the beliefs of the stockperson and possibly culture and tradition (Hemsworth, 2007). The stockperson also needs to correctly recognise disease and injury in order to provide appropriate intervention, and if necessary to promptly euthanase any debilitated or chronically sick animals (Winter, 2007).

Most on-farm welfare deficiencies do not appear to be the result of deliberate cruelty but appear to be related to the mental and physical health, financial situation and education of shepherds (Prettejohn, 1990). British sheep farmers themselves recognise that the limitations for good sheep welfare are the availability and quality of labour. Training farmers for competence in lambing may help improve ewe and lamb welfare (Scott, 2005a). However, the reduction in funding of training courses and reduced number of young people entering sheep farming is of concern (Winter and Fitzpatrick, 2008). The use of flock health plans may facilitate improvements in the productivity, economics, management and welfare of sheep (El Balaa and Marie, 2006) and may provide an opportunity for an increased level of veterinary input on sheep farms (Scott, 2005b).

The frequency of inspection and level of input provided by the stockperson will also depend on the type of management system, as hill or extensively managed sheep may not be gathered and inspected as frequently compared to sheep managed under more intensive conditions (Dwyer, 2009). Therefore, there may be a trade-off between each of the freedoms depending on the system of farming. Extensively-managed sheep are likely to have more freedom to express natural behaviour and have greater opportunity for exploration and grazing. However, these sheep can be exposed to extreme weather conditions, possible nutritional shortages and can receive less attention and inputs from the stockperson, so there are likely to be compromises for the other freedom criteria.

Gathering and handling of the individual and flock is a necessary aspect of sheep farming and the frequency depends on the type of management system. Handling has been identified as a welfare concern because of the fearful, anxious and frustrated behavioural reactions that sheep are perceived to experience (Le Neindre *et al.*, 1996). There is also a learned association between the quality of and skills of the handler so that rough handling can lead to an aversion for certain management procedures. The level of fear elicited by the handling of farm animals is affected by the behaviour and attitude of the stockperson (Hemsworth, 2007). Research has shown that young lambs can discriminate different stockpeople on the basis of their aversiveness and familiarity (Boivin *et al.*, 2001). Therefore, sheep that are not habituated to frequent handling may find handling and examination to be an aversive experience.

1.3 Developing indicators of sheep welfare

Following the identification of the key welfare concerns, the opinion of a group of experts can be used to identify indicators that are capable of assessing the relevant on-farm welfare issues. Welfare indicators then need to be tested on farms to identify that they are reliable and feasible for application under field conditions. This approach was used to develop animal-, resource- and management-based welfare indicators for dairy cows, pigs and laying hens by the BWAP (Whay *et al.*, 2003a) and Welfare Quality® (Botreau *et al.*, 2009) projects. Both projects used the level of observer agreement or 'reliability' to assess the validity of measures of farm animal welfare. However, the welfare criteria used to develop the animal welfare indicators differed across the two research projects. The Five Freedoms framework was used by BWAP, whereas the Welfare Quality® project developed a set of four welfare criteria based on good housing, good feeding, good health and appropriate behaviour (Botreau *et al.*, 2009).

In comparison to cattle, poultry and pigs, fewer indicators of sheep welfare have been developed and tested (Harkins, 2005; Napolitano *et al.*, 2009). Currently, the on-farm welfare of sheep in England and Wales is assessed as part of farm assurance schemes, such as the RSPCA and Soil Association, or for assessments with statutory welfare legislation. Currently, inspectors from Animal Health, an agency of the Department for the Environment, Food and Rural Affairs (Defra) use the codes of recommendations for the welfare of sheep (Defra, 2003) to assess the on-farm compliance with the Animal Welfare Act (2006) and Welfare of Farmed Animals (2007) regulations.

However, to date, few valid, reliable or feasible measures have been developed and tested to assess the on-farm welfare of sheep in the UK. Scientifically valid measures are needed to identify whether management changes have a positive or negative effect on the outcome of sheep welfare. In addition, the sheep industry needs to be able to demonstrate good welfare standards in order to maintain public confidence and to ensure that public money is not subsidising any farmers who maintain poor standards of animal welfare.

Expert opinion has been used to identify animal-based indicators for sheep transported by sea (Pines *et al.*, 2007), and extensively managed sheep (Goddard *et al.*, 2006). However, these indicators have not reportedly been tested in on-farm studies. More recently, five indicators used to assess the welfare of sheep managed in dairy flocks in Italy – body condition, integument alteration, claw overgrowth, dirtiness and skin lesions, were developed from measures of cattle welfare listed within the ANI protocol (Napolitano *et* *al.*, 2009). As indicators need to be sensitive to the current welfare issues for sheep, the modification of indicators that are used to assess the welfare of other farm animal species was not considered an appropriate approach for this thesis.

Previous studies, including BWAP and Welfare Quality®, have investigated the reliability and feasibility of animal welfare indicators as a means of evaluating the validity of the measures (Knierim and Winckler, 2009). Therefore, the reliability and feasibility were used as a means of examining the validity of on-farm indicators of sheep welfare. However, this thesis aimed to take a more thorough and extensive approach to validation of animal welfare indicators and therefore the animal-based indicators of sheep welfare was also evaluated in terms of their diagnostic performance (sensitivity and specificity) and in the context of their ability to detect seasonal and farm-level variation in the level of conditions associated with sheep welfare.

1.3.1 Validating indicators of animal welfare

A valid welfare indicator would be one that genuinely measures an animal's welfare status (Bracke *et al.*, 2004). Since welfare indicators are individual measures akin to diagnostic tests, principles used to validate diagnostic tests were used to assess the validity of indicators of animal welfare in this thesis. There are a number of different ways of establishing test validity, and therefore several approaches to evaluate the diagnostic performance of welfare indicators applied in this thesis.

For the purposes of diagnostic test evaluation, validity is generally defined as the ability of a test to produce correct test results (Greiner and Gardner, 2000). So, a welfare indicator could be conferred with criterion validity if the measure produced the same result as the reference test or 'gold standard' (Greiner and Gardner, 2000). Currently there is no 'gold standard' for the assessment of animal welfare (de Passillé and Rushen, 2005). In the absence of a reference test, the first step in determining the validity of a measure is to ascertain a consensus of expert opinion. By judging the suitability and relevance of the measures, expert opinion can provide consensual, content and face validity to the selection of animal welfare indicators (Abramson and Abramson, 2008; Scott *et al.*, 2001).

1.3.2 Using reliability as a measure of test validity

As a welfare indicator cannot be deemed to be 'valid' if it produces unreliable results (Hewetson *et al.*, 2006), the next step in the validation of animal welfare indicators is to evaluate the reliability of the test (Hewetson *et al.*, 2006; Kaler *et al.*, 2009). Reliability implies that observations are consistent between different observers and under variations in measurement conditions (Cronbach *et al.*, 1972). Consistent assessments by different observers are needed to demonstrate the fairness and robustness of any animal welfare measure. There are two ways of examining observer reliability: 1. intra-observer reliability (also known as test-retest reliability or repeatability) in which the same sample is repeatedly assessed by the same observer and 2. inter-observer reliability (also known as reproducibility) whereby different observers independently assess the same sample (Burkholder, 2000). Given the many terms that are found in the literature, the term 'reliability' will be applied throughout this thesis.

The reliability of indicators of sheep welfare was evaluated in this thesis using methods currently used in the field of animal welfare science. Kappa (κ) agreement has been used to evaluate the reliability of categorical or ordinal scoring indicators of cattle (Kristensen *et al.*, 2006), pig (Petersen *et al.*, 2004), horse (Burn *et al.*, 2009) and chicken welfare (Butterworth *et al.*, 2007). Essentially, κ assesses the degree of observed agreement compared to the agreement expected by chance (Sim and Wright, 2005). The type of κ selected depends on the number of observers involved - Fleiss's κ determines the reliability of multiple observers ($n > 2$) (Fleiss, 1981), whereas Cohen's kappa (Cohen, 1960) examines the reliability of paired assessments ($n = 2$). As κ assumes that all scoring disagreements are equally serious (Sim and Wright, 2005), it can also be useful to use Kendall's coefficient of concordance (Kendall's W) to examine the reliability of indicators that are scoring along ordinal (Sim and Wright, 2005) or continuous scales (Martin and Bateson, 2007).

The reliability of discrete (count) data can be evaluated using a reliability scale coefficient known as Cronbach's alpha (α) (Cronbach, 1947). This approach has been used to assess the reliability of welfare indicators of cattle (Herva *et al.*, 2009) and pigs (Munsterhjelm *et al.*, 2006). Essentially, α estimates the amount of variance in observer scores and the sample population to produce a value on a scale of 0 to 1 where α values close to 1 suggest a low level of variance (Cronbach *et al.*, 1972). An overall α value is not appropriate for studies where the same assessors do not examine the same sample population; instead a number of α values can be determined (Cronbach *et al.*, 1972). As validity is also concerned with the level of bias and measurement error, a useful way of examining the data is by using the approach of Bland and Altman (1986).

1.3.3 Validation of indicators in the absence of a gold standard

The test performance of any indicators developed in this thesis also needs to be examined under different farm conditions to identify whether the measures show predictive, convergent and construct validity and are responsive to seasonal and farm-level variation. Predictive validity is determined by identifying a connection between a welfare measure and an expected event or outcome. Whereas, measures that show a relationship with the underlying measure, for example, an association between negative mood and the level of lameness, could provide evidence of the construct validity of indicator tests (Abramson and Abramson, 2008). Convergent validity (or discriminant) is interested in the correlation between tests that measure the same welfare condition, for example the correlation between different methods of lameness assessment. Finally, the validity of an animal welfare measure can be examined in terms of its ability to measure a change (Abramson and Abramson, 2008), for example, the ability to detect seasonal variation in welfare indicator assessments.

1.3.4 Diagnostic sensitivity and specificity

In addition, diagnostic tests are frequently validated in terms of the level of diagnostic sensitivity (Se) – the proportion of animals that the test identifies as having a particular welfare condition; and specificity (Sp) – the proportion of animals the test identifies without the welfare condition (Greiner and Gardner, 2000). The diagnostic Se and Sp of a binary scoring test can be identified using cross-classification tables but this requires the use of a 'gold standard' test. This may be the reason why the diagnostic performance of animal welfare indicators has not been evaluated by other research studies. In this thesis the clinical examinations of an experienced assessor (the author) were selected as the reference test ('test standard observer') and were used to compare the test performance of other observers (Burn *et al.*, 2009; Hoehler, 2000). Accordingly, the Se and Sp of a group of observers was compared to the results of the 'test standard observer'. As clinical assessments and human observations involve a degree of subjectivity, the use of human observations as a gold standard may provide biased estimates of test performance (Bertrand *et al.*, 2005; Nielsen *et al.*, 2004). These issues motivated the development of statistical methods of analysis known as 'Tests in the Absence of a Gold Standard'

(TAG's), including Bayesian methods of analysis such as Latent Class Analysis (LCA) (Hui and Walter, 1980). LCA does not assume that the true condition of test subjects is known (Hui and Walter, 1980) but assumes 1. conditional independence given the disease or welfare condition; 2. Se and Sp are constant across different populations; and 3. the true (latent) prevalence of the welfare indicators differs across the study population (Hui and Walter, 1980). LCA has been previously applied to evaluate the diagnostic validity of a number of veterinary diagnostic tests (Bonde *et al.*, 2010; Nielsen *et al.*, 2004; Toft *et al.*, 2007a) and was therefore selected as the means of evaluating the diagnostic performance of animal-based indicators of sheep welfare in this thesis.

1.3.5 Design and conduct of validation studies

The concept of validity also encompasses the generalizability, applicability, and feasibility of the study findings (Rothwell, 2006). As well as being reliable, any welfare indicator developed in this thesis needs to be feasible for application under working farm conditions and applicable within the time and financial limits of a one-day visit.

In addition, whilst a study may confer test validity, it does not automatically follow that the study findings are generalisable because the sample population may be biased to a particular condition and may not be representative of the wider population (Greiner and Gardner 2000). Therefore, the guidelines of the Standards for Reporting of Diagnostic Accuracy (STARD) (Bossuyt *et al.*, 2003a; Bossuyt *et al.*, 2003b), Quality Appraisal of Reliability Studies (QAREL) (Lucas *et al.*, 2009) were used in this thesis to aid the design, conduct and reporting of validation study findings.

Important considerations in STARD and QAREL include the use of representative samples and observers, use of blinded observers, research setting, suitable intervals between repeat observations, clear scoring protocols, and the use of correct statistical analysis and interpretation (Lucas *et al.*, 2009). For the purposes of this thesis, study farms should be representative of the BSS system and measures should be relevant and applicable to everyday assessment conditions on sheep farms. Achieving a representative sample with a mixed prevalence of welfare indicators is difficult in field studies (Burn *et al.*, 2009), so some researchers impose welfare conditions on sample animals; for example maintaining sheep in very poor body condition (Calavas *et al.*, 1998) but this was not considered to be suitable for the field studies performed in this thesis. Furthermore, the assessors should be representative of those who are expected to apply them in the future (Lucas *et al.*, 2010).

As training, experience and occupation (Harkins, 2005; Kristensen *et al.*, 2006) and experience can affect the level of observer reliability it may be appropriate to study the reliability of a group of observers from a range of different backgrounds. The scientific literature does not appear to provide any consensus on the number of observers or number of visits required in validation and reliability studies. Walter *et al.*, (1993) propose the use of 3 assessors, although in practice a range of 2 (Napolitano *et al.*, 2009) to 56 observers has been used (Kristensen *et al.*, 2006).

The setting of reliability studies also requires consideration (Lucas *et al.*, 2009). As repeat on-farm assessments may not be feasible, it may be useful to use video clips as a means of evaluating observer reliability (Kaler *et al.*, 2009) and this approach might be beneficial for tests which are affected by the time stability of the measure of interest (Lucas *et al.*, 2009). Finally, the development of a clearly defined scoring system is also important since the scoring system (Tuyttens *et al.*, 2009) and the terminology used (Flower and Weary, 2009) can affect the level of observer agreement.

1.4 Indicators of sheep welfare identified by the literature review

The next step taken in this chapter was to use a review of the scientific literature to identify non-invasive indicators of sheep welfare that were considered to be valid, reliable and feasible for application under working farm conditions. Indicators identified by the literature review were then categorised as either animal- (physical, behavioural, anatomical and physiological measures), resource- or management-based indicators of sheep welfare and are discussed within the relevant category of the Five Freedoms framework. However, it is recognised that several indicators can be categorised into more than one Freedom area. The results of this review were then used to inform further studies presented in other chapters of this thesis.

1.4.1 Freedom from hunger and thirst – *by ready access to fresh water and a diet to maintain full health and vigour*

Body condition

Body condition scoring of sheep has been scientifically validated as a measure of internal body fat (Russel, 1969; Sanson, 1993) and is widely accepted as an indicator of previous nutritional management (Caldeira *et al.*, 2007). The technique was originally developed as a management tool so that changes in body condition could be detected, and adjustments in nutrition and body weight could be made (Jefferies, 1961). Body condition scoring of individual and groups of sheep also forms a routine part of veterinary clinical examinations (Lovatt, 2010) and flock health planning (Sargison and Scott, 2010). There are difficulties in weighing animals throughout the year and there can be a large amount of variation in the body weight of healthy animals (Russel, 1984). An advantage of body condition scoring is that it offers a way of comparing individuals within a flock and benchmarking of animals from different flocks (Russel 1984). Therefore, it has great potential as an on-farm indicator of sheep welfare (Caldeira *et al.*, 2007; FAWC, 1994).

The method for scoring the body condition of sheep, first described by Jefferies (1961) and later modified by Russel (1984), relies on manual palpation and assessment of the degree of muscle and fat cover over the *longissimus dorsi* muscle (eye muscle or loin). This region is used as it is the last area of the growing animal to develop and is thought to reflect rapid gains and losses in body fat (Jefferies, 1961). Alternative methods rely on palpation of the cover of fat over the tail (Sanson, 1993) or ribs (Shands *et al.*, 2009). At either end of the condition scale, a very thin or a very fat sheep can indicate potential compromise on welfare and health status (Jefferies, 1961). Ewe BCS during mid-gestation is a reliable indicator of ewe survival (Morgan-Davies, 2008). This is because ewes with BCS lower than 2 or higher than 3 are more susceptible to metabolic problems (Caldeira *et al.*, 2007), and an extremely low BCS during early pregnancy has deleterious effects on placental development and foetal growth (Muñoz, 2008). Very thin or emaciated ewes may be the consequence of poor feed utilisation due to chronic diseases such as molar tooth problems, parasitism, such as fasciolosis, or a wasting condition, such as Johnne's disease (Sargison and Scott, 2010).

As well as affecting ewe welfare, poor body condition also has potentially deleterious effects on the developing foetus and lamb survival (Tribe and Seebeck, 1962). Therefore, the presence of a large number of sheep of low body condition score may be used to indicate inadequate care and management (FAWC, 1994; Turner and Dwyer, 2007). Although body condition scoring has been scientifically validated as a measure of the internal body composition of sheep (Russel *et al.*, 1969), studies examining the reliability of the method have produced diverging results. For example, certain studies have demonstrated excellent levels of inter-observer reliability (Russel, 1969; Shands *et al.*, 2009) whilst others have identified considerable between-observer variation resulting in poor levels of reliability (Calavas *et al.*, 1998; Everitt, 1962; Harkins, 2005, Yates and Gleeson, 1975). It has been suggested that experienced assessors who are calibrated to the method of scoring, should perform body condition scoring of sheep (Evans, 1978; Milligan and Broadbent, 1974). Other studies have used different levels of scoring precision but the different analytical methods used to assess observer reliability can make it difficult to fully compare results from previous studies. Furthermore, details such as the number of assessors, assessor training and standardisation are also not always reported. In order to use body condition as a measure of sheep welfare, the method of condition scoring should be demonstrated to be valid, reliable and feasible. So, further studies addressing the reliability of this method appear to be needed.

Dehydration

The on-farm provision of water could be assessed by examining the number of water sources provided and by examining the quality and accessibility of the supply. As deprivation of water can lead to clinical signs of dehydration, this animal-based measure could be used to assess the freedom from thirst (Winter, 2007). Currently, no scientifically validated methods of measuring dehydration in sheep have been reported in the literature. The reliability of dehydration as a welfare indicator has been examined in working equids (Pritchard *et al.*, 2007). Studies found that the mucous membrane dryness and skin tent test were not deemed to be reliable indicators of dehydration. Instead, the quantity of water consumed, the number of water sources, and time spent drinking water were the most reliable measures (Pritchard *et al.*, 2008). Measures of packed, cell volume (PCV) could also be used to indicate hydration status, although this would require use of venepuncture and may not be feasible given the time, qualified personnel and resources needed for sampling.

1.4.2 Freedom from discomfort – *by providing an appropriate environment including shelter and a comfortable resting area*

Cleanliness scoring

The dirtiness of an animal can be used to indicate the quality of on-farm management, such as the hygiene and physical comfort of the environment. Recently, the reliability of dirtiness as a measure of sheep welfare has been evaluated on dairy flocks in Italy. Groups of sheep were observed from a distance of 2 metres and individual animals were classed as 'dirty' if they were observed with major splashing or distinct plaques of dirt over the hind-limbs and mammary glands (Napolitano *et al.*, 2009). An alternative method of cleanliness scoring for sheep is detailed within the Food Standards Agency (FSA) guidance on the cleanliness of animals presented for slaughter (FSA, 2007). This cleanliness scoring scale, based on a visual examination of the degree of dryness and contamination of the fleece, scores the cleanliness of an individual sheep on a scale from 0 (clean) – 4 (filthy) (FSA, 2007). The method appears to be a feasible on-farm measure, although the reliability of the scoring scale has not been reported.

Panting

The welfare of an animal can be compromised if it is unable to cope with its environment (Broom and Johnson, 1993). The welfare of sheep can be negatively affected under conditions of heat stress (Silanikove, 2000). Thermoregulatory responses to heat stress include reduced appetite, increased respiration rate, increased heart rate, panting and sweating and behavioural alterations, such as seeking shade or changing posture to aid heat loss (Silanikove, 2000). In sheep, the most important method of cooling occurs during expiration by the evaporation of water. Therefore, heat stress can be assessed using direct observations of sheep behaviour and signs of panting, which have been suggested to be feasibly assessed under extensive field conditions (Silanikove, 2000). Sheep managed in Mediterranean climates with no shade have a much greater respiration rate (RR) than those with access to shelter (Silinakove, 1987) and this work may suggest that the panting is a valid means of assessing thermal stress in sheep. Evidently a clear definition is needed to distinguish physiological panting following physical exertion from excessive panting associated with extreme thermal stresses and compromised welfare states. Lower levels of heat stress, indicated by a RR > 40 to 60 beats per minute (bpm) (Cockram, 2004; Silanikove, 2000), may reflect the physiological response associated with physical effort following gathering of the flock, whilst signs of excessive panting (RR 200 bpm) and open-mouthed breathing may indicate extreme thermal stress.

Body temperature

Rectal temperature is routinely used as part of clinical examination as a measure of health and disease. The method appears to be a valid means of reflecting core body temperature and could be used to measure the thermal comfort in sheep. Measurements of rectal temperature are easy to perform on individual animals, although it may not be a suitable means of assessing large groups of sheep. So, observations of respiration may be more appropriate for assessing the thermal comfort at the flock level (Silinakove, 2000). The on-farm assessment of thermal comfort also needs to consider the other extreme of temperature. Cold stress is also relevant to the on-farm welfare of sheep, and hypothermia is a well recognised cause of neonatal mortality (Dwyer, 2008b; Nowak, 1996). In this instance, direct observations of signs of trembling and shivering of lambs may identify lambs with hypothermia, as well as being part of the normal physiological and behavioural response to cold stress.

Resource-based indicators

Physical and thermal comfort can also be indicated by the quality and conditions of the environment. Good housing and management should prevent injuries and infectious diseases, provide physical and thermal comfort and allow animals to perform normal behaviours (EFSA, 2007). Resource-based indicators considered to be relevant and suitable to sheep farms include the location and type of housing, flooring type, bedding material, space allowance, pen design, fence, disease pens, stocking density, handling facilities, facilities for water and nutrition, ventilation systems, cooling systems, indoor temperature, relative humidity, air velocity, dust levels, light intensity, noise levels and concentrations of gases such as ammonia (EFSA 2007, Caropresе *et al.*, 2009).

1.4.3 Freedom from pain, injury and disease – *by prevention or rapid diagnosis and treatment*

Many measures of pain, injury and disease suggested in the scientific literature were based on the assessment of specific clinical signs of disease. The literature review previously identified that diseases such as footrot, mastitis, gastrointestinal parasitism and sheep scab were of particular concern for sheep welfare and so methods of assessing these conditions are further examined below.

Skin lesions

It has been suggested that a number of tests should be used to assess the presence of skin lesions in sheep. This is because the severity of certain skin lesions, such as sheep scab, may not be apparent on the basis of a clinical examination. Corke and Broom (1999) suggested that a clinical examination, assessment of behaviour and the response to a nociceptive stimulus, physiological assessment and immune function tests could be used to assess the presence of *Psoroptes ovis* infestation. Specific skin tests include examination of the size of the affected area, the sensitivity of the skin, and the degree of fleece loss (Corke and Broom, 1999). Sheep affected with skin lesions, such as sheep scab, show a significant increase in pruritus - mouthing and rubbing behaviours (Berriatua *et al.*, 2001; Corke and Broom, 1999). Pruritus can be examined by performing a 'nibble test' on individual sheep. A positive nibble reflex is indicated if a sheep shows nibbling and licking movement of the lips and tongue, extension of the head and neck or rhythmical movements of the head and body (D'Angelo *et al.*, 2007). These measures require the gathering and handling of individual sheep and methods that are capable of assessing the presence of skin lesions in groups of sheep, without the need for gathering or individual handling could offer clear benefits for the assessment of large, widely dispersed and extensively-managed flocks.

In a recent study, skin lesions - defined as swelling, wounds, scabs, and integument alterations, including skin damaged attributed to ectoparasitism, areas of wool loss and hyperkeratosis were assessed by observing groups of sheep from a distance of approximately 2 metres (Napolitano *et al.*, 2009). The method was reported to be feasible, but poor levels of inter-observer agreement were found. As the fleece can mask certain skin lesions, assessors were not able to reliably assess small lesions < 2 cm (Caroprese *et al.*, 2009; Napolitano *et al.*, 2009). This is important because lesion size may not be related to the severity of the welfare condition and a small lesion may alert to considerable welfare risk to the flock. Observing the behaviour of groups of sheep may offer a more reliable means of assessment. As animals affected with some skin lesions, including sheep scab, show a significant increase in pruritus - mouthing and rubbing behaviours (Berriatua *et al.*, 2001; Corke and Broom, 1999), a group of sheep could be observed for the presence of these behavioural signs.

Injuries and wounds

Traumatic injuries and wounds are considered to be very significant indicators of suffering and poor welfare status (Smidt, 1983). In particular, the frequency of traumatic injuries, location and severity of the lesions are important attributes to assess (Leeb *et al.*, 2001). Accordingly the presence of external injuries and skin lesions have been used as welfare indicators for dairy cattle, sows and broiler chickens in several on-farm welfare assessment schemes (Capdeville and Veissier, 2001; Leeb *et al.*, 2001). The presence of debilitating injuries has also received attention as a potential indicator of sheep welfare (Pines *et al.*, 2007). As wounds and injuries may be masked by the fleece, a close and careful assessment of the entire body of individual sheep could be used to locate small areas of myiasis (Farkas *et al.*, 1997), and skin and joint lesions (Gougoulis *et al.*, 2010).

Lameness

Generally, sheep lameness assessment is based on a combination of clinical observations and physical examination. As there are recognised differences in the interpretation of signs of sheep lameness (Kaler and Green, 2008b), a clear and agreed definition needs to be used before lameness is applied as an on-farm indicator of sheep welfare. A suitable definition might include the observation of any one of the following signs - alteration in the stride length, lifting or holding a limb, altered movement of joints, alteration in the time spent weight bearing or reluctance to move (Welsh *et al.*, 1993, Kaler *et al.*, 2009). Lameness scoring systems have also been developed for sheep including numerical rating scale (NRS), in which the degree of lameness observed is scored along a 5- or 6-point scale. Recently, Kaler *et al.*, (2009) tested a 6 point categorical lameness scoring scale by assessing video footage of individual sheep. The method produced high levels of inter- and intra-observer reliability, which may be due to the controlled conditions permitted by repeat assessment of video clips. Similar levels of reliability have not been found during field studies examining the reliability of NRS testing methods and a 5-point lameness scoring method was associated with poor levels of inter- and intra-observer reliability when tested under on-farm conditions (Welsh *et al.*, 1993, Harkins, 2005).

An alternative approach is to use a continuous method of lameness scoring using a visual analogue scale (VAS). In this method the assessor can score the level of lameness along a 100 millimetre horizontal line whereby one end of the line is labelled 'sound' and the opposite end of the line is labelled 'could not be more lame'. Research suggests that the VAS method provides higher specificity compared to NRS scales and may be a better means of diagnosing sheep as 'sound' (Welsh *et al.*, 1993).

The reliability of observer assessments of sheep lameness scoring methods may be influenced by the occupation, training and experience of the assessor. Harkins (2005) noted that poor inter-observer agreement was with associated occupational differences in the assessment of lameness scores between veterinary surgeons and stockpeople, suggesting that the reliability of any sheep lameness scoring system needs to be examined using a variety of different assessors. The feasibility of individual lameness assessment of sheep also requires consideration. Harkins (2005) suggested that assessment of a group of 25 sheep took 1 hour 10 minutes, so an NRS scoring system may not be suitable for assessing large flocks of sheep.

Recently, Conington *et al.*, (2008) reported that a binary scale for scoring lameness in sheep (present or absent) was more reliable than a continuous VAS method. A simpler assessment method with fewer scoring categories may provide higher levels of observer reliability and may be less time-consuming for on-farm assessments. Previous assessment methods have required handling and assessment of individual sheep. Given the difficulties for gathering and handling certain sheep managed under extensive conditions (Dwyer, 2009), the development of a method which permits the assessment of the gait of groups of sheep 'in the field' could offer a valuable tool to flock welfare inspections.

Foot lesions

A foot examination of individual sheep can be a useful means of identifying whether a flock is affected with foot lesions of welfare importance, such as footrot (Egerton and Roberts, 1971). Several footrot scoring systems, which score lesion severity using a NRS scale, have been described (Ley *et al.*, 1995, Welsh *et al.*, 1995). Whilst the assessment of inter-digital dermatitis and footrot have produced high levels of inter-observer reliability (Conington *et al.*, 2008), poorer levels of observer agreement have been reported when less severe foot lesions are examined (Harkins 2005). However, it may not be necessary to use a detailed categorical scoring system as a a binary foot lesion scale (present or absent) may provides sufficient information at the level of the individual animal (Nieuwhof *et al.*, 2008).

Mastitis

A method for assessing mastitis in sheep described by Harkins (2005) involved the examination and palpation of the udder while ewes were restrained in a standing position. Specific clinical signs that were recorded included the temperature, consistency, colour and appearance of the gland and presence of obvious abscesses and teat lesions. Fthenakis (2000) describes a similar method of assessment for cases of acute mastitis based on the quality of the milk, pain response, oedema and size of the udder, demeanour and appetite of the ewe. Additional measures such as Somatic Cell Counts (SCC's) are routinely used to assess the presence of mastitis in dairy ewes. However, the use of SCC has been removed from prior sheep welfare assessment protocols due to issues regarding the feasibility of acquiring samples from a large number of sheep that are not managed as milking flocks (Harkins, 2005). Pathological measures of inflammation including serum amyloid protein and acute phase protein may be useful indicators of mastitis (Winter and Fitzpatrick, 2008). Although, given the laboratory equipment and financial resources required, these measures may not be feasible to perform within the limits of a single on-farm assessment visit. In contrast, the clinical appearance and manual palpation of the mammary gland is a well-accepted practice, often performed by shepherds during the post-lambing and pre-tupping period, and is a relatively quick method to apply (Harkins, 2005).

Endoparasitism

Parasitic gastroenteritis (PGE) has been suggested to be an indicator of sheep welfare (Winter and Fitzpatrick, 2008). The condition can be associated with significant body weight loss, diarrhoea and death, or sheep may acquire natural immunity and so few or no clinical signs may be observed (Taylor *et al.*, 2007).

The level of dirtiness or 'daginess' of the breech and hindquarters is routinely used by sheep farmers in order to assess the health and welfare of their flock and to determine whether management interventions such as shearing the breech area ('crutching') or administration of anthelmintic treatments are required. Maintaining the cleanliness of

the breech is important because adherent faecal matter increases the risk of myiasis (blowfly strike) and is associated with mastitis in ewes (French *et al.*, 1996).
However, the degree of faecal soiling over the breech area or 'dag score' (French and Morgan, 1996) can also be a reflection of faecal consistency, so the cleanliness of the rear may also be affected by nutritional events such as seasonal fluctuations in the quality and composition of grazing pastures (Pollott *et al.*, 2004). As such, the dirtiness of the breech should not be used as an indicator of PGE but it may be a useful means of identifying diseases and events which present a risk to sheep welfare, including nutritional changes, myiasis and mastitis.
Clinical signs of diarrhoea and faecal staining are not features of all nematode parasites and signs of endoparasitism may be indicated by other animal-based outcomes. As an example, *Haemonchus cortortus* is associated with clinical signs of oedema, anaemia and ascites (Taylor *et al.*, 2007). In this instance, a validated scoring tool known as the FAMACHA© system, could be used to determine the degree of clinical anaemia and the risk of Haemonchosis in specific flocks (Kaplan *et al.*, 2004).
In practice, Faecal Egg Counts (FEC's) can be monitored on-farm by stockpeople and veterinary surgeons using portable laboratory equipment, such as the FECPAK system (McCoy *et al.*, 2005). However, a one-off FEC may not provide sufficient information to act as a measure as to whether the stockperson has taken appropriate action to prevent and control endoparasitic disease. The interpretation of FEC's can also be complicated by the fact that they do not indicate the total worm burden of the sheep and can be influenced by a number of factors including age, genotype and breed (Taylor *et al.* 2007). In addition, the level of flock information, the time and financial resources needed to inform interpretation was considered to preclude the use of FEC as an on-farm indicator of sheep welfare in this thesis. An alternative, management-based measure of endoparasitism could be based on the evaluation of farm medicine records and flock health plans (Gray, 2002) to assess whether appropriate action and control of PGE is being implemented on the farm.

Neonatal lambs

A lamb vigour scoring system has been described in which the ease of birth, activity level of lambs immediately post-partum (vigour) and suckling ability are assessed over the first few days of life (Macfarlane *et al.*, 2010). Lamb vigour scores could be used to select breeding stock with better lamb survival rates, although it is recognised that the expression of lamb suckling and maternal bonding can be affected by environmental, genetic and management influences, such as ewe health, nutrition and parity, and health and welfare of the lamb (Dwyer, 2008b; Macfarlane *et al.*, 2010). Whilst the method of Macfarlane and co-workers (2010) appears to have good face and content validity, the feasibility of assessment needs to be examined since it may not be possible to assess measures of sucking and birthing ease within the course of a single day assessment. Therefore, this system may be a more suitable tool for assessing the welfare of lambs in the immediate post-partum period.

Pain assessment

The experience of pain is widely viewed as a welfare concern for sheep and may be caused by routine farm practices such as mutilations, or specific diseases and injuries. Specific behaviours associated with pain that have been observed following castration and tail-docking of lambs include pacing, restlessness, writhing, rolling, licking castration site, stamping and lip curling (Molony, 1997; Thornton and Waterman-Pearson, 1999). Following an initial struggling behaviour period, the lamb becomes quieter and much stiller (Molony and Kent 1997). This may reduce the amount of inflammation and pain caused, and therefore, immobilisation may be an indicator of pain in sheep and lambs (Cockram, 2004; Wemelsfelder and Farish, 2004).

Other suggested signs of pain and discomfort in sheep include a reluctance to walk, altered posture, bruxism, neurological signs (Gougoulis *et al.*, 2010) and alterations in feeding (Villalba *et al.*, 2010), lying and resting behaviours (Berriatua *et al.*, 2001; Corke and Broom, 1999). In addition, the isolation of individual sheep from the rest of the flock may alert to the presence of pain, injury or disease (Lynch *et al.*, 1992).

Behavioural indicators of pain have been based on human judgements of the level of pain experienced by the animal (Molony and Kent, 1997; Thornton and Waterman-Pearson, 1999). Evidently, the assessment of the level of pain is likely to be based on personal beliefs, life experience and education of the assessor (Rutherford, 2002). Since humans and sheep share the same nociceptive signal pathways, it could be suggested that pain experienced by the nervous systems of these two species may be similar (Fitzpatrick *et al.*, 2006). However, the failure of humans to recognise sheep behavioural expressions of pain does not mean that sheep do not experience pain (Stubsjøen *et al.*, 2009) since sheep may mask the behavioural expression of pain in order to avoid detection by predators (Fitzpatrick *et al.*, 2006; Wemelsfelder and Farish, 2004). Alternative methods of assessing pain include measurements of hyperalgesia (Fitzpatrick *et al.*, 2006; Ley *et al.*, 1995; Welsh *et al.*, 1993), ear postural changes (Boissy *et al.*, 2011; Reefmann *et al.*, 2009), eye temperature and heart rate variability (Stubsjøen *et al.*, 2010; Stubsjøen *et al.*, 2009). Such tests are clearly valuable to pain research studies but may not be feasible for on-farm welfare inspections given the time, resources and potentially large numbers of sheep that need to be assessed.

Management-based indicators

Examination of on-farm records can offer a means of assessing the appropriateness of on-farm management practices and stockperson decisions. For example, as lamb survival is the outcome of ewe nutrition, lambing practices, housing environment and stress, records of lamb mortality could be viewed as an "integrated welfare measure" (Dwyer, 2008b). Additional management records that may useful for on-farm welfare assessments include disease incidence, culling rates, medicine use and abattoir reports. Evidently, a pre-requisite of the use of any type of record as a potential animal welfare indicator is that the information is accurate, up to date and reliable. Currently, welfare legislation in the UK requires that on-farm mortality records are maintained and so these records should be easy to access and feasible to obtain. However, the retrieval of farm health records and subsequent analysis can be a time and labour consuming task (Sørensen *et al.*, 2007). Furthermore, it has not been established whether farms that maintain good records are associated with good standards of sheep welfare and vice versa. Therefore it would be useful to examine the reliability of on-farm records and their correlation with animal-based measures of welfare.

For other welfare issues, it may be preferable to assess the quality of management practices, such as castration and tail-docking, by finding out which technique is used, the age of the lambs, and ascertain whether local anaesthesia and analgesia are routinely administered (Pritchard, 2008).

1.4.4 Freedom to express normal behaviour – *by providing sufficient space, proper facilities and company of the animal's own kind*

Behaviour is the direct result of intrinsic decisions of the animal and may be "the ultimate phenotype" of animal emotions (Dawkins 2004). As deviations from normal behaviour can reflect an acute and early response to a particular issue within the environment or with the animal itself (Broom 2003), behaviour can be used as an indicator of sheep welfare (Wemelsfelder and Farish 2004). Indeed, veterinary surgeons and stockpeople often rely on alterations in normal sheep behaviour to identify particular health and welfare conditions, for example, alterations in gait are used to identify lameness (Kaler *et al.,* 2009), and signs of pruritus can alert to a sheep scab infestation (Berriatua *et al.*, 2001, Corke and Broom 1999). It is clear from the literature that the interpretation and assessment of sheep behaviour require an understanding of the normal and natural behaviour of this species together with the effects of domestication and on-farm management on the behavioural expression of individual and groups of sheep.

Normal behaviour

The freedom to display normal behaviour raises the question as to whether animal welfare is compromised if a captive animal fails to show the full behavioural repertoire of a free-living member of its species (Dawkins, 1990). If sheep are highly motivated to perform a particular natural behaviour but are restricted, or lack the capacity for behavioural expression, the chronic stress produced is associated with a compromised state of welfare (Dwyer and Bornett, 2004). So, the expression of 'natural' behaviours – those performing in free living members of the species, may equate to good standards of animal welfare (Bracke and Hopster, 2006)

However, selection of animals for production and genetic traits, and adaptation of animals to the managed environment have also influenced the behavioural repertoire of production animals (D'Eath *et al.,* 2010). Since many housing systems are not capable of meeting the full behavioural requirements of farm animals the 'normal' behavioural expression of the domesticated housed sheep may be distinct from the 'normal' behaviour of wild, undomesticated sheep (Smidt, 1983).

Stereotypies are patterns of repetitive behaviour that are considered to be unnatural and have received a lot of attention as indicators of poor animal welfare (Broom, 1986; Mason and Latham, 2004; Smidt, 1983). However, there are conflicting views on the relevance and welfare implications of these behaviours. Some authors suggest that stereotypies may offer a means of coping with the captive environment (Broom and Johnson, 1993), others believe that they are expressed as a matter of habit, developed as a consequence of alterations in brain chemistry and expressed even following welfare improvements (Mason and Latham, 2004). So, the explanation of 'unnatural' behaviours can be extremely complicated. The interpretation of sheep behaviour is further complicated as it may be influenced by the age, physiological state (Dwyer, 2007), genotype (Dwyer and Lawrence, 2000), and gender of the animal.

Maternal bonding

The behavioural expression of the post-parturient ewe and development of a strong maternal bond is critical to lamb survival and welfare (Nowak, 1996). The ewe-lamb bond has been assessed using maternal behaviour scores (MBS) which grades the behavioural response of the ewe following handling and examination of her lamb(s) by the shepherd (O'Connor *et al.*, 1985). Studies have identified that older ewes and more fecund ewes provide a higher MBS and a positive correlation with lamb body weight and survival has been found (O'Connor *et al.*, 1985). However, a number of factors including genotype and breed affect the expression of maternal bonding, so it has been suggested that a simple scoring system may not capture the complexity of this behavioural expression (Dwyer, 2007).

Positive emotions

A qualitative approach to welfare assessment implies more than the absence of pain, distress or suffering and it questions whether an animal has a 'life worth living' (Wemelsfelder *et al.*, 2000). This approach recognises that sheep are sentient beings with emotional experiences – concepts that have been recognised within European law (EU Directive 86/609) and scientific research (Boissy *et al.*, *2011*; Greiveldinger *et al.*, 2007; Veissier *et al.*, 2009).

Positive emotions are also associated with coping, reward and goal-directed behaviours, and these types of behaviour could be used as indicators of positive welfare (Boissy *et al.*, 2007). Behaviours that may indicate positive emotional states in sheep include the expression of play behaviour, curiosity and calmness (Wemelsfelder and Farish, 2004). Curiosity and play are interpreted as positive indicators of enjoyment and relaxation and could be used as measures of good welfare. These behaviours are mainly identified in lambs by their frolicking, galloping and bucking with other lambs. Although less frequently observed, ewes and rams may express play behaviours at spring turnout (Wemelsfelder and Farish, 2004).

The expression of positive emotions is measured in the method known as Qualitative Behavioural Assessment (QBA). This is a whole-animal methodology that evaluates the quality of an animal's life by asking the observer to assess the expressive qualities of an animal's demeanour and body language using descriptors such as 'relaxed', 'anxious' or 'content' (Wemelsfelder, 2007). QBA has demonstrated face, construct and predictive validity through correlation with physiological measures (Wemelsfelder, 2007), quantitative ethograms (Rousing and Wemelsfelder, 2006) and measures of animal health (Brscic *et al.*, 2009). The method has achieved good levels of inter-observer reliability for welfare assessments of pigs (Wemelsfelder *et al.*, 2000 and 2001), and cattle (Rousing and Wemelsfelder, 2006) but to date, the reliability of sheep assessments has not been reported. Therefore the method deserves research as a potential tool in the on-farm assessment of sheep welfare.

1.4.5 Freedom from fear and distress – *by ensuring conditions and treatment which avoid mental suffering*

Most indicators of fear and distress identified by this literature review were concerned with cortisol and heart rate - measures that can be used to reflect the responses to an acute or chronic environmental stress (Smidt, 1983). In addition, behavioural measures that focus on quantitative measure of fear and distress and more qualitative measures, such as QBA, can be used to assess the overall wellbeing of an animal.

Cortisol

The stress experienced by an animal in order to cope with longstanding welfare compromise can result in a significant and severe depression in immunity (Dwyer and Bornett, 2004). Studies into the hypothalamic-pituitary-adrenal (HPA) axis have indicated that 'stress' leads to increases in the synthesis and secretion of corticosteroids, specifically cortisol (Dobson, 1990). Although other factors including adrenaline, noradrenalin, corticotropin releasing factor (CRF), adrenocorticotropic hormone (ACTH), glucocorticoids, progesterone and prolactin are released (Mellor and Stafford, 1999), cortisol has received most attention as an indicator of stress in sheep.

The release of cortisol is also related to states of arousal, so increased levels may occur if an animal attempts to escape from a dangerous situation or cope with a physiologically stressful situation (Cockram, 2005). Therefore, an increased cortisol value does not definitively indicate compromised states of welfare. In addition chronic stimulation of the

HPA can result in lower levels of circulating cortisol but may not necessarily correlate with good standards of animal welfare (Smith and Dobson, 2002).

The interpretation of cortisol values requires multiple samples so a one-off physiological measurement as part of an on-farm assessment visit may be of little value or significance to the actual welfare state of the sheep. The interpretation of physiological indicators is further complicated by the fact that the handling of animals, regularity of feeding (Dobson, 1990), timing of sampling, and the use of venepuncture (Caroprese *et al.*, 2006) may all affect cortisol release. Non-invasive methods of assessment, such as salivary and milk samples, may be more viable methods of assessing cortisol levels (Caroprese *et al.*, 2006). Although, there may still be issues regarding the feasibility of cortisol measures, given the labour, time and financial resources required for long-term monitoring (Smidt, 1983).

Heart rate

As elevations in heart rate (HR) have been associated with social isolation (Dwyer and Bornett, 2004), handling with dogs (Baldock and Sibly, 1990), and pain, the measure has been validated as an indicator of stress and distress in sheep. Whilst HR can be readily examined by palpation or auscultation of individual sheep, the handling and restraint required will affect resting HR (von Borell *et al.*, 2007). There are non-invasive means of assessing HR using subcutaneous electrodes that are attached to areas of shorn fleece in animals accustomed to the equipment (von Borell *et al.*, 2007). However, given the issues surrounding the need to assess large numbers of animals under working farm conditions, the length of time involved and the considerable variation of HR between individual animals (Rutherford, 2002), the measure was not considered to be a valid, reliable or feasible indicator of sheep welfare for this thesis.

Behaviour

Fearful behaviours are thought to inform other sheep of a potential or actual threat and so they can form part of useful defensive strategies. They can also be expressed during negative emotional states (Hemsworth, 2007; Wemelsfelder and Farish, 2004). The degree of fear expressed by sheep has been most frequently researched using the Novel Area Test (NAT) (Forkman *et al.*, 2007) and by measuring the distance sheep move from a fearful stimulus or situation – the ‘flight distance’ (Dwyer, 2004). Other fear tests available include the tonic immobility test, restraint test (RI), predator test, startle test and conditioned fear test (Forkman *et al.*, 2007). There are important issues regarding the validity (de Passillé and Rushen, 2005; Dwyer and Bornett, 2004; Hargreaves and Hutson, 1990) and reliability of fear testing methods (Boivin *et al.*, 2003; Forkman *et al.*, 2007). Factors such as assessor identity, location of testing and motivations other than fear, affect the results of fear tests (de Passillé and Rushden 2005). Flight distances can also be affected by the genotype of the animal (Dwyer 2004) and familiarity and frequency of human contact (Boivin *et al.*, 2003; de Passillé and Rushen, 2005). Therefore these measures may not appear suitable for assessing large groups of sheep on working farms under extensive management conditions or those performed by an assessor who is unfamiliar to the flock.

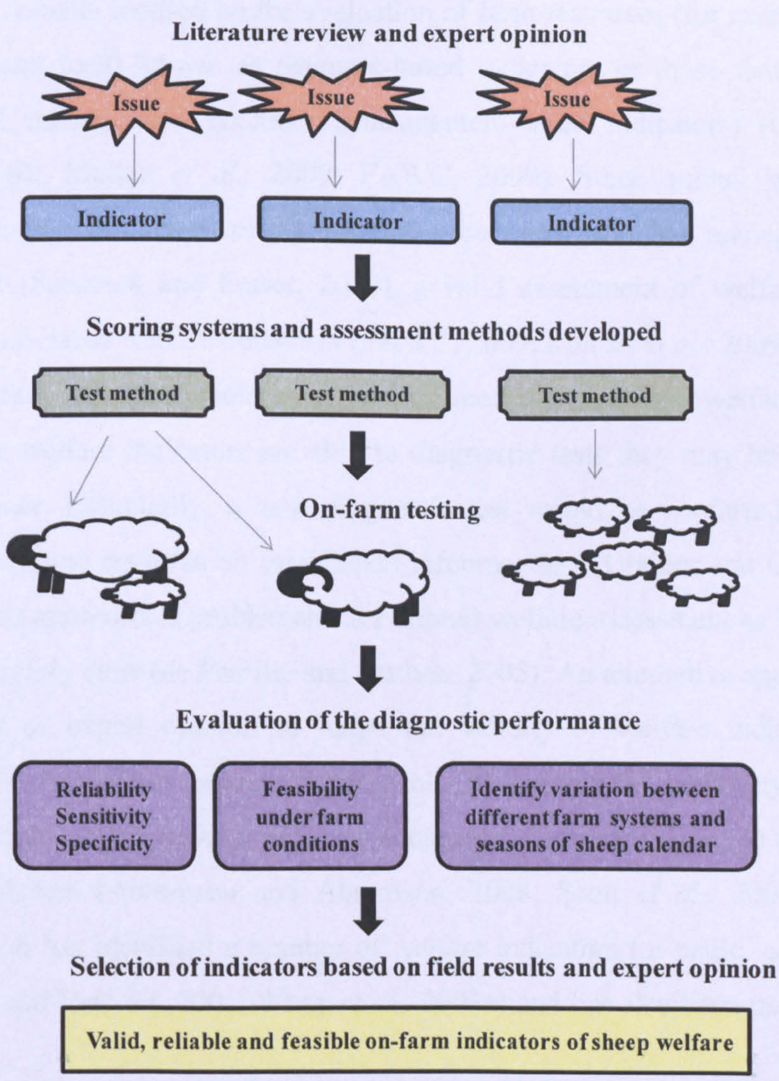
Other behavioural indicators of fear and distress in sheep welfare include alterations in locomotion, lying, rumination and appetitive behaviours (Berriatua *et al.*, 2001; Corke and Broom, 1999), vocalisation when isolated from the flock, foot stamping, and rearing (Cockram, 2004). Unlike other species, sheep do not frequently vocalise when injured or distressed and only demonstrate vocalisation at specific events – for example, the bleating that occurs between ewes and their offspring (Dwyer 2004) or the vocal communication produced by rams when approached by a ewe in oestrus (Boissy *et al.*, 2007, Wemelsfelder and Farish, 2004) - events that are associated with positive emotions. Therefore, vocalisation that occurs outside of these periods may indicate fear and distress in groups or individual sheep.

1.5 Outline of thesis

The objective of this thesis was to develop valid, reliable and feasible indicators for the on-farm assessment of sheep welfare. Indicators of sheep welfare were developed following the principles of animal welfare research and epidemiology, including the use of a scientific literature review, expert consultation and diagnostic test evaluation. The literature review presented in this chapter identified a number of welfare issues for sheep and several potential indicators of sheep welfare with face and content validity. For many of the identified indicators there is limited knowledge as to their practical relevance, feasibility or sensitivity as indicators of welfare for individual sheep and flocks. As there is no gold standard for assessment of sheep welfare, the validity of these measures needs to be examined to ensure that they are reliable and robust before being applied in any on-farm welfare assessments scheme (Figure 1.1).

The findings of the literature review presented in this chapter were used to inform an expert panel who identified a number of animal-, resource-, and management-based indicators (Chapter 2). An on-farm protocol was subsequently developed (Chapter 3) and the diagnostic test performance of each animal-based indicator was tested by observers of varying experience and training (Chapters 4 to 6). The feasibility and ability of animal-based measures to detect between-farm (Chapter 7) and seasonal variation (Chapter 8) was also examined during cross-sectional and longitudinal studies. Preliminary studies have also examined the reliability and feasibility of Qualitative Behavioural Assessment (QBA) (Chapter 9). Expert guidance on the use of the indicators together with preliminary cut-off points for acceptable and unacceptable levels of on-farm indicator assessments was provided (Chapter 10). Finally, a discussion of the findings and potential implications of the work presented in this thesis are discussed (Chapter 11).

Figure 1.1 Methodology for the development of indicators of sheep welfare



Chapter 2

VALIDATION OF WELFARE INDICATORS THROUGH A CONSENSUS OF EXPERT OPINION

2.1 Introduction

Consumer awareness and concern for farm animal welfare has provided the impetus for driving the provision of products from high welfare systems, assurance of farm animal welfare standards and demands for welfare labelling of food (Main *et al.*, 2001; Waiblinger *et al.*, 2001). In addition to non-regulatory purposes, welfare assessments are undertaken to assess compliance with legislative and regulatory standards (EFSA, 2007). Consequently, there is a need for valid, reliable and feasible measures of sheep welfare for use in on-farm welfare assessment systems which would be transparent and fair to both producers and consumers (Mullan *et al.*, 2009; FAWC, 2009). In the past, on-farm welfare assessments focused on the evaluation of farm resources (for example, provision of housing and feed) known as resource-based indicators or those that assess farmer policies and management practices (management-based indicators) (Capdeville and Veissier, 2001; Mullan *et al.*, 2009; FAWC, 2009). Since animal welfare can be considered to be the outcome of the interaction between genotype, management and the environment (Sørensen and Fraser, 2010), a valid assessment of welfare should also include animal-based welfare outcomes (FAWC, 2009; Mullan *et al.*, 2009).

A valid welfare indicator would be a genuine measure of animal welfare (Scott *et al.*, 2001). Since welfare indicators are akin to diagnostic tests they may be evaluated in a similar manner. Essentially, a new diagnostic test would be conferred as valid if it produced the same result as an established reference test (Greiner and Gardner, 2000). However, this approach is problematic for animal welfare assessment as a reference test does not currently exist (de Passillé and Rushen, 2005). An alternative approach is to use a consensus of expert opinion to judge the validity of welfare indicators that are incorporated into on-farm assessment protocols. By judging the suitability and relevance of each measure, expert opinion provides consensual and face validity to the selection of welfare indicators (Abramson and Abramson, 2008; Scott *et al.*, 2001). Previously, expert opinion has identified a number of welfare indicators for cattle, poultry and pigs (Capdeville and Veissier, 2001, Whay *et al.*, 2003a) and has also been used to identify a number of welfare issues for sheep (Cronin *et al.*, 2002, Waterhouse *et al.*, 2003, Pines *et al.*, 2007, Phillips *et al.*, 2009).

The methods for gathering the consensus of expert opinion vary, from using Delphi questionnaires (Whay *et al.*, 2003a, Phillips and Phillips, 2010), on-line surveys (Fernie *et al.*, 2008), small group meetings such as the Nominal Group Technique (NGT) (Delbecq *et al.*, 1975) to conference-style meetings (Glaser, 1980; NIH, 1990; Pines *et al.*, 2007). The NIH consensus development programme (NIH, 1990) is a well-recognised conference-style method that has been used in human medicine to identify numerous clinical and biotechnological issues (www.consensus.nih.gov). This method is relevant for developmental research studies as experts meet face-to-face enabling idea generation and open discussions (NIH, 1990). The NIH approach can involve a pre-meeting consultation, such as the use of questionnaires to give experts greater time to enlarge on the topic of interest (Glaser, 1980). Additionally, experts may be split into sub-groups during the conference to allow a more in-depth discussion of the relevant issues (Glaser, 1980).

The overall aim of this thesis was to establish valid, reliable and feasible indicators for the on-farm assessment of sheep welfare. In light of previous studies, the initial identification of sheep welfare indicators, which are sensitive to the current on-farm welfare issues, could be selected using the welfare concepts of the Farm Animal Welfare Council Five Freedoms (FAWC, 1994). Thereafter, an NIH approach (NIH, 1990) could allow a panel of experts to judge the validity (face and consensual) of sheep welfare indicators, which are included in on-farm assessment studies. Following this initial validation step, the internal validity (evaluation of systematic bias, reliability and diagnostic sensitivity and specificity) and feasibility of these indicators would be evaluated in field studies (Greiner and Gardner, 2000). In this manner, the ranking or exclusion of any indicators could be based on the evidence of their validity, reliability, and feasibility for use in on-farm welfare applications.

The objective of this chapter was to use a consensus of expert opinion to identify the current on-farm welfare issues for sheep in England and Wales. Secondly, the Five Freedoms framework was used to initially identify valid indicators of sheep welfare, which would be evaluated during on-farm studies.

2.2 Materials and methods

2.2.1 Expert panel selection

The objective of the expert panel selection process was to identify a broad panel of expert members (Garabed *et al.*, 2009). An 'expert' was defined as a person having a minimum of ten years experience of sheep farming in the UK sheep industry and/or professional achievements in industries and organisations allied to sheep farming, veterinary services and welfare research. Experts were short-listed by members of the research team to avoid overlap in areas of representation (Garabed *et al.*, 2009). Following approval by the Department for the Environment, Farming and Rural Affairs (Defra), 33 experts were formally invited to attend a one day expert panel meeting. Prior to their attendance, experts were given clear, written guidance on the objectives of the meeting and the required output of the expert panel – to produce a list of indicators of sheep welfare, which encompassed all aspects of the Five Freedoms and were sensitive to current on-farm welfare issues for sheep in England and Wales. In addition, experts were informed of their role at the meeting i.e. to validate the identification of potential indicators of sheep welfare. Where invited members did not respond to follow-up contact (n = 2), or were unable to attend (n = 3), a substitute member with similar expertise was selected and invited to attend. As a result, 33 members confirmed their attendance.

2.2.2 Identifying on-farm welfare issues for sheep using a postal worksheet

One month before the meeting, a one-page worksheet was distributed to all experts (electronic and postal hard copy) in order to identify potential on-farm welfare issues for the individual sheep. Each expert was asked to list a maximum of 10 issues considered to affect the on-farm welfare of 1. ewes (female sheep over one year-old), 2. rams (male sheep over one year-old), 3. growing and fat lambs (any sex, over six weeks to one year-old) and 4. young lambs (from birth to six weeks-old) in England and Wales. Experts were asked to consider the factors that affected the welfare at the level of the individual sheep. Anonymous responses, from completed worksheets, were reviewed and the suggested welfare issues were categorised into the relevant area(s) of the Five Freedoms framework (FAWC, 1994). The Five Freedoms detailed by the Farm Animal Welfare Council (FAWC) are accepted as a sound framework for the development of animal welfare indicators (Whay *et al.*, 2003a, FAWC, 2009). This information was summarised into a document that was distributed in electronic and paper versions to all members one week prior to the expert panel meeting (Figure 2.1).

2.2.3 Identifying on-farm welfare issues for sheep using an expert panel meeting

The expert panel meeting was held on 23 February 2008, and consisted of 30 members from welfare organisations and charities (n = 2), sheep welfare research (n = 5), welfare inspectors (n = 1), veterinary animal welfare advisors (n = 2), government advisors on animal welfare policies (n = 2), veterinary sheep specialists (n = 3), veterinary surgeons from general practice (n = 4), sheep farmers (n = 5), sheep industry services (n = 2), and sheep production consultants (n = 4). The meeting followed NIH guidelines (NIH, 1990), in which an introductory presentation on the objectives of the research project was given and members were informed of their specific role in the validation of welfare indicators.

The expert panel was comprised of members with different experiences and expertise in sheep health, welfare and production. Therefore, in line with NIH meetings (NIH, 1990), the author gave the expert panel a presentation on the current, scientific knowledge of on-farm welfare issues for sheep, identified through a literature review (Chapter 1). For the purposes of this study, welfare was defined as a factor, event or action that affected one or more of the Five Freedoms of an individual sheep (FAWC, 1994).

The panel was then divided into four focus groups (Figure 2.1). Each group focused on identifying the welfare issues for one of four production stages: 1. ewes, 2. rams, 3. growing and fat lambs, and 4. young lambs. The aim was to allow a balance of opinions within each group and a mix of members from the different representative areas was allocated into each group. A focus group comprised of seven or eight experts, plus a facilitator, who was a trained member of the project team (Glaser, 1980; Krueger and Casey, 2009). In addition, two non-participating recorders maintained accurate written minutes of group discussions. The welfare issues document, which categorised all the welfare issues identified from the pre-meeting worksheet into the Five Freedoms framework (FAWC, 1994), was distributed to the panel one-week ahead of the expert panel meeting. This provided the experts with more time to consider the welfare issues. At the meeting this document was used as the basis for focus group discussions. All panel members were given the opportunity to consider, discuss, modify and clarify the welfare issues listed in this document and were free to alter the categorisation of any welfare issue within the Five Freedoms framework. As well as guiding the identification of welfare indicators by the experts, this document was also one of the final outputs of the expert panel meeting - akin to the consensus of expert opinion statement produced in NIH conferences (NIH, 1990).

2.2.4 Identifying potential welfare indicators using an expert panel meeting

The next stage in the expert panel process was to identify valid indicators of sheep welfare based on the suggested welfare issues for sheep and the five freedoms framework (FAWC, 1994). Following the NIH process (NIH, 1990), the author gave a presentation to the panel on the current scientific knowledge of animal welfare indicators, including those developed for other species. Experts then returned to their focus groups (Figure 2.1) and were asked to consider and identify measures capable of assessing each of the issues listed within the Five Freedoms framework. Experts were asked to identify indicators that were animal-based (where possible), non-invasive and practicable for assessments performed under farm conditions. Once the focus groups had completed this task, the panel reconvened and each facilitator gave a presentation on the valid indicators identified by their group. All members of the panel were then asked to comment on the output of the other focus groups.

Following the guidance of NIH consensus methods, an official output for the consensus of expert opinion was produced after the meeting (NIH, 1990). This consisted of a final document listing the on-farm welfare issues and indicators for sheep as suggested by the expert panel. This document was electronically distributed to all experts within one week of the meeting. Experts were then requested to read the document and confirm that the indicators and issues were those suggested at the meeting. If the majority of experts agreed, then a consensus of opinion would be established (NIH 1990).

Figure 2.1 Methodology for ascertaining a consensus of expert opinion

Selection of expert panel members
Defined expert criteria and informed invited experts of the role of panel and meeting objectives

Expert panel pre-meeting consultation
Each expert asked to identify on-farm welfare issues for sheep and lambs using a postal worksheet
Welfare issues categorised into the FAWC five freedoms
A list of the suggested issues was distributed to all expert members ahead of the meeting

Expert panel meeting

Scientific presentation to panel
Meeting objectives and format of focus groups defined to the whole expert panel
Current scientific knowledge regarding the on-farm welfare issues for sheep presented

Focus group
Identified on-farm welfare issues for ewes

Focus group
Identified on-farm welfare issues for rams

Focus group
Identified on-farm welfare issues for growing lambs

Focus group
Identified on-farm welfare issues for young lambs

Scientific presentation to panel
Current scientific literature regarding development of animal welfare indicators presented

Focus group
Identified indicators for ewes

Focus group
Identified indicators for rams

Focus group
Identified indicators for growing lambs

Focus group
Identified indicators for young lambs

Presentation of focus group outputs
Facilitators presented the list of indicators identified by each focus group to the whole panel
Open discussion between all members of the expert panel

Post-meeting distribution of consensus output
List of indicators identified by all focus groups distributed to all experts following the meeting
Experts agreed on the final consensus output providing validity (face and consensual) to the identified indicators

On-farm validation of indicators
Evaluation of the diagnostic and external validity of indicators

2.3 Results

2.3.1 On-farm welfare issues for sheep

A total of 19 out of 33 worksheets were received (response rate of 58%). Experts suggested 193 potential welfare issues for sheep in England and Wales (Appendix A). Of these, 53 on-farm welfare concerns were identified for ewes, 45 issues for rams, 42 issues for growing and fat lambs, and 53 issues were raised for young lambs. Table 2.1 lists the welfare issues identified in the postal worksheet by at least 50% of experts ($n \geq 17$). The pre-meeting consultation found that lameness was consistently identified as an on-farm welfare issue for all four of the sheep production stages. For young lambs, mutilation procedures of tail-docking and castration were considered to be key welfare issues (Table 2.1). Poor body condition was identified as an on-farm welfare issue for both ewes and rams, whereas gastro-intestinal parasitism of growing lambs and dystocia in ewes were issues specific to these production stages.

Table 2.1 Welfare issues identified through the pre-meeting worksheet

Young lambs	Growing lambs	Ewes	Rams
Castration (63%)	Gastrointestinal parasitism (74%)	Poor body condition, poor nutrition (84%)	Poor body condition, poor nutrition (68%)
Tail docking (58%)	Lameness (53%)	Lameness (84%)	Lameness (63%)
Lameness, septic arthritis (53%)	Myiasis, maggot control (53%)	Dystocia (58%)	§

§ no additional issues suggested by >50% of respondents

At the panel meeting, experts identified and categorised a total of 193 welfare issues for ewes, rams and lambs into the relevant criteria of the Five Freedoms framework (FAWC 1994). Since all the experts indicated their agreement with the post-meeting output documents, a consensus of expert opinion on the welfare issues for sheep was achieved. Given the large number of issues identified by the panel, a summary of the welfare issues identified for each freedom criteria was produced (Table 2.2). The 193 issues were summarised into a list of general welfare issues, which covered all four production stages (n = 58). Where appropriate, issues specific to the welfare of ewes, rams, young and growing lambs were highlighted. This organisation provided 14 general welfare issues within the freedom from hunger and thirst, 14 within the freedom from discomfort, 11 within the freedom from pain, injury and disease, 8 for freedom to express normal behaviour, and 10 general welfare issues were produced under the freedom from fear and distress criterion. For example, within the freedom from pain, injury or disease, the expert panel suggested that a large number of diseases and health conditions were on-farm welfare issues for sheep and lambs. Accordingly, specific diseases and health conditions were grouped, for example contagious ovine digital dermatitis, foot rot and scald were summarised as 'infectious foot lesions' (shown in brackets in Table 2.2). Subsequently, the general disease conditions were grouped into an overall, general welfare issue, which was termed 'health status' (Table 2.2).

2.3.2 On-farm welfare indicators for sheep

The animal-, resource- and management-based indicators suggested by the panel are shown in Table 2.3. Of the animal-based indicators suggested, 10 were identified for ewes, 13 for rams, 11 for young lambs and 9 for growing and fat lambs. Many of the welfare indicators identified for each production stage were similar. Since there was overlap across the focus group outputs, all the suggested measures were presented as a single list of animal- (n = 26), resource- (n = 13), and management-based (n = 22) indicators. Whilst most indicators could be used to assess all four production stages, the assessment of time taken to stand and suck, along with an overall indicator of lamb thrift was specifically suggested for young lambs. Resource- and management-based indicators suggested by the four focus groups also showed similarity, so these indicators were also grouped into a single list (Table 2.3). Resource-based measures included the assessment of housing and associated facilities and provisions, for example space allowance and quality of bedding. The panel identified that castration and tail-docking procedures were painful procedures (Tables 2.1 and 2.2). However, experts suggested that it may be more feasible to ascertain whether these mutilations were practised, the method used, and the on-farm policy regarding analgesia and anaesthesia use, rather than measuring the pain responses of lambs (Molony and Kent, 1997).

Table 2.2 On-farm welfare issues for sheep and lambs identified by a consensus of expert opinion

Freedom	Welfare issues
Freedom from hunger and thirst	Provision of appropriate and adequate feed and forage Provision of a continuous supply of clean water Appropriate body condition for production stage and purpose Colostrum management (immunoglobulin concentration, adequate volume)[YL] Management and nutrition of orphan lambs (artificial rearing methods) [YL] Neonatal lamb behaviour (time taken to stand and suck)[YL] Maternal behaviour (ewe-lamb bonding)[YL] Culling policy Management of dietary change (bought-in animals, weaning[YL, GL]) Selection of animals suited to management system (able to thrive on grass) Management flexible to changing animal requirements (parturient ewe nutrition[E]) Appropriate stocking density General health status Dental health (condition of molar and incisor teeth)
Freedom from discomfort	Provision of shelter and shade for grazed animals Provision of lie-back area for grazed animals (grazing root crops) Clean, dry environmental and pasture conditions Appropriate fleece cover for turnout post-shearing Use of appropriate breed or genotype (ability to withstand climatic conditions) Stocking density for housed and grazed animals Provision and quality of bedding in housing Flooring surface and hygiene Management of hypothermia Provision of appropriate winter housing Provision of exercise for housed animals Year-round care and inspection Fitting and maintenance of equipment (correct fitting of harness/raddle[R]) Appropriate rearing for intended management and purpose
Freedom from pain, injury or disease	Year round inspection (cast sheep) Appropriate intervention at lambing (management of dystocia)[E, YL] Transport of pregnant ewes and lame sheep Presence of horns (physical injuries and blindness) Appropriate body conformation (double muscling) Predation[YL] Reproductive management (mis-mating)[R, E, GL] Quality of equipment and buildings (design and maintenance of housing and handling facilities, correct fitting of raddle and harness[R]) Management practices associated with pain (use of electro-ejaculator[R], vasectomy[R], lambing intervention[E, YL], ear tagging and notching, castration and tail-docking[YL, GL]) Health status (presence of specific clinical signs and diseases including lameness, joint disease, infectious foot lesions, nutrient deficiencies and toxicities, metabolic disease, skin lesions, reproductive disease, inheritable

	defects, endo- and ecto-parasitism, dental health, mastitis, infectious disease, ocular conditions, respiratory disease) Disease prevention and control (vaccination, anthelmintic control, appropriate action and treatment of disease, stockperson inspection, veterinary input, culling policy, bio-security measures
Freedom to express normal behaviour	Appropriate intervention at lambing[E, YL] Reproductive management practices (use of artificial insemination) Space allowance: opportunity for exercise for indoor-housed animals Provision of environmental enrichment for housed animals Group size and composition (overcrowding, mixing unfamiliar sheep) Isolation of individual sheep (sick animals) Artificial rearing of lambs [YL] Maternal behaviour (ewe-lamb bonding, mis-mothering)[E, YL]
Freedom from fear and distress	Appropriate intervention at lambing[E, YL] Artificial insemination Intensive finishing systems[GL] Quality of handling: stockperson skills and quality of handling facilities Method and timing of weaning Use of lamb adopter (tethering of ewes[E]) Transport of pregnant ewes and lame sheep Group size and composition (stocking rates, ewe to ram ratios) Ability of humans to recognise and interpret sheep behaviour Use and control of dogs (dog worrying, use of aggressive farm dogs)

[R] welfare issue specific to rams,
[E] welfare issue specific to ewes
[GL] welfare issue specific to growing lambs,
[YL] welfare issue specific to young lambs

Experts also considered the feasibility of performing the assessments under working farm conditions. Several difficulties, including the location and gathering of extensive flocks, handling of pregnant ewes and gathering of ewes with lambs at foot, were identified. Experts therefore suggested that the nature of the sheep production year should be considered before assessment visits were undertaken. Following distribution of the post-meeting document, two panel members made additional comments which were duly noted and recorded. These comments did not affect the list of identified indicators (Table 2.3), but they helped to inform the on-farm evaluation studies. Experts confirmed the validity of the indicators listed in the post-meeting consensus document. Therefore, a consensus of expert opinion provided validity to the selection of welfare indicators that would later undergo validation in on-farm studies.

Table 2.3 Animal, resource and management-based indicators of sheep welfare identified by a consensus of expert opinion

Animal-based indicators	Resource-based indicators	Management-based indicators
Body condition	Water provision, access and quality	Mortality and culling records
Live weight	Quality and quantity of food	Abattoir feedback
General demeanour	Trough space	Medicine records
Alertness to approach in the field	Space allowance in housing	Disease records
Atypical behaviours	Flooring type	Movement records
Separation from the flock	Provision and quality of bedding	Faecal egg count records
Play behaviours	Presence of an isolation area	Growth rate records
Rumination	Good fencing and farm boundaries	Scanning records
Qualitative behavioural assessment	Sward height	Policy for hypothermia management
Panting	Stocking density at grazing	Ear tagging policy and quality
Cleanliness (rear and belly)	Provision of shelter and shade	Observation of specific management procedures (tagging, castration)
Time spent standing	Presence of handling facilities	Observation of daily farmer tasks
Gait/lameness	Presence of a lie back area	Assessment of handling skills
Skin condition		Castration and tail-docking policy
Fleece condition and wool cover		Stockperson awareness of on-farm disease
Urolithiasis		Judgement of appropriate on-farm action
Tail length		Current flock health plan
Eye condition		Mixing of horned and un-horned sheep
Nasal discharge		Mixing of different ages and sizes
Ear integrity (tagging injuries)		Presence of carcasses with live sheep
Biochemical measures		Farmer self-assessment
Rumen fill		Reproductive management and policy
Presence of in-growing horns		
Body injuries and wounds		
Time to stand and suck [YL]		
Thriving lamb assessment [YL]		

[YL] Indicators specific to young lambs

2.4 Discussion

In the absence of a reference test (or gold standard) for welfare assessment (de Passillé and Rushden, 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; Whay *et al.*, 2003a). This is because expert opinion is considered to provide both face and consensual validity to the welfare indicators (Abramson and Abramson, 2008; Scott *et al.*, 2001). The study presented here is the first to determine valid indicators for the on-farm welfare assessment of sheep using a NIH consensus approach (NIH, 1990).

The NIH methodology provided validation of welfare indicators that were identified by expert opinion. The consensus method appears to be suitable for other animal health and welfare research applications, for example agreeing on standard laboratory techniques or defining a specific animal welfare condition. Before applying this method, there are several factors that need to be considered. Firstly, the definition of who is an 'expert' and what is their level of expertise (Spoolder *et al.*, 2003). Secondly, expert opinion can differ according to occupation (Bracke *et al.*, 2008; Fernie *et al.*, 2008). Therefore, a biased view of welfare may be ascertained by using an unbalanced panel with experts from specific occupations (Vang, 1986). Furthermore, consensus methodologies make assumptions on the quality of the expert panel's decisions (Fink *et al.*, 1984). Expert opinion may have little in common with widely known scientific facts or reasoning and specialists may concentrate on a small and specific issue instead of considering the whole picture (Vang, 1986). Expert opinion should alter as scientific knowledge advances, although experts can ignore welfare research findings and provide opinion based solely on their personal experiences (Main *et al.*, 2003).

In view of these issues and in the absence of a reference test for animal welfare, a combined approach using a scientific literature review (Chapter 1) and expert opinion guided the identification of valid welfare indicators in this study. The advantage of employing expert opinion, over the opinion of the study researchers, is that a group of experts would be expected to possess greater diversity in experience and knowledge of the subject under discussion, which is beneficial to an idea-generation process.

As a number of concerns are recognised with the use of expert opinion, these factors were accounted for in both study design and conduct. In particular, the following considerations were made: 1. clear definition of experts (Garabed *et al.*, 2009), 2. selection of experts according to level of expertise, 3. balanced composition of members in terms of expertise in sheep welfare and occupation, 4. quality of decisions, by using a sound consensus method (NIH, 1990), 5. clear communication of meeting objectives and fulfilment of distinct tasks (Kynn, 2008), and 6. presentation of scientific knowledge to establish an educated and informed panel.

The Delphi method has been suggested as a feasible way of collecting expert opinion by avoiding conflicts between individual opinions and thereby minimising bias of opinion (Delbecq *et al.*, 1975; Whay *et al.*, 2003a). However, Delphi questionnaires can be criticised on the grounds of poor response rates and long response times (Hsu and Sanford, 2007). Furthermore, postal questionnaire methods are unsuitable for research requiring personal and direct communications (Fink *et al.*, 1984). Here, face-to-face Delphi questionnaire methods, recently used to identify individual farmer concerns for sheep welfare (Phillips and Phillips, 2010) may prove useful. Individual interviews were not considered to be appropriate for this study, instead, a method that elicited the opinion of a number of experts during a one-day meeting and encouraged discussion and ideas-generation was required. As a result, other face-to-face consensus methods including the Nominal Group Technique (NGT) (Delbecq *et al.*, 1975), Improved Nominal Group Technique (INGT) (Fox, 1989) and the NIH method (NIH, 1990) were examined.

The selected NIH method allows researchers to present the most current, scientific knowledge and concepts to experts attending a consensus-style meeting (NIH, 1990). NIH can include a pre-meeting consultation to prevent a 'bottleneck' in group discussions and also to reduce the meeting time required (Fox, 1989). Focus groups, as used in NIH meetings, also allow for idea generation and exchange of expertise and opinions (Fitzpatrick and Boulton, 1994; Krueger and Casey, 2009). Therefore, in this study, each focus group was comprised of seven to eight members, guided by trained facilitators (Glaser, 1980) to encourage full contribution and minimise over-domination by any member (Delbecq *et al.*, 1975; Krueger and Casey, 2009). To reduce opinion biases, groups were mixed in terms of occupation, experience and expertise, and this also prevented the group from deferring to the opinion of any one individual (Krueger and Casey, 2009; Kynn, 2008).

There is a lack of a definition on the establishment of a consensus in decision-making processes (Scott and Black, 1991). The NIH method, takes the 'majority rule' approach (NIH, 1990) to the identification of consensus, so, one disadvantage is that minority or extreme views may be ignored. The goal of this study was to identify broad expert agreement, so it may be appropriate to discount very polarised or minority opinions

(Scott and Black, 1991). The final step in NIH methods allows experts to comment and deliberate on the scientific evidence before finally offering their judgement (NIH, 1990). The post-meeting distribution of the output of this study permitted individual experts to review the welfare indicators further and confirm their agreement of the list of welfare indicators created at the expert panel meeting. As a result, this approach provided both face and consensual validity to the sheep welfare indicators identified by the expert panel. This study identified a number of on-farm welfare issues affecting sheep. Whilst experts were asked to specifically identify issues that were relevant to farms in England and Wales, many of the welfare issues raised are applicable to other countries, systems and stages of production, such as transport and lairage (Cronin *et al.*, 2002; Phillips and Phillips, 2010; Waterhouse *et al.*, 2003). Furthermore, this is the first study to validate potential welfare indicators that consider: 1. sheep of all ages (neonates to adult sheep), 2. all stages of on-farm production (birth, lactation, weaning, growth, reproduction), and 3. the diversity in sheep management (intensive, indoor-managed flocks to large, extensive systems). Pre-meeting consultation also found that both lameness and poor body condition were consistently identified as sheep welfare issues. Interestingly, the welfare issues suggested by this expert panel concur with the expert opinion regarding the welfare issues for extensively managed sheep (Waterhouse *et al.*, 2003), and for sheep managed in Australia (Cronin *et al.*, 2002; Phillips and Phillips, 2010; Phillips *et al.*, 2009).

In contrast to other studies (Cronin *et al.*, 2002; Phillips and Phillips, 2010; Phillips *et al.*, 2009; Rousing *et al.*, 2007), the suggested welfare issues were not ranked in terms of their importance. Ranking can be used to create an overall welfare index or to identify a set number of measures that are used to assess on-farm welfare standards (Rousing *et al.*, 2007). However, the creation of an overall welfare index (Fernie *et al.*, 2009) or overall welfare assessment system for sheep was not the objective of this study. Instead the aim was to identify a number of valid welfare measures, essentially individual diagnostic tests, which would undergo further validation in on-farm studies.

A fundamental aspect of this study was to view welfare from the animal's point of view by encompassing both physical and mental elements as detailed in the FAWC Five Freedoms (1994). Since each of the Five Freedoms was deemed to be of equal importance for the welfare of the individual sheep, ranking of the issues was not considered to be appropriate for this initial stage in the research project. Indeed, ranking of measures at this stage risked omitting a welfare issue that was specific to a certain time point or to a particular farm management system. It was intended that any ranking, synthesis or exclusion of the suggested indicators would be based on the results of diagnostic validity, reliability and feasibility studies.

Prior to this study, indicators for the on-farm assessment of sheep in England and Wales had not been identified by expert opinion. Previous work has identified potential animal-based welfare indicators for sheep transported by sea (Pines *et al.*, 2007) and for organic sheep managed in Italy (Napolitano *et al.*, 2009). Animal-based indicators have received increasing attention for their value in on-farm welfare assessments (Whay *et al.*, 2003b, Knierim and Winckler, 2009). A combination of animal-, resource-, and management-based indicators may be most appropriate for on-farm welfare assessments (Capdeville and Veissier, 2001). The panel, therefore, generated a range of animal-based, resource-based and management-based indicators that they considered to be practical under farm conditions and the limits of a one-day assessment period.

Many of the animal-based indicators suggested by this expert panel were focused on measures of flock health, for example, body condition scoring (BCS), lameness scoring, and skin lesion assessment. Whilst BCS was identified as a useful welfare assessment tool, and has been scientifically validated as a measure of the body fat content of sheep (Russel, 1969), the panel recognised that BCS was not an appropriate indicator for young and growing animals. The panel also suggested that fewer categories of condition scores may be more appropriate for welfare assessment purposes. For example, sheep could be categorized as simply 'fit', 'fat' or 'thin'. Another specific indicator suggested by the panel included the use of an overall young lamb thriving score in order to categorise lambs as either "thriving" or "ill-thriven". This composite indicator was designed by the panel members to provide an overall welfare score based on the assessment of a number of health and welfare measures such as posture, demeanour, standing ability, and also specific health-based measures such as abdominal fill, gait, body condition and the presence of any ocular abnormalities. Experts also identified the need to assess animals managed at specific times during the sheep production calendar. They also emphasised the importance of including artificially-reared lambs in any young lamb welfare assessments.

Experts suggested that an assessment of sheep behaviour was a useful welfare assessment tool and that behavioural separation of individual animals from the flock and/or the expression of a dull, depressed demeanour clearly indicated a health and/or welfare issue for the individual sheep. In addition, the use of a whole-animal behavioural approach to sheep welfare assessment was proposed. The expert panel identified that Qualitative Behavioural Assessment (QBA); a holistic approach to the assessment of animal behavioural expression (Wemelsfelder and Lawrence, 2001), should also be included in the list of animal-based indicators.

Overall, the panel's suggestions closely followed those of another expert panel (Waterhouse *et al.*, 2003) who proposed that on-farm issues for sheep welfare may be considered in terms of the severity, duration and intensity of the effect on welfare. In this project, the intensity was considered to be the number of sheep affected by a specific indicator. There is a lack of definition between individual and flock welfare indicators (Waterhouse *et al.*, 2003), but in this study, experts were asked to consider welfare issues that were important for the individual sheep. Whilst it was clear from the outputs of each focus group that experts identified issues and indicators that were relevant to the welfare of the individual sheep, they suggested that the interpretation of indicator assessments could be based on a flock-level approach. Indeed, the panel in this study suggested that the flock prevalence of indicators was the most valid and feasible way of performing on-farm assessments of the animal-based indicators.

Experts were informed that any indicators initially suggested by the panel would undergo further validation in the next stage of the project, when diagnostic test evaluation of the indicators would be performed in field studies. This approach permitted the study of both individual and group-level methods of assessing sheep. Results of on-farm testing of indicators would be presented to the expert panel at subsequent meetings in order to allow experts to make an informed and science-based decision on the interpretation of individual- and flock-based welfare assessments. The panel also suggested that, where appropriate, the severity and duration of the welfare indicators could be considered in any scoring systems that were developed by the authors.

As this is the first step in identifying valid indicators of sheep welfare, not all the animal-based indicators suggested by the panel may be applied during field testing. Experts suggested that biochemical measures, such as plasma cortisol levels, could be used for on-farm assessment of sheep welfare, although they also suggested that measures may not be feasible for a whole flock assessment. Interestingly, experts did not suggest any non-invasive biochemical measures such as faecal cortisol levels. Accordingly, biochemical measures were not incorporated in any subsequent field testing of the welfare indicators.

Overall, the indicators suggested by the panel broadly fell into 1. assessments of grazing, housing and on-farm facilities, 2. assessments of farm records, and, finally, 3. discussions with farmers to ascertain on-farm management practices. The effect of season and production system should be considered in the application of resource-based indicators, for example, sheep may only be housed during short and specific periods and often receive intermittent supplementary feeding. For a large proportion of the sheep production calendar, the main resource for sheep in the UK may be considered to be the provision of grazing. Therefore, the main resource-based indicator would be the quality and the quantity (sward height) of the grazing provided.

2.5 Conclusion

The objective of the work presented in this chapter was to identify valid indicators for the on-farm assessment of sheep welfare. This study followed good practice guidelines regarding meeting design, conduct of focus groups, and selection of experts in order to use a consensus of expert opinion to validate potential indicators of sheep welfare. Whilst the NIH approach is widely used in human medicine, it has not previously been reported for use in expert consultation processes for animal health and welfare research purposes. A clear advantage for expert consultations that require direct and face-to-face communication of a number of members is the balanced, multi-disciplinary, science-based and informed approach offered by the NIH technique. Experts identified a large number of on-farm welfare concerns for sheep and also suggested a range of potential animal-, resource- and management-based indicators and the method employed here may be appropriate for the development of welfare indicators for other species. This expert validation process was a first step in the identification of valid indicators of sheep welfare. The next step is to investigate the diagnostic validity (reliability, sensitivity and specificity) and feasibility of these measures during on-farm studies.

Chapter 3

GENERAL METHODOLOGY

3.1 Introduction

This chapter describes the scoring scale and assessment method for animal-, resource- and management-based indicators of sheep welfare, which were developed following consultation with the scientific literature (Chapter 1), a panel of experts (Chapter 2), and testing during on-farm pilot studies.

3.2 On-farm protocol

The on-farm protocol applied during an on-farm visit, including assessments of animals, housing, facilities and discussion with the farmer are described below. The outcome of each assessment was manually recorded on the appropriate charts shown in Appendix B.

3.2.1 Definition of sample animals

Sheep and lambs were defined according to their age. Adult sheep were defined as both male and female sheep aged one year or over. Sheep were aged according to the appearance of their incisor teeth. The presence of two central permanent incisors or further permanent incisors was used to age an animal as one year and over. Growing animals included both male and female animals that were over 6 weeks of age and under one year-old - determined by the presence of temporary incisor teeth. Young lambs were defined as lambs aged 6 weeks and under.

3.2.2 Selection of sample animals

The sample of sheep and lambs presented for assessment was selected by the farmer. To guide selection of sample animals, each participating farm was provided with a written protocol, in which the number of animals requested was determined according to the aim and type of study and the feasibility for assessment during a one-day visit.

3.3 Adult sheep and growing lamb welfare indicators

3.3.1 Indicators assessed by group observation

The observer stood at the barrier of the field or paddock or housing to observe the behaviour of the sample group, with minimal disturbance, for a period of 5 minutes. After this initial observation period the observer entered the group assessment area. The observer walked around the sheep and encouraged the sample group to walk away in as quiet a manner as possible. By walking the group around the assessment area, the observer assessed and recorded the total number of individual sheep observed with the following conditions:

Lameness

The assessment of lame sheep was performed when sheep were walking and not running within the observation area. A 'sound' sheep was defined as one that bears weight evenly on all four feet. Lameness was defined as the observation of any or a combination of the following clinical signs: visible nodding of head in time with short stride, grazing on knees, uneven gait, arcing of the back during locomotion, non-weight bearing on affected limb when standing, extreme difficulty rising, reluctance to move once standing and more than one limb affected by any of these signs (Kaler *et al.*, 2009).

Dull demeanour

Behavioural separation from the rest of the group, lowered head carriage, and reduced responsiveness to the approach of the observer were signs that were used to assess the presence of dull demeanour.

Excessive panting

The presence of sheep showing signs of panting with an obvious and active abdominal effort to respiration, with or without an open mouthed appearance was recorded as excessive panting.

Coughing

The number of sheep heard or observed with any one or a combination of the following signs was recorded: coughing in a paroxysmal fashion, sheep showing signs of respiratory distress including obvious abdominal effort associated with breathing or noisy breathing such as wheezing. A single cough that may occur as part of a normal protective reflex when grazing was not included in the assessment of coughing.

Skin irritation

Behavioural signs of pruritus (skin irritation) included any one or a combination of the following signs: rubbing and scratching along walls, posts or other objects, restlessness, stamping of feet and biting, or nibbling at areas of the body (Berriatua *et al.,* 2001).

Wool loss

The presence of visible wool tags and/or any area of wool loss were recorded.

Dirty rear

Soiling or adhesion of the perineum, tail and hindquarters by soil, mud or faeces. This included dried as well as freshly contaminated areas.

Dirty belly

Soiling or adhesion of the ventral abdomen (belly) by mud or faeces. This included dried as well as freshly contaminated areas.

3.2.2 Indicators assessed by individual examination

The sample group was then gathered by the farmer to a holding area for assessment of individual animal indicators. Each individual sheep was assigned a numeric identification, given by the order in which they were moved into the assessment pen. The stage of production for each individual sheep; ram, ewe or growing lamb was recorded. A mobile handling system (Harrington handling system, B & P Engineering: York) was available to facilitate individual examination and manual turning of sheep. In other farms, on-farm handling facilities and hurdling pens were used for individual sheep assessment. To reduce the effect of behavioural isolation on a single sheep, two animals were held in the individual assessment pen at any assessment period. Each sheep was individually assessed using the indicators described below (Table 3.1).

Lameness

As the individual sheep entered the assessment area the gait was evaluated to identify any clinical signs of lameness (previously defined in group observation). The observer walked the sheep around the pen to examine the gait in both directions, and recorded the location of assessment and the quality of the flooring surface for lameness assessment purposes.

Dull demeanour

The general demeanour of individual sheep was also assessed as the sheep approached the examination area. Individual sheep were assessed using the description previously outlined under the group observational method.

The individual sheep was then restrained and held quietly by placing a hand under the jaw to allow the observer to make the necessary assessments.

Eye condition

An eye condition was deemed to be present if any one of the following signs was observed – an eye was held partially or fully closed (blepharospasm), corneal opacity (keratitis), presence of an ocular discharge (muco-purulent, or haemorrhagic discharge), conjunctivitis, entropion (inversion of the lower eyelid).

Nasal discharge

The presence of a visible nasal discharge (mucoid, purulent, or haemorrhagic discharge).

Tooth disease

With the sheep restrained in a standing position, the lower lip was carefully manipulated to expose the incisors of the lower jaw. Any missing permanent incisors – often observed as 'gaps' between adult incisors or broken incisors (colloquially termed 'broken mouthed') were assessed as incisor loss. Following this assessment, the molar or cheek teeth were assessed by squeezing the index finger and thumb together and running both hands along the lower mandible. The thickness, sharpness, length and position of the molar teeth on both sides of the face were assessed. The palms of both hands were placed along the cheeks and the area was palpated to detect any displacement of the molars. The upper cheek teeth normally overhang the lower molars but in advanced molar disease the teeth will move outwards and swelling or altered face symmetry may be observed (Ridler and West, 2010). Very sharp edges to teeth and/or thickening of the mandible or palpable bony changes can occur with molar disease. The presence of sharp molar edges, extreme discomfort on examination of cheek tooth areas, palpable mandibular swellings and bony growths was recorded as molar tooth disease.

Coughing

Coughing was scored as present if sheep were heard to repeatedly cough or cough in paroxysmal fashion during the period of the physical examination. Obvious difficulty with breathing, for example, increased abdominal effort associated with the rhythmic movements of breathing or wheezing associated with respiratory movements were also included within the individual examination indicator of coughing.

Ear lesion

Examination of the outline and regularity of the ears was performed visually without the need for handling of the ears in many instances. The presence of tears, rips, or inflammation and infection was visually assessed.

In-growing horns

In horned animals, the tips and edges of the horns were examined. Horns may overgrow and curl over onto the sides of the skull causing skin wounds or they may physically impinge on the surface of the eye. Any wounds or breaks in the skin caused by horn penetration or contact with the surface of the skull or if overgrown, curled horns were physically threatening the eye were recorded as an in-growing horn.

Body condition score

Body condition was assessed using the Russel (1984) 6-point scoring scale. Using both hands the lumbar vertebrae and transverse processes were manually palpated. An assessment of the sharpness and prominence of the spinal process together with coverage over the *longissimus dorsi* (loin) and degree of fat cover was made by pressing the fingertips underneath the ends of the lumbar processes to assess the amount of muscle. The scoring scale is fully detailed in Table 3.1.

Fit – Fat - Thin

The assessment of body condition using the fit – fat – thin system was also based on the Russel (1984) scoring descriptions described above. The term 'fit' was equivalent to body condition scores between 2 and 4. 'Thin' was categorised as body condition below a score of 2, and 'fat' was categorised as body condition above a score of 4.

Skin irritation

The indicator skin irritation was assessed using the nibble test (D'Angelo *et al.*, 2007). This was performed by rubbing the fingertips on the skin of the sheep along the lumbar, flank and shoulder regions. A positive response and presence of skin irritation was interpreted as positive if the animal showed head and neck extension, nibbling and chewing movements associated with head and tongue movements after manually stimulation.

Wool loss

Assessment of fleece cover was made in the shorn and unshorn sheep and an area of wool loss of any size was recorded.

Skin lesion

An assessment of the integrity of the skin was made in the standing and turned sheep. Skin lesions and the presence of any maggots in the integument (skin of head and body and hoof) were scored and recorded. Skin lesions included abscesses, areas of scabby, scaly, flaking skin and moist, erythematous areas. Yellow discoloration of the fleece and presence of any odours were also used to alert to the presence of a skin lesion. In the fully fleeced sheep, the observer ran their hands through the wool and areas of the wool were parted in order to examine the integrity of skin.

Injuries and wounds

Injuries and wounds were simultaneously assessed along with skin lesion and wool loss indicators. The skin of the entire body and head was examined for signs of injury such as wounds, bruises, cut and scratches and scored according to Table 3.1.

Short tail length

With the individual sheep remaining in a standing position, the length of the tail was assessed according to the Welfare of Farmed Animals (England) regulations 2007 (S.I. 2007 No.2078) in which a short tail is defined as one that "does not cover the anus in males or the vulva in females".

Dirty rear

Furthermore the cleanliness of the rear area was scored along a 4-point scale according to the descriptors outlined in Table 3.1. Dirt was defined as dried on or fresh contamination by soiling (mud or faecal). The rear area was defined to include the perineum, the superficial and medial aspects of the gluteal region to the top of the hind limbs.

Sheep were then cast and turned over to examine the under belly, feet, udder and pre-putial body regions. Ewes in early stages of pregnancy (during and one month following the tupping period) and heavily pregnant ewes (4 - 5 months gestation) were not turned.

Dirty belly

Whilst restrained in an upended position, the cleanliness of the ventral abdomen (belly) of the sheep was scored along a 4-point scale according to the descriptors outlined in Table 3.1. Dirt was defined as dried on or freshly contaminated soiling of the belly by mud or faeces.

Mastitis (ewes only)

In the ewe both mammary glands were palpated for areas of thickening and hard masses with or without signs of active inflammation: heat, discomfort and engorgement of the glands and teats. Scoring was based on the presence of mastitis and the number of glands affected (Table 3.1).

Crystals (rams and ram lambs only)

In rams, the pre-putial area and end of penis was extruded and examined for the presence of uroliths – defined as a gritty, sand-like material.

Dirty legs

Leg cleanliness was scored along a 4-point scale according to the descriptors outlined in Table 3.1. Dirt was defined as dried on or freshly contaminated soiling of the rear by faeces or mud. A summary judgement based on the assessment of all four limbs was made by the observer.

Joint swelling

Joint swelling was assessed as the presence of swelling, heat and obvious discomfort in any limb joints above the pedal joint. This indicator included recording the presence of lesions such as osteoarthritis and septic arthritis ('joint-ill'). Scuffing of the carpus or stifles was not recorded as joint swelling.

Foot lesion

Progressing from the leg assessment, each foot was individually examined. The two claws of each foot were separated to examine the inter-digital space. The integrity of the horn of the sole, the hoof wall, white line and the coronary band was then examined. The presence of any foot lesion was recorded; defined in further details as below. Assessment of foot lesions required minimal use of foot trimmers in certain cases in order to remove overgrown horn.

Significant foot lesion

Where any foot lesion was present, the observer made a subjective assessment as to whether the lesion was considered to result in lameness for the individual sheep.

White line lesion (WL)

Assessment of this lesion required specific assessment of the white line area - the junction between the sole and wall horn, visible as a pale line. The presence of a white line lesion included separation and detachment of the white line ('shelly hoof'), impaction and infection and was recorded as present if one or more feet were affected (Winter, 2004b).

Inter-digital dermatitis

The inter-digital skin of each foot was examined for inter-digital dermatitis ('scald') - a raw to white, moist hairless area, progressing to inflammation, infection and necrosis of the inter-digital skin (Winter, 2004b).

Foot rot

Foot rot was defined as separation of the horn of the hoof, beginning at the junction of the skin and horn, near the heel, through to invasion of the sole with separation of insensitive and sensitive laminae (Egerton and Roberts. 1971).

Contagious Ovine Digital Dermatitis (CODD)

CODD was defined as a small, ulcerated region around the coronary band and included loosening of the claw through to the total loss of the hoof capsule and presence of a raw stump of sensitive lamina tissue (Winter, 2004b).

Other foot lesions

A **toe granuloma** was recognised as a strawberry-like growth of proud flesh which is found at the tip of the toe and may be covered with loose horn. The interdigital space was palpated and examined for the presence of **inter-digital growths,** which were defined as masses of variable sizes, located within the inter-digital cleft of the foot. A **pedal joint abscess** was diagnosed as the presence of heat, swelling and possible infection of the pedal joint. Discharging tracks of pus visible along the coronary band may be observed and the affected foot can appear to splay away outwards (Winter, 2004). Where it was not possible to define or diagnose a specific foot lesion, the observer recorded the lesion as 'unknown' or 'unidentifiable'.

Myiasis

The presence of maggots anywhere on the skin of the body or feet (myiasis) was recorded. The size of the affected area was included in the maggot scoring system (Table 3.1).

Table 3.1 Welfare indicators for adult sheep and growing lambs assessed by individual examination

Welfare indicator	Welfare issue	Description of scoring system
Demeanour	General ill-health, poor welfare	0 = Bright, alert, responsive 1 = Dull, depressive, reduced responsiveness
Eye condition	Neglected treatment of an eye condition	0 = No discharge or ocular abnormality 1 = Presence of ocular discharge and/or abnormality
Nasal Discharge	Respiratory disease	0 = No nasal discharge 1 = Presence of abnormal nasal discharge
Tooth disease	Broken mouth, molar disease	0 = No evidence of incisor loss or molar disease 1 = Loss of incisors (broken mouth) 2 = Molar disease: thickening or swelling of mandible, sharp irregular molar teeth 3 = Both incisor loss and molar disease
Coughing	Respiratory disease	0 = No bouts of coughing 1 = Repeated bouts or paroxysmal coughing
Ear lesion	Poor ear tagging technique	0 = No damage, tears, swelling of deformity of one or both ears 1 = Presence of damage, torn, ripped, infected of deformity of one or both ears
In-growing horns	Neglect of in-growing horns	0 = No in-growing horns 1 = Horn(s) penetrating into the skin or skull or physically threatening the cornea
Dirty belly	Cleanliness of environment	0 = Clean: may be minor splashing with mud or faeces 1 = Dirty: discreet to solid plaques of dirt or mud on belly 2 = Dirty: discreet to solid plaques of faeces on belly 3 = Filthy: belly very heavily clagged with mud or faeces
Dirty rear	Gastro-intestinal disease	0 = Clean: may be minor splashing with mud or faeces 1 = Dirty: discreet to solid plaques of dirt or mud on belly 2 = Dirty: discreet to solid plaques of faeces on belly 3 = Filthy: belly very heavily clagged with mud or faeces
Dirty legs	Cleanliness of environment	0 = Clean: may be minor splashing of leg(s) with mud or faeces 1 = Dirty: discreet to solid plaques of dirt or mud on leg(s) 2 = Dirty: discreet to solid plaques of faeces on leg(s)

		3 = Filthy: leg(s) very heavily clagged with mud or faeces
Mastitis	Mastitis (ewes)	0 = No evidence of mastitis in any gland 1 = One gland affected by mastitis 2 = Both glands affected by mastitis
Crystals	Urolithiasis (rams and ram-lambs)	0 = No crystals or grit-like substance evident in prepuce or tip of penis 1 = Crystals or grit-like substance present in prepuce or tip of penis
Tail length	Compliance with welfare legislation	0 = Appropriate tail length: tail covers anus in males or vulva in females 1 = Inappropriate tail length: tail does not cover anus in males or vulva in females
Wool loss	Ectoparasite, skin or neurological condition	0 = No wool loss observed 1 = Area of wool loss observed
Skin irritation	Ectoparasite, skin or neurological condition	0 = No response to nibble test 1 = Positive response to nibble test: extension of head and neck, mouthing, or nibbling
Skin lesion	Ectoparasite, skin or neurological *condition*	0 = No skin lesions observed 1 = Presence of a small single lesion: size of a 50 pence coin or less 2 = Presence of multiple small skin lesions: each lesion size 50 pence coin or less 3 = Presence of a single area: approx hand sized area (10x5cm) covering body/head 4 = Presence of multiple hand sized areas: (10x5cm) covering body/head 5 = Presence of skin lesions all over body/head: diffuse, generalised distribution 6 = Presence of a single or multiple ($n < 5$) : size of a 5 pence coin or less
Injuries and wounds	Poor handling, management, maintenance of building and equipment, aggressive dogs	0 = No injuries, wounds or scratches observed 1 = Superficial scratches or superficial cuts less than 5 cm long on body/ head 2 = Superficial scratches or superficial cuts greater than 5cm long on body/head 3 = Healing or healed wounds: dried scab covering wound 4 = Presence of a single open wound on the body or head 5 = Presence of multiple open wounds on body or head
Body condition score	Inappropriate body condition – emaciated or obese sheep	0 = Extreme emaciation, impossible to detect any muscle of fat between the skin and bone 1 = Spinous processes are prominent and sharp. Transverse processes are also sharp; fingers pass easily under the ends and it is possible to feel between each process. Eye muscles are shallow with no fat cover 2 = Spinous processes still feel prominent but smooth, individual processes feel like fine corrugations. Transverse processes are smooth and rounded, possible to pass the fingers

		under the ends. Eye muscle areas of moderate depth but little fat cover 3 = Spinous processes are detected only as small elevations; smooth and rounded, felt only with pressure. Transverse processes are smooth and well covered, firm pressure is required to feel them. Eye muscle of moderate depth with some fat cover 4 = Spinous processes can just be detected, with pressure as a hard line between the fat covered areas. Transverse processes cannot be felt. Eye muscles are full and have a thick covering of fat. 5 = Spinous and transverse processes cannot be detected even with firm pressure and there is depression between the layers of fat where the spinous processes would normally be felt. Eye muscle is very full with very thick fat cover and of fat deposits over the rump and tail
Fit – fat – thin	Inappropriate body condition – emaciated or obese sheep	Thin = Body condition score below 2 (as above) Fit = Body condition score 2 to 4 Fat = Body condition score over 4
Lame	Lameness	0 = Sound 1 = Lame
Foot lesion	Lameness	0 = No foot lesion observed in any foot 1 = Presence of a foot lesion in one or more feet
Significant foot lesion	Lameness	0 = No foot lesion considered to result in lameness 1 = Presence of a foot lesion considered to result in lameness
White line	Lameness	0 = No white line lesion observed 1 = Presence of white line lesion in one or more feet
Scald	Lameness	0 = No scald observed 1 = Presence of scald in one or more feet
Foot rot	Lameness	0 = No foot rot observed 1 = Presence of foot rot in one or more feet
CODD	Lameness	0 = No CODD observed 1 = Presence of CODD in one or more feet
Other foot lesion	Lameness	1 = Toe granuloma identified on clinical examination 2 = Interdigital growth (hyperplasia) identified on clinical examination 3 = Pedal abscess identified on clinical examination 4 = Foot lesion present of unknown or unidentifiable diagnosis

Joint swelling	Joint disease, lameness	0 = No joint swelling, discomfort, heat or pain observed 1 = Presence of joint swelling, discomfort, heat or pain
Myiasis	Myiasis (blowfly strike)	0 = No maggots observed on body, head or feet 1 = Presence of maggots in a single area: size of 50 pence coin or less 2 = Presence of maggots in multiple areas: each areas size of 50 pence coin or less 3 = Presence of maggots in a single area: approx area hand size (10x5cm) 4 = Presence of maggots in multiple areas: each area approx hand sized (10x5cm) 5 = Presence of maggots throughout body: diffuse, general distribution

3.4 Young lamb welfare indicators

To prevent mis-mothering and disturbance of ewe-lamb bonding behaviour, young lambs managed at foot with ewes were not gathered at any stage of the assessment. Freshly born animals and young lambs under 6 hours old were not examined. Observations were performed by standing outside of individual pens or groups or by walking around fields at a distance sufficient to allow assessment of demeanour and lameness but with minimal disturbance to the group. On the day of a visit, farmer consent was confirmed and the location of the lambing sheds and fields of ewes and young lambs was identified. The observer additionally recorded: 1. location of the lamb assessments – indoor or outdoor environment, 2. method of rearing – with a non-tethered or tethered ewe (physical restraint of the ewe to minimise aggressive behaviour and maternal rejection of the lamb), or artificial rearing (lambs not reared with a ewe), 3. management system – individual pen or groups of animals, and 4. approximate age of lamb (0 – 3 days old, >3 – 7 days old, >1 – 6 weeks). Each sample lamb was individually assessed using the 11 indicator tests described below and summarised in Table 3.2.

Play behaviour

An assessment of the level of play behaviour was made by observing lambs undisturbed for 1 minute. During that period any play behaviours: frolicking, jumping, mutual grooming, playing with objects, such as feed troughs or forage was assessed. The level of play behaviour was subjectively assessed and recorded on a visual analogue scale (VAS) by making a mark along a 100 millimetre line labelled at one end 'no play behaviour' and the opposite end 'could not be more playful' (Figure 3.1). The perceived level of play behaviour observed was indicated by drawing a vertical line along this scale.

Figure 3.1 Visual analogue scoring scale for play behaviour assessment

No play behaviour   Could not be more playful

Stimulation

Stimulation was assessed as the responsiveness of the lamb to observer presence. In group situations or resting lambs this may have required the observer to whistle, wave or touch lambs to assess awareness and response to stimulation. An unresponsive lamb was one that did not show any behavioural reaction and was not alert to any stimulus or activity in its surroundings. This indicator was not assessed on sleeping animals.

Demeanour

Demeanour was judged according to the general appearance and behaviour of the lamb. Dull demeanour was described as a lamb that appeared to be depressed and withdrawn from the ewe or other lambs.

Shivering

The presence of shivering, trembling of the musculature, or quivering of the body of the lamb was recorded.

Lambs reared in individual pens occasionally required lifting out of the pen for a short period to assess standing ability, lameness and some of the indicators outlined below.

Standing ability

The standing ability of the lamb was scored by observing the movement of lambs around the pen area or field. A lamb that stood easily was one that was observed to bear weight equally on all legs without collapsing or swaying. Weakness of legs was defined as a lamb that was able to rise to a standing position but could not maintain this position without swaying or collapsing. A recumbent lamb was unable to rise to a standing position.

Lameness

Assessment of lameness included the observation of any clinical signs of lameness, including a three-legged gait, holding a foot off the ground, an obvious head-nod in time with movement of a limb or clinical signs of septic arthritis (joint ill) including the presence of an enlarged, inflamed joint, and the observation of a stiff, stilted gait.

Posture

The posture of lambs was assessed as 'hunched' if lambs showed arching of the backbone with a tucked up abdomen.

Abdominal fill

The degree of abdominal fill was assessed visually and, where necessary, through gentle palpation of the abdomen. This was specifically performed to determine the presence of a bloated abdomen.

Body condition

Body condition was assessed by examining the cover of fat and muscle over the ilial wings (hips) and spinal vertebrae (backbone). In lambs with woolly fleeces or wrinkled skins, it was necessary to gently palpate the skeleton to assess the degree of cover. The ilial crest may be distinguished as a raised area but does not appear to be sharp or prominent in a lamb of 'appropriate' body condition. 'Inappropriate' body condition was assessed as the presence of a sharp, prominent skeleton with no fat or muscle cover.

Eye condition

The presence of an eye condition included the observation of any one or a combination of the following signs: blepharospasm (holding the eye tightly closed), presence of an ocular discharge (purulent, mucoid or haemorrhagic), corneal opacity, conjunctivitis, and inversion of the lower eyelid (entropion).

Salivation

Lambs were examined for the presence of excess salivation around the lips ('watery mouth'). Slight frothing of milk around the lips, which may occur in lambs following sucking, was not scored as salivation.

Table 3.2 Young lamb welfare indicators

Indicator	Welfare issue assessed	Description of scoring system
Play behaviour	Overall welfare	Scoring possible anywhere along a 0-100mm visual analogue scale – a line labelled at one end: 'no play'(0) and the opposite end was labelled 'could not be more playful' (100)
Demeanour	General health, poor welfare	0 = Bright, alert demeanour 1 = Dull, depressed demeanour
Stimulation	General health, poor welfare	0 = Responsive to stimulation 1 = Unresponsive to stimulation
Shivering	Hypothermia	0 = No shivering or trembling 1 = Active shivering or trembling
Standing	Lameness, general health	0 = Stands easily, without difficulty 1 = Weak, unstable when standing 2 = Cannot stand, recumbent
Posture	General health, lameness, inappropriate or insufficient nutrition	0 = No evidence of hunched and 'tucked-up' posture 1 = Hunched or tucked up posture
Abdomen	Inappropriate or insufficient nutrition Poor management of orphan lambs	0 = Appropriate abdominal fill 1 = Abdomen distended, ballooning of abdomen and flank 2 = Hollowing of abdomen; hollowed out appearance to flank
Lameness	Lameness, septic arthritis	0 = Sound; no obvious lameness, stiffness or joint swelling 1 = Lame; obvious signs of lameness, stiffness or joint swelling
Body condition	Insufficient nutrition, starvation, appropriate body condition	0 = Hip bone not prominent, good cover over backbone: body condition appropriate, 1= Prominent hip bone, poor cover over backbone: body condition inappropriate
Eye condition	Entropion, neglected treatment of eye condition	0 = No signs of ocular disease 1 = Presence of an ocular discharge, corneal opacity, conjunctivitis or entropion
Salivation	Infectious disease, inadequate colostrum *E.coli* septicaemia (watery mouth)	0 = No shivering or trembling 1= Shivering or trembling present

3.5 Resource-based indicators

To assess resource-based indicators (Table 3.3) the test standard observer performed visual assessments and observations throughout the farm visit. In addition, a tape measure, ruler and clear container were required for assessing certain resources.

3.5.1 Assessment of fields and grazing areas

If sample sheep were presented in a field area, an assessment of the grazing area was performed as follows. The presence of a shelter and shade from adverse weather conditions included the presence of hedgerows, stone walls, fences, trees.

Grazing assessment

The type of pasture (grass, soil, rough hill area, stubble turnip field) was recorded and an assessment of the availability of a well-drained area to permit sheep to lie down; particularly applicable to stubble grazed animals, was performed.

Grazing boundaries

Field boundaries and fences were assessed to determine their condition and maintenance.

Stocking density

A visual assessment of stocking density was made by the observer in terms of the appropriateness of the size of the area. In addition the number of sample sheep and size of the field (acres) as reported by the farmer was also recorded so that the stocking density was calculated (number of sheep / size of field).

Water supply

A visual assessment of the presence of a water supply in the field, for example a water trough, automatic drinker, water buckets or stream, was recorded. The accessibility of this water resource was then visually assessed to verify that sheep could assess the water and a further measure of the cleanliness was made. Using the clear container, the colour and transparency of a sample of water was visually evaluated.

Sward height assessment

The sward height of the area where sample sheep were presented was measured. The assessor walked across the field in a 'W' shape to take in as wide an area of the field as possible. Hedge areas, gateways and other areas that were not representative of the field were avoided. The assessor paced each step, and at every second pace a reading of the length of the blade of grass was made. Stems, weeds and flower heads were not measured. A ruler was held at the level of the ground and the blade length, to the nearest 0.5 centimetre (cm) was recorded. A total of 40 individual sward height measurements were determined for each field assessment area. The mean cm value of the 40 measures was determined to be the sward height of the grazing area (Stewart *et al.*, 2001).

3.5.2 Assessment of housing

Space allowance

In housed animals, space allowance and stocking density was determined by counting the number of animals in individual pens or groups and measuring the length and width of 3 pens. A visual assessment of freedom to move was made by assessing whether animals could move freely or if movement appeared to be restricted by over-crowding.

Lighting conditions

The appropriateness of the level of lighting to permit inspection of sheep was visually assessed.

Flooring

The type of flooring in the housing and the type and presence of any bedding was recorded. Where bedding was provided, it was assessed in terms of the quality and quantity provided to allow for physical comfort and cleanliness of housed animals. The wetness of bedding was assessed by pressing the wellington-booted foot down onto the bedding ('squelch test') and categorising the amount of water pooling and dryness in 5 separate areas of bedding (Tadich *et al.*, 1998).

Water supply

The accessibility was evaluated in terms of non-obstruction to water supply and the production stage/age of sheep. A clear container was used to collect a sample of water from the supply and this was visually assessed to determine whether the sample was clear and fresh or murky with or without contamination from debris and faeces.

Food supply

Presence of food in food troughs or racks in front of animals (including concentrates and forage) was assessed. The assessor made a visual observation as to the appropriateness of the height of feed racks and whether animals could access the food source provided. The forage offered (silage, hay, haylage or straw) was visually assessed for the presence of moulds and the observer smelled 4 hand-sized samples of forage to detect any foul odours.

3.5.3 Assessment of farm facilities

During the walk around the farm, the presence of foot bathing facilities was examined. The presence of gates and hurdles, roll over crush or specific sheep handling system; including specific handling facilities for rams, was recorded. A general assessment regarding the maintenance and order of equipment and tools used in the management of sheep; such as clippers, handling crush, foot baths, was made. This indicator included observation of sharp, uneven, broken edges in areas of housing and/or handling facilities. Throughout the on-farm visit, the assessor remained observant to the presence of dead bodies or rotting carcases lying around the farm and recorded the number of carcasses present amongst live sheep.

3.6 Management-based indicators

3.6.1 Data capture form

A brief farmer interview (approximately 10 minutes) was performed by the author during the day of assessment to complete a data capture form to ascertain background information regarding flock and management details (Appendix B).

3.6.2 Records

The following management-based indicators required evaluation of specific on-farm records:

Mortality records

Farmers were asked if they maintained mortality records. Following affirmation, the observer requested a brief examination of the records. Mortality records were classed as 'useful' if information regarding the reason of death (culled or suspected/confirmed diagnosis), date and further details such as production stage, was recorded.

Medicine records

A similar request regarding medicine records and their examination was made. Medicine records were recorded as 'useful' if information regarding reason for medical treatment or death (suspected/confirmed diagnosis) and names and dates of medical treatments.

Disease records

Farms were also asked about the presence of disease records. Disease records were recorded as 'useful' if information regarding the date and suspected or confirmed diagnosis of disease and action or investigation undertaken were maintained.

Flock health plan

Where present, farmers were asked for their permission to examine these records and where a flock health plan (FHP) was maintained, farmers were asked whether there was any veterinary involvement with this plan. Again consent for brief examination of the FHP was requested and where possible, the topics listed below were recorded if present:

- Lameness treatment and prevention
- Scab treatment and prevention
- Obstetric advice
- Tail docking and castration policy
- Vaccination policy
- Planning for emergency conditions, movement restrictions, fire, flooding
- Policy for hypothermia treatment during the lambing period

Feeding policy

Farmers were asked whether any written feed plan was available and whether any policy for specific winter ration/supplementary winter feeding was made.

Appropriate health control and action

The indicator 'evidence of appropriate health control and action' was made at the completion of the visit. The assessment of the indicator required the assessor to make a subjective judgement on the awareness of the farmer as the level of on-farm disease and/or welfare issues and in light of any evidence of appropriate action to address any health and welfare issues, such as use of isolation pens, animals individually identified as being treated, recently opened health treatments and drugs in medicine cabinet labelled with recent dispensing date. This was scored as: 'no disease observed', or 'disease present with appropriate action', or 'disease present but inappropriate action taken'.

Table 3.3 Resource-based indicators of sheep welfare

Assessment	Indicator	Welfare issue assessed	Description of scoring system
Grazing	**Presence of facilities to deal with weather**	Presence of shelter and shade	Presence / Absence
	Good boundaries and fences	Injuries and wounds	Presence / Absence
	Freedom to move in field	Over-crowding	Lots of room / adequate / over-crowded
	Stocking density on grazing	Over-crowding	Number of sheep in grazing area (acres)
	Stocking density in housing	Over-crowding	Number of sheep in housed area/pen Width and length of pen (metres)
	Sward Height (cm)	Insufficient grazing	Average of 40 measures of grass length (cm)
	Lie-back area provided	Comfort of lying area	Presence / Absence
	Grazing provided	Appropriate nutrition	Presence / Absence (grassland pasture, rough grazing, root crops)
Housing	**Freedom to move inside housing**	Over-crowding	Lots of room / adequate / over-crowded
	Adequate lighting for inspection	Compliance with legislation	Presence / Absence
	Flooring type	Comfortable lying area	Concrete / wire mesh / slats / other
	Provision of bedding	Comfortable lying area	Presence / Absence
	Bedding sufficient for comfort & hygiene	Clean, dry and comfortable lying area	Presence / Absence
	Squelch test	Cleanliness of housing	Dry / Squelchy underfoot / Filthy and wet

Food and water	**Provision of water source**	Appropriate nutrition	Presence / Absence
	Water supply accessible	Appropriate nutrition	Presence / Absence
	Water quality	Appropriate nutrition	Clear and fresh / murky / filthy with faecal contamination
	Presence of food in troughs/racks	Appropriate nutrition	Presence / Absence
	Feed accessible	Appropriate nutrition	Presence / Absence
	Forage of sufficient quality	Appropriate nutrition	Presence / Absence
Facilities	**Presence of foot bathing facilities**	Farmer practices	Presence / Absence
	Presence of isolation area	Farmer practices	Presence / Absence
	Presence of sheep handling equipment	Handling facilities and skills	Presence / Absence
	Equipment well maintained and in good order	Farmer policies and practices	Presence / Absence

3.7 Observer population

3.7.1 Selection of observers

A pool of 9 observers was recruited from the University of Liverpool, School of Veterinary Science. Observers were selected according to their prior involvement with the research project – staff members of the research project, or students undertaking undergraduate research projects. Accordingly, the observer pool comprised, pre-clinical undergraduate veterinary (n = 1) and bio-veterinary (n = 3) science students, and veterinary surgeons – defined as members of the Royal College of Veterinary Surgeons (RCVS) (n = 5) and sheep welfare research team. Observers were selected to include those with a range of prior experience in sheep health, welfare and production assessments and with different occupations. Observers were classified as 'experienced' if they had regularly applied sheep health and welfare assessments in the year prior to the study (Table 3.4). Observers who did not meet these criteria were classified as inexperienced. For the purposes of this study, observer occupation was recorded as either "vet", defined as a member of the Royal College of Veterinary Surgeons (RCVS), or "non-vet" if they did not meet this criterion.

3.7.2 Observer training

Observer 1 (the author), an experienced veterinary assessor who developed the on-farm indicator protocol and provided observer training, was designated as the 'test standard observer' (reference standard). Trained observers were provided with an on-farm protocol detailing the assessment of each indicator, and were requested to familiarise themselves with all indicator scoring systems and assessment methods. In addition to theoretical guidance through the on-farm protocol manual, all trained observers received on-farm training. For assessments of the adult and growing sheep indicators, trained observers (n = 6) attended a one-day on-farm practice session at the University of Liverpool sheep farm on 2nd July 2008, prior to the commencement of on-farm studies. During this training session, the test standard observer demonstrated the application of group and individual indicators and observers had the opportunity to practise using the recording charts. Specific training regarding identification of foot lesions using cadaver specimens and photographic images was used in the training of undergraduate student observers (observers 3, 4 and 7).

For the young lamb welfare indicators, four observers (observers 1, 3, 8 and 9) conducted assessments. Observers 3, 8 and 9 received training from the test standard observer. This included both theoretical training in the on-farm protocol and viewing video clips of lambs expressing dull demeanour and play behaviours, and a one-day on-farm training session was performed in February 2009 on 2 lambing flocks to clarify indicator assessments and to provide all observers with experience of the range of young lamb indicator scores.

Table 3.4 Observer population

Observer	Training	Experience	Category
1	Trained	Experienced	Vet
2	Trained	Inexperienced	Vet
3	Trained	Inexperienced	Non-vet
4	Trained	Inexperienced	Non-vet
5	Untrained	Inexperienced	Vet
6	Untrained	Experienced	Vet
7	Trained	Experienced	Non-vet
8	Trained	Experienced	Vet
9	Trained	Inexperienced	Non-vet

3.8 Study farm population

3.8.1 Farm recruitment

A database of 79 veterinary practices listed in the RCVS 2005 register as providing sheep health services and located within an 80 mile radius of the University of Liverpool, School of Veterinary Science, Leahurst campus was established. Veterinary practices were contacted by letter, electronic mail, facsimile, and telephone and asked to obtain the informed consent of sheep farming clients to supply their farm contact details to the author. Following a 2 month period, the recruitment strategy was broadened to include independent sheep consultants known to the research team, members of the Sheep Veterinary Society (SVS) and National Sheep Association (NSA) and the recruitment zone was also extended to a 150 mile radius. In order to test the indicators on farms in which a high proportion of sheep are affected by health and welfare problems, the Department for the Environment, Food and Rural Affairs (Defra) was approached in July 2008 to assist with farm recruitment. Following approval in April 2009 for involvement of Animal Health – an agency of Defra, an outline of the research proposal and farmer recruitment literature was circulated to divisional Animal Health Offices in Northern England and North Wales. In addition, the author gave an oral presentation to trading standards inspectors, veterinary surgeons and Animal Health officers in the North of England during July 2008 to request their assistance with farm recruitment.

3.8.2 Inclusion criteria

Eligible farms were defined as having either young lambs, growing lambs, ewes or rams present. Farm type was categorised as 1. lowland, 2. upland or, 3. hill flocks (Pollott and Stone, 2006) and farm purpose was classed as commercial, pedigree, or hobby. A commercial sheep farm was defined as a farm whose objective is the production of lambs for meat consumption that are finished on-farm or sold as stores and this category included farms that bred their own flock replacements. A pedigree farm was defined as a farm whose prime objective was the production of purebred breeding stock of known genetic history. Hobby farms were defined as farms or 'smallholdings' of an area < 50 acres and < 60 sheep or lambs; these farms were not reliant on income from sheep sales or meat production and may provide a recreational activity for their owners, akin to 'pet' or companion animals. All eligible farms were entered into a database and each was provided with details of the research project. Within 2 weeks of sending the letter, the author telephoned the farmers to confirm their interest in the study. Prior to recruitment, all farms that were identified as willing to participate were visited in person by the author to inform farmers of the nature of study, and the confidentiality and the data. The criteria for inclusion were the consent to participate and the distance of the farm from Leahurst (≤ 150 miles). Farms that did not meet these criteria were excluded.

3.8.3 Study farm selection

This approach identified 52 sheep farms, of which 50 farms met the inclusion criteria. The study farms were located across Northern and Central England and North to Mid Wales and encompassed a range of pure-breeds and cross-breeds (Table 3.5). Each farm was categorised into the farm stratification types of lowland (n = 26), upland (n = 12) and hill (n = 12). Farms were also classed as to their purpose into commercial non-pedigree (n = 43), commercial pedigree (n = 3), hobby (n = 5) (Table 3.6).

Table 3.5 Sheep breeds within the study population

	Sheep breeds	
Balwen Welsh Mountain	Jacob	Welsh half-bred
Welsh Mountain	Masham	Lincoln longwool
Welsh Mule	Welsh hill speckled face	Derbyshire Gritstone
North of England Mule	North Country Cheviot	Greyfaced Dartmoor
Beulah	Scottish Blackface	Hebridean
Swaledale	Berrichon du cher	Beltex
Rouge de l'ouest	Cambridge	Charollais
Texel	Shropshire	Bleu de Maine
Suffolk	Hampshire Down	Merino
Border Leicester	South Down	Welsh improved
Blue-faced Leicester	Badger-face	Lleyn

3.9 Study biosecurity protocol

The observer wore waterproof protective clothing and wellington boots throughout the course of farm assessments. Disposable examination gloves were worn, frequently changed between handling animals, and were always changed when potentially infective animals or tissues were handled. No eating or drinking in the animal handling area was permitted. Following completion, sheep handling equipment, clothing, boots and vehicle tyres were cleaned until all visible signs of contamination were removed. This was followed by disinfection with iodine (FAM 30, Evans Vanodine) at dilution rates and contact times according to manufacturer's instructions.

3.10 Ethical protocol

All studies described in this thesis were approved by the University of Liverpool Ethics Committee (reference RETH000287). On the day of a visit, farmer consent was given before any assessments were performed. In the event of any on-farm health or welfare issues that became apparent during the course of the farm visit, the consent of the farmer was also obtained to allow the author to contact the regular veterinary surgeon.

Table 3.6 Study farm population

Farm ID	Farm type	Farm purpose	Location	Assessment type
1	Lowland	Commercial	Cheshire	Adult sheep and young lambs
2	Lowland	Pedigree	Cheshire	Adult sheep and young lambs
3	Lowland	Commercial	Cheshire	Adult sheep and young lambs
4	Lowland	Commercial	Cheshire	Adult sheep
5	Lowland	Commercial	Cheshire	Adult sheep
6	Hill	Hobby	Denbighshire	Adult sheep and young lambs
7	Lowland	Commercial	Staffordshire	Adult sheep and young lambs
8	Lowland	Commercial	Lancashire	Adult sheep and young lambs
9	Lowland	Hobby	Cheshire	Adult sheep
10	Lowland	Commercial	Cheshire	Adult sheep and young lambs
11	Upland	Commercial	Lancashire	Adult sheep
12	Lowland	Commercial	Cheshire	Adult sheep
13	Lowland	Commercial	Cheshire	Adult sheep and young lambs
14	Upland	Commercial	Denbighshire	Adult sheep
15	Lowland	Commercial	Lancashire	Adult sheep and young lambs
16	Upland	Commercial	Denbighshire	Adult sheep and young lambs
17	Upland	Commercial	Denbighshire	Adult sheep and young lambs
18	Hill	Commercial	Gwynedd	Adult sheep and young lambs
19	Hill	Commercial	Denbighshire	Adult sheep and young lambs
20	Hill	Commercial	Gwynedd	Adult sheep and young lambs
21	Hill	Commercial	Gwynedd	Adult sheep and young lambs
22	Hill	Commercial	Gwynedd	Adult sheep and young lambs
23	Upland	Commercial	Lancashire	Adult sheep and young lambs
24	Lowland	Commercial	Cheshire	Adult sheep
25	Lowland	Commercial	Cheshire	Adult sheep
26	Lowland	Commercial	Cheshire	Adult sheep
27	Lowland	Pedigree	Cheshire	Adult sheep
28	Upland	Commercial	Powys	Adult sheep
29	Hill	Commercial	Powys	Adult sheep
30	Lowland	Commercial	Cheshire	Adult sheep
31	Lowland	Commercial	Lancashire	Adult sheep
32	Hill	Commercial	North Yorkshire	Adult sheep
33	Upland	Commercial	County Durham	Adult sheep
34	Hill	Commercial	County Durham	Adult sheep
35	Upland	Commercial	County Durham	Adult sheep
36	Upland	Commercial	County Durham	Adult sheep
37	Hill	Commercial	Gwynedd	Adult sheep and young lambs
38	Upland	Commercial	Powys	Adult sheep
39	Upland	Pedigree	Powys	Adult sheep
40	Hill	Commercial	Conwy	Adult sheep and young lambs
41	Lowland	Commercial	Staffordshire	Young lambs
42	Lowland	Commercial	Staffordshire	Young lambs
43	Lowland	Commercial	Cheshire	Young lambs
44	Lowland	Hobby	Cheshire	Young lambs
45	Lowland	Commercial	Staffordshire	Young lambs
46	Lowland	Hobby	Cheshire	Young lambs
47	Lowland	Commercial	Merseyside	Young lambs
48	Upland	Commercial	Lancashire	Young lambs
49	Lowland	Commercial	Shropshire	Young lambs
50	Hill	Hobby	Gwynedd	Young lambs

Chapter 4

VALIDATION OF GROUP OBSERVATIONS AS INDICATORS OF SHEEP WELFARE

4.1 Introduction

In the UK, sheep farming can vary according to the season and the events of the farming calendar, from extensive periods of management during the summer to the intensive management that occurs during the winter and lambing season (FAWC, 1994). For a large part of the sheep production cycle, the quantity and quality of grazing may be considered to be the main resource input to flock health and welfare. Therefore, the use of direct observations of animal-based outcomes, such as sheep behaviour, may be a better reflection of the true welfare status of the animal and, compared to the use of resource-based measures, could be more widely applied within on-farm welfare assessment protocols (Main *et al.*, 2007).

Since the behaviour of sheep is affected by their familiarity with handling, their stage of production, and physical features of the environment (Lynch *et al.*, 1992; Dwyer, 2009), the use of welfare indicators employed to assess the welfare of other farm animal species, such as avoidance distance (Napolitano *et al.*, 2005), may not be appropriate. Sheep are generally managed in groups and require gathering and handling to facilitate individual examination. A means of assessing the welfare of individual sheep with minimal disturbance and without the requirement for gathering and individual handling may be of value. Behavioural observations of groups of sheep are routinely employed by shepherds who perform brief, visual scans of their flocks to identify any individuals with a behaviour that differs from the rest of the flock. This approach appears to require few resources apart from the observational skills and knowledge of the assessor.

A number of indicators that are assessed by observing the behaviour of a group of sheep have been judged to be valid measures of sheep welfare (Chapter 2). However, the validity of these measures in terms of their on-farm test performance needs to be examined on a variety of sheep flocks by a range of different observers. The objective of this chapter was therefore to investigate the inter- and intra-observer reliability and the degree of observer bias of indicators of sheep welfare that were assessed by counting the number of individual sheep in the group observed with a specific animal-based outcome.

4.2 Materials and Methods

4.2.1 Study population

During July to November 2008, each welfare indicator was tested on a sample of 35 study farms, including lowland (n = 15), upland (n = 10), hill (n = 10) and commercial farms comprised of non-pedigree (n = 30), commercial pedigree (n = 3) and hobby (n = 2) flocks, recruited as described in Chapter 3.

4.2.2 Observer population

Eight observers, further described in Chapter 3, were selected to participate in farm visits according to their availability during July – November 2008. Each observer performed independent assessments of indicators by counting the number of individual animals affected by each indicator (Chapter 3).

Inter-observer study

Each indicator was tested in an inter-observer study on a total of 2406 adult sheep and growing lambs from 35 farms. A median group size of 68 sheep (range 24 – 120) was assessed on each farm. As it was not possible for all observers to perform assessments on all study farms, a varying combination of 2 – 3 observers of differing occupation, training and experience performed indicator assessments (Table 4.1). Observer combinations A to D included observers with a range of training and experience (Chapter 3) and only observer combinations E and F consisted of both trained and experienced observers.

Intra-observer study

The intra-observer reliability of the test standard (observer 1) was examined by repeatedly assessing a total of 88 sheep on 4 commercial, lowland farms selected according to the feasibility for repeat assessments. Sheep were managed in groups of 19 – 24. Each individual sheep was examined twice within a 24 hour period with a 5-hour interval between repeat assessments to minimise any alteration in indicator outcomes. Once the first assessment was completed, all recording sheets were sealed in an envelope, which was not examined until the completion of the study. Following completion of the first visit, the observer left the farm and performed unrelated tasks.

Table 4.1 Observer population

Observer combination	Observers	*n* farms	*n* sheep
A	1, 3, 4	23	1606
B	1, 3, 6	1	92
C	1, 4, 5	2	118
D	1, 3, 7	3	209
E	1, 7, 8	2	122
F	1, 7	4	259

4.2.3 Statistical analysis

Data was analysed in Stata version 10 (StataCorp LP, College Station, Texas). As different observer combinations performed assessments of different proportions of the study population (Table 4.1), an overall reliability value was not appropriate for each indicator. Instead, the proportion of sheep (%) assessed with each indicator, standardised Cronbach's alpha (α), and logistic regression evaluated each specific observer combination.

The overall % of animals assessed with each indicator was calculated for each observer in each observer group and compared to the overall proportion of sheep observed by test standard observer – the 'reference proportion'. The proportion of sheep determined by the remaining observers of the observer combination (A to F) was expressed as the 'second and third observer proportion'.

Cronbach's alpha (α) was used to measure the level of between and within observer reliability (Cronbach, 1951) and was interpreted according to Nunnally and Bernstein (1994) in which values > 0.70 were deemed to be 'reliable'. Cronbach's α was not calculated when counts of group indicators of zero, or ≤ 2 on-farm assessments occurred (Tables 4.2 and 4.3).

In addition, systematic differences in observer assessments were evaluated using binomial logistic regression models. The effect on the outcome of each group indicator (number of sheep) was examined in fixed effects models, in which farm identity was maintained as a fixed effect and observer identity was maintained as a covariate. Wald tests were used to examine whether significant differences (p-values ≤ 0.05) in indicator assessments were attributed to specific observer combinations. Observer assessments were also examined for the presence of bias, as described in Bland and Altman, 1986. The number of indicator counts (n sheep) recorded by each observer was also graphically represented to examine for evidence of observer bias.

4.3 Results

The level of inter-observer reliability for each observer combination was presented in terms of the proportion (%) of the sample assessed with each indicator, Cronbach's α and logistic regression in Table 4.3 and Figure 4.2. The results of the intra-observer study are presented in Table 4.2 and Figure 4.1.

Demeanour

Inter-observer assessments showed little difference in the proportion of sheep assessed with dull demeanour. Cronbach's α values ≥ 0.95 were found in both inter- and intra-observer test studies. As dull demeanour was only identified by observer combination A, it was not possible to explore this reliability result further.

Excessive panting

Excessive panting was not observed on any of the sample farms by any observer and therefore no α values or logistic regression model results could be provided.

Coughing

The inter-observer reliability study (Table 4.3) produced conflicting results regarding coughing (α 0.54 – 1.00), varying from poor agreement (combination A) to perfect agreement (combination D). A single animal recorded as coughing by one observer provided the poor agreement identified in observer combination A. Coughing was not observed during the intra-observer study so no reliability estimates were produced.

Skin irritation

Skin irritation was identified only by observer group F who achieved an α of 1.00. All other observer combinations (A – E) showed perfect agreement as they did not record any sheep with signs of skin irritation. Given the low proportion of sheep observed with signs of skin irritation no further results can be presented here.

Wool loss

Group assessment of wool loss, observed by the observer combination A, was identified with an α of 0.14. Logistic regression analysis identified that the assessments of observer 3 were significantly different to observer 1 (p 0.025).

Lameness

A range of α values were identified for the group assessment of lameness (0.76 – 1.00). Figure 4.1 illustrates that most observers counted a similar number of lame sheep and counts differed maximally by approximately 2 sheep. The assessments of the test standard observer were also interpreted as 'reliable', only changing slightly ($\leq$ 2 %) due to the observation of one additional lame animal (Table 4.3). However, further evaluation of inter-observer assessments using logistic regression and graphical distributions, identified considerable differences in the number of lame sheep assessed by observer combination A. This was attributed to the observations of observer 3 who appeared to count more lame sheep compared to observers 1 (p 0.001). Logistic regression analysis also identified that an inexperienced and untrained assessor in observer combination C (observer 5) significantly differed from the test standard (Table 4.3).

Dirty rear

All observer assessments produced Cronbach's α values from 0.96 to 1.00. With the exception of combination D and E, the proportion values suggested that between-observer assessments of rear cleanliness were highly consistent, for example, the proportion only ranged 19.25 – 22.68 % in observer combination A. In spite of this, logistic regression suggested that significant differences in group assessment of dirty rear occurred in combination A. In this combination, it was apparent that observer agreements were different on 15 out of 23 farms and were attributed to observer 3 (p 0.01), who identified fewer cases of dirty rears compared to observer 1. During the intra-observer study, the test standard observer assessments only changed slightly ($\leq$ 2 %) during the repeat assessment of lameness and dirty rear – this represented a difference in the count of one animal (Table 4.2).

Dirty belly

Dirty belly appeared to be 'reliable' according to the Cronbach's and co-workers (1972) interpretation of α (Table 4.3). As this indicator was only recorded by 2 observer combinations (A and F), further evaluation of test validity was not possible. No significant differences were identified by logistic regression analysis for most study farm assessments, although, observer 3 showed significant differences in the cleanliness assessment of the belly, when compared to the test standard observer (p 0.023).

Table 4.2 Intra-observer reliability of indicators assessed by group observation

Indicator	Cronbach's *α*	First % proportion (95 % CI)	Repeat % proportion (95 % CI)	Logistic regression (p-value)
Dull demeanour	1.00	1.14 (-2.76 – 5.03)	1.14 (-2.76 – 5.03)	[b]
Skin irritation	[a]	0	0	[b]
Excessive panting	[a]	0	0	[b]
Wool loss	1.00	2.27 (-4.66 – 9.20)	2.27 (-4.66 – 9.20)	[b]
Lameness	0.99	17.05 (-3.59 – 37.68)	15.91 (-1.10 – 32.92)	[b]
Coughing	[a]	0	0	[b]
Dirty rear	0.97	28.41 (-3.76 – 60.56)	27.27 (-2.90 – 57.44)	[b]
Dirty belly	[a]	0	0	[b]

[a] zero observations or insufficient observations for meaningful estimates

[b] no significant difference between observations ($p \geq 0.05$)

Table 4.3 Inter-observer reliability of indicators assessed by group observation

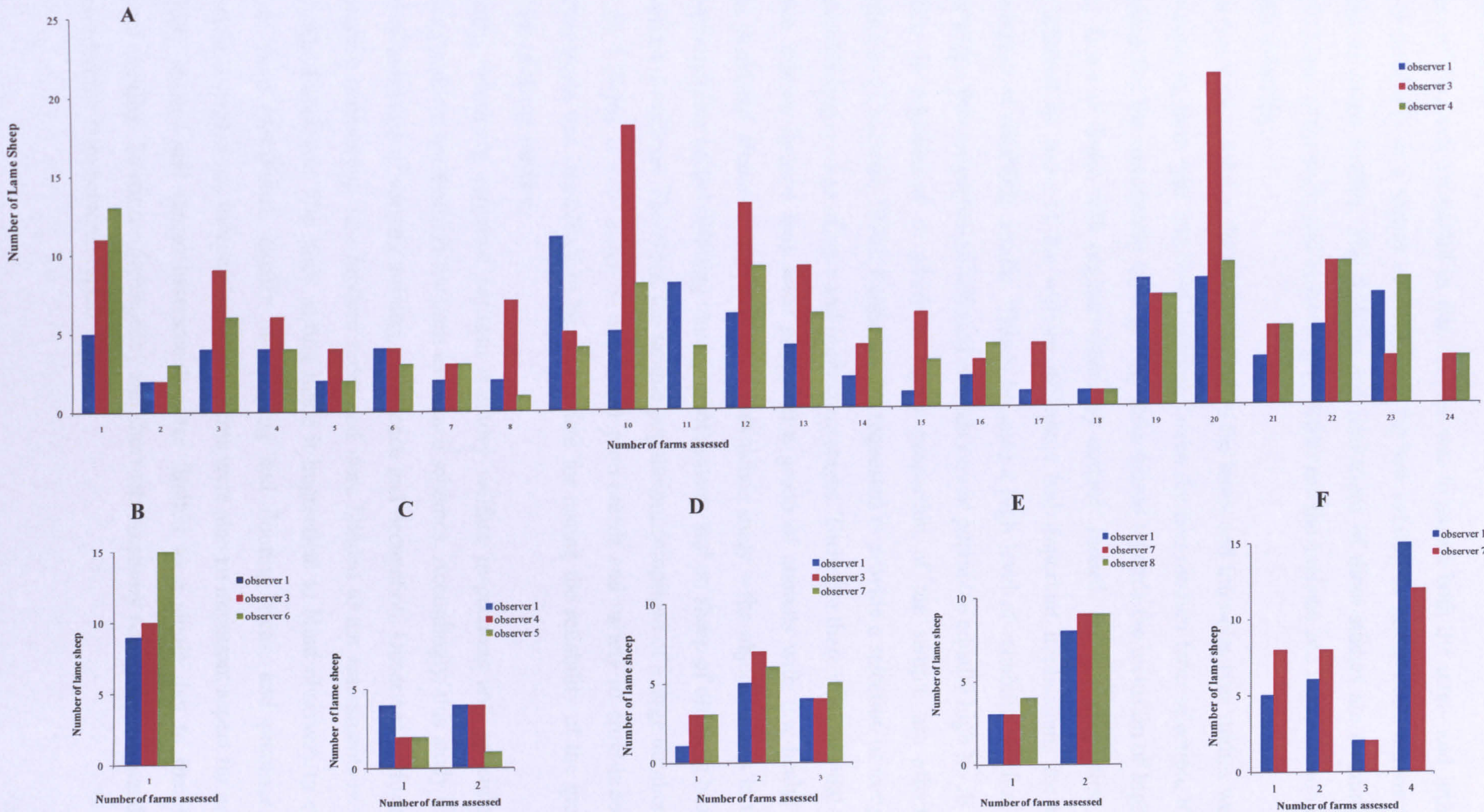
Indicator	Observer Combination	Test standard proportion[§] (%) (95 % CI)	Second observer proportion (%) (95 % CI)	Third observer proportion (%) (95 % CI)	Cronbach's α	Logistic regression (p-value)
Demeanour	A	0.13 (-0.06 – 0.31)	0.13 (-0.06 – 0.31)	0.19 (-0.03 – 0.41)	0.95	[b]
	B	0	0	0	[a]	[b]
	C	0	0	0	[a]	[a]
	D	0	0	0	[a]	[b]
	E	0	0	0	[a]	[b]
	F	0	0		[a]	[b]
Excessive panting	A	0	0	0	[a]	[b]
	B	0	0	0	[a]	[a]
	C	0	0	0	[a]	[a]
	D	0	0	0	[a]	[a]
	E	0	0	0	[a]	[a]
	F	0	0		[a]	[b]
Skin irritation	A	0.06 (-0.07 – 0.19)	0.00	0.00	[a]	[b]
	B	0	0	0	[a]	[a]
	C	0	0	0	[a]	[a]
	D	0	0	0	[a]	[a]
	E	0	0	0	[a]	[a]
	F	0.39 (-0.82 – 1.59)	0.39 (-0.82 – 1.59)		1.00	[b]
Wool loss	A	0.06 (-0.07 – 0.19)	0.18 (-0.1 – 0.38)	0.00	0.14	0.025
	B	0	0	0	[a]	[a]
	C	0	0	0	[a]	[a]
	D	0	0	0	[a]	[a]
	E	0	0	0	[a]	[a]
	F	0	0		[a]	[a]

Table 4.3 continued

Indicator	Observer Combination	Test standard proportion[§] (%) (95 % CI)	Second observer proportion (%) (95 % CI)	Third observer proportion (%) (95 % CI)	α	Logistic regression (p-value)
Lameness	A	6.06 (4.16 – 7.96)	9.81 (6.19 – 13.43)	7.50 (5.24 – 9.76)	0.76	0.001
	B	9.78 ([a])	10.87 ([a])	16.30 ([a])	[a]	[a]
	C	6.78 (0.94 – 12.62)	5.08 (-20.83 – 31.00)	2.54 (-6.04 – 11.12)	0.89	0.003
	D	4.74 (-1.80 – 11.28)	6.64 (2.27 – 11.00)	6.64 (2.27 – 11.00)	0.99	[b]
	E	9.02 (-14.89 – 32.92)	9.84 (-21.92 – 41.59)	10.66 (-8.13 – 29.44)	1.00	[b]
	F	10.81 (2.79 – 24.41)	11.58 (2.91 – 20.25)		0.94	[b]
Coughing	A	0.25 (-0.05 – 0.55)	0.00	0.06 (-0.07 – 0.18)	0.54	[b]
	B	0.00	0.00	0.00	[a]	[a]
	C	0.00	0.00	0.00	[a]	[a]
	D	0.47 (-1.66 – 2.61)	0.47 (-1.66 – 2.61)	0.47 (-1.66 – 2.61)	1.00	[b]
	E	0.00	0.00	0.00	[a]	[a]
	F	0.00	0.00		[a]	[a]
Dirty rear	A	22.68 (13.89 – 31.49)	19.25 (11.74 – 26.75)	21.06 (20.04 – 29.06)	0.97	0.01
	B	0	0	0	[a]	[a]
	C	0	0	0	[a]	[a]
	D	12.80 (-12.68 – 38.39)	12.79 (-7.06 – 32.65)	7.58 (-7.27 – 22.43)	0.98	[b]
	E	6.56 (14.61 – 27.73)	2.38 (-30.75 – 35.52)	4.10 (-35.17 – 43.37)	1.00	[b]
	F	8.88 (3.78 – 13.96)	7.33 (0.78 – 13.89)		0.98	[b]
Dirty belly	A	2.38 (-1.83 – 6.58)	1.43 (-1.34 – 4.19)	2.00 (-0.51 – 4.51)	0.98	0.04
	B	0	0	0	[a]	[a]
	C	0.00	3.39 (-36.76 – 43.54)	0.00	[a]	[a]
	D	0	0	0	[a]	[b]
	E	0	0	0	[a]	[b]
	F	3.09 (-6.14 – 12.32)	3.09 (-6.14 – 12.32)		1.00	[b]

[a] zero or insufficient observations for meaningful estimates, [b] no difference between observations ($p \geq 0.05$)

Figure 4.1 Comparison of group assessment of lameness by observer combination



4.4 Discussion

The aim of the study presented in this chapter was to assess both the inter- and intra-observer reliability as a means of evaluating the test validity of group observations as measures of sheep welfare. The findings and limitations of these studies are examined below in terms of the study and observer population and the analysis and interpretation of indicator reliability.

Factors that were crucial to farm recruitment to the inter- and intra-observer studies were farmer consent, farm type and farm location. Given the non-random farm selection, it is recognised that the recruitment strategy may have biased towards the inclusion of higher welfare farms or those with regular veterinary contact. Indeed, the low proportion of sheep affected by most of the welfare indicators had important implications for the interpretation of reliability results. This is because a high level of reliability elicited in studies with a low proportion of affected animals cannot guarantee equally high levels of reliability in populations in which a higher proportion of the sample are affected (Feinstein and Cicchetti, 1990). Farmers were requested to provide a selection of sample animals including a range of ages and production stages. Therefore there was potential for selection bias as farmers may have presented a group of animals with few health or welfare problems. However, this was not a prevalence study – the objective was to test observer reliability across differing management systems and on sheep of differing health and welfare conditions. Therefore, the sample population, comprised of a large number of sheep (n = 2406), diverse range of breeds (31 pure-breeds and variety of cross-breeds) and farm types was considered to be appropriate for testing the reliability of the group measures of sheep welfare.

Currently, veterinary surgeons perform statutory welfare inspections and agricultural assessors perform assessments for farm assurance schemes. Accordingly, this study used a pool of observers of varying training, experience and occupation. Observers performed independent assessments in a random order and were blinded to the assessments of the test standard observer. The study setting made it impractical to blind observers to cues such as farm cleanliness, quality of handling and flooring areas, and presence of equipment or medicines. Independent assessments were also an important aspect for intra-reliability studies and repeat assessments were limited to a single day to maintain indicator stability. To reduce dependency and short-term memory recall, the test standard observer left the farm between repeat visits.

In this study, the same observers did not perform all group assessments, so the provision of an overall α was not appropriate (Cronbach *et al.*, 1972). Instead, reliability of specific observer combinations produced several reliability coefficients (Cronbach *et al.*, 1972) which allowed the effect of different observers to be further scrutinised. An α value ≥ 0.70 was used as an initial cut-off for reliability (Nunnally, 1994), although higher values (≥ 0.90) may be more appropriate when applying indicators in formal welfare inspections. It was evident from the results of this study that Cronbach's α was not only affected by the low number of sheep observed with any of the conditions, but it was also influenced by the number of assessments made by each observer group. Since some indicators (demeanour, wool loss, skin irritation, coughing and dirty belly) were not identified on each farm, it is not possible to provide reliability values for all observer combinations. It was also apparent that Cronbach's approach required a minimum of 2 farm assessments, which is why B observer group provided insufficient observations for α analysis. Given the low proportion of sheep affected with any welfare condition and the varying number of observer assessments, alternative approaches to examining reliability were undertaken in this study.

A 'test standard' observer' (gold standard) (Burn *et al.*, 2009) was used to compare the overall % proportion assessed by each observer and allowed evaluation of agreement when α could not be calculated. Logistic regression also provided a means of assessing whether sample group assessments were affected by observer identity. Overall, both logistic regression and proportion results showed that most observer assessments only differed by a farm-level prevalence of approximately 2 %.

A possible explanation for the poor inter-observer agreement for the assessment of wool loss may be the practicality of examining and observing the entire surface area of each sheep from a distant observation point. Experience from this study suggests that close flocking of a large number of sheep restricted the ability to fully observe the degree of wool cover on each individual sheep of the group. In contrast, intra-observer testing identified that group assessment of wool loss was reliable. This may have been a reflection of the smaller sample assessed (range 20 – 22 sheep), or a different observational quality of the test standard observer.

Evidently, the ability to reliably count and record the number of sheep affected by each group indicator is also an important observer quality and skill. According to both α and proportion results, all observers reliably assessed group lameness. However, logistic regression identified that some disagreement with observer 3, in observer combination A.

Cronbach's α values did not appear to offer any additional useful information regarding observer reliability. Instead, graphical representation of the number of lame sheep assessed by each observer provided a rapid and clear means of examining observer differences. These study results identified that observer 3 appeared to show double-counting of the number of lame sheep. However, this appeared to be a temporary issue and did not occur in other observer combinations (D, E) and may have reflected data recording errors, differences in the quality of on-farm assessments conditions, or lack of experience in the assessment of lame sheep. The ability to accurately count lame sheep may be affected by the number of lame sheep in the group and the total group size assessed - factors that both require further elucidation. Categorical and continuous scoring of lame sheep has been associated with poor reliability (Harkins, 2005; Welsh *et al.*, 1993). More recently, good levels of reliability for individual lameness assessment of sheep were achieved by observers who assessed lameness using video-images of individual sheep (Kaler *et al.*, 2009). However, the use of video-footage is not comparable to assessments performed under farm conditions, and was not a feasible approach for assessing large numbers of sheep within a group situation (Chapter 3). Instead, a binary scoring system was developed that covered a range of lameness definitions including those of Kaler *et al.*, 2009. This simplified method may have been responsible for the reliability of on-farm lameness assessment found in this chapter.

In this study, ordinal cleanliness scoring systems of the abdominal and breech areas that covered a range of severity and types of contamination (mud and faeces) were found to be 'reliable' in both inter- and intra-observer studies, concurring with the findings of Napolitano *et al.*, (2009). Interestingly, the assessment of dirty rear produced good agreement even when over 20 % of sheep were observed with this indicator.

Overall, higher levels of reliability for all indicator assessments were achieved during the intra-observer studies of the test standard observer. As the test standard observer provided training and developed indicator protocols, they were expected to have been more familiar with the observational methods and scoring descriptors. The higher level of agreement may also have been attributed to the smaller sample size (n = 20 – 24) which permitted a closer examination of individual sheep. Previous work has identified that the training of observers can be important for attaining high levels of observer reliability (Kristensen *et al.*, 2006), therefore both trained and untrained assessors were included in this study. Although, experienced and trained observers (combinations E and F) achieved the highest degree of inter-observer agreement, this study did not provide sufficient data to determine a clear association between observer reliability and observer training and experience. Therefore, additional studies examining the effect of training and experience are needed to provide further evidence of the test validity of these observational measures of sheep welfare.

4.5 Conclusion

As well as potential inclusion in future statutory on-farm welfare assessments, indicators assessed by group observation could be feasibly applied by sheep farmers and veterinary surgeons as part of proactive flock health and welfare planning. The studies presented in this chapter suggested that group observation of individual sheep was a reliable test for the assessment of dirty rear and lameness in adult and growing sheep. Whilst dirty belly, demeanour and skin irritation appeared to show promising levels of reliability in both inter- and intra-observer studies, the low level of these conditions on study farms meant that the interpretation of reliability was limited. Additional work, examining observer reliability on farms with a higher proportion of sheep affected by these conditions, would be useful in enhancing the interpretation of indicator reliability reported here. The effect of the sample size, farm conditions and effect of observer training and prior experience also requires elucidation.

Chapter 5

VALIDATING INDICATORS OF SHEEP WELFARE ASSESSED USING AN INDIVIDUAL EXAMINATION

5.1 Introduction

Stockpeople and veterinary surgeons frequently assess the behaviour and physical appearance of individual sheep as a means of judging the welfare and productivity of the flock. Some of the animal-based measures used, for example body condition or dental condition, cannot be assessed by observing a group of sheep from a distance. As a result, groups of sheep may be gathered and handled to facilitate a closer inspection and physical examination of individual animals at various stages of the production cycle.

Welfare indicators applied during a physical examination of individual sheep are used to diagnose the presence or absence of a particular welfare condition. Therefore, the principles used to evaluate diagnostic tests are relevant to the validation of indicators of sheep welfare developed by this thesis. Ideally the performance of a diagnostic test would be examined against a reference standard or 'gold standard' (Greiner and Gardner, 2000). However, there is currently no reference standard for animal welfare assessment (de Passillé and Rushen, 2005). Therefore, in line with previous research (Burn *et al.*, 2009), the approach taken in this chapter was to assume that the assessment findings of an experienced observer were the 'gold standard' or reference result.

A valid welfare indicator needs to be consistently applied i.e. reliable when applied by observers of differing occupations with varying levels of training and experience. In addition, an animal welfare indicator needs to have good levels of diagnostic sensitivity (Se) – be capable of correctly identifying sheep with a particular welfare condition, and diagnostic specificity (Sp) – be capable of correctly identifying sheep without a particular welfare condition. Therefore, the objective of this chapter was to investigate the level of observer agreement, and the diagnostic Se and Sp of a range of assessors as a means of evaluating the test validity of indicators assessed by an individual sheep examination.

5.2 Materials and Methods

5.2.1 Study population

A study population of 38 farms, recruited as described in Chapter 3, comprising lowland (n = 16), upland (n = 11) and hill (n = 11) flocks was asked to provide a sample of 70 sheep including adult sheep (aged > 1 year) and growing lambs (aged > 6 weeks < 1 year). A sample of sheep from each farm was assessed by 8 observers (observers 1 to 8), described in Chapter 3, according to their availability during July to December 2009.

5.2.2 Welfare indicator assessments

Each individual indicator test (n = 29) was designed to identify specific welfare concerns for growing lambs and adult sheep (Chapter 2). Following a group observation and assessment of 8 group indicators sample (Chapter 4), the sample of sheep was gathered to a holding area and each sheep was individually examined using 29 indicator tests described in Chapter 3. Each individual sheep was assigned a unique numeric identification, given by the order they moved into the assessment pen.

Inter-observer assessments

A group of observers assessed a total of 1146 adult sheep and growing lambs during July to November 2008. Observer 1 (the author) was designated the 'test standard observer'. The assessment findings of this observer were used as the reference standard against which the test performance of other observers was compared. The test standard observer performed assessments on all 1146 sheep but it was not feasible for observers 2 to 8 to participate in all study farm visits. Therefore, inter-observer assessments were performed by a varied group of 2 to 3 observers (Table 5.1). Each observer independently assessed 30 sheep which were selected using a pre-determined random number system described in Altman (1994).

Intra-observer assessments

For the intra-observer study, 88 sheep from 4 commercial lowland flocks were individually examined by the test standard observer, twice within a 24 hour period during November to December 2009. A 5 hour delay between repeat assessments was selected to minimise any alteration in indicator outcomes. Once the first assessment was finished, all recording sheets remained sealed in an envelope and were not examined until the end of the study. Following completion of the first visit the observer left the farm and performed unrelated tasks in order to reduce any memory recall.

Table 5.1 Observer population for indicators assessed by individual exam

Observer identity	n farms	n sheep
1, 3, 4	24	720
1, 3, 6	1	30
1, 4, 5	2	60
1, 3, 7	3	90
1, 7, 8	2	60
1, 7	4	120
1, 2, 3	3	90

5.2.3 Statistical analysis

Individual welfare indicator scoring scales consisted of categorical, ordinal and binary data (Chapter 3). Data was analysed using Minitab version 15.1 (Minitab, Inc, State College, PA) and Stata version 10 (StataCorp LP, College Station, Texas).

The overall level of inter-observer reliability of multiple observer assessments ($n > 2$) was determined by Fleiss's kappa (κ) (Fleiss, 1981). The level of paired agreement between the test standard observer and each observer was examined in terms of the percentage (%) of agreement and Cohen's κ (Cohen, 1960). All κ results were interpreted according to Fleiss (1981), whereby values ≥ 0.75 suggested excellent levels of agreement, κ of 0.40 - 0.75 indicated fair to good agreement, and $\kappa \leq 0.40$ was poor agreement. Kendall's coefficient of concordance, also known as Kendall's W (Kendall and Smith, 1939), was also used to evaluate the degree of inter- and intra-observer agreement of ordinal scoring indicators (tooth disease, cleanliness of the legs, belly and rear, mastitis, body condition score, injury and wounds and myiasis).

The Se and Sp of each observer were determined using a classical approach and latent class analysis (LCA) according to the type of study. Both methods evaluate the Se and Sp of binary tests, and so categorical and ordinal indicators were dichotomised. The classical approach, using cross-classification table, was used to provide the Se and Sp (with 95 % confidence intervals) of the repeat assessments of the test standard observer. In this approach, the results of the first assessment were assumed to be the reference result. LCA was used to estimate the Se and Sp of each observer participating in the inter-observer study. LCA was performed using a random-effects model in OpenBUGS software (Lunn *et al.*, 2009). Markov Chain Monte Carlo (MCMV) sampling was used to obtain the joint posterior distribution of the model. The random effects model assumed that observers were selected from a normal distribution. The first 10,000 samples were discarded as burn-in and the subsequent 10,000 iterations were used for posterior inference (Toft *et al.*, 2007b). Visual assessment of time-series plots and Gelman-Rubin diagnostic plots were used to assess MCMC chain convergence (Toft *et al.*, 2007b). The Se and Sp of each observer were provided with 95 % posterior credibility intervals (PCI) - the Bayesian analogue of confidence intervals (Bonde *et al.*, 2010). In addition, LCA used the Se and Sp results of the inter-observer study to predict the Se and Sp of 'new' observers – randomly selected observers who may be expected to apply the welfare indicators in the future.

The difference between the scores of each observer and the test standard observer was graphically represented to identify any evidence of scoring bias. The difference between the first (reference result) and second assessment of the intra-observer study were examined in the same way.

5.3 Results

The diagnostic performance of each indicator, in terms of the level of observer agreement (reliability), Se and Sp is described in turn below. Table 5.2 presents the overall level of inter-observer reliability as determined by Fleiss's κ. The agreement between each observer (2 – 8) and the test standard was then compared in terms of % agreement (Table 5.3), Cohen's κ (Table 5.4) and Kendall's W (Table 5.5). The test performance of the test standard observer is described in Table 6. Graphical representations of the differences in observer scoring are provided in Figures 5.1 and 5.2. The diagnostic Se and Sp of each observer of the inter-observer study and the predicted performance of 'new' observers was determined by LCA (Table 5.6). The Se and Sp of the test standard was analysed by a classical approach to test evaluation (Table 5.7).

Demeanour

Overall, a Fleiss's κ of 0.85 was found although paired assessments of the test standard and each observer found that κ values ranged from 0.40 – 1.00. Se and Sp ≥ 0.98 were identified for all observers of the inter-observer study but repeat assessments of the test standard observer produced lower κ (0.37) and Se (0.50).

Eye condition

Percentage agreement was over 96.00 % and Fleiss's κ was 0.54, although κ agreement with the test standard ranged from 0.00 (no agreement) to 0.89 ('excellent' agreement). Observers could clearly identify animals without an eye condition (Sp 1.00) but were less sensitive at identifying those with an eye abnormality (Se ≥ 0.62).

Nasal discharge

Overall, the inter-observer study found that observers were reliable (Table 5.2) but there were insufficient observations during the intra-observer study to provide any meaningful test results

Tooth disease

Both inter- and intra-observer studies produced κ ≥ 0.50, although, there was variation in observer reliability (78.33 – 96.67 % and κ 0.31 – 0.65) Kendall's W suggested that observers could identify the presence or absence of a tooth abnormality (0.71 – 0.85) but it was clear that some observers had difficulty with the assessment of molar abnormalities (Se 0.37 – 0.86).

Coughing

Coughing achieved a Fleiss κ of 0.63, although only 3 assessors recorded sufficient assessments (Tables 5.3 and 5.4). Whereas, poor levels of observer agreement (κ 0.31) and Se (0.33) were found for the repeat assessments of the test standard observer.

Ear lesions

Fleiss's κ (0.66) indicated that, overall, observers were reliable, although Table 5.4 clearly identified that there was observer variation in the agreement of ear lesion assessments.

In-growing horns

No observations of in-growing horns were made during the inter- and intra-observer studies, therefore no results of the reliability, Se and Sp of this indicator can be provided.

Dirty belly

Whilst observer 5 showed high levels of agreement with the test standard observer (κ 0.97) other assessors were not as reliable at scoring belly cleanliness (Tables 5.4 and 5.7). Kendall's W results suggest that most scoring discrepancies occurred between 'dirty' and 'filthy' scores (Table 5.6).

Dirty legs

Reliability, Se and Sp results showed that cleanliness assessments of the legs proved to be inconsistent across all assessors (Tables 5.4, 5.5 and 5.7).

Dirty rear

A Fleiss κ of 0.36 was found and paired agreement with the test standard varied from < 0.00 (observer 5) to 0.77 (Table 5.4). Kendall's W suggests that scoring differences were limited to disagreement between 'dirty' and 'filthy' scores (Table 5.5).

Mastitis

Both the test standard and observer 7 appeared to be reliable (98.44 %, κ 0.79) and specific (Se 1.00) when performing mastitis assessments. However, this level of test performance was not maintained by other observers (Table 5.4).

Crystals

No observations of crystals (visible uroliths) were made during either the inter-observer or intra-observer study, therefore no results of the reliability, Se or Sp of this indicator can be provided.

Tail length

Tail length was a reliable indicator during inter-observer studies (Fleiss $\kappa \geq 0.70$) and, with the exception of observer 6, all observers achieved Se $\geq$ 0.82 and Se $\geq$ 0.99.

Wool loss

Assessment of wool loss in individual sheep produced a Fleiss's κ of 0.80, which was corroborated by consistently high estimates of Se and Sp (Table 5.6).

Skin irritation

Few observers recorded the presence of pruritus (Table 5.3) and all observers of the inter-observer study were identified with lower levels of Se (0.13 – 0.40).

Skin lesion

Both the inter- and intra-observer studies provided κ 0.42 but there were considerable differences in observer reliability ranging from κ < 0.00 – 0.88. Interestingly, LCA identified that observers had higher Se for lesions under 50p size but lower Se for the assessment of diffuse lesions (Table 5.6).

Injuries and wounds

Fleiss's κ 0.38 ('poor' agreement) was found. Only sufficient observations of the 'healing wound' score were made and all observers produced a Sp of 1.00. With the exception of the intra-observer study, Se estimates were very low (0.06 – 0.37) for the assessment of healing wounds.

Body condition scoring

Overall, a Fleiss's κ of 0.46 was found although there was variation in observer assessments (Table 5.4). Figures 5.1 and 5.2 suggested that most disagreements followed a normal distribution and occurred between a single score. By contrast, observer 6 consistently assessed body condition as being one unit higher than the test standard observer (Figure 5.1). Se and Sp is described under Fit-Fat-Thin.

Fit-Fat-Thin

This broader system of body condition assessment provided higher levels of inter- (κ 0.63) and intra-observer agreement (κ 0.87). Interestingly, observer 7 agreed with the test standard in 279 out of 280 assessments (99.26 %) but achieved a κ of 0.00. Most observers could reliably assess sheep as being 'fit' (BCS 2 – 4) (Tables 5.7 and 5.8), whereas the 'thin' score (BCS < 2) had much lower Se.

Lameness

Overall a good level of reliability, Se and Sp were produced for the gait assessment of individual sheep (Tables 5.4 – 5.8).

Foot lesion and significant foot lesion

Se and Sp levels ≥ 0.73 were achieved for diagnosing the presence or absence of any foot lesion. With the exception of observer 5, judgements of lesion 'significance' provided high Se and Sp.

Specific foot lesions

Inter- and intra- observer agreement for the identification of CODD achieved κ values $\geq$ 0.70. Good levels of observer agreement (Table 5.4) and relatively high Se and Sp (Tables 5.6 and 5.8) were also found for WL, inter-digital dermatitis and footrot.

Joint swelling

This indicator produced Sp ≥ 0.99 and $\kappa \geq 0.77$ in both inter- and intra-observer studies.

Myiasis

Although a Fleiss $\kappa \geq 0.70$ was identified, there were very few observations of myiasis in the inter-observer study and no sheep with evidence of myiasis were observed during the intra-observer study.

Table 5.2 Overall level of inter-observer reliability determined by Fleiss's κ

Indicator	κ	95 % CI	Interpretation
Demeanour	0.85	0.76 - 0.92	Excellent
Eye condition	0.54	0.39 - 0.62	Good-fair
Nasal discharge	0.54	0.39 - 0.62	Good-fair
Tooth disease	0.50	0.47 - 0.53	Good-fair
Coughing	0.63	0.31 - 0.83	Good-fair
Ear lesion	0.66	0.62 - 0.68	Good-fair
Dirty belly	0.62	0.62 - 0.63	Good-fair
Dirty legs	0.43	0.39 - 0.46	Good-fair
Dirty rear	0.36	0.35 - 0.39	Poor
Mastitis	0.44	0.42 - 0.45	Good-fair
Tail length	0.71	0.60 - 0.75	Good-fair
Wool loss	0.80	0.59 - 1.00	Excellent
Skin irritation	0.46	0.44 - 0.57	Good-fair
Skin lesion	0.42	0.28 - 0.48	Good-fair
Injuries & wounds	0.38	0.24 - 0.52	Poor
Body condition score	0.46	0.43 - 0.48	Good-fair
Fit-Fat-Thin	0.63	[a]	Good-fair
Lameness	0.66	0.61 - 0.68	Good-fair
Foot lesion	0.48	0.45 - 0.53	Good-fair
Significant foot lesion	0.56	0.54 - 0.60	Good-fair
White line	0.47	0.46 - 0.47	Good-fair
Inter-digital dermatitis	0.49	0.35 - 0.63	Good-fair
Foot rot	0.49	0.37 - 0.65	Good-fair
CODD	0.72	0.71 - 0.77	Good-fair
Other foot lesion	0.60	0.47 - 0.68	Good-fair
Joint swelling	0.77	0.73 - 0.80	Excellent
Myiasis	0.77	0.53 - 1.00	Excellent

[a] Insufficient observations to provide meaningful estimate

Table 5.3 Percentage (%) agreement with the test standard observer

Indicator	Percentage (%) agreement by observer identity						
	2	3	4	5	6	7	8
Demeanour	[a]	99.67	99.74	[a]	[a]	100.00	95.64
Eye condition	[a]	99.56	99.35	98.33	[a]	99.63	96.67
Nasal discharge	[a]	99.56	99.35	98.33	[a]	99.63	96.67
Tooth disease	87.21	86.09	87.53	78.33	96.67	86.62	90.00
Coughing	[a]	99.67	99.61	100.00	[a]	[a]	[a]
Ear lesion	94.19	83.90	88.72	91.67	93.33	92.94	98.33
Dirty belly	87.21	91.07	86.64	98.33	[a]	87.73	80.00
Dirty legs	83.72	96.03	92.87	50.00	[a]	87.36	93.33
Dirty rear	75.58	67.48	77.17	38.33	80.00	87.36	95.00
Mastitis	97.30	93.95	95.00	73.33	93.33	99.23	[a]
Tail length	97.67	97.46	97.80	93.33	96.67	100.00	[a]
Wool loss	[a]	99.78	100.00	[a]	[a]	100.00	[a]
Skin irritation	90.70	98.79	[a]	[a]	[a]	[a]	[a]
Skin lesion	88.37	94.16	96.50	96.67	90.00	97.03	98.33
Injuries & wounds	97.67	97.57	97.28	96.67	[a]	99.26	100.00
Body condition score	60.47	64.50	68.22	51.67	40.00	78.07	63.33
Fit-Fat-Thin	93.02	97.13	95.59	85.00	66.67	99.26	100.00
Lameness	94.19	95.48	95.59	76.67	96.67	95.54	100.00
Foot lesion	88.37	72.88	74.19	65.00	63.33	85.87	91.67
Significant foot lesion	93.02	93.05	95.20	86.67	100.00	98.14	96.67
White Line	93.02	71.11	72.63	68.33	66.67	78.07	85.00
Inter-digital dermatitis	97.67	98.13	99.35	96.67	[a]	99.63	[a]
Foot rot	98.84	98.90	98.96	98.33	96.67	98.88	[a]
CODD	100.0	99.12	99.48	100.00	[a]	97.03	100.00
Other foot lesion	98.84	99.23	98.57	91.67	96.67	97.77	100.00
Joint Swelling	[a]	99.56	99.61	[a]	[a]	99.26	100.00
Myiasis	97.67	100.00	100.00	[a]	[a]	[a]	[a]

[a] Insufficient observations to provide meaningful estimate

Table 5.4 Inter-observer reliability with the test standard determined by Cohen's κ

Indicator	κ by observer identity						
	2	3	4	5	6	7	8
Demeanour	[a]	0.40	0.50	[a]	[a]	1.00	1.00
Eye condition	[a]	0.60	0.28	0.00	[a]	0.89	0.48
Nasal discharge	[a]	0.60	0.28	0.00	[a]	0.89	0.48
Tooth disease	0.31	0.49	0.51	0.44	0.65	0.64	0.50
Coughing	[a]	0.66	0.57	1.00	[a]	[a]	[a]
Ear lesions	0.75	0.34	0.28	0.38	-0.03	0.52	0.00
Dirty belly	0.13	0.60	0.53	0.97	[a]	0.72	0.54
Dirty legs	-0.05	0.15	0.11	0.02	[a]	0.65	0.86
Dirty rear	0.32	0.36	0.51	-0.02	0.63	0.49	0.77
Mastitis	0.49	0.30	0.61	0.41	-0.03	0.83	[a]
Tail length	0.79	0.72	0.80	0.31	0.00	1.00	[a]
Wool loss	[a]	0.66	1.00	[a]	[a]	1.00	[a]
Skin irritation	0.50	0.47	[a]	[a]	[a]	[a]	[a]
Skin lesion	0.13	0.36	0.58	0.00	-0.02	0.49	0.88
Injuries and wounds	0.33	0.21	0.39	0.00	[a]	0.66	1.00
Body condition	0.45	0.44	0.52	0.27	-0.08	0.57	0.29
Fit-Fat-Thin	0.82	0.74	0.65	0.38	0.00	0.00	1.00
Lameness	0.41	0.66	0.69	0.17	0.00	0.77	1.00
Foot lesion	0.57	0.46	0.48	0.21	0.26	0.65	0.83
Significant foot lesion	0.22	0.47	0.61	0.17	1.00	0.89	0.86
White line	0.63	0.42	0.46	0.28	0.34	0.53	0.70
Inter-digital dermatitis	-0.01	0.25	0.73	0.78	[a]	0.66	[a]
Foot rot	0.00	0.58	0.55	0.00	0.00	0.39	[a]
CODD	1.00	0.55	0.75	1.00	[a]	0.68	1.00
Other foot lesion	0.66	0.63	0.47	0.56	0.00	0.72	1.00
Joint swelling	[a]	0.50	0.72	[a]	[a]	0.85	1.00
Myiasis	-0.01	1.00	1.00	[a]	[a]	[a]	[a]

[a] Insufficient observations to provide meaningful estimate

Table 5.5 Inter-observer reliability of ordinal scoring indicators

Indicator	Kendall's W by observer identity						
	2	3	4	5	6	7	8
Tooth disease	0.71	0.80	0.79	0.85	[a]	0.84	0.78
Dirty belly	[a]	0.84	0.81	0.99	[a]	[a]	[a]
Dirty legs	0.47	0.58	0.58	0.50	[a]	0.83	[a]
Dirty rear	0.73	0.76	0.79	0.44	0.86	0.78	0.91
Mastitis	[a]	0.66	0.85	0.79	[a]	0.91	[a]
Skin lesion	0.64	0.71	0.80	0.50	0.47	0.74	0.94
Body condition	0.86	0.80	0.86	0.79	0.67	0.83	0.71
Myiasis	[a]	1.00	1.00	[a]	[a]	[a]	[a]

[a] Insufficient observations to provide meaningful estimate

Figure 5.1 Differences in observer scores of sheep body condition

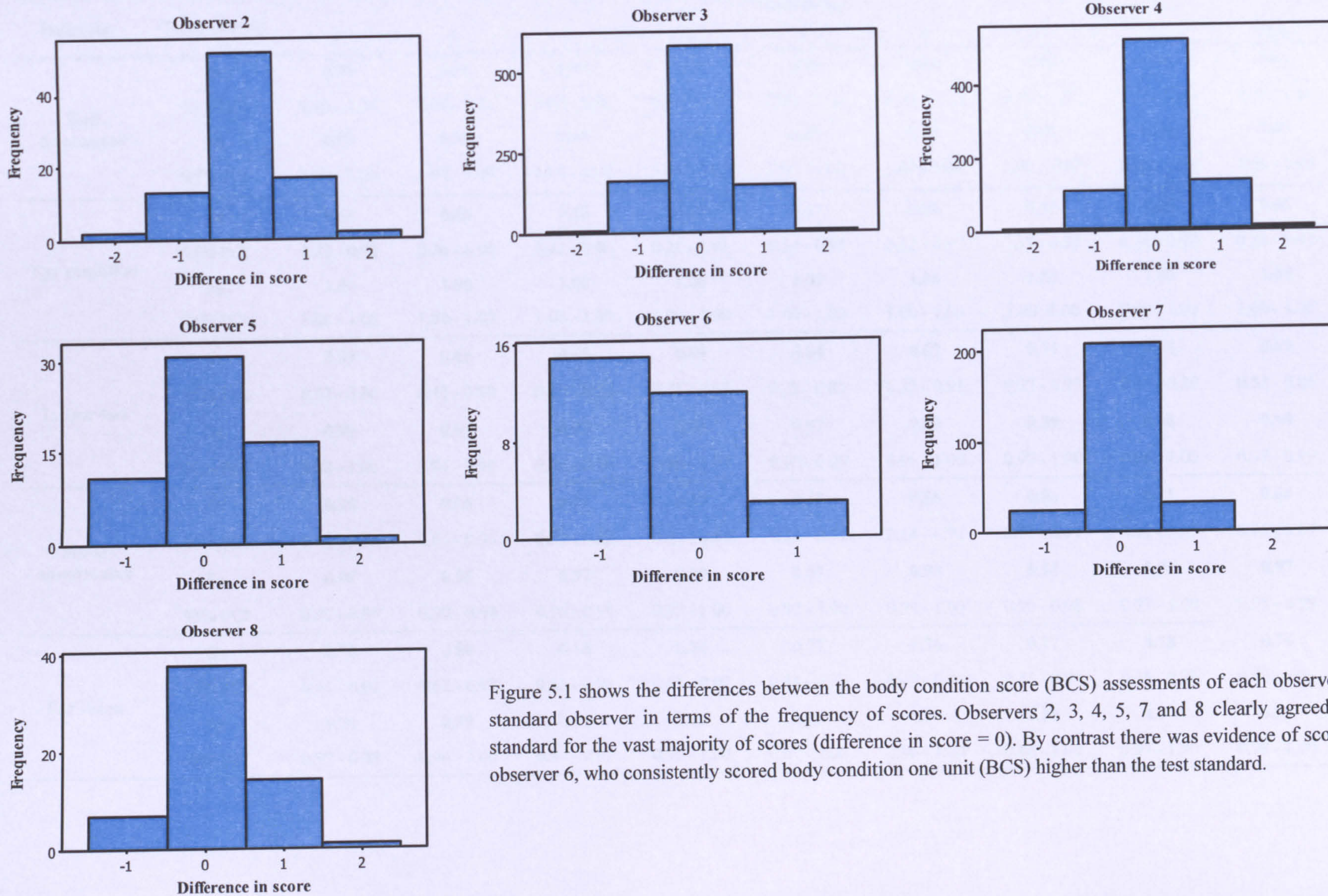


Figure 5.1 shows the differences between the body condition score (BCS) assessments of each observer and the test standard observer in terms of the frequency of scores. Observers 2, 3, 4, 5, 7 and 8 clearly agreed with the test standard for the vast majority of scores (difference in score = 0). By contrast there was evidence of scoring bias with observer 6, who consistently scored body condition one unit (BCS) higher than the test standard.

Table 5.6 Latent class evaluation of the Se and Sp of each individual sheep welfare indicator determined during the inter-observer study

Indicator	Test estimate	Observer								
		1	2	3	4	5	6	7	8	new
Dull demeanour	**Se**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**	**0.99**	**0.99**
	95% PCI	0.96 - 1.00	0.92 - 1.00	0.88 - 1.00	0.89 - 1.00	0.92 - 1.00	0.78 - 1.00	0.95 - 1.00	0.95 - 1.00	0.96 - 1.00
	Sp	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**	**0.99**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Eye condition	**Se**	**0.64**	**0.65**	**0.69**	**0.62**	**0.67**	**0.66**	**0.67**	**0.65**	**0.66**
	95% PCI	0.33 - 0.92	0.26 - 0.95	0.41 - 0.96	0.25 - 0.91	0.34 - 0.93	0.32 - 0.95	0.37 - 0.92	0.34 - 0.92	0.38 - 0.92
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 -1.00	0.99 - 1.00	1.00 - 1.00
Incisor loss	**Se**	**0.69**	**0.68**	**0.65**	**0.69**	**0.64**	**0.67**	**0.74**	**0.68**	**0.69**
	95% PCI	0.52 - 0.86	0.43 - 0.90	0.45 - 0.82	0.52 - 0.87	0.28 - 0.85	0.32 - 0.91	0.57 - 0.97	0.44 - 0.89	0.52 - 0.86
	Sp	**0.99**	**0.97**	**0.98**	**0.99**	**0.97**	**0.99**	**0.99**	**0.98**	**0.98**
	95% PCI	0.98 - 1.00	0.94 - 0.99	0.96 - 0.99	0.98 - 0.99	0.90 - 0.99	0.96 - 1.00	0.97 - 1.00	0.96 - 1.00	0.97 - 0.99
Molar abnormality	**Se**	**0.75**	**0.55**	**0.71**	**0.49**	**0.37**	**0.58**	**0.86**	**0.71**	**0.64**
	95% PCI	0.63 - 0.86	0.06 - 0.96	0.59 - 0.82	0.37 - 0.63	0.12 - 0.71	0.14 - 0.92	0.71 - 0.97	0.30 - 0.99	0.35 - 0.87
	Sp	**0.98**	**0.95**	**0.97**	**0.99**	**0.97**	**0.98**	**0.94**	**0.97**	**0.97**
	95% PCI	0.97 - 0.99	0.90 - 0.99	0.96 - 0.99	0.97 - 1.00	0.92 - 1.00	0.94 - 1.00	0.90 - 0.98	0.93 - 1.00	0.95 - 0.99
Ear lesion	**Se**	**0.78**	**0.80**	**0.78**	**0.74**	**0.77**	**0.76**	**0.77**	**0.78**	**0.78**
	95% PCI	0.61 - 0.92	0.62 - 0.97	0.61 - 0.93	0.45 - 0.92	0.51 - 0.95	0.49 - 0.94	0.58 - 0.94	0.55 - 0.96	0.60 - 0.93
	Sp	**0.98**	**0.99**	**0.95**	**1.00**	**0.99**	**0.99**	**1.00**	**0.99**	**0.99**
	95% PCI	0.97 - 0.99	0.96 - 1.00	0.94 - 0.97	0.99 - 1.00	0.97 - 1.00	0.96 - 1.00	0.98 - 1.00	0.97 - 1.00	0.98 - 1.00

Table 5.6 continued

Indicator	Test estimate	Observer 1	2	3	4	5	6	7	8	new
Dirty belly: mud score	**Se**	**0.85**	**0.75**	**0.74**	**0.67**	**0.91**	**0.79**	**0.85**	**0.85**	**0.82**
	95% PCI	0.91 - 0.78	0.97- 0.24	0.87 - 0.61	0.78 - 0.55	0.99 - 0.81	0.98 - 0.35	0.93 - 0.76	0.94 - 0.74	0.65 - 0.92
	Sp	**0.98**	**0.99**	**0.98**	**0.95**	**0.98**	**0.99**	**0.91**	**0.78**	**0.97**
	95% PCI	0.78 - 0.91	0.24 - 0.97	0.61 - 0.87	0.55 - 0.78	0.81 - 0.99	0.35 - 0.98	0.76 - 0.93	0.74 - 0.94	0.91 - 1.00
Dirty legs: mud score	**Se**	**0.61**	**0.48**	**0.20**	**0.15**	**0.78**	**0.46**	**0.84**	**0.94**	**0.60**
	95% PCI	0.38 - 0.87	0.01 - 1.00	0.06 - 0.44	0.02 - 0.37	0.19 - 1.00	0.00 - 1.00	0.73 - 0.93	0.82 - 1.00	0.12 - 0.97
	Sp	**1.00**	**0.99**	**0.99**	**0.96**	**0.53**	**0.99**	**0.92**	**0.98**	**0.98**
	95% PCI	1.00 - 1.00	0.97 - 1.00	0.98 - 1.00	0.94 - 0.98	0.40 - 0.67	0.96 - 1.00	0.88 - 0.97	0.90 - 1.00	0.89 - 1.00
Dirty legs: faecal score	**Se**	**0.33**	**0.15**	**0.14**	**0.18**	**0.18**	**0.22**	**0.22**	**0.20**	**0.17**
	95% PCI	0.08 - 0.83	0.00 - 0.43	0.02 - 0.33	0.04 - 0.42	0.00 - 0.61	0.01 - 0.93	0.02 - 0.81	0.00 - 0.76	0.02 - 0.45
	Sp	**1.00**	**0.88**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	0.99 - 1.00	0.80 - 0.94	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	0.99 - 1.00	1.00 - 1.00	0.99 - 1.00	1.00 - 1.00
Dirty rear: mud score	**Se**	**0.54**	**0.16**	**0.98**	**0.62**	**0.81**	**0.70**	**0.95**	**0.91**	**0.82**
	95% PCI	0.34 - 0.75	0.00 - 0.79	0.87 - 1.00	0.40 - 0.83	0.00 - 1.00	0.00 - 1.00	0.70 - 1.00	0.47 - 1.00	0.02 - 1.00
	Sp	**0.99**	**0.99**	**0.84**	**0.97**	**0.54**	**0.99**	**0.94**	**0.99**	**0.96**
	95% PCI	0.99 - 1.00	0.96 - 1.00	0.81 -0.87	0.96 - 0.99	0.41 - 0.66	0.94 - 1.00	0.90 - 0.97	0.97 - 1.00	0.86 - 1.00
Dirty rear: faecal score	**Se**	**0.83**	**0.82**	**0.55**	**0.81**	**0.58**	**0.79**	**0.76**	**0.82**	**0.76**
	95% PCI	0.75 - 0.90	0.60 - 0.99	0.47 - 0.63	0.72 - 0.89	0.16 - 0.89	0.57 - 0.96	0.56 - 0.93	0.56 - 0.99	0.59 - 0.92
	Sp	**0.96**	**0.92**	**0.98**	**0.91**	**0.96**	**0.94**	**0.98**	**0.97**	**0.96**
	95% PCI	0.94 - 0.98	0.84 - 0.97	0.97 - 0.99	0.88 - 0.94	0.91 - 0.99	0.83 - 1.00	0.95 - 1.00	0.93 - 1.00	0.93 - 0.99

Table 5.6 continued

Indicator	Test estimate	Observer								
		1	2	3	4	5	6	7	8	new
Tail length	Se	**0.88**	**0.88**	**0.76**	**0.89**	**0.84**	**0.71**	**0.87**	**0.82**	**0.85**
	95% PCI	0.77 - 0.97	0.67 - 1.00	0.61 - 0.88	0.77 - 0.97	0.45 - 1.00	0.06 - 0.95	0.66 - 1.00	0.33 - 0.99	0.67 - 0.97
	Sp	**1.00**	**0.99**	**0.99**	**0.99**	**0.95**	**0.99**	**1.00**	**1.00**	**0.99**
	95% PCI	0.99 - 1.00	0.97 - 1.00	0.99 - 1.00	0.98 - 1.00	0.88 - 0.99	0.97 - 1.00	0.99 - 1.00	0.98 - 1.00	0.98 - 1.00
Mastitis	Se	**0.67**	**0.53**	**0.50**	**0.72**	**0.62**	**0.57**	**0.58**	**0.57**	**0.61**
	95% PCI	0.50 - 0.81	0.18 - 0.81	0.33 - 0.71	0.53 - 0.88	0.46 - 0.81	0.22 - 0.85	0.30 - 0.82	0.15 - 0.87	0.38 - 0.79
	Sp	**1.00**	**1.00**	**0.98**	**1.00**	**0.98**	**0.99**	**1.00**	**1.00**	**1.00**
	95% PCI	0.99 - 1.00	0.98 - 1.00	0.97 - 1.00	0.99 - 1.00	0.90 - 1.00	0.95 - 1.00	0.99 - 1.00	0.98 - 1.00	0.98 - 1.00
Wool loss	Se	**1.00**	**0.97**	**1.00**	**1.00**	**0.97**	**0.98**	**1.00**	**0.97**	**0.99**
	95% PCI	0.99 - 1.00	0.00 - 1.00	0.99 - 1.00	0.99 - 1.00	0.16 - 1.00	0.95 - 1.00	0.99 - 1.00	0.00 - 1.00	0.99 - 1.00
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	0.99 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Skin irritation	Se	**0.29**	**0.50**	**0.40**	**0.13**	**0.25**	**0.28**	**0.19**	**0.25**	**0.24**
	95% PCI	0.09 - 0.61	0.27 - 0.71	0.17 - 0.69	0.00 - 0.51	0.00 - 0.76	0.00 - 0.90	0.00 - 0.64	0.00 - 0.80	0.00 - 0.59
	Sp	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	0.93 - 1.00	0.99 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Skin lesion: ≤ 50p size	Se	**0.78**	**0.56**	**0.70**	**0.80**	**0.71**	**0.72**	**0.76**	**0.77**	**0.75**
	95% PCI	0.50 - 1.00	0.00 - 0.95	0.41 - 0.94	0.54 - 1.00	0.06 - 1.00	0.03 - 1.00	0.43 - 1.00	0.42 - 1.00	0.45 - 1.00
	Sp	**1.00**	**1.00**	**0.98**	**1.00**	**0.99**	**0.97**	**0.99**	**0.99**	**0.99**
	95% PCI	0.99 - 1.00	0.98 – 1.00	0.97 - 0.99	0.99 - 1.00	0.98 - 1.00	0.87 - 1.00	0.97 - 1.00	0.98 - 1.00	0.98 - 1.00

Table 5.6 continued

Indicator	Test estimate	Observer								
		1	2	3	4	5	6	7	8	new
Skin lesion: ≤ hand-size	Se	**0.53**	**0.51**	**0.52**	**0.54**	**0.51**	**0.51**	**0.54**	**0.55**	**0.53**
	95% PCI	0.27 - 0.80	0.15 - 0.82	0.27 - 0.80	0.29 - 0.80	0.16 - 0.82	0.15 - 0.81	0.26 - 0.86	0.27 - 0.88	0.26 - 0.80
	Sp	**1.00**	**0.97**	**0.99**	**0.99**	**1.00**	**0.99**	**1.00**	**1.00**	**0.99**
	95% PCI	0.99 - 1.00	0.91 - 1.00	0.98 - 1.00	0.98 - 1.00	0.98 - 1.00	0.98 - 1.00	0.99 - 1.00	0.98 - 1.00	0.98 - 1.00
Skin lesion: diffuse	Se	**0.40**	**0.40**	**0.34**	**0.34**	**0.38**	**0.37**	**0.36**	**0.43**	**0.38**
	95% PCI	0.06 - 0.80	0.07 - 0.93	0.06 - 0.67	0.06 - 0.76	0.06 - 0.88	0.06 - 0.89	0.06 - 0.79	0.06 - 0.96	0.06 - 0.79
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Healing wounds	Se	**0.37**	**0.13**	**0.06**	**0.23**	**0.51**	**0.22**	**0.15**	**0.19**	**0.18**
	95% PCI	0.18 - 0.62	0.00 - 0.52	0.00 - 0.17	0.09 - 0.43	0.09 - 1.00	0.00 - 0.92	0.01 - 0.55	0.00 - 0.78	0.01 - 0.51
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Fit (BCS 2 – 4)	Se	**0.99**	**0.99**	**1.00**	**0.99**	**0.95**	**0.69**	**1.00**	**1.00**	**0.99**
	95% PCI	0.99 - 1.00	0.96 - 1.00	0.99 - 1.00	0.98 - 1.00	0.87 - 0.99	0.51- 0.85	0.99 - 1.00	0.97 - 1.00	0.95 - 1.00
	Sp	**0.79**	**0.76**	**0.76**	**0.73**	**0.68**	**0.73**	**0.67**	**0.75**	**0.74**
	95% PCI	0.67 - 0.91	0.60 - 0.92	0.63 - 0.87	0.58 - 0.85	0.36 - 0.85	0.36 - 0.93	0.20 - 0.88	0.45 - 0.94	0.55 - 0.87
Fat (BCS > 4)	Se	**0.89**	**0.82**	**0.83**	**0.77**	**0.69**	**0.78**	**0.76**	**0.78**	**0.80**
	95% PCI	0.76 - 0.98	0.65 - 0.95	0.70 - 0.92	0.61 - 0.90	0.33 - 0.91	0.22 - 0.98	0.20 - 0.97	0.23 - 0.98	0.56 - 0.93
	Sp	**0.99**	**0.99**	**1.00**	**0.99**	**0.94**	**0.69**	**1.00**	**1.00**	**0.99**
	95% PCI	0.99 - 1.00	0.95 - 1.00	0.99 - 1.00	0.98 - 1.00	0.86 - 0.99	0.52 - 0.83	0.99 - 1.00	0.98 - 1.00	0.94 - 1.00

Table 5.6 continued

Indicator	Test estimate	Observer 1	2	3	4	5	6	7	8	new
Thin (BCS < 1)	Se	**0.34**	**0.36**	**0.37**	**0.38**	**0.36**	**0.36**	**0.35**	**0.37**	**0.36**
	95% PCI	0.18 - 0.53	0.18 - 0.61	0.20 - 0.57	0.23 - 0.61	0.14 - 0.65	0.15 - 0.61	0.14 - 0.58	0.18 - 0.68	0.19 - 0.56
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00
Lameness	Se	**0.91**	**0.86**	**0.89**	**0.91**	**0.89**	**0.88**	**0.90**	**0.91**	**0.90**
	95% PCI	0.83 - 0.96	0.49 - 0.95	0.80 - 0.95	0.82 - 0.97	0.72 - 0.98	0.62 - 0.98	0.82 - 0.97	0.82 - 0.99	0.81 - 0.96
	Sp	**0.99**	**0.98**	**0.98**	**0.99**	**0.86**	**0.98**	**0.97**	**0.99**	**0.98**
	95% PCI	0.98 - 0.99	0.94 - 1.00	0.98 - 0.99	0.98 - 0.99	0.76 - 0.94	0.94 - 1.00	0.95 - 0.99	0.96 - 1.00	0.95 - 0.99
Foot lesion	Se	**0.89**	**0.75**	**0.80**	**0.88**	**0.73**	**0.76**	**0.94**	**0.90**	**0.85**
	95% PCI	0.84 - 0.93	0.52 - 0.90	0.75 - 0.84	0.84 - 0.91	0.59 - 0.85	0.53 - 0.92	0.89 - 0.98	0.80 - 0.97	0.74 - 0.93
	Sp	**0.93**	**0.95**	**0.78**	**0.73**	**0.80**	**0.80**	**0.83**	**0.95**	**0.88**
	95% PCI	0.88 - 0.97	0.88 - 1.00	0.73 - 0.84	0.66 - 0.80	0.47 - 0.99	0.56 - 0.96	0.73 - 0.93	0.86 - 1.00	0.73 - 0.97
Significant foot lesion	Se	**0.91**	**0.55**	**0.84**	**0.89**	**0.26**	**0.85**	**0.90**	**0.78**	**0.79**
	95% PCI	0.81 - 0.99	0.10 - 0.96	0.69 - 0.95	0.74 - 0.99	0.03 - 0.64	0.43 - 1.00	0.75 - 0.99	0.51 - 0.96	0.47 - 0.96
	Sp	**0.98**	**0.99**	**0.96**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
	95% PCI	0.97 - 0.99	0.96 - 1.00	0.94 - 0.97	0.96 - 0.99	0.97 - 1.00	0.96 - 1.00	0.98 - 1.00	0.97 - 1.00	0.97 - 1.00
White line lesion	Se	**0.84**	**0.82**	**0.81**	**0.90**	**0.86**	**0.84**	**0.84**	**0.87**	**0.85**
	95% PCI	0.79 - 0.89	0.62 - 0.92	0.76 - 0.85	0.86 - 0.94	0.77 - 0.93	0.72 - 0.93	0.78 - 0.90	0.76 - 0.95	0.79 - 0.90
	Sp	**0.91**	**0.97**	**0.80**	**0.74**	**0.59**	**0.79**	**0.84**	**0.92**	**0.85**
	95% PCI	0.87 - 0.95	0.91 - 1.00	0.75 - 0.85	0.67 - 0.81	0.31 - 0.86	0.53 - 0.96	0.74 - 0.93	0.80 - 1.00	0.68 - 0.97

Table 5.6 continued

Indicator	Test estimate	Observer								
		1	2	3	4	5	6	7	8	new
Inter-digital dermatitis	**Se**	**0.91**	**0.88**	**0.90**	**0.94**	**0.93**	**0.90**	**0.91**	**0.89**	**0.92**
	95% PCI	0.68 - 1.00	0.00 - 1.00	0.59 - 1.00	0.73 - 1.00	0.71 - 1.00	0.01 - 1.00	0.59 - 1.00	0.00 - 1.00	0.71 - 1.00
	Sp	**1.00**	**0.99**	**0.99**	**1.00**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**
	95% PCI	0.99 - 1.00	0.98 - 1.00	0.98 - 0.99	0.99 - 1.00	0.94 - 1.00	0.98 - 1.00	0.99 - 1.00	0.98 - 1.00	0.99 - 1.00
Footrot	**Se**	**0.95**	**0.93**	**0.95**	**0.93**	**0.93**	**0.93**	**0.94**	**0.93**	**0.94**
	95% PCI	0.79 - 1.00	0.60 - 1.00	0.79 - 1.00	0.62 - 1.00	0.58 - 1.00	0.59 - 1.00	0.70 - 1.00	0.59 - 1.00	0.75 - 1.00
	Sp	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
	95% PCI	0.99 - 1.00	0.99 - 1.00	0.99 - 1.00	0.99 - 1.00	0.99 - 1.00	0.98 - 1.00	0.99 - 1.00	0.98 - 1.00	0.99 - 1.00
CODD	**Se**	**0.81**	**0.81**	**0.83**	**0.85**	**0.84**	**0.83**	**0.86**	**0.84**	**0.84**
	95% PCI	0.81 - 0.92	0.84 - 0.99	0.85 - 0.95	0.86 - 1.00	0.86 - 1.00	0.86 - 1.00	0.87 - 1.00	0.86 - 1.00	0.86 - 0.99
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.97**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	0.97 - 0.99	1.00 - 1.00	1.00 - 1.00
Toe granuloma	**Se**	**0.75**	**0.72**	**0.75**	**0.72**	**0.75**	**0.74**	**0.72**	**0.74**	**0.79**
	95% PCI	0.51 - 0.97	0.36 - 0.97	0.49 - 0.98	0.46 - 0.96	0.49 - 0.98	0.47 - 0.98	0.47 - 0.96	0.47 - 0.98	0.56 - 0.97
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	0.99 - 1.00	0.99 - 1.00	1.00 - 1.00	0.99 - 1.00	0.98 - 1.00	0.99 - 1.00	0.99 - 1.00	0.99 - 1.00	0.99 - 1.00
Joint swelling	**Se**	**0.78**	**0.76**	**0.71**	**0.78**	**0.76**	**0.76**	**0.79**	**0.77**	**0.77**
	95% PCI	0.57 - 0.90	0.37 - 0.94	0.35 - 0.86	0.52 - 0.93	0.43 - 0.95	0.44 - 0.94	0.56 - 0.93	0.51 - 0.92	0.54 - 0.88
	Sp	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	95% PCI	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00	1.00 - 1.00

Table 5.7 Intra-observer reliability of the test standard observer assessments of individual sheep welfare indicators

Indicator	% agreement	κ (95% CI)	Kendall's W
Demeanour	93.18	0.37 (0.03 – 0.76)	[b]
Eye condition	[a]	[a]	[b]
Nasal discharge	[a]	[a]	[b]
Tooth disease	90.91	0.52 (0.37 – 0.69)	0.87
Coughing	95.45	0.31 (-0.17 – 0.94)	[b]
Ear lesion	[a]	[a]	[b]
Dirty belly	96.59	0.55 (-0.06 – 0.02)	[a]
Dirty legs	95.45	0.02 (0.10 –1.00)	[a]
Dirty rear	80.68	0.56 (0.38 – 0.63)	0.80
Mastitis	98.44	0.79 (0.40 – 1.00)	0.99
Tail length	[a]	[a]	[b]
Wool loss	98.86	0.79 (0.40 – 1.00)	[b]
Skin irritation	[a]	[a]	[b]
Skin lesion	94.32	0.42 (0.26 – 0.66)	[b]
Injuries & wounds	98.86	0.90 (0.88 – 1.00)	0.95
Body condition score	65.91	0.44 (0.31 – 0.56)	0.80
Fit-fat-thin	98.86	0.87 (0.58 – 1.00)	[b]
Lame	90.91	0.72 (0.54 – 0.90)	[b]
Foot lesion	76.14	0.49 (0.30 – 0.68)	[b]
Significant foot lesion	90.91	0.55 (0.29 – 0.82)	[b]
White line	90.91	0.79 (0.26 – 0.67)	[b]
Inter-digital dermatitis	98.86	0.79 (0.40 – 1.00)	[b]
Foot rot	95.45	0.48 (0.04 – 0.91)	[b]
CODD	98.86	0.70 0.40 – 1.00)	[b]
Other foot lesion	[a]	[a]	[a]
Joint swelling	98.86	0.79 (0.40 – 1.00)	[b]
Myiasis	[a]	[a]	[a]

[a] Insufficient observations to provide meaningful estimate
[b] Binary scoring indicator

Figure 5.2 Differences in the repeat assessment of individual sheep indicators by the test standard observer

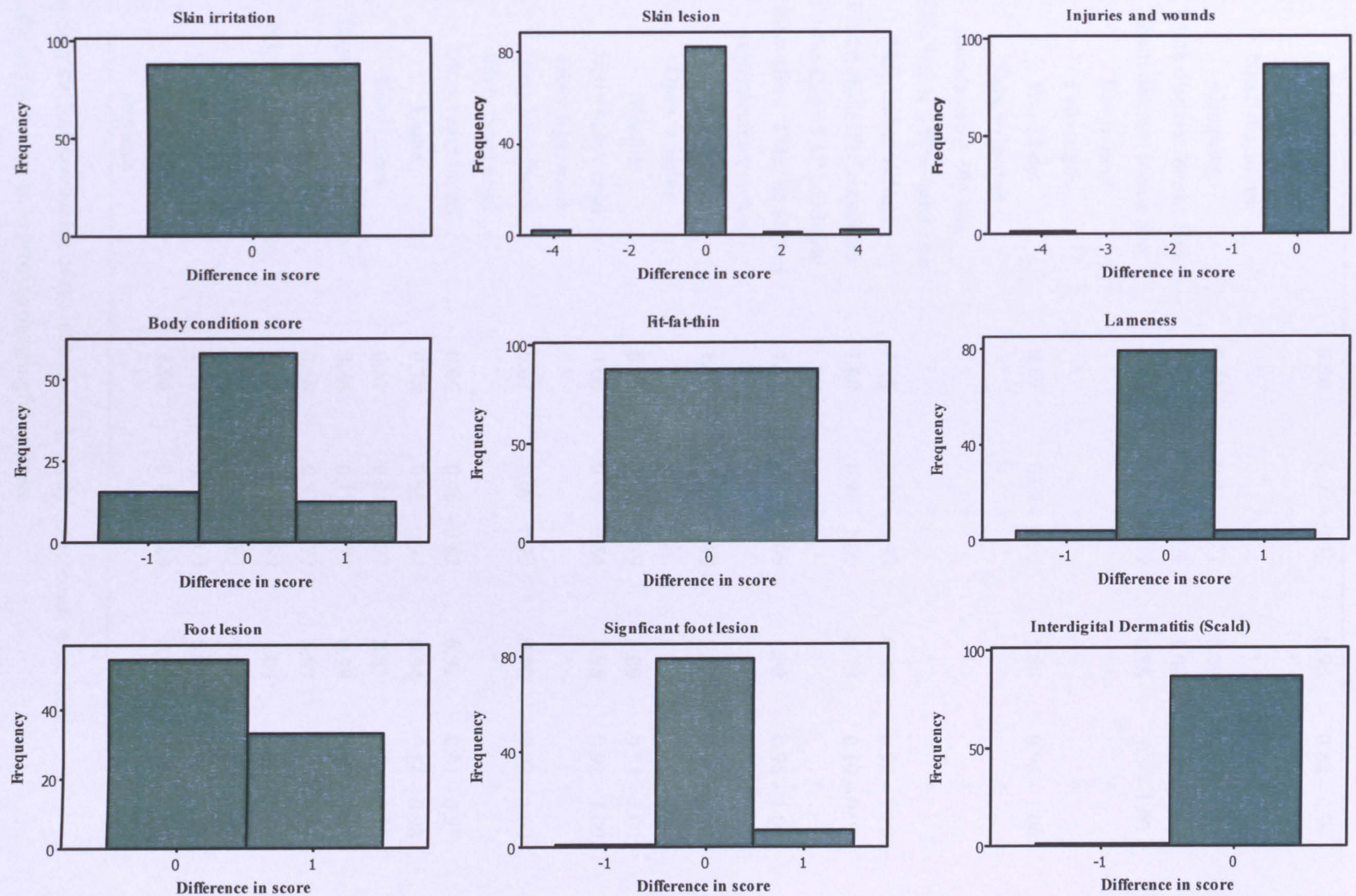


Table 5.8 Se and Sp of the repeat assessments of individual sheep indicators by the test standard observer §

Indicator	Se	95 % CI	Sp	95 % CI
Demeanour	**0.50**	0.07 – 0.93	**0.95**	0.88 – 0.99
Eye condition	[a]	[a]	[a]	[a]
Nasal discharge	[a]	[a]	[a]	[a]
Coughing	**0.33**	0.84 – 0.91	**0.99**	0.94 – 1.00
Tooth disease: incisor loss	**0.50**	0.01 – 0.99	**1.00**	0.96 – 1.00
Tooth disease: molar loss	**0.44**	0.14 – 0.79	**0.98**	0.9 – 1.00
Ear lesions	[a]	[a]	[a]	[a]
Tail length	[a]	[a]	[a]	[a]
Wool Loss	**0.67**	0.09 – 0.99	**1.00**	0.96 – 1.00
Skin irritation	[a]	[a]	[a]	[a]
Skin lesion ≥ 50p size	[a]	[a]	[a]	[a]
Skin lesion ≤ 50p -hand size	[a]	[a]	[a]	[a]
Skin lesion: diffuse	**0.75**	0.19 – 0.99	**0.98**	0.92 – 1.00
Fit-fat-thin: 'Fit' condition	**1.00**	0.96 – 1.00	**0.75**	0.19 – 0.99
Fit-fat-thin: 'Fat' condition	[a]	[a]	[a]	[a]
Fit-fat-thin: 'Thin' condition	**1.00**	0.40 – 1.00	**1.00**	0.96 – 1.00
superficial scratches	[a]	[a]	[a]	[a]
Healing wounds	**1.00**	0.03 – 1.00	**1.00**	0.96 – 1.00
Open wounds	[a]	[a]	[a]	[a]
Mastitis	**0.67**	0.09 – 0.99	**1.00**	0.94 – 1.00
Dirty belly: mud	**0.00**	0.00 – 0.84	**0.98**	0.92 – 1.00
Dirty legs: mud	[a]	[a]	[a]	[a]
Dirty legs: faeces	**0.67**	0.09 – 0.99	**0.98**	0.92 – 1.00
Dirty rear: mud	[a]	[a]	[a]	[a]
Dirty rear: faeces	**0.66**	0.46 – 0.82	**0.92**	0.81 – 0.97
Lame	**0.78**	0.52 – 0.94	**0.94**	0.86 – 0.98
Foot Lesion	**0.61**	0.44 – 0.77	**0.87**	0.74 – 0.95
Significant Foot Lesion	**0.46**	0.19 – 0.75	**0.99**	0.93 – 1.00
White Line	**0.58**	0.37 – 0.77	**0.87**	0.76 – 0.94
Inter-digital dermatitis	**0.67**	0.09 – 0.99	**1.00**	0.96 –1.00
Foot rot	**0.40**	0.05 – 0.85	**0.99**	0.94 – 1.00
CODD	**1.00**	0.16 – 1.00	**0.99**	0.94 – 1.00
Joint Swelling	**1.00**	0.16 – 1.00	**0.99**	0.94 – 1.00
Myiasis	[a]	[a]	[a]	[a]

§ assuming the first assessment of the intra-observer study was the reference result

[a] Insufficient observations to provide meaningful estimate

5.4 Discussion

The objective of this chapter was to evaluate the diagnostic test validity of 29 animal-based indicators of sheep welfare assessed by an individual examination. The test performance of each indicator was assessed by 8 observers on a sample of 1146 adult sheep and growing lambs from 38 farms. Overall a low number of sheep were observed with many of the welfare conditions, including pruritus and myiasis, and this is likely to have affected the interpretation of the test validity of some indicators. In spite of this, many indicators produced fair-good levels of reliability and high levels of diagnostic specificity. The lower level of diagnostic sensitivity and reliability identified for some indicators may have been influenced by the study population, scoring system, observer characteristics, and quality of the assessment conditions. These factors are considered in the discussion that follows.

Farms were selected according to their consent to participate, location, management type and the feasibility for assessment within the limits of a one-day farm visit. Given the non-random selection of farms, it is likely that the study population may be biased towards farms of higher welfare status or those with regular contact with a veterinary surgeon. However, the objective of the on-farm studies was to investigate the reliability, Se and Sp of the indicators, rather than infer conclusions regarding the on-farm welfare of the general sheep population. Therefore the test performance of each indicator was examined on sheep of different breeds, from a range of geographical locations and managed under a variety of farm systems (Chapter 3).

To maintain independent assessments, observers were not provided with any clinical or production information prior to the farm visits. In addition, sampling bias was reduced by using a random number identifier to pre-select 30 sheep for the inter-observer study. In order to reduce the amount of handling and casting of the same animals during the multiple observer assessments, one observer held the sheep whilst other observers performed their examinations. Therefore, if one observer elicited a withdrawal response on examination of the hoof, this reaction was not easy to hide. So the handler may have been alerted to the presence of a particular condition and this may have affected the level of observer reliability produced. However, it was not feasible to have additional labour during this study so it was not possible to blind observers to this type of information. Due to the study setting it was also not practical to blind observers to cues or stimuli, such as farm cleanliness, quality of handling area and the presence of equipment or medicines which may have alerted them to certain health or welfare conditions in the sample group. Repeat assessments were performed on the same day for reasons of farmer convenience and indicator stability. The test standard observer was blinded to previous scores, assessed sheep in a random order and undertook unrelated work between visits so as to reduce any dependency of the measurements. In the absence of any 'gold standard' for the validation of clinical measures and signs of sheep behaviour, a 'test standard' observer was used as the reference test for comparison (Burn *et al.*, 2009). Since LCA identified that this observer produced higher Se and Sp estimates compared to other observers, this was considered to be an appropriate reference test for evaluating the diagnostic performance of each observer.

The time stability of indicators was considered in the design of the intra-observer study. This was because it is recognised that biological variability can affect the test-retest results of diagnostic tests (Lucas *et al.*, 2010). The absence of an animal welfare reference test made this difficult to estimate, so, in accordance with previous on-farm reliability studies (Kristensen *et al.*, 2006) intra-observer repeat assessments were limited to a single day. In spite of the short interval, it was apparent that biological variability occurred between repeat visits. For example, there was a difference in the number of sheep assessed with dull demeanour. Whilst this difference did not appear to affect the level of overall intra-observer reliability, the test Se was relatively low. The variation across first and second assessments could be attributed to missed observations. However, the small sample size and good assessment conditions for the assessment of demeanour, suggested this was unlikely. Experience from the intra-observer study suggested that the repeated gathering and examination of individual sheep affected the demeanour of sheep with a pre-existing health or welfare condition. So, a longer interval between individual examinations of sheep, for example a 2 – 3 day period, may be advisable for future reliability studies.

Following the Quality of Reporting of Reliability studies (QAREL) guidelines (Lucas *et al.*, 2010), the test performance of each welfare indicator was analysed using methods that are in current use in animal welfare science. Percentage (%) agreement was selected as it provides a good starting point for understanding observer reliability. However, it does not take into account the amount of agreement that occurs due to chance alone (Sim and Wright, 2005). Therefore, kappa (κ), a chance-corrected method - a well recognised means of evaluating the reliability of animal welfare indicators was also used as a means of examining the level of observer agreement.

The interpretation of κ values requires consideration of the prevalence of the welfare condition as a low prevalence can provide artificially low estimates of reliability (Feinstein and Cicchetti, 1990). Despite a high % agreement, a low κ value was produced for the assessment of injuries and wounds. In this case the low level of observer agreement was likely to be related to the low level of injuries and wounds observed in the study population. Therefore, poor levels of reliability could have been found because κ takes account of the fact that chance agreement is high when a low proportion of the population is affected by a condition (Sim and Wright, 2005). The understanding of reliability is also affected by the subjective interpretation of κ agreement. The results presented in this chapter can be interpreted according to the suggestions of Fleiss (1981), although it is unknown whether a $\kappa \geq 0.75$ really does equate to 'excellent' agreement, as this is based on a subjective interpretation of the data.

Another affect of the low proportion of sheep with conditions associated with poor welfare was that no meaningful estimates of test performance could be provided for some animal-based indicators. For example, no observations of visible uroliths (crystals) and in-growing horns were made during the cross-sectional study. These conditions have been identified as important welfare issues for sheep (Chapter 2) so the absence of these conditions in this study might indicate that these measures are not good tests for identifying these specific welfare issues. However, it is more likely that these results reflect the true level of urolithiasis and in-growing horns in the study population.

In spite of the low proportion of many welfare conditions and issues with interpretation, κ analysis identified that the overall reliability of many individual sheep indicators could be interpreted as 'fair to good' including the assessment of lameness, body condition scoring and eye condition. Particular indicators identified with excellent levels of reliability included demeanour, wool loss, tail length, joint swelling and myiasis. Given that κ is influenced by the level of the welfare condition in the study population, it is suggested that the test performance of these indicators could be even greater if the indicators were tested on a sample with a greater level of sub-optimal welfare conditions, such as lameness, emaciation or ocular abnormalities.

A useful addition to κ agreement analysis was the graphical representation of differences in observer scores. This approach identified that few disagreements occurred in the inter- and intra-observer studies. There were few differences in body condition scores (BCS) provided by most observers of the inter-observer study. Interestingly, most observers disagreed over the mid-range of scores i.e. between BCS 2 and 3. So, the use of a broader scoring system, such as the fit-fat-thin indicator, in which 'fit' covers BCS 2 and 3 may prove useful as the method was found to provide a higher level of observer agreement compared to the scoring system of Russel (1984). Graphical representation also identified evidence of observer bias. It was clear that observer 6 scored body condition one unit higher than the test standard observer and the possible reasons are discussed below.

Many of the individual indicators tested in this chapter are comparable to those performed as part of a veterinary clinical examination and routine stock inspection. It was therefore important to investigate whether these indicators were reliably applied between different types of assessors. Certain observer characteristics, such as training, experience and occupation can influence the level of observer agreement (Kristensen *et al.,* 2006), therefore, both trained and untrained assessors and those with a variety of experience in assessing the health and welfare of individual sheep were included.

Whilst this chapter cannot provide strong evidence of the effect of training or experience on reliability, there are some data trends that suggest trained observers achieved higher levels of agreement compared to untrained observers. For example, observers 5 and 6 had not received training in the application of these indicators and so were unlikely to be aware of the clearly defined and legal criteria for the assessment of a short tail. The lack of training may therefore have resulted in the lower level of reliability, Se and Sp of these observers.

The level of experience may also have influenced the test performance of study observers. For example, observers 7 and 8 (experienced and trained) produced higher levels of agreement compared to observers 3 and 4 (inexperienced and trained). Despite these findings, experience alone may not predict high levels of inter-observer reliability. For example, observer 6 - an experienced but untrained veterinary assessor, had a tendency to score body condition as one BCS higher than the test standard observer. Observer 6 only performed a single farm assessment (n = 30 sheep), and it is possible that other observers may have become calibrated to the scoring system during additional farm visits. Whilst there is insufficient data to provide a clear conclusion, it is suggested that the poorer test performance of observer 6 may be a result of the lack of training and standardisation to the Russel (1984) scoring system or the familiarity of the observer with examining sheep with poorer body condition.

Unlike previous studies (Harkins, 2005), high levels of agreement were also consistent with the category of 'vet' for observers 2, 5 and 6. These observers had a lower level of agreement with the test standard observer for several indicator assessments. By contrast, observer 7 - a trained, experienced, non-vet assessor, achieved good levels of reliability for many indicators. The lower level of reliability for observers 2, 5 and 6 may have reflected the fact that the 'vet' category included non-clinical lecturers and researchers who may not have been very familiar with the welfare assessment of sheep.

The indicators of 'foot lesion' and 'significant foot lesion' were therefore included to allow inexperienced or non-clinical assessors to record the presence of any foot lesion without needing to provide a specific diagnosis. With the exception of the untrained observers (5 and 6), 'foot lesion' was a highly reliable indicator, which may have reflected missed observations or a different interpretation of foot lesion scores. On-farm experience suggested that white line lesions (WL) were commonly observed in the study population. So, the poorer reliability and Sp may reflect a failure to recognise white line separation as a specific foot lesion. In addition, the quality of assessment conditions may have affected the ability to clearly identify specific foot lesions. During the intra-observer study most indicators were found to be reliable and specific but there was considerable variation over the assessment of the foot lesion indicator. This was largely due to differences between the first and second assessment of WL. This appeared to be due to the quality of assessment conditions during the second assessment. The fading light levels during the winter period of assessment hampered the ability to distinguish mild cases of WL separation when individual examinations were performed outdoors under natural light conditions.

Closer examination of foot lesion scoring revealed that most disagreement occurred between the assessment of inter-digital dermatitis ('scald') and footrot. As there is overlap in the pathology and epidemiology of these two foot conditions (Egerton, 1971) it may be more appropriate to combine these conditions into a single lesion scoring system as suggested by Conington *et al.*, (2008). The higher level of reliability achieved for the diagnosis of contagious ovine digital dermatitis (CODD) might be attributed to the clear scoring definition and ease of distinguishing cases of CODD from other foot lesions. It might also suggest that foot lesions such as CODD which are associated with severe foot pathology are consistently identified but less obvious lesions, such as WL separation, may be less readily observed (Harkins, 2005).

The number of scoring categories and the descriptive terms used for some of the indicators may have affected the level of observer agreement. Therefore, modifications and reductions of certain indicator scoring systems might be beneficial. This may be particularly useful for the assessment of mastitis, cleanliness of the abdomen and breech, skin lesions and injuries and wounds. Despite the good level of overall reliability achieved, it was evident that some observers experienced difficulty in assessing whether one or both mammary glands were affected by mastitis and this may have led to scoring discrepancies. There were also difficulties in assessing the presence of mastitis in freshly weaned ewes so the palpation of mammary glands congested with milk may have been mistakenly identified as mastitis. This experience could inform the timing or interpretation of welfare assessments. If a large number of recently weaned ewes are assessed with mastitis it may be sensible to conduct a follow-up assessment. In addition, mastitis might be more consistently identified if the indicator scoring scale was altered from an ordinal to a binary system i.e. presence of mastitis in one or more mammary gland(s). Indeed, the dichotomisation of mastitis scores for the evaluation of test Se and Sp identified that observers were good at identifying sheep without mastitis so this appears to be an appropriate modification of the scoring scale.

For other indicators, such as cleanliness scoring of the belly and rear, the low level of inter-observer reliability may have reflected issues with distinguishing difference in the categorical scores. Scoring difference lay between a score unit i.e. between 'dirty' and 'filthy' scores. Therefore, improving the cleanliness scoring definitions and amalgamating these categories into a single score may provide more consistent assessments. Results from this study suggested that leg cleanliness was not a reliable, sensitive or specific test and it may be sensible to exclude this indicator from subsequent studies.

The results of this chapter also corroborate previous research which found that poor levels of reliability were associated with the assessments of skin lesions in sheep (Napolitano *et al.*, 2009). This may be because the presence of the fleece can mask skin lesions, making the identification of small lesions particularly difficult. Consequently, Napolitano and co-workers (2009) suggested only scoring shorn sheep and skin lesions over 2 centimetres (cm) in size. Whilst this approach may improve the level of test reliability it does not appear to be a credible way of assessing sheep as part of 'spot-check' or year-round welfare inspections. This approach could also risk missing skin lesions of great welfare importance such as sheep scab and lice (van den Broek and

Huntley, 2003). In this study, the estimation of the size of skin lesions was based on 'hand-size' (5 x 10 cm) and so a variation in the hand-size of the observer may have led to measurement variability thus leading to poor observer agreement. Given the difficulties in assessing small lesions and the potential for measurement errors it may be sensible to refine the scoring system for skin lesion assessment and reduce the number of scoring categories.

This was the first study to evaluate the validity of animal-based welfare indictors in terms of the Se and Sp of observer assessments. Both LCA and classical approaches to diagnostic test evaluation found that the majority of indicators assessed by individual exam had high Sp but lower Se. This suggests that observers were better at identifying sheep that were not affected by a particular welfare condition and so these tests might be less likely to penalise farmers with good standards of sheep welfare. However, there is always a trade-off between the level of diagnostic Se and Sp and a test with a lower Se might miss some animals with important welfare conditions. However, the indicators need to be tested on a population experiencing a higher proportion of welfare issues before any conclusive statements can be made.

The LCA method was a useful approach for evaluating test performance in the absence of a 'gold standard' and appears to be highly relevant for analysing the validity of animal welfare indicators. LCA also provided another advantage over classical evaluation of test performance by predicting Se and Sp of 'new' observers – unknown assessors who may apply these indicators in the future. As a result, poor predicted levels of Se and Sp were identified for leg cleanliness, skin irritation, injuries and wounds, and skin lesions. These results might be used to suggest that these indicators are unsuitable for on-farm welfare schemes. However, it was apparent that low Se of these indicators was likely to be attributed to the low level of many of the indicator scores, and the predicted test performance could be higher if future assessors tested the full scoring systems on a sheep population with a higher proportion of sub-optimal welfare conditions.

The simultaneous assessment of 29 indicators per sheep produced abundant information on the health and welfare of individual sheep but it also increased the potential for mis-recording of assessments, which may have affected the level of test reliability, Se and Sp reported in this chapter. Selecting indicators with good reliability, Se and Sp would reduce the number of tests that need to be applied to each sheep and could reduce the amount of time needed to perform assessments and recordings. However, discarding 'poor' indicators at this stage could result in the loss of indicators of valid sheep welfare issues. It is clear that there are additional criteria that can be used to select valid indicators of sheep welfare. Indicator tests also need to be capable of identifying between-farm and seasonal differences in animal-based outcomes of sheep welfare and these aspects of validity need to be addressed in further on-farm studies.

5.5 Conclusion

Whilst there are data trends to suggest that experienced and trained assessors achieved higher levels of reliability, Se and Sp, the effect of observer training and experience on the level of test performance were not fully investigated in this study and further studies are needed in order to provide conclusive results. The low level of sub-optimal welfare conditions, such as myiasis and pruritus, affected the test validation of some indicators but it also meant that the observational skills of the assessors were thoroughly tested. The ability of observers to identify certain welfare conditions on a population with a low level of health and welfare suggests that higher levels of reliability, Se and Sp could be achieved if the tests were applied to a population of sheep with a higher proportion of welfare issues. As well as being reliable, sensitive and specific, the welfare indicators developed in this thesis should be feasible, robust to the range of farming systems and responsive to seasonal changes in sheep welfare. Therefore, further validation of these indicators needs to be investigated before key or "iceberg" indicators of sheep welfare can be selected.

Chapter 6
VALIDATING INDICATORS OF YOUNG LAMB WELFARE

6.1 Introduction

The indicators developed in this thesis need to be capable of assessing the welfare of sheep throughout the on-farm production cycle. Therefore, valid indicators for assessing the on-farm welfare of young lambs – defined as lambs aged 6 weeks and under, were developed following a consultation of the scientific literature (Chapter 1) and a consensus of expert opinion (Chapter 2). These approaches identified a range of on-farm welfare issues for lambs including starvation, hypothermia and the presence of infectious and inheritable diseases.

Young lambs are managed in a variety of ways, from intensive, indoor-lambing flocks which require a high input of labour and resources to extensive lambing systems in which lambs may be inspected less frequently and are exposed to more natural conditions, including extremes of temperature and predation (Dwyer, 2008b). In addition, young lambs may be reared with their birth ewe or, in cases of ill-health, death, maternal rejection or multiple births, young lambs may be fostered onto another ewe that is tethered, to prevent aggressive behaviours and rejection of the fostered lamb. Alternatively, orphan lambs can be reared on a substitute milk replacer by bottle feeding individual lambs or using automatic feeding systems used to rear groups of orphan lambs. As the on-farm welfare issues for young lambs may be influenced by the method of rearing and on-farm management, any indicators developed in this thesis need to be valid tests that can be applied to a wide range of lambing flocks.

A number of animal-based indicators, which relied on observations of specific clinical signs and behaviours (Chapter 3), were developed to assess the welfare of individual lambs. As animal welfare indicators are applied to diagnose specific welfare conditions, the principles of diagnostic test evaluation can be used to investigate whether different observers have similar levels of diagnostic ability. The objective of this chapter was therefore to evaluate the test validity of animal-based indicators of young lamb welfare in terms of the reliability, sensitivity (Se) and specificity (Sp) of observer assessments.

6.2 Materials and methods

6.2.1 Study population

During February to April 2009, 17 flocks categorised as lowland (n = 12), upland (n = 2), and hill (n = 3) were recruited, as described in Chapter 3, and selected on the basis of their farm type, period of lambing, farm location and consent to participate.

6.2.2 Observer population

Four trained observers (Table 6.1) were selected from an observer pool, previously described in Chapter 3. Observer 1 (the author) was designated the 'test standard observer' and the findings of this observer were used as the reference results in order to compare the test performance of observers 3, 8 and 9. Each observer independently tested the 11 non-invasive indicators, which were assessed by performing physical examinations and behavioural observations of individual lambs (Chapter 3).

Inter-observer study

For the inter-observer study the test standard observer independently performed assessments of 966 young lambs (aged $\leq$ 6 weeks) with a varied combination of an additional 1 to 2 observers (Table 6.1). The number of lambs examined on any one occasion ranged between 30 – 90 lambs (median 59) and were selected by the test standard. 50.5% of lambs were managed in individual pens, 28.7% were managed indoors in groups and 25.9% were managed outdoors. 86.1% were observed as being reared with a non-tethered ewe, 3.3% were observed to be reared with a tethered ewe and 4.9% were classed as orphan lambs.

Intra-observer study

For the intra-observer study, the test standard observer examined 81 lambs from 2 lowland farms in North-West England. The sample population comprised young lambs housed in individual pens and marked with a unique numeric identifier and those observed to be reared with a non-tethered ewe (93.8%) and orphan lambs (6.2%). Each lamb was examined twice within a twenty-four hour period. A five hour interval between repeat assessments was selected to minimise any alteration in indicator outcomes. Following completion of the first assessment, the test standard observer left the farm and undertook unrelated work to reduce short-term memory recall. The results of the first assessment visit were sealed so that the observer remained blinded to these results.

Table 6.1 Observer population for the inter-observer study of young lamb indicators

Observer combination	n farms assessed	n lambs assessed
1 and 3	1	53
1 and 8	1	50
1, 3 and 8	2	100
1, 3 and 9	12	703
1, 3, 8 and 9	1	60

6.2.3 Statistical analysis

Reliability data was analysed using Stata version 10 (StataCorp LP, College Station, Texas). To permit cross-tabulation of observer assessments all categorical indicators were evaluated using a binary scoring system. Therefore, abdominal fill was re-coded as normal (0) or hollow and bloated scores (1). Standing ability was re-coded as stands freely (0) or (1) weak on legs and/or unable to stand. Inter-observer assessments of demeanour, stimulation, shivering, standing ability, posture, body condition, abdominal fill, lameness, eye condition and salivation were assessed using Fleiss's kappa (κ) (Fleiss, 1981). Cohen's κ (Cohen, 1960) and percentage agreement (%) was used to assess the paired agreement for each test observer (2, 8 and 9) with the test standard observer, and the repeat assessments of the test standard observer. All κ values were interpreted according to Fleiss (1981), whereby values $\geq$ 0.75 suggested excellent levels of agreement, κ of 0.40 - 0.75 indicated fair to good agreement, and κ $\leq$ 0.40 was poor agreement. Graphical representation of scoring differences between the assessment of each observer and the test standard observer were also examined for evidence of bias.

The reliability of play behaviour assessments was evaluated using Kendall's coefficient of concordance (W) and interpreted on a scale of 0 (no agreement) to 1 (perfect agreement) as described by Martin and Bateson (2007). Differences in play behaviour scores of observers 3, 8 and 9 and the test standard observer were graphically represented using Bland-Altman plots (1986), in which the mean play behaviour scores recoded by each observer were plotted against observer differences in play behaviour scores. Differences between observer scores of play behaviour scores were tested using Pittman's test of variance (Pittman, 1939).

The Se and Sp of indicators assessed during the inter-observer study was analysed using Latent Class Analysis (LCA), as previously described in Chapter 5. For the intra-observer study, the results of the first assessment were assumed to be the reference result and the Se and Sp of the test standard observer was evaluated using a classical approach using a cross-classification table (Greiner and Gardner, 2000).

6.3 Results

Overall inter-observer reliability results, determined by Fleiss's κ are shown in Table 6.2. The paired assessments between each observer (3, 8 and 9) and the test standard are provided by percentage (%) agreement (Table 6.3) and Cohen's κ (Table 6.4). Results of the intra-observer study are presented in Table 6.5. The Se and Sp of all observers as well as future, unknown ('new') observers were evaluated by LCA (Table 6.5). Graphical representations of scoring differences during the inter- and intra-observer studies are shown in Figures 6.1 and 6.2. The results of each indicator are presented below.

Demeanour

Overall a Fleiss κ of 0.54 was achieved. Some variation in agreement with the test standard occurred (κ 0.42 – 0.58), although few disagreements in observer scores occurred (Figure 6.1). The test standard achieved κ 0.93 for the repeat assessment of demeanour. High Se (0.75 – 0.85) and Sp (0.98 – 1.00) was found for all observer assessments (Table 6.5).

Response to stimulation

Fleiss's κ 0.54 was found, although the Se of observer 9 (Se 0.42) was much lower than other observers (Table 6.4). The test standard showed higher levels of reliability, Se (1.00), and Sp (0.97) in comparison to other observers.

Standing ability

Both observer 3 and the test standard achieved excellent levels of reliability for assessment of standing ability. High Se was also apparent in both intra- and inter-observer studies (Tables 6.5 and 6.6).

Shivering

Shivering achieved a Fleiss κ of 0.55, although there was variation between observers. For example, results suggested that observer 9 was less reliable (κ 0.40) and had lower Se (0.37) compared to observer 3 (κ 0.75, Se 0.74). By contrast, the test standard observer produced consistently high levels of Se (Tables 6.5 and 6.6).

Posture

Posture appeared to be reliable between different observers (Fleiss's κ 0.45). Observer 8 and test standard achieved higher Se ($\geq$ 0.83) than observers 3 (0. 56) and 9 (0.62).

Body condition

High levels of reliability, Se and Sp were noted for body condition assessments performed by the test standard (κ1.00 Se and Sp 1.00) and observers 3 (κ 0.71, Se 0.8) and 9 (κ 0.76, Se 0.9). By contrast a lower Se (0.38) was found for observer 8, although Figure 6.1 indicated that few scoring disagreements had actually occurred.

Lameness

Paired observer assessments of lameness assessments provided κ agreement between 0.70 – 0.80). Observers 3 and 9 showed no disagreement with the assessment of the test standard observer (Figure 6.1). LCA provided Se estimates $\geq$ 0.7, although only the test standard achieved a Se of 0.8 (Table 6.6).

Abdominal fill

Abdominal fill appeared to be a reliable, sensitive and specific test for most observers. However, the results suggested that observer 9 had a lower level of diagnostic ability (κ 0.39 and Se 0.39) compared to other assessors.

Eye condition

Assessment of eye condition was reliable between (κ 0.72) and within observers (κ 1.00) with very few disagreements in both inter- and intra-observer studies (Figures 6.1 and 6.2). In addition, LCA identified that eye condition had consistently good levels of Se ($\geq$ 0.86) and Sp ($\geq$ 0.99) identified across all observers.

Salivation

Observers 1, 3 and 8 produced intra- and inter-observer kappa agreement > 0.70 (Tables 6.4 and 6.5). The test standard provided perfect Se and Sp (1.00) but there were insufficient observations of lambs with salivation during the inter-observer study to produce any meaningful estimation of the Se and Sp of each observer (Table 6.6).

Play behaviour

The reliability of play behaviour assessments achieved by observer 3 was W 0.64 and a W 0.69 for observer 9. The repeat assessments of the test standard observer provided W 0.71. Bland-Altman plots indicated that the mean of observers 3 and 9 assessments did not greatly differ, but there was significant variation in play behaviour scores (Figure 6.3). This variation was also identified by Pittman's test of variance ($p < 0.001$). Observer 8 did not record sufficient observations of play behaviour, so no further analysis could be performed.

The LCA approach predicted a Se $\geq$ 0.77 for all indicators, except for posture (0.67) and shivering (0.64) when applied by 'new' i.e. unknown and randomly selected assessors who may perform assessments in the future. This approach also predicted high levels of Sp for all indicators ($\geq$ 0.98).

Table 6.2 Overall inter-observer reliability of young lamb indicators determined by Fleiss's κ

Indicator	κ	95% CI	Interpretation
Demeanour	**0.54**	0.45 – 0.59	Good-fair
Stimulation	**0.54**	0.45 – 0.55	Good-fair
Shivering	**0.55**	0.35 – 0.66	Good-fair
Standing ability	**0.57**	0.52 – 0.63	Good-fair
Posture	**0.45**	0.34 – 0.48	Good-fair
Body condition	**0.72**	0.60 – 0.74	Good-fair
Abdominal fill	**0.44**	0.42 – 0.47	Good-fair
Lameness	**0.68**	0.53 – 0.69	Good-fair
Eye condition	**0.72**	0.63 – 0.77	Good-fair
Salivation	**0.71**	0.54 – 1.00	Good-fair

Table 6.3 Percentage agreement of each observer with the test standard for assessments of young lamb indicators

Indicator	% agreement by observer identity		
	3	8	9
Demeanour	97.27	96.67	96.59
Stimulation	98.03	98.57	98.43
Shivering	99.77	[a]	99.59
Standing	98.36	98.56	96.72
Posture	97.57	98.24	96.94
Body condition	97.11	97.65	97.47
Abdominal fill	97.76	100.00	97.96
Lameness	98.78	98.51	99.07
Eye condition	97.32	97.08	95.90
Salivation	100.00	99.33	100.00

[a] Insufficient observations to produce meaningful estimate

Table 6.4 Inter-observer reliability with the test standard observer for young lamb indicators determined by Cohen's κ

Indicator	Observer identity	κ (95 % CI)	Interpretation
	3	**0.55** (0.39 – 0.71)	Good-fair
Demeanour	8	**0.52** (0.21 – 0.83)	Good-fair
	9	**0.44** (0.27 – 0.62)	Good-fair
	3	**0.70** (0.57 – 0.86)	Good-fair
Standing ability	8	**0.66** (0.30 – 1.00)	Good-fair
	9	**0.58** (0.43 – 0.74)	Good-fair
	3	**0.75** (0.41 – 1.00)	Excellent
Shivering	8	**1.00** (1.00 – 1.00)	Excellent
	9	**0.40** (-0.14 – 0.94)	Good-fair
	3	**0.72** (0.42 –1.00)	Good-fair
Stimulation	8	**0.56** (0.34 – 0.79)	Good-fair
	9	**0.56** (0.34 – 0.79)	Good-fair
	3	**0.50** (0.32 – 0.69)	Good-fair
Posture	8	**0.76** (0.50 – 1.00)	Good-fair
	9	**0.50** (0.32 – 0.67)	Good-fair
	3	**0.71** (0.60 – 0.82)	Good-fair
Body condition	8	**0.49** (0.07 – 0.92)	Good-fair
	9	**0.76** (0.66 – 0.87)	Excellent
	3	**0.60** (0.41 – 0.79)	Good-fair
Abdominal fill	8	**1.00** (1.00 – 1.00)	Excellent
	9	**0.39** (0.12 – 0.66)	Good-fair
	3	**0.70** (0.53 – 0.87)	Good-fair
Lameness	8	**0.72** (0.42 – 1.00)	Good-fair
	9	**0.81** (0.67 – 0.95)	Excellent
	3	**0.76** (0.66 – 0.86)	Excellent
Eye condition	8	**0.84** (0.69 – 0.99)	Excellent
	9	**0.66** (0.54 – 0.78)	Good-fair
	3	**1.00** (1.00 – 1.00)	Excellent
Salivation	8	[a]	[a]
	9	**1.00** (1.00 – 1.00)	Excellent

[a] Insufficient observations to produce meaningful estimate

Table 6.5 Test performance of the repeat assessments of young lamb indicators by the test standard observer [§]

Indicator	% agreement	κ (95% CI)	Se (95% CI)	Sp (95% CI)
Demeanour	98.77	0.93 (0.81 – 1.00)	0.89 (0.51 – 0.99)	1.00 (0.95 – 1.00)
Stimulation	97.53	0.89 (0.73 – 1.00)	0.82 (0.66 – 1.00)	1.00 (0.94 – 1.00)
Shivering	100.00	1.00 (1.00 – 1.00)	1.00 (0.25 – 1.00)	1.00 (0.96 – 1.00)
Standing ability	98.77	0.94 (0.92 – 1.00)	1.00 (0.66 – 1.00)	1.00 (0.95 – 1.00)
Posture	98.77	0.85 (0.56 – 1.00)	0.75 (0.19 – 0.99)	1.00 (0.95 – 1.00)
Body condition	100.00	1.00 (1.00 – 1.00)	1.00 (0.48 – 1.00)	1.00 (0.95 – 1.00)
Abdominal fill	98.77	0.79 (0.40 – 1.00)	0.67 (0.09 – 0.99)	1.00 (0.95 – 1.00)
Lameness	100.00	1.00 (1.00 –1.00)	1.00 (0.16 – 1.00)	1.00 (0.95 – 1.00)
Eye condition	100.00	1.00 (1.00 –1.00)	1.00 (0.75 – 1.00)	1.00 (0.95 – 1.00)
Salivation	98.77	0.90 (0.71 – 1.00)	0.83 (0.36 – 1.00)	0.99 (0.95 – 1.00)

§ The first assessment of the intra-observer study was assumed as the reference result

Table 6.6 Latent class analysis of the Se and Sp of young lamb indicators

Indicator	Observer identity	Se (95 % PCI)	Sp (95 % PCI)
Demeanour	1	0.75 (0.58 – 0.89)	0.98 (0.97 – 0.99)
	3	0.85 (0.69 – 0.99)	1.00 (0.99 – 1.00)
	8	0.77 (0.50 – 0.96)	0.98 (0.96 – 1.00)
	9	0.70 (0.47 – 0.86)	1.00 (1.00 – 1.00)
	New	0.78 (0.52 – 0.56)	0.98 (0.88 – 1.00)
Stimulation	1	0.55 (0.40 – 0.70)	1.00 (1.00 – 1.00)
	3	0.74 (0.57 – 0.89)	1.00 (1.00 – 1.00)
	8	0.72 (0.34 – 0.95)	1.00 (1.00 – 1.00)
	9	0.30 (0.18 – 0.45)	1.00 (1.00 – 1.00)
	New	0.77 (0.48 – 0.99)	0.98 (0.86 – 1.00)
Standing ability	1	0.82 (0.62 – 0.94)	1.00 (0.99 – 1.00)
	3	0.80 (0.61 – 0.91)	1.00 (0.99 – 1.00)
	8	0.81(0.59 – 0.96)	0.99 (0.98 – 1.00)
	9	0.80 (0.60 – 0.90)	0.99 (0.97 – 0.99)
	New	0.80 (0.60 – 0.82)	0.99 (0.98 – 1.00)
Lameness	1	0.80 (0.60 – 1.00)	1.00 (0.99 – 1.00)
	3	0.73 (0.54 – 0.87)	1.00 (0.99 – 1.00)
	8	0.73 (0.47 – 0.91)	1.00 (0.99 – 1.00)
	9	0.76 (0.59 – 0.91)	1.00 (0.99 – 1.00)
	New	0.76 (0.56 – 0.96)	1.00 (0.99 – 1.00)
Body condition	1	0.84 (0.70 – 0.95)	0.99 (0.99 – 1.00)
	3	0.80 (0.66 – 0.91)	0.99 (0.99 – 1.00)
	8	0.38 (0.07 – 0.80)	0.99 (0.99 – 1.00)
	9	0.90 (0.76 – 0.99)	0.99 (0.99 – 1.00)
	New	0.74 (0.21 – 0.97)	0.99 (0.99 – 1.00)
Posture	1	0.75 (0.42 – 1.00)	0.99 (0.99 – 1.00)
	3	0.56 (0.30 – 0.82)	0.99 (0.98 – 1.00)
	8	0.70 (0.38 – 0.99)	0.99 (0.98 – 1.00)
	9	0.62 (0.36 – 0.87)	0.99 (0.98 – 1.00)
	New	0.67 (0.32 – 1.00)	0.99 (0.99 – 1.00)
Abdominal fill	1	0.96 (0.56 – 1.00)	0.99 (0.99 – 1.00)
	3	0.98 (0.75 – 1.00)	0.99 (0.98 – 1.00)
	8	0.98 (0.78 – 1.00)	0.99 (0.99 – 1.00)
	9	0.39 (0.12 – 0.71)	0.99 (0.99 – 1.00)
	New	0.91 (0.00 – 1.00)	0.99 (0.98 – 1.00)
Shivering	1	0.85 (0.41 – 1.00)	1.00 (1.00 – 1.00)
	3	0.56 (0.23 – 0.90)	1.00 (1.00 – 1.00)
	8	0.58 (0.01 – 1.00)	1.00 (1.00 – 1.00)
	9	0.37 (0.23 – 0.81)	1.00 (1.00 – 1.00)
	New	0.64 (0.00 – 1.00)	0.99 (1.00 – 1.00)
Eye condition	1	0.89 (0.80 – 0.89)	0.99 (0.98 – 0.99)
	3	0.87 (0.75 – 0.88)	0.99 (0.99 – 1.00)
	8	0.89 (0.79 – 0.89)	0.99 (0.98 – 1.00)
	9	0.86 (0.73 – 0.87)	0.99 (0.98 – 0.99)
	New	0.88 (0.77 – 0.97)	0.99 (0.99 – 1.00)

Figure 6.1 Differences in observer scores of the demeanour, body condition, lameness and eye condition of young lambs

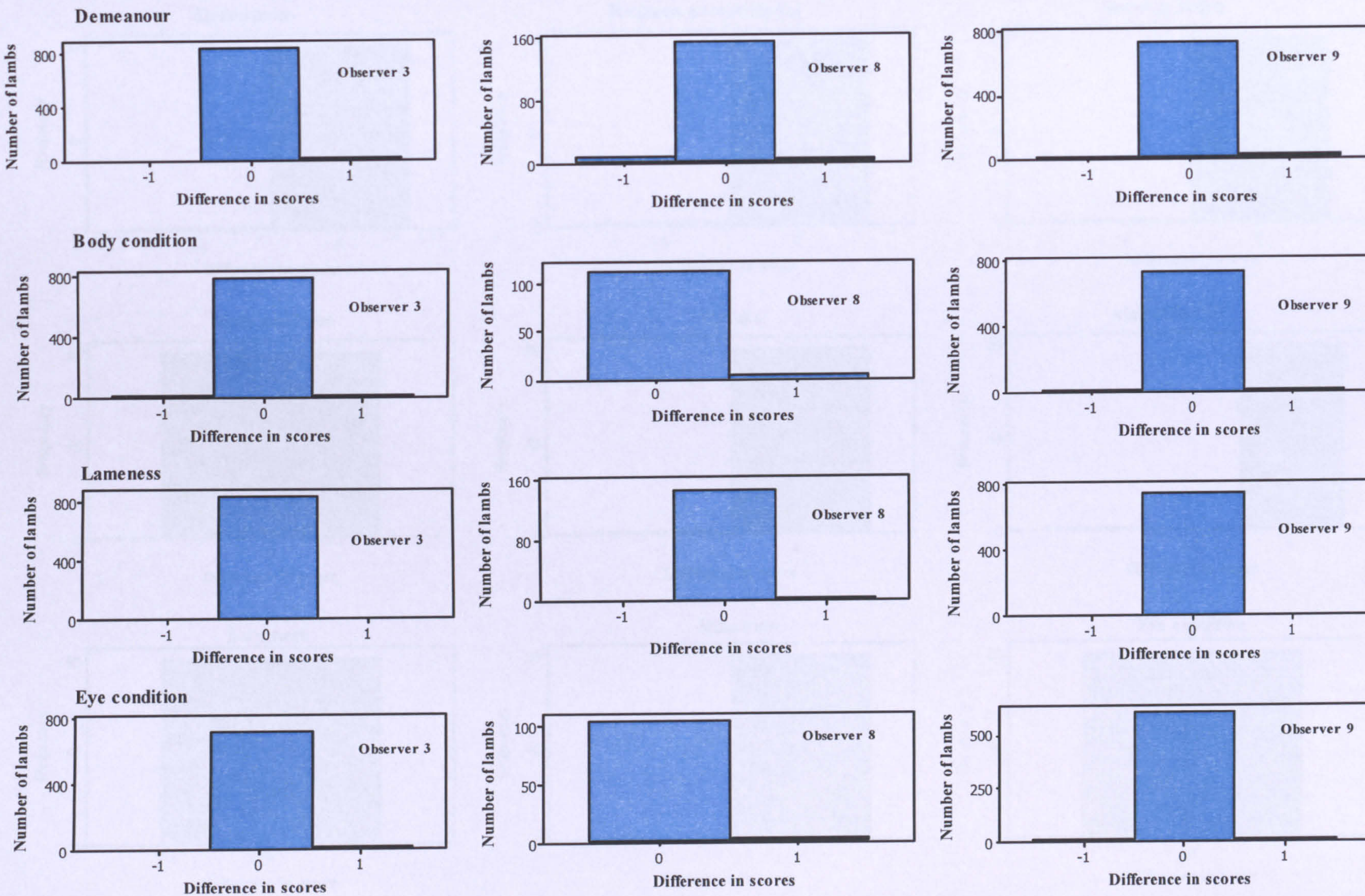


Figure 6.2 Differences in the repeat assessments of the test standard observer for young lamb welfare indicators

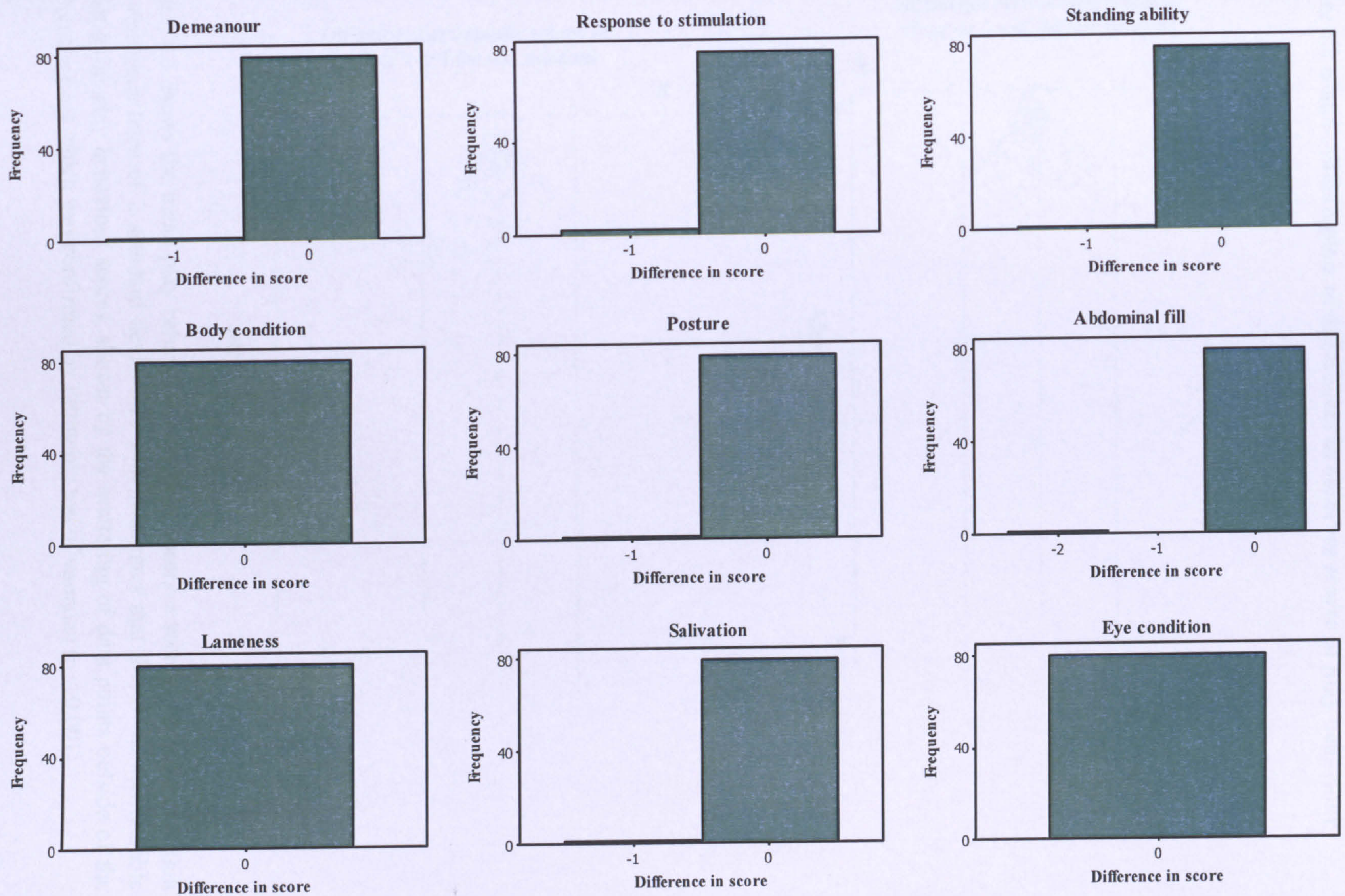
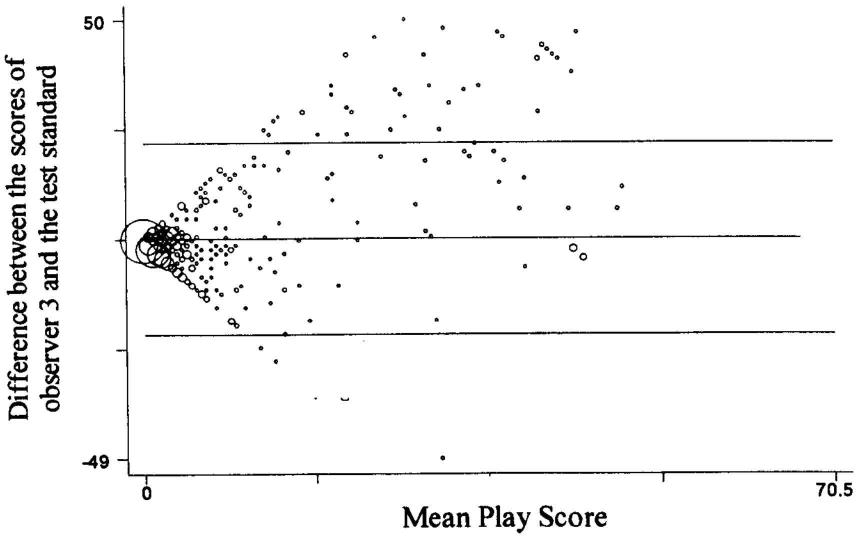


Figure 6.3 Bland-Altman plots of differences in observer scores of play behaviour

Observer 3



Observer 9

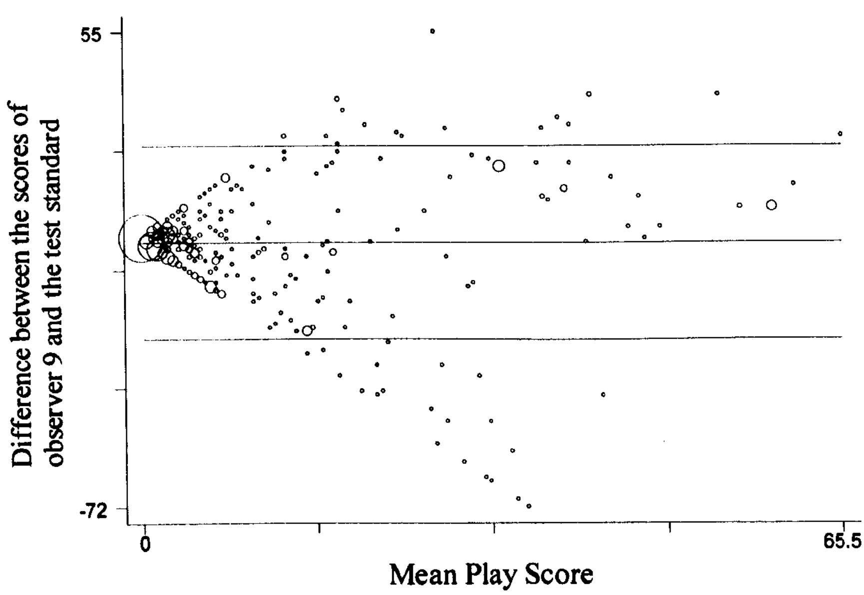


Figure 6.3 shows the mean play behaviour score and observer scoring differences within the reference lines of 2 standard deviations. This illustrates that there was considerable variation in play behaviour scores, shown by the scattering of data points outside of the references lines which was confirmed by Pittman's test of variance ($p < 0.001$).

6.4 Discussion

The objective of the studies presented in this chapter was to investigate the diagnostic test validity of 11 indicators of young lamb welfare that were assessed on the basis of an individual examination. These animal-based indicators were developed following consultation of the scientific literature and expert opinion (Chapters 1 and 2) as proxy measures of the welfare status of a lamb, used to identify specific conditions such as eye abnormalities, septic arthritis and malnutrition (Chapter 3). Principles used to examine the validity of diagnostic tests were used in this thesis to investigate the test performance of each indictor. Accordingly, the reliability, Se and Sp of a group of observers was examined against a test standard observer. Overall, good levels of diagnostic validity for most animal-based indicators of young lamb welfare were achieved during the inter- and intra-observer studies. However, the diagnostic test performance of some indicators may have been influenced by the study population, observer population, indicator scoring scale and analysis of test validity – factors that are considered in the following discussion.

The 17 lambing flocks were selected according to their consent to participate, location, management type and the feasibility for assessment within the limits of a one-day farm visit. Given the non-random selection of farms, this population may be biased towards farms of higher welfare status or those with regular contact with a veterinary surgeon. However, the objective of this chapter was to investigate the reliability, Se and Sp of the indicators, rather than infer conclusions regarding the on-farm welfare of the general sheep population. Therefore, the test performance of each indicator was examined on lambs of different breeds, reared under different systems and managed in both indoor and outdoor conditions.

All observers performed conditionally independent assessments and were blinded to historical and clinical information regarding the extent of on-farm welfare conditions. However, due to the study setting it was not possible to blind observers to additional cues such as farm cleanliness and hygiene, and presence of lambing equipment or medicines, which may have alerted observers to the presence of certain on-farm welfare issues. As biological variation may have affected the outcome of some measures, the repeat assessments of the test standard observer were limited to a single day.

The method of analysing the test performance of the young lamb welfare indicators may also have influenced the level of reliability and Se and Sp achieved in the on-farm studies.

The test performance of each indicator was examined using methods of analysis that are currently used in animal welfare research. Percentage (%) agreement was selected as it provides a good starting point for understanding observer reliability and revealed a high level of observer agreement. However, this approach does not take into account the amount of agreement that occurs due to chance alone (Sim and Wright, 2005). Therefore, kappa (κ), a method routinely used to assess the reliability of animal welfare indicators (Burn *et al.*, 2009; Kaler *et al.*, 2009) was selected to evaluate the overall level of observer agreement and the agreement with the test standard observer ('gold standard'). Although, it was evident that the low level of young lambs observed with any one of the welfare conditions may have affected the interpretation of the level of observer agreement determined by κ analysis (Feinstein and Cicchetti, 1990).

In the absence of a reference standard for animal welfare assessment, the author was selected as the test standard observer. Additional observers were then compared (3, 8 and 9) against this standard. Results clearly showed that the test standard achieved the highest levels of reliability for indicator assessments in both inter- and intra-observer studies. Since this observer developed the assessment methods and trained other assessors, they were expected to have a greater level of experience and understanding of indicator case definitions. The consistency in intra-observer assessments may also be due to fewer cases of misidentified sample lambs and the ease of assessing lambs in individual pens compared to assessments performed in outdoor environments.

As well as being reliable, the welfare indicators should have acceptable levels of diagnostic Se and Sp. Therefore, the test validity of young lamb welfare indicators was also examined in terms of the Se and Sp of each observer. LCA was a useful method for evaluating test performance as it did not require a comparative reference standard (Hui and Walter, 1990). As LCA identified that the highest test performance was achieved by the test standard observer this suggested that the use of this assessor as the 'gold standard' and provider of training was appropriate for this study. LCA also provided another advantage over classical evaluation of test performance by predicting Se and Sp of 'new' observers – unknown assessors who may apply these indicators in the future.

The level of Se is determined by examining how many animals with the condition of interest are detected by the test. In this chapter both the test standard observer and other study observers did not identify a large number of lambs with thin body condition or abnormal abdominal fill and this is suggested to be the reason why a low Se was obtained for these indicator tests. However, it was evident that there were sufficient observations of lambs with lameness, eye condition and dull demeanour and this provided good levels of observer agreement, good Se and a high level of Sp for these indicator tests. In common with the measures for adult sheep and growing lamb welfare (Chapter 5), the results of this chapter suggest that young lamb indicators have higher levels of Sp, suggesting that they were better at identifying farms with good welfare standards. As the diagnostic test estimates were affected by the low proportion of lambs with most of the welfare conditions, it is likely that these indicators may have greater levels of diagnostic Se when tested on a population of lambs with a higher proportion of conditions associated with poor welfare. Whilst it may have been preferable to have selected a study population with a greater range of indicator scores (Hoehler, 2000), this approach was not feasible for this study since the on-farm welfare status of young lambs was previously unknown.

Shivering, was used as an indicator of hypothermia but as this condition was rarely observed in the study population, this resulted in a low diagnostic Se. Given the clearly observable signs of shivering, the low number of lambs observed was considered to be an indication of the true level of these conditions on the 17 study farms rather than being due to missed observations. In addition, the interpretation of the relevance of shivering as an indicator of welfare has not been elucidated by this study. Although stockpeople and veterinary surgeons routinely assess the thermal comfort and health of young lambs by identifying signs of shivering, this behaviour also plays a role in the physiological response to tolerance of colder environments. Therefore, the observation of shivering in young lambs may not always be linked to deficits in on-farm management or the provision of specific resources, such as shelter.

In addition to the features of the study population, the test performance of the indicators presented in this chapter, may also have been affected by the observer characteristics, such as occupation, previous experience and the level of training. A pool of 4 trained observers of different occupations and with varying levels of experience was included in this study to investigate whether different types of assessor could consistently apply the welfare measures. However, the low number of sup-optimal welfare conditions and the small number of observers used in this study meant that no conclusive findings on the effect of experience or occupation can be presented. There were some interesting data trends which are described, but further work is needed to ascertain the effect of observer experience and training on the level of diagnostic ability.

Veterinary surgeons - observer 8 and the test standard observer, were considered to be the most experienced assessors. Both observers achieved an excellent level of reliability and good Sp for many indicators, including the assessment of posture. By contrast, lower levels were produced by the inexperienced and non-veterinary observers 3 and 9. This may suggest that posture is a clinical sign that is readily used by veterinary assessors or may reflect the level of experience in examining the health and welfare of young lambs. This example of variation in test performance may indicate that there are also differences in the observational skills of different types of farm assessors.

However, the category of experienced or veterinary observer did not always result in a higher level of test performance. High levels of reliability, Se and Sp were produced for body condition assessments by observers 1, 3 and 9. The lower reliability and Se of observer 8 may suggest that factors other than occupation affected the level of observer agreement reported here.

Ocular abnormalities were clearly well recognised by all observers. As inexperienced observers were previously unfamiliar with the assessment of eye conditions, the level of Se and Sp identified by LCA may indicate the value of on-farm training and discussion of scoring scales that were used to train the observers. With the exception of observer 9, the assessment of abdominal fill produced good inter-observer agreement and high Sp. This may have been due to difficulties with assessing abdominal fill as, following a large feed, enlargement of the sub-lumbar fossa can occur in healthy lambs. Therefore, there was the potential for misclassification of indicator scores and additional training could also be valuable for this indicator. It may also be preferable to alter the scoring system to a binary scale in order to classify lambs with 'normal' and 'abnormal' abdominal fill. This simpler scoring system would still cover the welfare concerns regarding orphan lambs that ingest excess amounts of milk substitute (bloated abdominal) or starvation of any lambs (hollow abdominal fill).

The use of a smaller number of indicators and simpler scoring systems may have been the reason why the young lamb welfare measures achieved considerably higher levels of reliability, Se and Sp compared to adult sheep and growing lamb indicator (Chapter 5). Previously, ordinal lameness scoring systems of sheep have achieved poor levels of observer reliability (Harkins, 2005; Welsh *et al.,* 1993). Therefore, a simple binary scoring system was developed to distinguish between 'sound' and 'lame' animals (Chapter 3). The high level of observer reliability achieved in the present study may be attributed to the clear and simple scoring system used to assess lameness in young lambs.

In addition, it was much easier to perform an individual gait examination in young lambs in comparison to the individual lameness assessment of sheep in later stages of the production cycle. Unlike adult sheep and growing lambs, young lambs did not appear to react as strongly to the presence of observers and handling. Evidently, the smaller body size meant that handling and gait assessment of individual lambs was easier and quicker to perform compared to lameness scoring of older animals. It is also suggested that young lambs may not have a learned aversion to handling and certain management procedures, so they may not mask painful conditions in the same way as adult sheep (Fitzpatrick *et al.,* 2006).

The level of test performance may also have been affected by the severity of the welfare condition observed in young lambs. Septic arthritis produces severe pathological changes in synovial joints and is perceived to be a very painful condition (Angus, 1991). The pathology results in stiff, joint swellings and severe gait abnormalities and the severe clinical signs demonstrated by affected lambs may be the reason why observers could consistently identify lame lambs. It is also possible that observers have a different perception of the effects of certain conditions for neonatal and young lamb welfare, compared to the effects they may have on animals in the later stages of production.

Additional guidance on young lamb assessments was provided as it was recognised that demeanour may not be as readily identified in lambs compared to adult sheep. This is because it can be difficult to determine the difference between a healthy, sleeping lamb and a somulosed, dull, depressed lamb of poor welfare status. Therefore, the responsiveness of young lambs to stimuli, such as movement or palpation by the assessor, appeared to be a particularly useful tool for assessing the demeanour and responsiveness of indoor-housed lambs. On-farm experience suggests that it would be possible to reduce demeanour and response to stimulation into a single indicator by combining the scoring descriptions. In this way lambs could be scored as bright, alert and responsive to stimulation, or dull, depressed and unresponsive to stimulation.

Many of the indicators tested in this chapter have been focused on physical signs of poor health and welfare, for example the presence of lameness or an ocular abnormality. An advantage of including play behaviour was that it assessed positive aspects of lamb welfare and was easy to assess in housing and outdoor environments. The failure to observe this behaviour during a one-minute sampling period did not provide any meaningful interpretation of lamb welfare and the brevity of the assessment period may be the reason why there was variation in observer scores. The interpretation of play behaviour may also be complicated by other factors such as management, breed, age and genotype. Therefore, given the issues with scoring and variation in observer agreement, play behaviour was not tested as an indicator of young lamb welfare in subsequent studies of this thesis.

6.5 Conclusion

Overall this study identified that trained assessors achieved good levels of test validity for most indicators of young lamb welfare. However, the sensitivity of most of the indicators was affected by the low level of welfare conditions such as hypothermia and starvation in the study population. Given that the tests were capable of detecting the conditions on a sample population with few welfare issues, it is likely that the indicators would perform even better if applied to a young lamb population with a higher level of sub-optimal welfare conditions. Therefore, a valuable step in the further development of these indicators would be to evaluate the diagnostic performance of these measures in a sample population with a greater proportion of on-farm welfare issues.

Chapter 7

VALIDATING THE ABILITY OF INDICATORS TO DETECT FARM VARIATION IN SHEEP WELFARE

7.1 Introduction

The overall aim of this thesis was to develop valid, reliable and feasible indicators of sheep welfare that were robust enough to be applied under working farm conditions. If a test is to be both useful and feasible, it is essential that it is evaluated under field conditions against a known reference test or 'gold standard' and not solely validated under controlled, laboratory conditions. In the absence of a reference test for animal welfare assessment (de Passillé and Rushen, 2005) the author was assumed to be the 'test standard' (Burn *et al.*, 2009). This approach was used to investigate the validity of animal-based measures of sheep welfare in terms of the reliability, sensitivity and specificity of each indicator (Chapters 4 – 6).

It is known that the proportion of sheep affected by conditions such as lameness (Kaler and Green, 2009a), myiasis (French *et al.*, 1994a) and sheep scab (Cross *et al.*, 2010) varies between different flocks. Therefore, animal-based indicators should be validated in terms of their ability to detect the underlying between-farm variation in the proportion of sheep affected by specific health and welfare conditions. In this thesis, the study population was selected on a convenience basis but selection was such that the study farms were representative of the British Sheep stratification system (Chapter 3). As such, the validity of the indicators also needs to be addressed in the context of asking whether the measures are valid when applied under different management systems and a range of on-farm conditions.

Therefore, the objective of this chapter was to investigate the feasibility of animal-, resource- and management-based indicators and to test the hypothesis that, given this known variation, the animal-based measures are capable of detecting between-farm differences in the level of welfare conditions, such as lameness, thin body condition and mastitis, on a range of farms.

7.2 Materials and methods

7.2.1 Assessment of welfare indicators

The investigation was a cross-sectional study in which 50 farms, further described in Chapter 3, were each visited once during the period July 2008 to May 2009. The animal-based indicators were tested on a total of 4848 animals including 3167 adult and growing sheep and 1681 young lambs. This sample population comprised a range of pure-, half- and cross-breeds of sheep (Chapter 3). The indicators of adult and growing sheep were assessed on a range of 24 – 120 sheep (median 69 sheep) on each study farm using two methods of observation; group observation and individual animal examination (Chapter 3). The age distribution of the adult sheep and growing lamb sample population was 1514 animals (47.8%) > 4 years, 1051 (33.2%) > 1 – 3 years, and 554 (17.5%) 3 – 7 months.

The young lamb indicators were assessed on a range of 30 – 90 lambs (median 59) using an individual examination, as described in Chapter 3. Assessments were carried out on a total of 538 lambs managed outdoors (32%), 426 were housed indoors in groups (25.3%), and 717 managed in an individual lambing pen (42.7%). 1494 lambs (89%) were observed on the day of the visit as being reared with a ewe, 65 lambs were found to be reared with a ewe in a head-yoke (3.9 %) and 122 lambs were classed as orphan lambs (7.3 %). The age distribution of the young lamb sample population – estimated according to farmer reports and records was 0 – ≤ 3 days old (35%), 4 – 7 days old (32 %), and >1 – 6 weeks old (31%).

In addition, resource- and management-based indictors described in Chapter 3 were assessed by measuring aspects of housing and grazing facilities and performing a brief farm interview. All indicator assessments were performed by the author.

7.2.2 Statistical analysis

Data was analysed using Stata version 10 (StataCorp LP, College Station, Texas). The farm-level proportion (%) of each animal-based indicator was graphically represented and standard descriptive statistics were used to describe the outcome of resource- and management-based welfare indicator assessments. Proportion values were calculated with 95% confidence intervals (CI) using Huber–White robust standard error estimates to account for farm-level clustering (Kirkwood and Sterne, 2003). The correlation between different assessment methods for the same indicator, for example, the assessment of lameness by group observation and individual examination was examined graphically and using Spearman's Rank Correlation Coefficient (rho).

7.3 Results

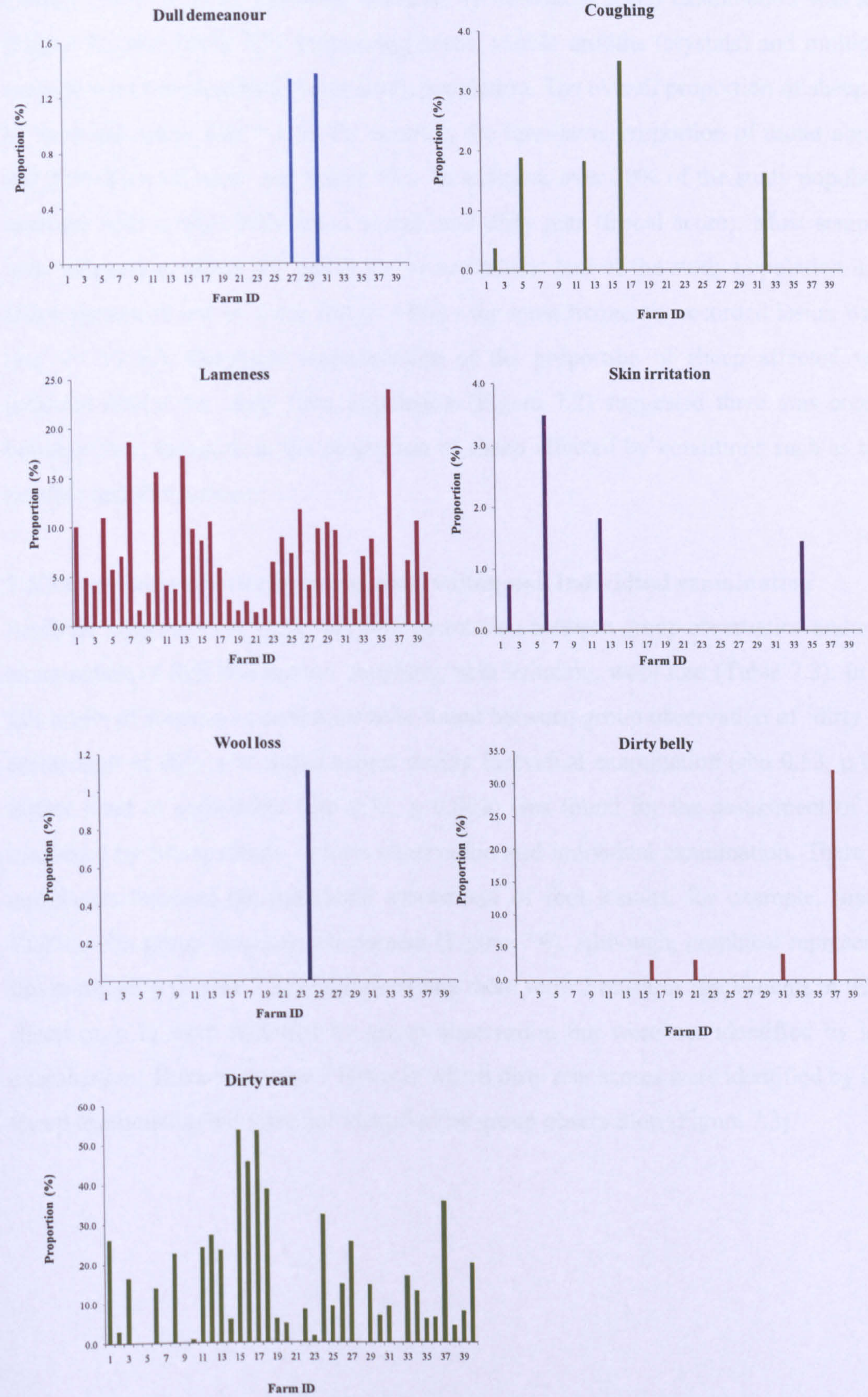
7.3.1. Indicators assessed by group observation

Overall, a low level of all group indicators was observed across the study population (Figure 7.1 and Table 7.1). No observations of excessive panting were made and the proportion of dull demeanour, coughing and skin irritation on all study farms was below 0.5%. Less than 5 % of sheep were identified with wool loss and less than 10 % of the sample population was identified as 'lame'. Over half of the study population (54.4%) was assessed as having a 'dirty rear'. Graphical representation of the farm-level prevalence of each indicator (Figure 7.2) suggested that there was considerable between-farm variation in the proportion of sheep observed with each indicator outcome.

Table 7.1 Proportion of group indicator scores across study population

Indicator	Proportion (%)	Robust 95% CI
Dull demeanour	0.06	- 0.03 – 0.17
Coughing	0.19	0.02 – 0.37
Excessive panting	0	
Skin irritation	0.27	-0.05 – 0.60
Wool loss	4.88	-2.01 – 11.76
Lameness	7.12	5.49 – 8.76
Dirty rear	54.42	29.93 – 78.91
Dirty belly	1.35	-0.72 – 3.41

Figure 7.1 Proportion of indicators assessed by group observation on each study farm



7.3.2 Indicators assessed by individual sheep examination

Overall, a low level of indicators assessed by individual sheep examination was identified (Figure 7.2 and Table 7.2). In-growing horns, visible uroliths (crystals) and multiple, open wounds were not identified in this study population. The overall proportion of sheep affected by most indicators was < 1 %. By contrast, the farm-level proportion of molar abnormality and dirty leg mud score was nearly 5%. In addition, over 20% of the study population was assessed with a dirty belly (mud score), and dirty rear (faecal score). Most sample sheep were assessed as either 'fit' (BCS 2 – 4) and almost half of the study population had a foot lesion present in one or more feet (< 44%) - the most frequently recorded lesion was 'white line' (< 50 %). Graphical representation of the proportion of sheep affected with each indicator across the study farm population (Figure 7.2) suggested there was considerable between-farm variation in the proportion of sheep affected by conditions such as lameness, mastitis and foot lesions.

7.3.3 Correlation between group observation and individual examination

Analysis suggested that there was poor correlation between group observation and individual examination of dull demeanour, coughing, skin irritation, wool loss (Table 7.3). In contrast, fair levels of statistical correlation were found between group observation of 'dirty rear' and assessment of dirty rear faecal scores during individual examination (rho 0.53, p 0.001). A higher level of correlation (rho 0.72, p 0.001) was found for the assessment of lameness measured by two methods - group observation and individual examination. There was also correlation between the individual assessment of foot lesions, for example, foot rot and CODD with group lameness assessment (Figure 7.4). Although, graphical representation of this correlation (Figure 7.3) highlighted that there were 7 on-farm assessments in which lame sheep ($n \geq 1$) were recorded by group observation but were not identified by individual examination. There were also 7 farms in which dirty rear scores were identified by individual sheep examination but were not identified by group observation (Figure 7.3).

Table 7.2 Proportion of individual sheep welfare indicator scores across study population

Indicator	Mean proportion (%)	Robust 95% CI
Dull demeanour	0.55	0.24 – 0.87
Eye condition	0.71	0.21 – 1.22
Nasal discharge	0.71	0.21 – 1.22
Tooth disease – incisor loss	3.73	2.46 – 5.01
Tooth disease – molar abnormality	6.89	3.43 – 10.35
Tooth disease – incisor and molar abnormality	3.22	1.37 – 5.06
Coughing	0.29	0.05 – 0.50
Ear lesion	4.06	1.24 – 6.89
In-growing horns	0	
Dirty belly – mud score	13.35	4.83 – 21.87
Dirty belly – faecal score	0.16	-0.02 – 0.35
Dirty legs – mud score	5.33	-0.68 – 11.34
Dirty legs – faecal score	1.10	0.41 – 1.80
Dirty rear – mud	3.28	-0.27 – 6.84
Dirty rear – faecal	17.67	12.90 – 22.45
Dirty rear – filthy faecal	2.14	0.88 – 3.40
Mastitis – single gland	3.96	2.10 – 5.80
Mastitis – both glands	1.51	0.58 – 2.43
Crystals	0	
Short tail length	3.70	0.74 – 6.67
Wool loss	0.36	0.11 – 0.61
Skin irritation	0.84	-0.34 – 2.07
Multiple skin lesions ≤ 5p size	0.32	-0.16 – 0.82
Single skin lesion > 5p < 50 p size	1.33	0.58 – 2.08
Multiple skin lesions ≥ 50 p size < hand-size	1.53	0.57 – 2.48
Single hand-size skin lesion of hand-size	0.55	0.29 – 0.87
Multiple hand-size skin lesions	0.49	0.17 – 0.81
Diffuse skin lesion	0.32	-0.17 – 0.82
≤ 5 superficial scratches	0.26	0.07 – 0.44
> 5 superficial scratches	0.03	-0.03 – 0.10
Healing wound(s)	1.49	0.31 – 2.67
Single open wound	0.58	0.18 – 0.99
Multiple open wounds	0.00	
Body condition score 1	0.81	0.18 – 1.44
Body condition score 2	26.59	20.75 – 32.44
Body condition score 3	52.28	47.23 – 57.23
Body condition score 4	15.31	12.01 – 18.61
Body condition score 5	5.01	1.73 – 8.29
Fit-Fat-Thin – 'thin' score	0.71	0.12 – 1.31
Fit-Fat-Thin – 'fat' score	4.58	1.42 – 7.75
Foot lesion	48.02	38.87 – 75.42
Significant foot lesion	5.69	3.84 – 7.54
Lame	6.18	4.26 – 8.09
White line lesion	43.92	34.68 – 53.15
Scald	1.01	0.30 – 1.71
Footrot	1.27	0.59 – 1.95
CODD	1.01	0.36 – 1.66
Toe granuloma	1.11	0.61 – 1.60
Joint swelling	0.46	0.10 – 0.69
Myiasis	0.33	-0.11 – 0.70

Figure 7.2 Proportion of individual sheep indicator scores on each study farm

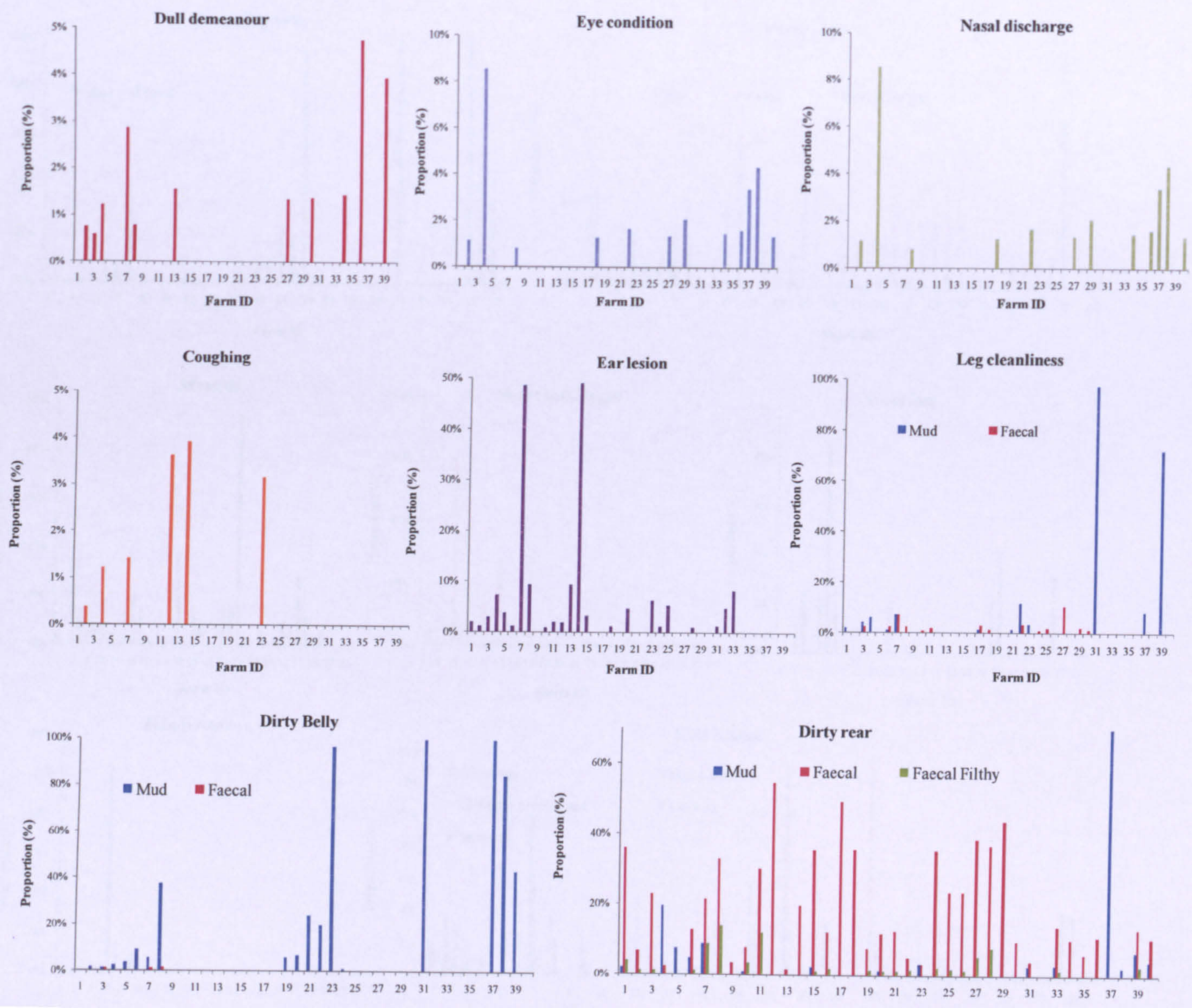


Figure 7.2 continued

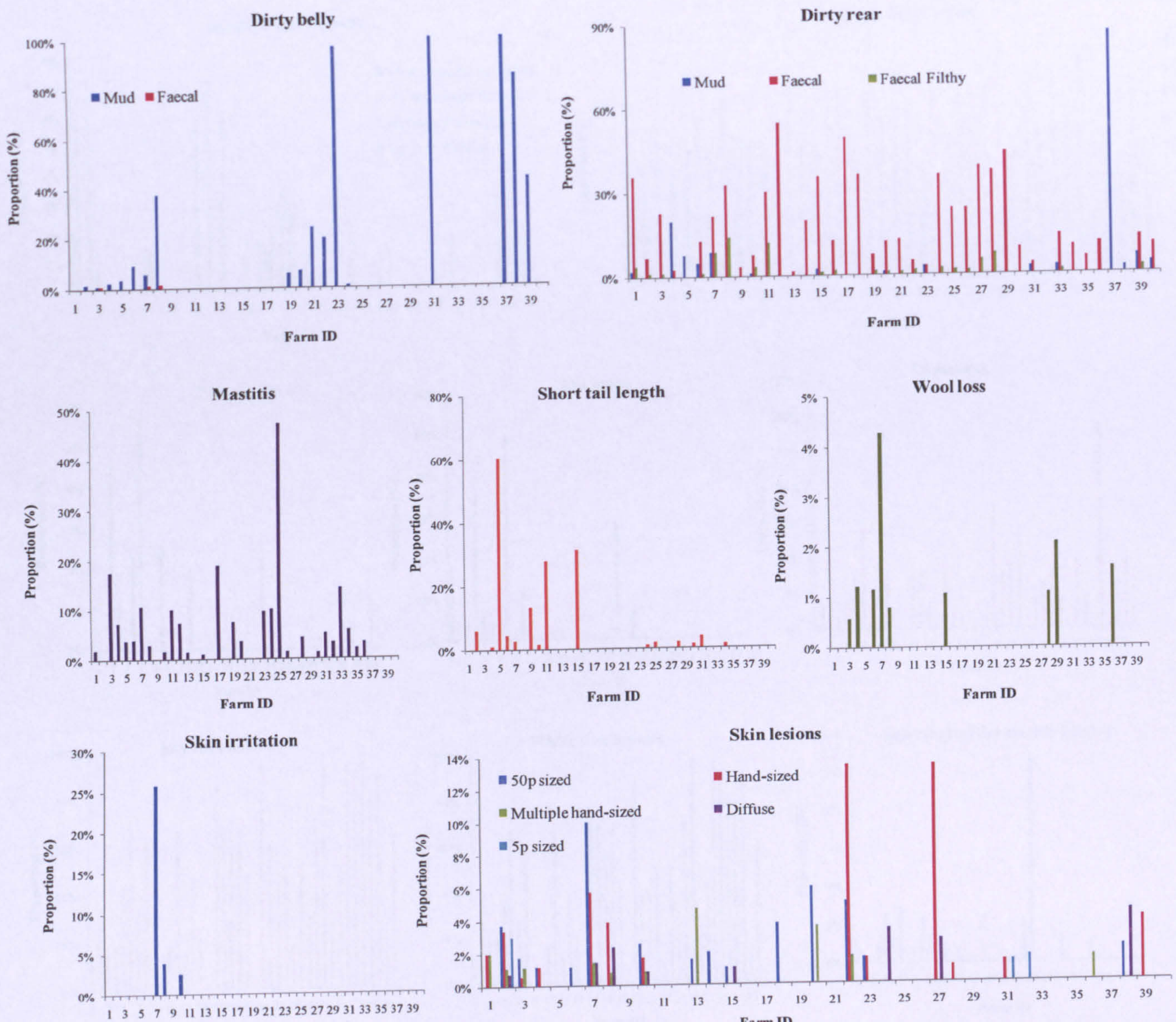


Figure 7.2 continued

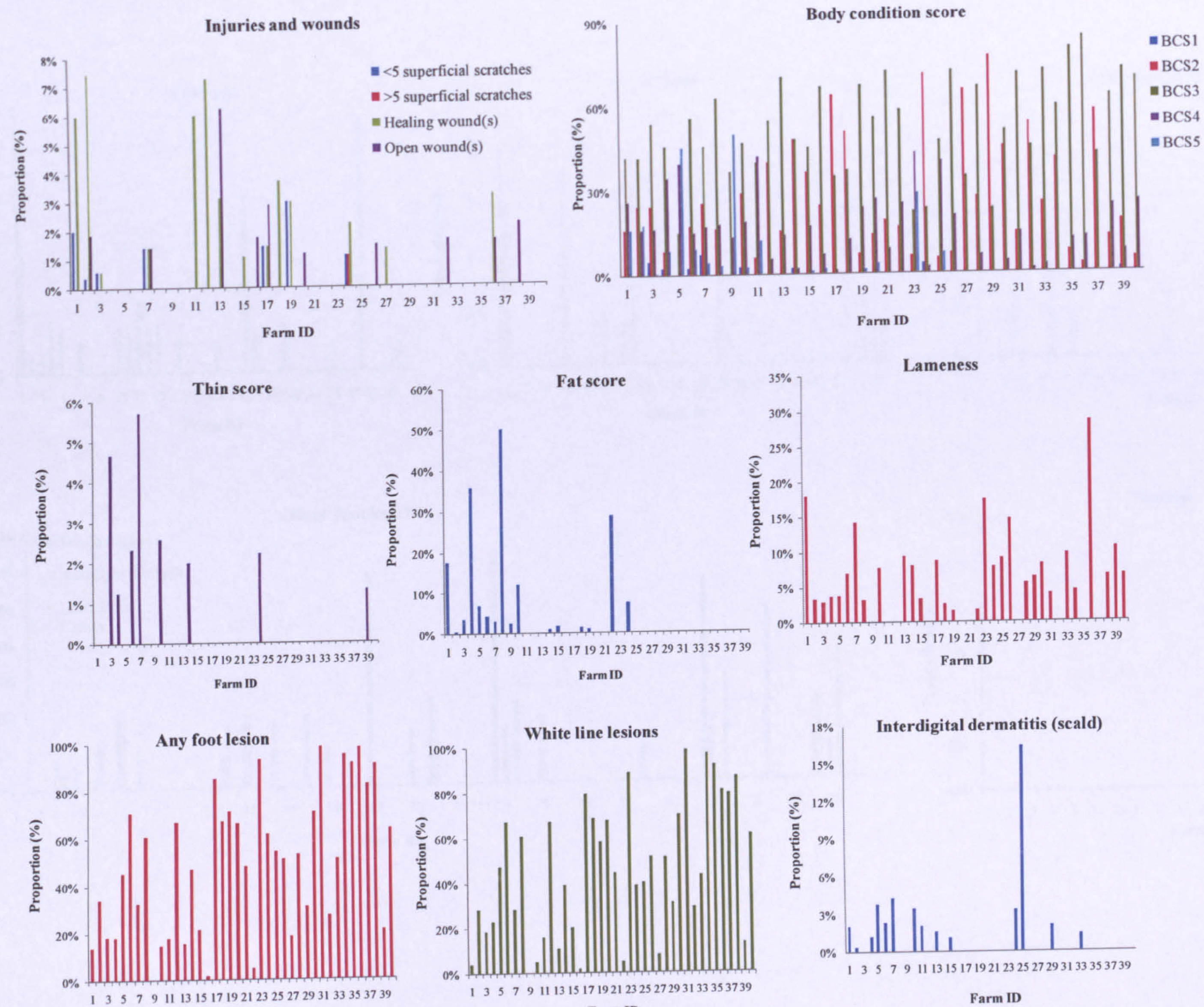


Figure 7.2 continued

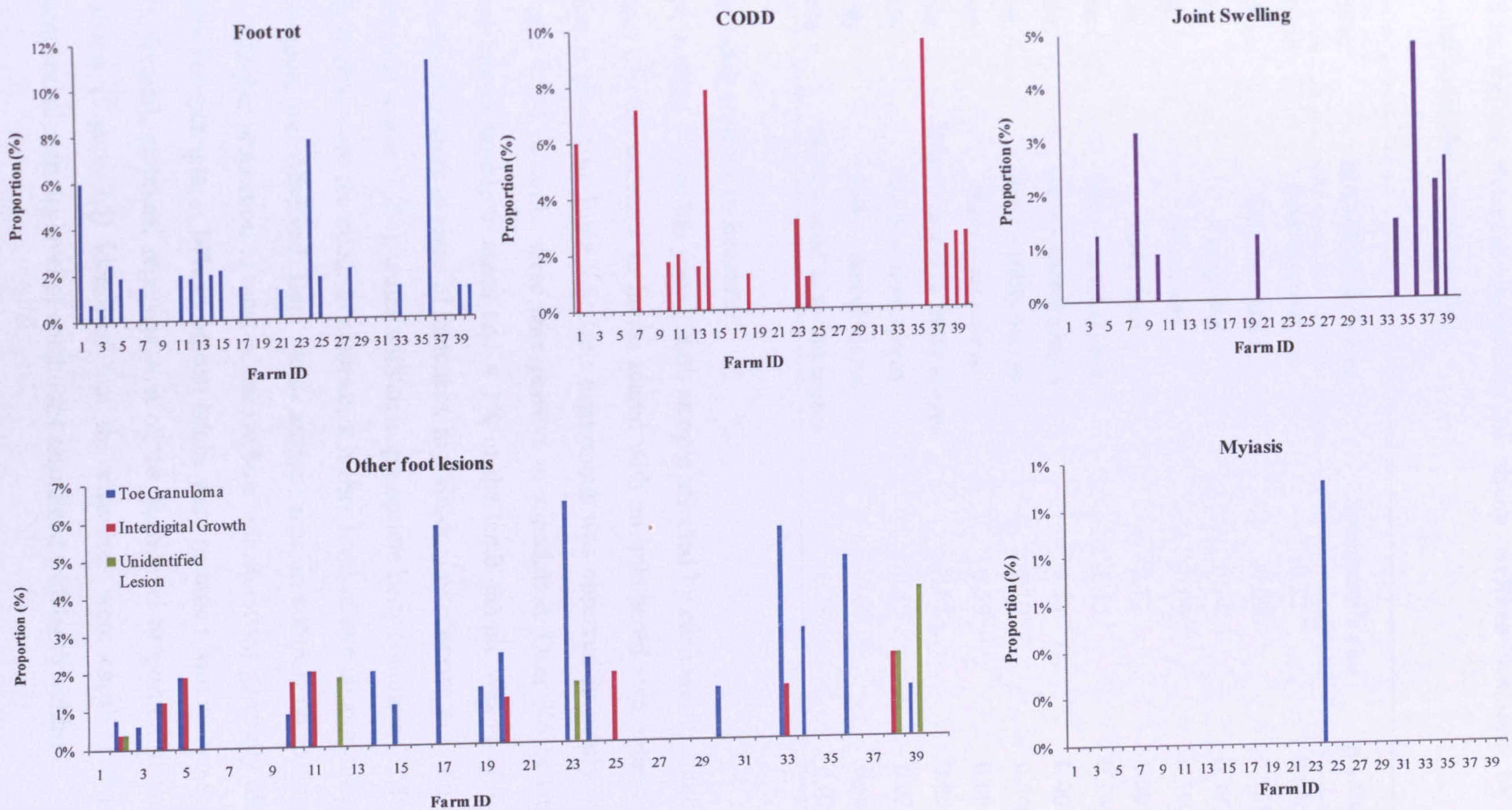


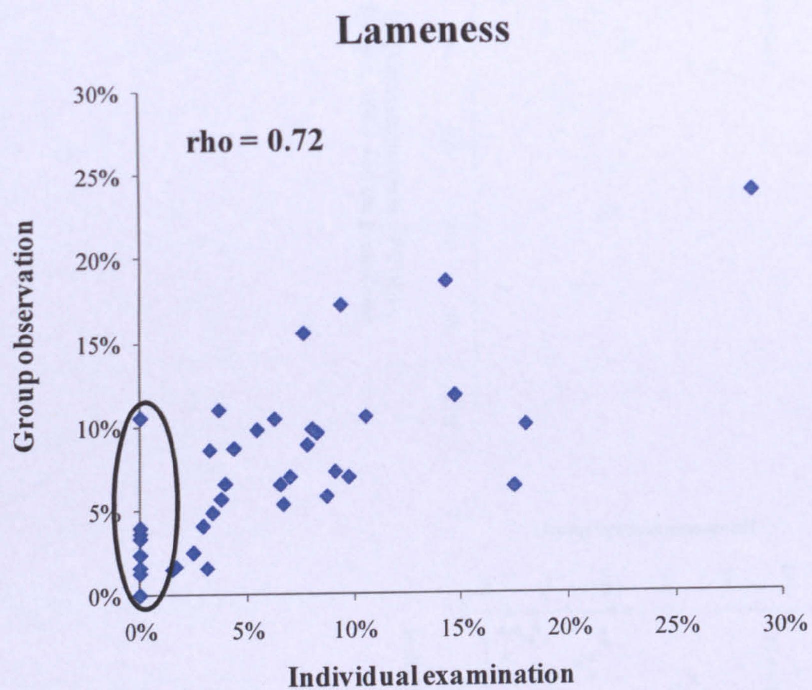
Table 7.3 Correlation between indicators of sheep welfare assessed by group observation and individual examination

Group indicator	Individual indicator	Spearman's rho	p-value
Dull demeanour	Dull demeanour	0.32	0.046
Skin irritation	Skin irritation	-0.09	0.561
Coughing	Coughing	0.14	0.397
Wool loss	Wool loss	-0.09	0.599
Lameness	Lameness	0.72	0.001
Dirty rear	Rear – mud scores	-0.15	0.365
Dirty rear	Rear – faecal scores	0.53	0.001
Dirty rear	Rear – filthy scores	0.10	0.540
Dirty rear	Rear – all scores	0.55	0.001
Dirty rear	Rear – faecal & filthy scores	0.51	0.001
Dirty belly	Belly – mud scores	0.35	0.027
Dirty belly	Belly – faecal scores	-0.09	0.561
Dirty belly	Belly – mud & faecal scores	0.35	0.027

7.3.4 Young lamb welfare indicators

The mean proportion (%) of the young lamb sample affected by each welfare outcome and the proportion of each outcome in lambs reared with an untethered ewe, tethered ewe or orphan lambs is shown in Table 7.4. Dull demeanour was observed in nearly 3% of the population and < 2% of lambs were unresponsive to stimulation. Over 3% of lambs were weak on standing or unable to stand and < 2% of the lamb sample was identified with a hunched/tucked-up posture or signs of lameness. In addition, few observations of bloated and hollow abdominal scores (1.2%), lambs with an inappropriate body condition (3.5%) or signs of shivering (0.6%) were recorded. In contrast, a higher level of eye abnormalities (> 5%), such as entropion, was observed. Data trends suggest that, in comparison to lambs reared with a ewe, a higher proportion of most of the welfare indicators was generally identified in orphan lambs For example, > 14% of orphan lambs were recorded with an eye abnormality (Table 7.4). Overall, graphical representation of the farm-level proportion of young lamb indicator scores (Figure 7.5) identified that the measures were capable of identifying between-farm variation in the level of conditions associated with lamb welfare.

Figure 7.3 Correlation between sheep welfare indicators assessed by group observation and individual examination



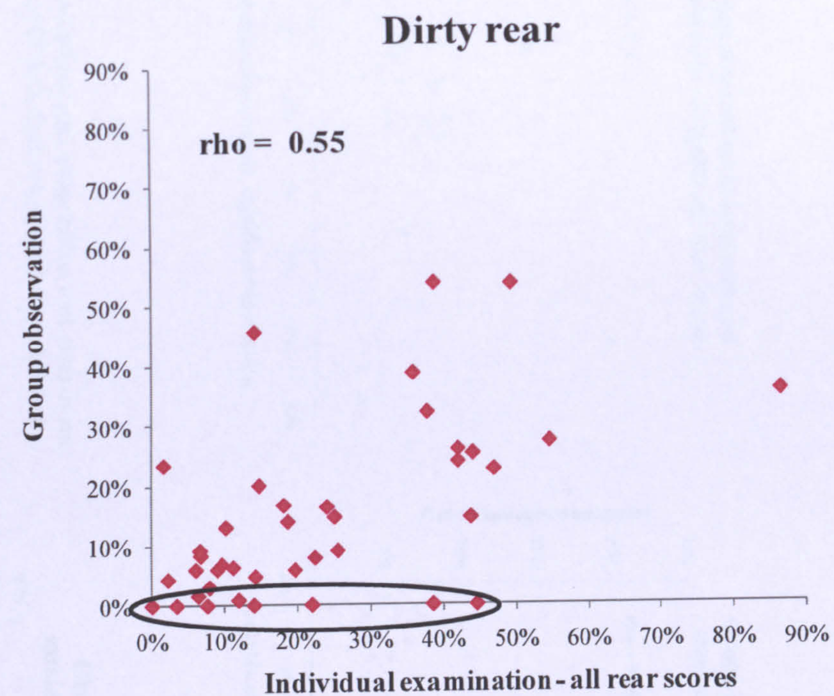


Figure 7.3 shows that there were farms in which the percentage of lame sheep assessed on each farm differed according to the method of assessment. Clear differences in the percentage of lame sheep assessed by group observation and individual examination are shown within the black oval (left). This highlighted 7 farm assessments in which lame sheep were identified by group observation but not by individual gait assessment. Similarly there were 7 farms in which group observation did not identify the same percentage of rear dirtiness as per individual examination, outlined by the black oval (right).

Figure 7.4 Correlation of group lameness assessment with foot lesion examination

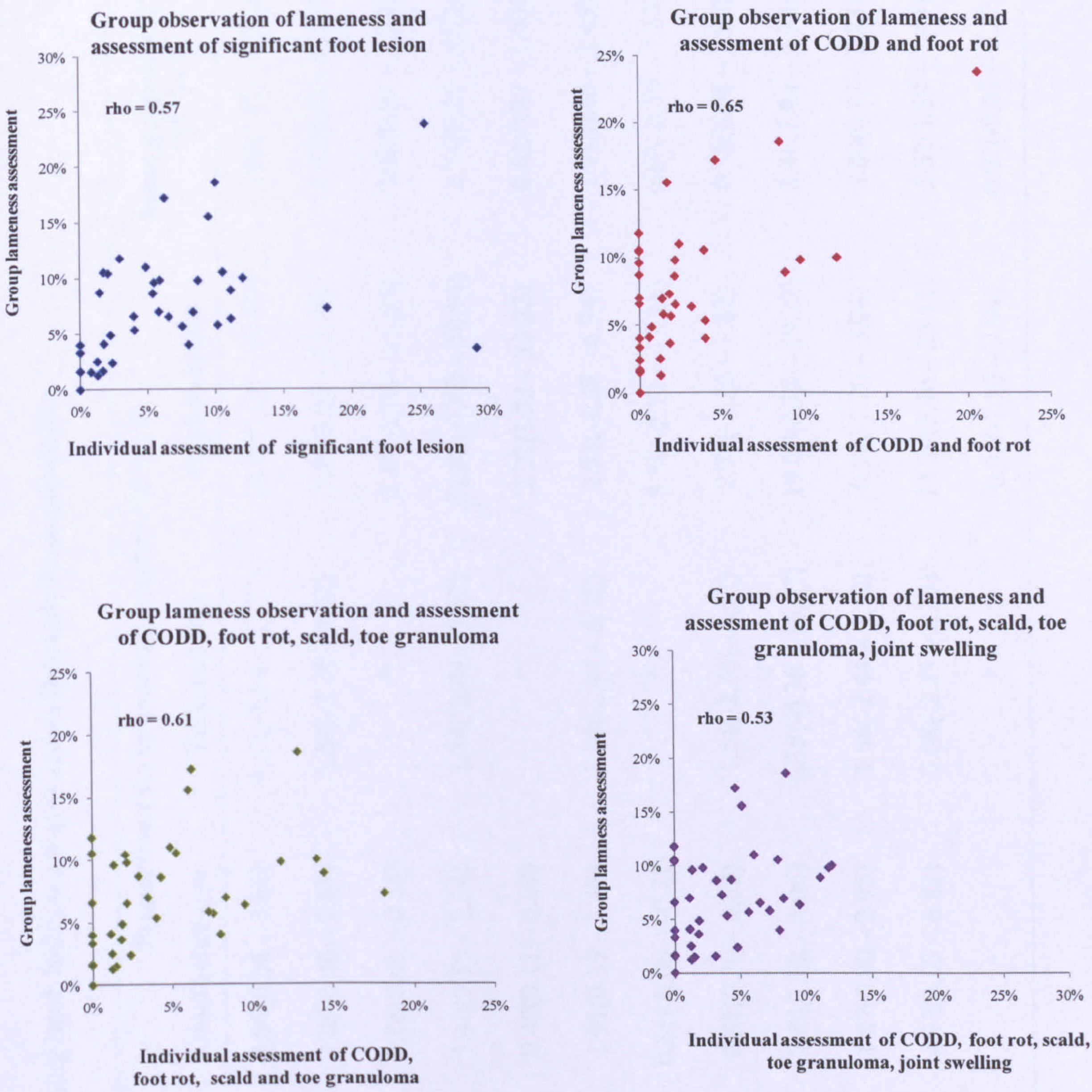
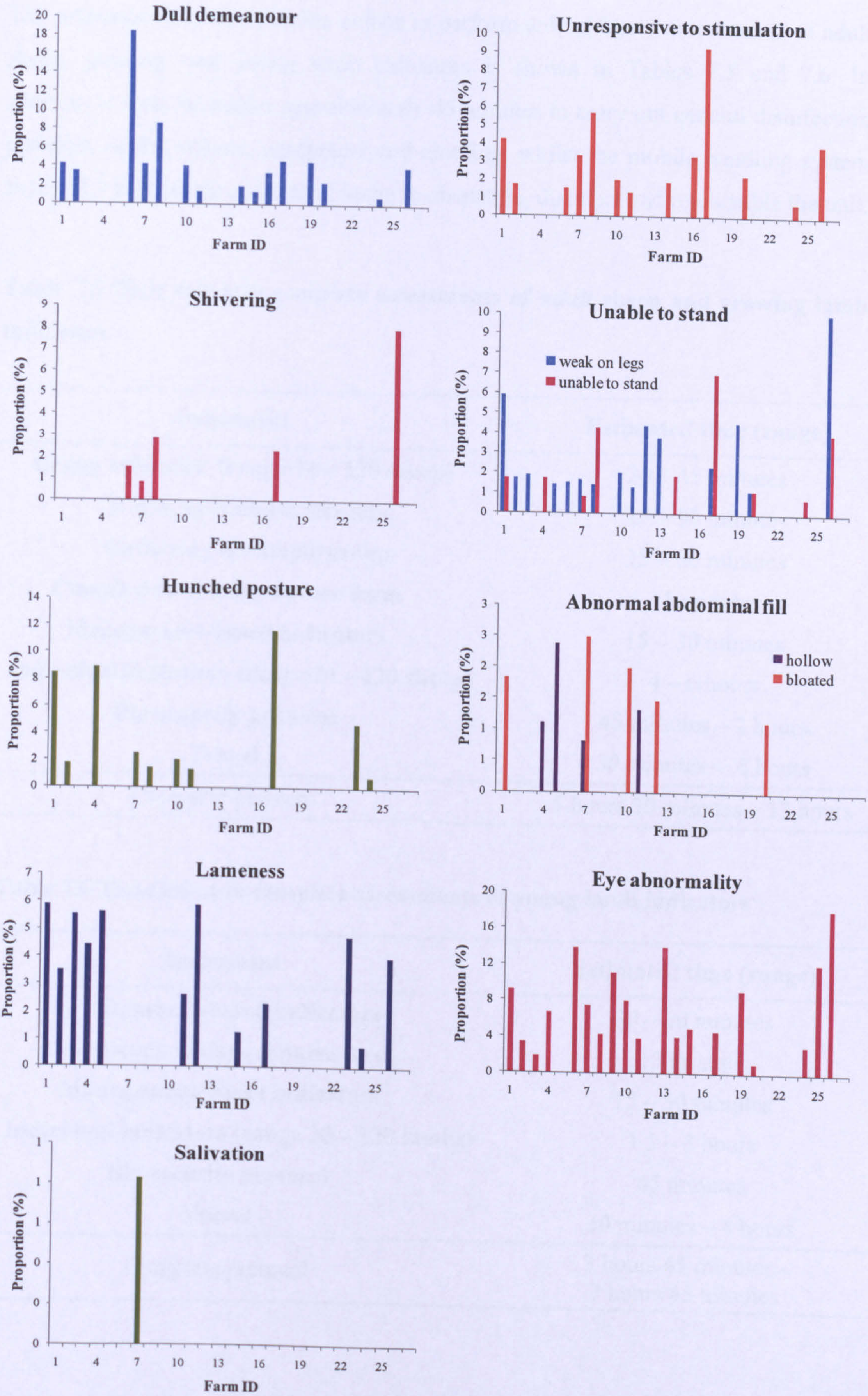


Table 7.4 Proportion of young lamb indicator scores across the study farm population

Indicator	Proportion (%) by rearing category (95 % CI)			Mean proportion(%) (95 % CI)
	Untethered ewe	Tethered ewe	Orphan lambs	
Dull demeanour	1.99(0.34 – 3.65)	4.62(-4.54 – 13.77)	10.00(1.20 – 18.80)	2.62(1.02 – 4.22)
Unresponsive to stimulation	1.06(0.46 – 1.67)	1.54(-2.44 – 5.52)	3.59(2.25 – 17.75)	1.67(0.81 – 2.52)
Shivering	0.28(0.01 – 0.55)	a	5.45(-1.08 – 11.99)	0.62(0.03 – 1.21)
Standing ability – unstable	1.40(0.55 – 2.24)	1.54(-2.44 – 5.52)	21.82(-7.03 – 50.94)	2.74(0.26 – 5.22)
Standing ability – recumbent	0.40(0.11 – 0.69)	a	7.27(1.32 – 13.22)	0.83(0.27 – 1.40)
Tucked-up posture	1.91(0.39 – 3.43)	1.54(-2.44 – 5.52)	2.73(-1.29 – 6.74)	1.95(0.46 – 3.45)
Hollow abdominal fill	0.31(-0.08 – -0.71)	a	5.56(-2.13 – 13.24)	0.69(-0.19 – 1.57)
Bloated abdominal fill	0.31(0.30 – 0.60)	1.54(-2.44 – 5.52)	2.78(-3.26 – 8.82)	0.55(0.08 – 1.02)
Body condition	2.84(1.62 – 4.06)	6.15(-4.36 – 16.67)	10.00(0.78 – 19.22)	3.46(1.94 – 4.98)
Lameness	1.87(0.80 – 2.95)	1.54(-2.44 – 5.52)	0.92(-1.20 – 3.04)	1.80(0.77 – 2.83)
Eye abnormality	4.92(3.03 – 6.82)	3.08(-2.23 – 8.38)	14.29(3.46 – 25.11)	5.52(3.59 – 7.44)
Excessive salivation	a	a	0.93(-1.10 – 2.94)	0.07(-0.07 – 0.22)

Figure 7.5 Proportion of young lamb indicators on each study farm



7.3.5 Feasibility of indicator assessments

The estimated time taken by the author to perform animal-based assessments of adult sheep, growing, and young lamb indicators is shown in Tables 7.5 and 7.6. In addition, it took the author approximately 45 minutes to carry out general disinfection practices on the vehicle, equipment and clothing, whilst the mobile handling system required 3 to 4 people and took 2 hours to dismantle, disinfect and reassemble the unit.

Table 7.5 Time taken to complete assessments of adult sheep and growing lamb indicators

Assessment	Estimated time (range)
Group indicator (range 24 – 120 sheep)	20 – 45 minutes
Resource-based indicators	15 – 90 minutes
Gathering of sample group	15 – 60 minutes
Completion of data capture form	15 minutes
Management-based indicators	15 – 30 minutes
Individual indicators (range 24 – 120 sheep)	4 – 6 hours
Bio-security protocol	45 minutes – 2 hours
Travel	30 minutes – 6 hours
Complete protocol	5 hours 30 minutes – 13 hours

Table 7.6 Time taken to complete assessments of young lamb indicators

Assessment	Estimated time (range)
Resource-based indicators	30 – 60 minutes
Completion of data capture form	15 minutes
Management-based indicators	15 – 30 minutes
Individual indicators (range 30 – 120 lambs)	1.5 – 3 hours
Bio-security protocol	45 minutes
Travel	30 minutes – 4 hours
Complete protocol	3 hours 45 minutes – 7 hours 45 minutes

7.3.6 Resource-based indicators

Sward height

Each sward height assessment took approximately 20 – 30 minutes to record 40 individual measurements of the sward. The mean sward height recorded per farm during the period July – November 2008 was 8.26 centimetres (range 4.19 – 20. 48). In order to avoid disturbing ewe-lamb bonding, no sward height assessments were performed during young lamb assessment visits.

Water supply

Over half of study farms were assessed as providing a source of water for grazing sheep (60%) and 88% of these were deemed to be accessible for sheep. Clean water was provided on most farms (71%), there was some debris or contamination on a few farms (29%) but no water source was classed as 'filthy' (Chapter 3).

Assessment area

Over 20% (n = 11) of flocks were reportedly moved to the area of assessment 2 days prior to the study visit. On 61% (n = 28) of farms the sample had been previously grazing the area for a period of 1 – 4 weeks. On a few farms (n = 5), sheep had remained in the assessment area from between 3 – 5 months.

7.3.7 Management-based welfare indicators

Completion of the data capture form relied on the shepherd or farmer being on the farm at the time of the on-farm assessment. A brief farmer interview was not feasible on 4 farms during the lambing period (January to April 2009), and so the management-indicators were tested on 46 out of the 50 farms (92%).

Farm assurance

83% of the study farms reported that they belonged to a farm assurance scheme, including organic and non-organic certification schemes.

Reproductive management

The majority of rams were bought at an auction (71%). Other sources included private vendors (50%) or home-bred ram replacements (18%). These rams were managed at a ram to ewe ratio of 1: 50. However, the reproductive outcome of farm management practices was not easy to evaluate as lambing % records proved difficult to obtain. Instead, farmers reported a mean scanning % of 165% (range 100 – 247%).

Farm labour

Most flocks (83%) were managed by one full-time shepherd or farmer. A maximum of 3 people were employed to care for sheep on 3 hill and upland farms which covered several holding areas and had flock sizes ranging from 950 – 1250 ewes. Over two-thirds of study farms (68%) employed additional labour during the lambing season.

Lambing practices

Nearly 70% of farmers reported there were no dystocia problems at lambing time. Despite this, over 92% reported that vaginal prolapses were regularly observed. On 47 % of farms these prolapses were retained using a plastic spoon or harness. Post-lambing, young lambs were routinely castrated and tail-docked on 75% of farms at 2 days–old (range 1 - 56 days). Elastrator rings were the most widely used method of tailing and castrating (96%).

Lamb production

Most of the study farms were commercial flocks (n = 46), producing prime lambs for slaughter (76%) at a mean live weight of 40 kg (range 25 – 45 kg). Only a small number (8%) were retained as stores or kept on the farm as breeding replacements (15%).

Preventive health management

Over two-thirds (65%) of farms reported that they would request veterinary attention for the treatment of an individual animal and 50% of farmers reported that a veterinary surgeon had performed at least one flock health visit in the previous year. A small number (8%) reported that they did not seek the advice of their local veterinary surgeon for flock health and welfare matters. Overall, there was considerable variation in the number of farm visits by a veterinary surgeon. Whilst 3 farms reported that there had been no visit in the previous year, a single hobby flock reported 12 visits had occurred.

Routine foot bathing was performed 4 – 5 times a year by 68% of study farms and less than half of farms routinely performed foot trimming (45%). Ectoparasiticides were applied to control scab, lice and myiasis in most study flocks – 85% of farmers reported they applied a topical agent twice yearly. Crutching was also used by 88% of farmers as a means of control and prevention of ectoparasitism.

Over 97% of study farms administered *Clostridial* vaccination, slightly fewer vaccinated against *Pasteurella* (90%) and 48% used Enzootic abortion vaccine. Few farmers administered Toxoplasma (31%), orf (27%), or foot rot vaccines (8 %). When questioned about biosecurity practices, 39 % of farms reported they quarantined bought-in sheep, although the period ranged from 1 week (7 %), 2 weeks (36%), 3 weeks (7%), and 4 weeks (14%) to 6 weeks (14%). A single hill farm reported that bought-in rams remained separate from the rest of the flock for a 6 month period.

Farmer perception of lameness and skin condition

The perceived prevalence of on-farm lameness varied across the study population - 17% of farmers suggested 1% prevalence, 20% perceived 5% of the flock were lame, and a further 17% suggested a flock lameness prevalence of 10% at the time of the study visit. The rest of the interviewed farmers did not provide an estimate of the lameness prevalence in sheep on their farms. Over 80% of farmers suggested there were no skin lesions in their flock, including abscesses, lice, sheep scab and myiasis.

7.4 Discussion

The objective of this chapter was to investigate whether the welfare indicators developed in this thesis were valid and feasible for on-farm application i.e. that they worked on farms with different welfare issues and varying management practices. Accordingly, the author tested the indicators on 3167 adult and growing lambs and 1681 young lambs from 50 flocks during a cross-sectional study. This approach evaluated the ability of the animal-, resource- and management-based indicators to identify differences in the level of sheep and lamb welfare conditions when tested under different farm conditions.

The pre-requisite for informed consent of recruited farms and extensive farmer co-operation meant that random sampling approaches were considered unfeasible. Instead, a non-random, convenience sample of farms was taken, whereby farms were selected on the basis of their consent to participate, location and stratification type. The involvement of veterinary practices in the recruitment strategy may have introduced responder bias as certain veterinary surgeons may have been more likely to respond. In addition, this approach may also have introduced selection bias since farms with regular or good contact with veterinary surgeons and sheep consultants, and those willing to participate in an on-farm welfare research study may have been more readily recruited.

Given the extensive locations of sheep and feasibility for assessment, farmers provided a selection of sheep for assessment, so it was also possible that a biased selection of sample animals with few health or welfare conditions may have been presented. Therefore, the proportion values provided in this chapter are only applicable to this study population and cannot be used as prevalence estimates of conditions affecting sheep or lamb welfare on farms across England and Wales and this was not the objective of this thesis.

As most sample sheep were moved to the area of group assessment a few days prior to the farm visit, certain resource-based assessments of grazing areas, such as measurements of sward height, cannot be considered to be truly representative of the grazing at the time of the visit. Therefore, sward height was not considered to be a valid measure of sheep welfare due to the pre-planned nature of these inspection visits. In contrast, the measurement of other resource-based indicators, such as water supply, provision of feed, was easy to perform and required limited time and resources and was considered to reflect everyday conditions for housed and grazing sheep.

The data capture form consisted of a brief interview to ascertain on-farm management practices according to the opinion and reports of the farmer. The interview was conducted on 46 out of the 50 study farms, as it was not possible to ascertain background information from all farmers during the lambing period. The data capture form relied on farmer reports of management practices and flock performance and was therefore open to recall bias (Abramson and Abramson, 2008). Few farms provided evidence of scanning figures or records of management practices and although results were maintained as confidential, farmers may not have reported the actual on-farm practices to an external, on-farm welfare assessor. Farmers were asked also to suggest estimates of flock lameness and skin conditions. It is recognised that the method of lameness assessment and definition of lameness was not clarified and this may account for the low estimates of lameness as suggested by some farmers. It may have been more appropriate to have clarified the definition of lameness and the farmer's perception of the level of lameness in the sample group and to have compared the perceived level of lameness with the outcome of a group lameness assessment.

Management-based questions also uncovered interesting results, such as the perceived problem of vaginal prolapses, limited veterinary input to flock health planning, high sheep to shepherd ratio and differences in bio-security practices. These results could be used to inform farmer training and industry initiatives in order to improve welfare of both the sheep and the shepherd through targeted management practices and greater veterinary support. However, given the potential issues with the validity of some resource- and management-based indicators, animal-based measures may offer a more valid, reliable and feasible means of assessing certain sheep welfare issues.

The results of this cross-sectional study suggest that the animal-based indicators capture this farm and breed variation. Accordingly, the hypothesis under investigation – that the indicators were capable of detecting the between-farm variation in the proportion of sheep affected by conditions associated with sheep welfare, was accepted.

Overall, a low level of most animal-based indicators of sheep and lamb welfare were present in the sample population. In particular, few sheep were identified with sub-optimal welfare conditions such as sheep scab, mastitis, myiasis, and emaciation. Similarly, for young lambs, very few cases of watery mouth, recumbency or hypothermia (shivering) were recorded. In addition, the presence of crystals and in-grown horns was not identified on any study farm. These results may suggest that the tests were not an effective measure of urolithiasis or in-growing horns. However, it is more likely that this was a true reflection of the absence of these conditions in the study population.

For other indicators, the low level of sheep observed with particular welfare issues may have been affected by the timing of the farm visit. In certain instances, farmers reported that they had identified and treated cases of myiasis some weeks prior to the on-farm visit. This may have accounted for the few observations of myiasis that were identified. Examination of the skin and fleece did however identify lesions that were suggestive of healed or treated cases of myiasis. So, evidence of blowfly strike can still be identified using these animal-based welfare indicators.

As well as reflecting the true level of skin lesions, injuries and wounds in the study population, the low proportion of sheep affected by these indicator outcomes may have been influenced by other factors. The presence of a dense fleece affected the ability to assess small, superficial skin conditions and injuries (Napolitano *et al.*, 2009). The ability to fully examine the entire body of the sheep also required good lighting conditions and handling facilities.

The ability of these indicators to perform on a sample population with a low proportion of sheep with poor welfare provides further evidence of their validity. As tests were capable of identifying farm-level variation in a population with few affected animals it is likely that they would perform even better if applied to sheep on farms with a higher proportion of health and welfare issues.

It was also important to evaluate the validity of these indicators in terms of their feasibility for application under working farm conditions. Therefore, the time and resources required to assess the indicators was examined across a range of management and farming conditions. Briefly it took approximately 30 minutes to perform a group observation of 70 sheep and a further 4.5 hours to assess each individual sheep within this group. Following an undisturbed assessment of group behaviour, the group observational method required the observer to quietly walk around the group to allow a closer inspection of individual sheep within the group. On some farms the sample group was widely dispersed over extensive hill terrains and on other farms the sheep appeared to be highly responsive to the presence of the assessor and moved a considerable distance away from the point of observation. In these cases, a group observation may have been facilitated by the use of shepherding dogs or use of an all-terrain vehicle to improve the ease of assessing sheep that were widely dispersed or located in rugged and steep terrains.

The author handled and examined all individual sheep and assistance with examination and the turning of sheep, particularly of rams, was only required on a few farms due to the quality of the on-farm handling facilities and the size and behaviour of the sheep. The need for additional labour for the handling of sheep may not be as important for statutory or voluntary assessment purposes as there may be a pre-requisite for farmer involvement which was not required in this study.

The type and quality of the farm facilities available for handling and inspecting individual sheep varied considerably, from the use of hurdles for making a pen, provision of a turning crate or use of the mobile handling unit owned by the research project. Given the amount of labour and time needed to assemble, use and disinfect the mobile unit, this handling system does not appear to be feasible for use in routine on-farm inspections. Instead, assessors may need to rely on on-farm handling systems and the assistance of the farmer. On-farm experience suggested that the ease and speed of individual sheep assessments was facilitated by gathering the sheep into a holding pen and then performing individual sheep examinations within a well-lit, square- or rectangular-shaped pen with a non-slip floor that was free from debris. However, the standardisation of assessment conditions was not within the scope of this research project.

As well as performing assessments, the author manually recorded all individual indicator scores. This part of the study was estimated to take between 1 to 2 minutes per animal. For future applications, it may be possible to reduce the amount of recording time by using hand-held electronic recording devices or having an assistant to prompt and record each score (Pritchard *et al.*, 2005).

The results show that there was an association between group observation and individual examination of rear and belly cleanliness but it is unclear whether this correlation would exist for a sample population with higher level of belly dirtiness. Group observation also relies on the ability to count the number of individual sheep affected by each indicator and this may be not be as feasible if larger group sizes (n > 120) are assessed. On some farms, it was not always possible to walk closely among the group of sheep and may explain the lack of correlation between rear cleanliness assessment by group observation and individual examination that occurred on 7 study farms.

The correlation between group observation and individual examination provided evidence of the construct validity of the method of lameness assessment used in this thesis. Correlation of the infectious foot lesions, including foot rot and CODD, with group lameness assessment further validated the group observational method of lameness assessment. On a small number of farms (n = 7), sheep were identified as lame on group assessment but not by individual examination, which may have reflected the quality of the individual examination area as gait assessment proved to be difficult to perform in poorly lit, straw-bedded and circular assessment areas. In addition, isolation from the group may have affected the behaviour of individual sheep. It is known that sheep may mask painful condition such as lameness (Fitzpatrick *et al.*, 2006; Stubsjøen *et al.*, 2009) so this may account for the difficulty in assessing lameness in individual sheep. Also, the method of individual gait assessment relied on the ability to quietly walk the sheep around the assessment pen, which was challenging when very responsive or 'flighty' sheep were examined and may have been affected by the genotype, breed or age of the sheep (Dwyer and Lawrence, 2000).

The lack of a correlation between group observation and individual examination of demeanour, coughing and skin irritation may have reflected the low proportion of sheep observed with these indicators. For wool loss the lack of correlation may also be affected by the ability to perform close observation of the fleece during group observation. It can be difficult to assess the wool cover over the whole body when sheep flock together. Therefore, an assessment of wool loss and skin lesions using an individual, physical examination may be appropriate on some farms.

This cross-sectional study also identified some trends in young lamb indicator outcomes. Over 5% of lambs were observed with an ocular abnormality – in most cases this was diagnosed as entropion. This was considered to be an on-farm welfare issue because the condition can be easily and effectively treated, so neglecting to treat this condition may suggest that shepherds do not always examine the eyelids of young lambs or they do not consider entropion to be an important welfare issue. It may also be a reflection of the labour and management demands at lambing time which may result in delayed treatment of certain conditions. Training of farmers to recognise particular behaviours and clinical signs, such as entropion, may therefore be a useful tool for improving the on-farm health and welfare of young lambs. Entropion appears to be a heritable defect so a high level of lambs with this condition on certain study farms may suggest that the condition also affects breeding animals. A single ram has the potential to effect the welfare of a large number of offspring, so the results of this study could be used to inform management decisions, such as the selection of breeding stock that are free from visible and inherited defects. However, further investigations would be needed to address the reasons behind the level of specific welfare conditions observed in this study population.

The study also examined the feasibility of performing indicator assessments. Inspections of 70 young lambs required approximately 2 hours and this shorter period is likely to have reflected the smaller number of tests and the less physically demanding nature of assessments. All indicators could be applied to indoor-housed lambs but it was not feasible to assess measures requiring close inspection, including body condition, eye condition, salivation and abdominal fill in all lambs that were managed in outdoor environments. Indicators that could be consistently applied to lambs in all study flocks were lameness, the ability to stand, demeanour, response to stimulation and posture. Therefore, it may be more appropriate to tailor the selection of young lamb indicators to the on-farm management system. Although it should be recognised that using a reduced set of indicators for outdoor-reared lambs could mean that conditions such as watery mouth, entropion, thin body condition or abnormal abdominal fill could be missed by this approach.

As the expert panel identified welfare concerns associated with artificial rearing systems (Chapter 2), the indicators were specifically tested on orphan lambs. Interestingly, the cross-sectional study results suggested that orphan lambs were generally assessed with a higher proportion of conditions associated with poor welfare. Whilst maternal behaviour, care and protection can be advantageous for lamb survival (Dwyer, 2008b) this study cannot be used to identify an association between artificial rearing methods and poor lamb welfare. This is because the category of orphan lambs could have been biased towards sick or ill-thriving lambs which may have been removed from the ewe in order to receive further care and attention. However, for the purposes of the study, the inclusion of a range of lamb categories showed that indicator tests such as lameness, demeanour and posture were capable of assessing variation between lambs reared and managed under different farming conditions. An alternative measure would be to assess the relative number of orphan lambs to the number of live ewes. This is because the presence of a large number of artificially-reared lambs may be indicative of a number of on-farm welfare issues including poor nutritional management of parturient ewes, diseases such as acute mastitis or a high rate of ewe mortality.

7.5 Conclusion

Many animal-based indicators of sheep and lamb welfare were found to be valid tests that could be feasibly applied on a wide range of farm management systems. Overall, few sheep in the study population were identified with welfare issues such as pruritus and myiasis. The ability of animal-based indicators to perform under these conditions suggests that the indicators could be considerably better if applied on farms with a greater number of sheep and lambs affected by sub-optimal health and welfare conditions. It would also be useful to examine if these tests are capable of detecting seasonal variation in conditions that have an impact on welfare as this could provide further evidence of the validity of the animal-based measures.

Chapter 8

VALIDATING THE ABILITY OF INDICATORS TO DETECT SEASONAL VARIATION IN SHEEP WELFARE

8.1 Introduction

The Five Freedoms framework has been used in this thesis to develop indicators of animal welfare that are sensitive to the current on-farm welfare issues for sheep. In the UK, sheep farming systems can be very diverse, ranging from extensive management during the summer period to more intensive periods of production during the lambing season (FAWC, 1994). Whilst extensively-managed sheep have the freedom to express a greater repertoire of behaviours, they are exposed to more extreme climatic conditions and are less frequently inspected compared to more intensively-managed or housed sheep. Consequently, these periods of sheep production can be associated with untreated or chronic disease and starvation (Dwyer, 2009). The level of resource input and management intervention depends on the time of the year and so there can be seasonal variation in the concerns for sheep welfare.

Seasonal, climatic variation can also produce marked differences in food quality and nutrient availability (Dwyer, 2008a). These changes can be reflected in animal-based outcomes such as the alteration in body condition which occurs over the course of the sheep production cycle (Russel, 1984). Seasonal variation in environmental and climatic conditions can also affect the risk of diseases, such as cutaneous myiasis (French *et al.*, 1994a), tick-borne fever (Lees and Milne, 1951) and footrot (Whittington, 1995), which can have a negative impact on sheep welfare.

Therefore, as well as being capable in identifying between-farm variation in the proportion of sheep that are affected by conditions that impact on sheep welfare (Chapter 7), the animal-based indicators under investigation in this thesis should be responsive to seasonal changes in sheep welfare. The objective of the study presented in this chapter was to investigate the hypothesis that, given there is known seasonal variation, the indicators are capable of detecting seasonal differences in the proportion of sheep observed with animal-based outcomes, such as lameness and thin body condition.

8.2 Materials and methods

8.2.1 Study population

A population of 12 study farms, previously recruited as described in Chapter 3, was selected to participate in a one-year prospective longitudinal study from 9 May 2009 – 14 April 2010. These farms were selected according to their consent to participate, farm stratification type and farm location. Each study farm was coded with unique numeric identity (Table 8.1) and categorised as either a lowland (n = 6), upland (n = 1), or hill flock (n = 5).

Table 8.1 Longitudinal study population

Farm ID	Farm type	Location	Farm purpose	Farm assured
1	Lowland	Cheshire	Commercial non-pedigree	No
3	Lowland	Cheshire	Commercial non-pedigree	No
6	Hill	Clwyd	Hobby	No
7	Lowland	Staffordshire	Commercial non-pedigree	No
8	Lowland	Lancashire	Commercial non-pedigree	Yes
10	Lowland	Cheshire	Commercial non-pedigree	No
13	Lowland	Cheshire	Commercial non-pedigree	No
19	Upland	Denbighshire	Commercial non-pedigree	Yes
20	Hill	Gwynedd	Commercial non-pedigree	Yes
21	Hill	Gwynedd	Commercial non-pedigree	Yes
22	Hill	Gwynedd	Commercial non-pedigree	Yes
23	Hill	Gwynedd	Commercial non-pedigree	Yes

At each study visit, the author assessed the entire flock if the flock size was ≤ 70, or, if the flock size was > 100, it was assumed that there was homogeneity within the flock and a sample group of approximately 70 sheep was selected *ad hoc* by the farmer. Each farm was assessed six times within the one-year study period, at intervals of approximately 60 days (Table 8.2). A total of 5740 adult ewes, rams, growing lambs were assessed. The number of sheep presented by the farmer for assessment varied according to the study visit. A median sample size of 77 adult sheep and growing lambs (range 24 – 137) was assessed on each study farm using group observation and individual examination during visits 1 – 5. By contrast, a median sample size of 79 adult sheep and growing lambs was assessed by group observation (range 30 – 108) and 22 sheep were individually examined (range 14 – 47) during visit 6 (Table 8.3).

Table 8.2 Details of longitudinal visits

Visit	Study period	Season	Production stage
1	May – June 2009	Spring/Summer	Post-lambing
2	July – August 2009	Summer	Weaning
3	September – October 2009	Autumn	Tupping
4	November – December 2009	Autumn/Winter	Early pregnancy
5	January – February 2010	Winter	Mid-pregnancy
6	March – April 2010	Spring	Lambing

8.2.2 Animal-based welfare indicator assessments

At each sampling visit (n = 6), a number of sheep was presented by the farmer. All animals were assessed by the author using the animal-based indicators assessed by group observation and individual examination, as described in Chapter 3. Following the findings of Chapters 5 and 7, the scoring systems of indicators of tooth disease, mastitis, individual rear and belly cleanliness, skin lesions, injuries and wounds and myiasis were reduced (Appendix B). As the aim of this study was to identify the ability of animal-based indicators to identify seasonal variation, no resource-based indicator assessments were reported.

Table 8.3 Sample size for longitudinal study visits

Assessment method	Number of sheep by longitudinal study visit						*n* total
	1	2	3	4	5	6	
Group observation	1182	1133	990	780	709	946	**5740**
Individual exam	1182	1133	990	780	709	283	**5077**

8.2.3 Statistical analysis

Data was analysed using Stata version 10 (StataCorp LP, College Station, Texas). For the purposes of this longitudinal study, 'visit proportion' was defined as the proportion (%) of sheep identified with each indicator at each study visit (n = 6), and 'mean proportion' was defined as the mean % of sheep assessed with each indicator over the 6 study visits. Proportion values were calculated using Huber–White robust standard error estimation to account for farm-level clustering (Kirkwood and Sterne, 2003). Binomial logistic regression modelling was employed to investigate any seasonal variation in the proportion of sheep affected with each animal-based indicator. Modelling was not attempted for indicators with a mean proportion < 3 %. Logistic regression models were fitted with the binary outcome variable being the presence or absence of the welfare condition. Farm identity was included as a random effect to account for farm-level clustering (McDermott *et al.*, 1994). Time was offered to the model as a composite of four sine and cosine functions (harmonic regression) to allow modelling of seasonal periodicity (Stolwijk *et al.*, 1999). The functions were defined as follows, where t = day of study period (day 1 was 9 May 2009):

$x1 = \cos(2\pi t/365)$, $x2 = \text{sine}(2\pi t/365)$, $x3 = \cos(4\pi t/365)$, $x4 = \text{sine}(4\pi t/365)$

To examine the ability of indicators to identify seasonal variation, the predicted coefficient (β) from the model and 95 % confidence intervals (CI) (produced using only the time covariates), was plotted for each indicator against time i.e. month of the study period.

8.3 Results

The proportion of each animal-based welfare indicator assessed at each study visit is provided in Tables 8.4, 8.6 and 8.8. Graphical representations of the seasonal variation in adult sheep and growing lamb welfare indicators are shown in Figures 8.1 and 8.2.

8.3.1 Indicators of adult sheep and growing lamb welfare assessed by group observation

Throughout the longitudinal study, few sample sheep (< 2%) were observed with dull demeanour, wool loss, pruritus, and coughing and excessive panting was not observed during any study visit (Table 8.4). No modelling of this longitudinal data was performed and therefore any seasonal variation in these indicators could not be explored.

Lameness was observed on all study farms and a mean proportion of 13.5 % was recorded (range 9.3% – 17.5%). The method of group observation was capable of identifying seasonal differences in the proportion of lame sheep and modelling suggested that a greater level of lameness was evident during the summer and early winter (Figure 8.1). The method of group observation also consistently identified sheep with dirty rears at each study visit (Table 8.4) and a higher level of dirty rears (22.1%) was observed during the spring period (Figure 8.1). Results of cleanliness assessments of the ventral abdomen also suggested that the group observation method was capable of identifying seasonal variation. The proportion of sheep observed with dirty bellies ranged from 0% during the summer to 39.2% during the winter.

8.3.2 Indicators of adult sheep and growing lambs assessed by individual exam

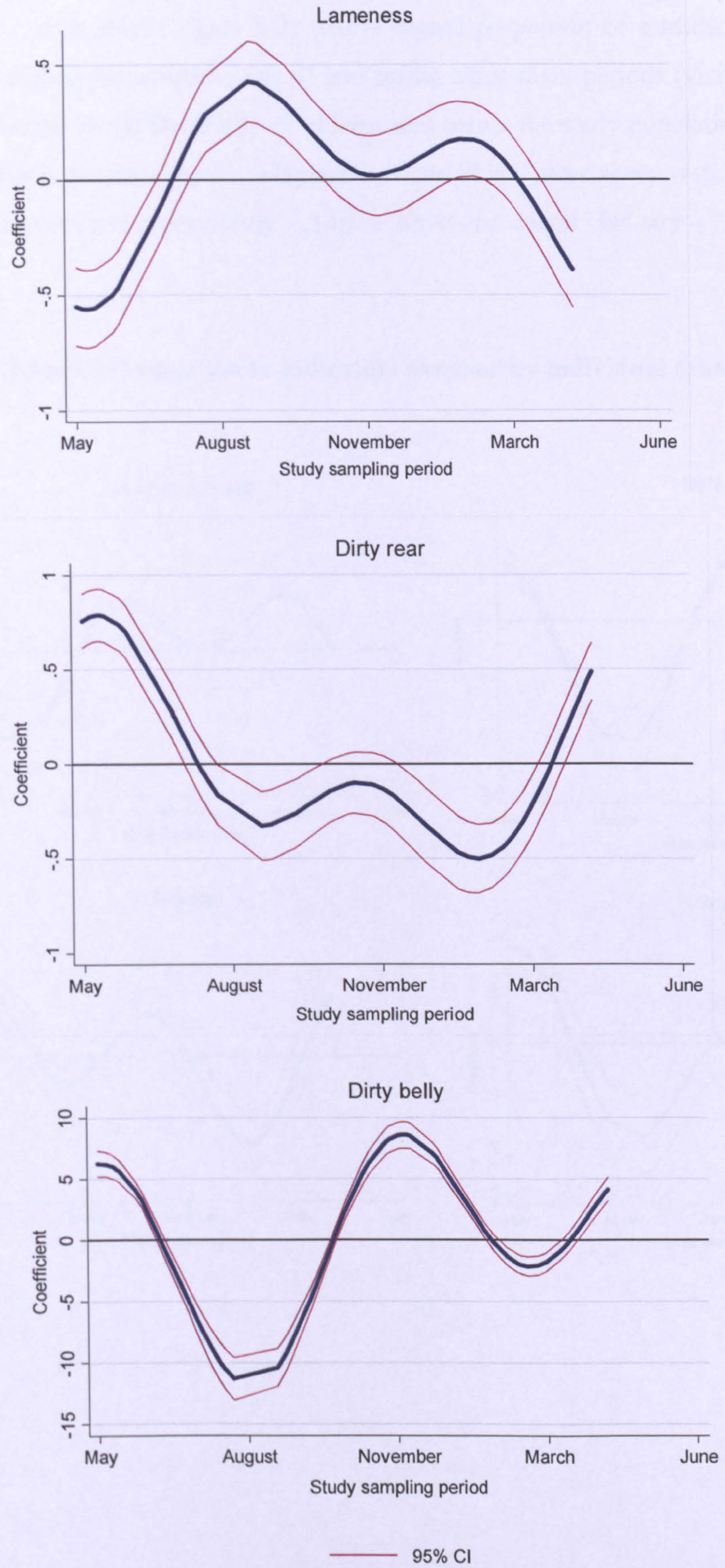
The proportion of sheep assessed with each individual indicator is shown in Table 8.5. Overall, less than 3% of the sample population was observed with dull demeanour, an eye condition, ear lesion, nasal discharge, incisor loss, coughing, filthy belly score, tail length, wool loss, pruritus, skin lesions, injuries and wounds, thin (BCS < 2) or fat (BCS > 4) body condition, specific foot lesions including CODD and toe granulomas, joint swellings or myiasis. In agreement with previous studies (Chapters 5 and 7), no observations of in-growing horns or uroliths (crystals) were recorded during the longitudinal study. As the overall proportion of these particular conditions was below 3%, seasonal variation was not investigated by logistic regression modelling.

Seasonal variation in tooth condition was identified and the proportion of sheep assessed with a molar abnormality varied from 3% (July – August) to 6.9% (September – October). Harmonic regression analysis also identified seasonal variation in the outcome of cleanliness scoring with dirty belly scores ranging from 0 – 24.6%. In agreement with group observation findings, dirty belly scores peaked in the winter (visit 4). By comparison, dirty (17.8%) and filthy rear scores (8.1%) peaked during the spring period (Figure 8.2). Noticeable differences in the body condition of the sample population were also evident during. Overall few thin (BCS < 2) or obese (BCS > 4) sheep were identified within the longitudinal study population. Most sample sheep were assessed with a body condition that was designated as 'fit for purpose' (BCS 2 – 3). However, both Table 8.4 and Figure 8.2 illustrate that the body condition of this sample population did alter across the year-long sampling period. A loss in body condition was observed in the late summer, when sample animals were more likely to be assessed as thin (BCS < 2).

Table 8.4 Proportion of indicators assessed by group observation during the longitudinal study

Indicator	Proportion (%) by longitudinal study visit (95 % CI)						Mean (%) proportion
	1	2	3	4	5	6	
Dull demeanour	**0.17** (-0.07 – 0.41)	**0.71** (0.02 – 1. 40)	**1.62** (0.01 – 3.21)	**0.64** (-0.78 – 2.06)	**3.10** (0.04 – 6.17)	**3.70** (-0.65 – 8.05)	**1.53** (0.30 – 2.77)
Excessive panting	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Coughing	**0**	**0.18** (-0.08 – 0.43)	**0**	**0.13** (-0.16 – 0.41)	**0.28** (-0.11 – 0.67)	**0.11** (-0.12 – 0.33)	**0.10** (0.00 – 0.21)
Pruritus	**0**	**0.15** (-0.10 – 0.39)	**0.05** (-0.07 – 0.17)	**0**	**0**	**0.29** (-0.18 – 0.77)	**0.08** (0.01 – 0.16)
Wool loss	**1.86** (-0.65 – 4.37)	**0.18** (-0.09 – 0.44)	**0.40** (-0.07 – 0.88)	**0**	**0.85** (0.13 – 1.56)	**2.64** (-0.12 – 5.41)	**1.03** (-0.08 – 2.13)
Lameness	**9.31** (1.78 – 16.83)	**13.33** (8.57– 18.09)	**17.47** (3.30 – 31.65)	**12.95** (6.57 – 19.32)	**17.49** (7.61 – 27.37)	**12.26** (4.31 – 20.22)	**13.50** (7.18 – 19.82)
Dirty rear	**22.07** (10.92 – 33.32)	**12.00** (8.40 – 15.61)	**9.39** (1.82 – 16.97)	**8.61** (0.54 – 16.67)	**7.79** (1.03 – 14.56)	**13.85** (7.23 – 20.46)	**12.69** (8.29 – 17.08)
Dirty belly	**11.59** (-6.20 – 29.39)	**0**	**0.10** (-0.12 – 0.33)	**39.18** (6.16 – 72.20)	**1.13** (0.66 – 2.91)	**0.42** (-0.50 – 1.35)	**7.94** (2.22 – 13.66)

Figure 8.1 Seasonal variation in indicators assessed by group observation



Following this period, body condition improved throughout autumn – winter and towards spring (March – April 2010). There also appeared to be seasonal variation in the assessment of mastitis (Figure 8.2) with a higher proportion of mastitis (> 4%) being observed during the summer (visit 4) and spring assessment periods (visit 6). Individual gait assessment found that the level of lameness across the study population ranged from 6.6% – 16.8%. Seasonal peaks in lameness occurred and more sheep were observed to be lame during the late summer (July – August 2009) and winter (January – February 2010).

Figure 8.2 Seasonal variation in indicators assessed by individual examination

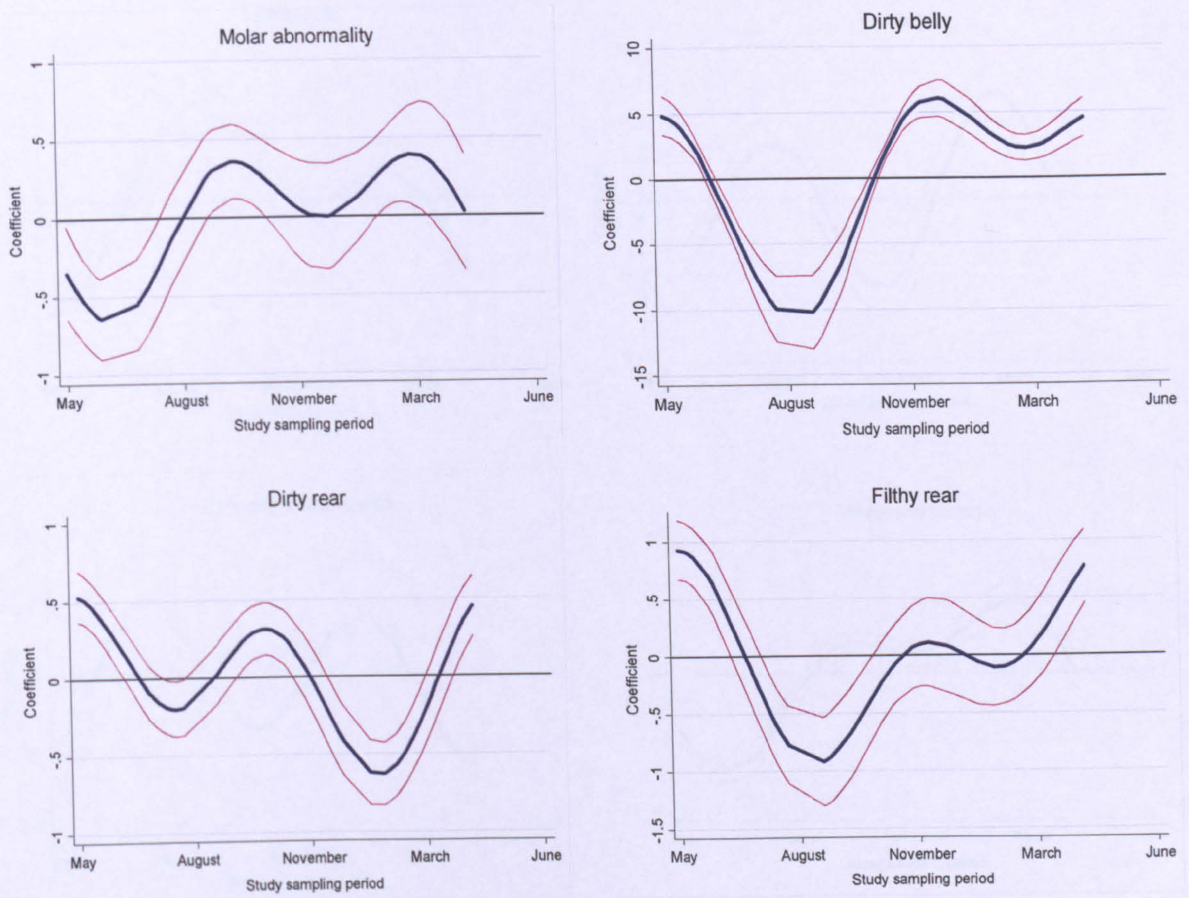


Figure 8.2 continued

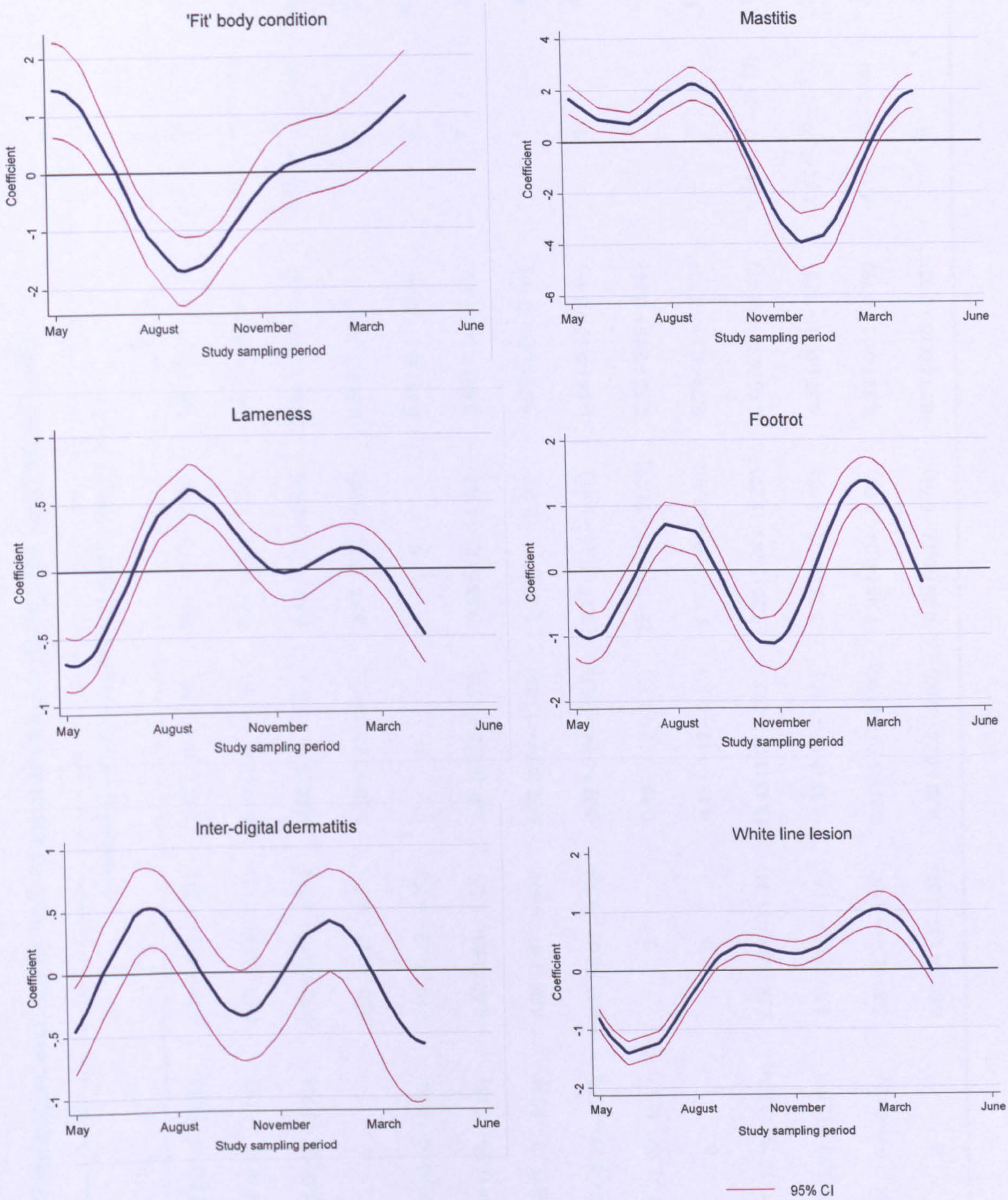


Table 8.5 Proportion of indicators assessed by individual examination during the longitudinal study

Indicator	Proportion (%) by longitudinal study visit (95 % CI)						Mean (%) proportion
	1	2	3	4	5	6	
Dull demeanour	**0.08** (-0.09 – 0.26)	**0.88** (-0.05 – 1.81)	**1.12** (0.16 – 2.08)	**0.51** (-0.36 – 1.39)	**1.69** (-0.26 – 3.65)	**1.77** (-0.18 – 3.72)	**0.85** (0.24 – 1.46)
Eye condition	**0.59** (0.06 – 1.12)	**0.79** (-0.56 – 2.15)	**0.92** (0.38 – 1. 45)	**0.64** (-0.04 – 1.32)	**2.82** (-0.46 – 6.11)	**2.47** (-0.15 – 5.10)	**1.12** (0.53 – 1.71)
Nasal discharge	**0.42**(-0.09 – 0.94)	**0.71** (-0.12 – 1.53)	**0.82** (0.03 – 1.60)	**0.64** (-0.31 – 1.59)	**0.14** (-0.17 – 0.45)	**0.35** (-0.36 – 1.06)	**0.55** (0.17 – 0.93)
Active ear lesion	**0**	**0.09** (-0.12 – 0.29)	**0.10** (-0.13 – 0.33)	**0.13** (-0.14 – 0.40)	**0.56** (-0.14 – 1.27)	**0**	**0.14** (0.01 – 0.26)
Inactive ear lesion	**0.08** (-0.09 – 0.26)	**0.09** (-0.12 – 0.29)	**0**	**0**	**0.14** (-0.17 – 0.46)	**0**	**0.06** (-0.01 – 0.13)
Incisor loss	**3.64** (1.80 – 5.48)	**0.62** (0.00 – 1.23)	**3.46** (-0.65 – 7.57)	**1.03** (0.13 – 1.92)	**3.67** (1.21 – 6.12)	[a]	**2.52** (1.17 – 3.56)
Molar disease	**3.38** (2.40 – 4.37)	**3.00** (1.07 – 4.93)	**6.92** (0.10 – 13.73)	**5.38** (2.41 – 8.36)	**6.35** (3.74 – 8.96)	[a]	**4.80** (2.87 – 6.73)
Coughing	**0.08** (-0.09 – 0.27)	**0.35** (-0.09 – 0.80)	**0.31** (-0.04 – 0.65)	**0.13** (-0.16 – 0.41)	**0.14** (-0.17 – 0.45)	**0**	**0.20** (0.04 – 0.35)
Dirty belly	**6.77** (-1.59 – 15.13)	**0**	**0.10** (-0.12 – 0.33)	**24.62** (-2.47 – 51.70)	**2.52** (-0.10 – 5.18)	[a]	**6.07** (0.95 – 11.19)
Filthy belly	**0**	**0**	**0.10** (-0.12 – 0.33)	**8.72** (-5.96 – 23.40)	**0.79** (-0.55 – 2.13)	[a]	**1.55** (-0.90 – 4.01)
Dirty rear	**17.77** (10.79 – 24.74)	**8.38** (5.34 – 11.43)	**13.43** (7.03 – 19.83)	**9.23** (3.24 – 15.22)	**8.46** (4.46 – 12.47)	**9.54** (2.92 – 16.17)	**11.76** (9.47 – 14.04)
Filthy rear	**8.12** (3.72 – 12.52)	**1.77** (0.67 – 2.86)	**2.53** (0.67 – 3.60)	**2.31** (-0.07 – 4.68)	**4.37** (-0.21 – 8.95)	**1.41** (0.09 – 2.73)	**3.75** (2.10 – 5.39)
Mastitis	**6.04** (2.99 – 9.10)	**3.87** (1.26 – 6.48)	**7.23** (-3.50 – 17.96)	**0.40** (-0.24 – 1.05)	**0.32** (-0.15 – 0.79)	**4.78** (2.33 –7.24)	**3.69** (1.07 – 6.30)
Short tail length	**0**	**0.09** (-0.10 – 0.28)	**0.20** (-0.26 – 0.67)	**0.26** (-0.13 – 0.64)	**0.14** (-0.16 – 0.45)	**0**	**0.12** (0.02 – 0.22)

Table 8.5 continued

Indicator	Proportion (%) by study visit (95 % CI)						Mean (%) proportion
	1	2	3	4	5	6	
Wool loss	**5.41** (0.33 – 11.16)	**0.62** (-0.33 – 1.57)	**0.71** (-0.01 – 1.44)	**0.38** (-0.07 – 0.84)	**2.82** (-0.06 – 5.70)	**2.47** (-0.99 – 5.94)	**2.13** (0.50 – 3.76)
Skin irritation	**0.08** (-0.10 – 0.27)	**0.09** (-0.12 – 0.29)	**0.10** (-0.13 – 0.33)	**0**	**0.99** (-0.49 – 2.46)	**0.71** (-0.89 – 2.30)	**0.24** (-0.03 – 0.50)
Skin lesion small	**0.25** (-0.04 – 0.55)	**3.00** (-0.30 – 6.30)	**2.85** (-1.53 – 7.23)	**0.13** (-0.16 – 0.42)	**1.13** (-0.19 – 2.45)	**1.16** (-0.20 – 2.53)	**1.53** (0.28 – 2.87)
Skin lesion hand	**0.17** (-0.08 – 0.43)	**1.15** (0.42 – 1.87)	**0.41** (-0.12 – 0.94)	**0.39** (-0.02 – 1.03)	**0.85** (-0.48 – 2.17)	**0**	**0.56** (0.25 – 0.87)
Skin lesion diffuse	**0**	**0**	**0.20** (-0.25 – 0.66)	**0**	**0**	**0**	**0.04** (-0.05 – 0.13)
Superficial scratches	**0.17** (-0.19 – 0.53)	**0.08** (-0.11 – 0.29)	**0.10** (-0.12 – 0.33)	**0**	**0**	**0.18** (-0.48 – 1.26)	**0.10** (0.00 – 0.20)
Healing wound(s)	**0.51** (-0.02 – 1.03)	**1.85** (0.32 – 3.39)	**2.75** (0.44 – 5.06)	**1.54** (0.10 – 2.25)	**4.37** (1.52 – 7.23)	**0**	**2.00** (1.02 – 2.83)
Open wound(s)	**0.59** (-0.47 – 1.66)	**0.44** (-0.36 – 1.24)	**0.31** (-0.05 – 0.66)	**0.13** (-0.15 – 0.40)	**0.56** (-0.38 – 1.51)	**0**	**0.39** (-0.02 – 0.81)
BCS 1	**0.15** (-0.19 – 0.49)	**1.10** (0.15 – 2.06)	**0.52** (-0.62 – 1.66)	**0.43** (-0.11 – 0.97)	**0.86** (-0.06 – 1.77)	**1.06** (-0.23 – 2.35)	**0.64** (0.10 – 1.17)
BCS 2	**30.29** (11.85 – 48.74)	**39.53** (22.60 – 56.45)	**13.52** (5.96 – 21.09)	**9.33** (3.07 – 15.59)	**15.12** (6.84 – 23.40)	**44.17** (30.20 – 58.14)	**22.76** (14.18 – 31.34)
BCS 3	**54.85** (38.24 – 71.47)	**39.37** (28.14 – 50.60)	**62.81** (50.21 – 75.41)	**58.68** (41.17 – 76.19)	**58.35** (42.89 – 73.80)	**54.06** (40.82 – 67.31)	**55.17** (47.81 – 62.52)
BCS 4	**14.71** (5.32 – 24.09)	**19.06** (11.23 – 26.88)	**23.02** (7.67 – 38.36)	**31.13** (9.70 – 52.57)	**25.68** (5.35 – 46.00)	**0.71** (-0.85 – 2.27)	**21.17** (8.57 – 33.77)
BCS 5	**0**	**0.94** (-0.78 – 2.67)	**0.13** (-0.18 – 0.44)	**0.43** (-0.52 – 1.38)	**0**	**0**	**0.27** (-0.05 – 0.58)
Thin (BCS < 2)	**0.08** (-0.11 – 0.28)	**1.59** (-0.68 – 3.86)	**4.07** (-4.01 – 12.22)	**0.38** (-0.06 – 0.83)	**0.85** (-0.06 – 1.76)	**1.06** (-0.23 – 2.35)	**1.40** (-0.32 – 3.12)

Table 8.5 continued

Indicator	Proportion (%) by study visit (95 % CI)						Mean proportion (%)
	1	2	3	4	5	6	
Fit (BCS 2 – 3)	**99.92** (99.72 – 100)	**97.88** (95.39 – 100)	**95.83** (87.46 – 100)	**99.23** (98.33 – 100)	**99.15** (98.24 – 100)	**98.94** (97.65 – 100)	**98.40** (96.62 – 100)
Fat (BCS > 4)	**0**	**0.53** (-0.49 – 1.55)	**0.10** (-0.11 – 0.32)	**0.38** (-0.47 – 1.24)	**0**	**0**	**0.20** (-0.05 – 0.45)
Lameness	**6.60** (2.98 – 10.23)	**13.15** (7.81 – 18.50)	**16.79** (0.23 – 33.35)	**11.04** (6.11 – 15.97)	**15.09** (4.58 – 25.60)	**8.83** (3.11 – 14.55)	**12.04** (5.69 – 18.39)
White line	**14.53** (3.39 – 25.67)	**19.51** (11.54 – 27.47)	**38.25** (16.30 – 60.20)	**37.82** (18.91 – 56.74)	**32.51** (-1.66 – 66.68)	[a]	**26.71** (14.70 – 38.71)
Inter-digital dermatitis	**2.35** (0.75 – 3.95)	**5.65** (3.21 – 8.09)	**2.95** (1.45 – 4.45)	**2.50** (1.15 – 3.85)	**7.39** (-0.23 – 17.09)	[a]	**3.63** (2.51 – 4.75)
Footrot	**1.13** (0.33 – 1.93)	**5.83** (2.15 – 9.51)	**2.54** (1.00 – 4.08)	**2.10** (-0.32 – 4.52)	**17.73** (-0.42 – 35.89)	[a]	**3.67** (1.93 – 5.42)
CODD	**0.17** (-0.09 – 0.44)	**0.09** (-0.11 – 0.29)	**7.93** (-8.17 – 24.04)	**1.05** (-0.58 – 2.68)	**8.37** (-5.33 – 22.08)	[a]	**2.50** (-2.38 – 7.37)
Toe granuloma	**1.31** (0.46 – 2.15)	**0.44** (-0.07 – 0.96)	**0.61** (-0.20 – 1.42)	**0.13** (-0.16 – 0.43)	**0.49** (-0.49 – 1.47)	[a]	**0.66** (0.22 – 1.11)
Joint swelling	**0.62** (-0.06 – 1.30)	**0.71** (0.03 – 1.39)	**1.02** (0.10 – 1.94)	**0.79** (-0.40 – 1.98)	**2.37** (0.36 – 4.37)	[a]	**0.97** (0.28 – 1.66)
Myiasis	**0.08** (-0.10 – 0.27)	**0.25** (-0.11 – 0.29)	**0.20** (-0.23 – 0.63)	**0**	**0**	[a]	**0.08** (-0.04 – 0.20)

[a] Indicator was not performed during the assessment of individual sheep

8.4 Discussion

The objective of the longitudinal study presented in this chapter was to investigate the hypothesis that, given the known seasonal variation in the level of sheep welfare conditions, the animal-based indicators were capable of detecting variation in indicator scores over the seasons of the sheep production cycle. Accordingly, the welfare indicators were applied during a one-year longitudinal study on a convenience sample population of 12 flocks, selected on the basis of their informed consent and co-operation to participate. This non-random selection of farms will have introduced the potential for referral, responder and selection bias and therefore, the outcomes of welfare indicators assessments on these flocks may not have been representative of the on-farm welfare of sheep from all farms within the British Sheep Stratification (BSS) system.

Given the large flock sizes and the varied and dispersed locations of sheep, it was not considered feasible for the author to select the sample animals on flocks with > 100 animals. Therefore, a convenience sampling approach was taken in which sample sheep were selected by the farmer. Whilst it may be argued that this approach could have introduce selection bias as animals with few health or welfare issues may have been presented for assessment, there was no evidence to suggest that there was any deliberate bias in sample selection. Furthermore, because of the difficulties in identifying individual sheep and management changes in the grouping of the flock, it was not always feasible to select the same sheep at repeat visits. Consequently, an assumption of this study was that on-farm management affected the whole flock. Therefore, repeat sampling of sheep from the same farm was used to examine farm-level variation in indicator scores over the period of the longitudinal study. As such it was recognised that the variation in indicator scores may not be solely due to seasonal changes. However, it was not the objective of this study to identify the on-farm standards of the 12 study farms and therefore no inferences regarding the reasons behind variation in indicator scores over the period of the longitudinal study can be made. For the purposes of this study, the ability of indicators to detect changes over the time course of the annual sheep production cycle was considered to provide further evidence of the validity of the animal-based measures.

As the expert panel suggested that type and breed of sheep may affect the outcome of on-farm welfare assessments (Chapter 2), the animal-based welfare measures were tested on a range of breeds, ages and types of sheep. To minimise any disturbance of lambing behaviour and maternal bonding, all ewes were not individually examined or turned during the final study visit. Therefore, a thorough assessment of the limbs, feet, ventral abdomen and mammary glands could not be performed at this visit. Although all the animal-based indicators could be applied to rams throughout the course of the study, it was apparent that rams were not always available for assessment at every study visit. This was likely to be due to the varying location of rams, which were not always kept near to the farm holding over the summer or lambing period. In the author's experience, rams are often afflicted with a number of welfare conditions and it is possible to speculate that rams may not have received the same level of year-round attention as ewes and lambs. Since there can be difficulties in physically examining rams and rams may be at different locations from the rest of the flock, these are important considerations that need to be addressed in the planning and conduct of future on-farm welfare assessments.

Several animal-based measures, including mastitis, lameness, cleanliness, body condition and footrot were capable of identifying seasonal variation in outcomes of sheep welfare. Clinical signs of mastitis peaked at lambing time and during the late summer period, suggesting that the indicator was capable of detecting cases of acute and chronic mastitis (Winter, 2001). On Farm 3, over 40% of the sample population was identified with mastitis during visit 2, which may have been due to the large number of cull ewes that were presented for examination. These results suggest that the indicator was capable of detecting seasonal differences in the proportion of sheep affected by mastitis. There is currently limited knowledge regarding the epidemiology of mastitis in sheep managed under British farming systems, although it may be suggested that management factors, such as the cleanliness of housing (Caroprese, 2008), and age distribution of the flock (Mork *et al.*, 2007), may be factors that are involved.

Greater levels of wool loss, identified by both group and individual exam, were observed during the late spring (visit 1) and lambing period (visit 6). In many flocks the springtime shedding of the fleece appeared to concur with the physiological reduction in wool production (Doney and Smith, 1961). Therefore, wool loss identified during the spring period may not necessarily be due to pathological causes. On some farms, the wool loss might also have been associated with the level of activity and behaviour of young lambs - a considerable amount of wool loss could be attributed to young lambs jumping onto the backs of ewes.

However, the observation of wool loss during the spring period should not be dismissed as solely a natural or physiological phenomenon as it may alert the observer to the presence of skin lesions such as sheep scab or lice infestation. The nibble test, performed on individual sheep, can be a useful aid in identifying pruritic skin conditions (D'Angelo *et al.*, 2007). However, it was not possible to individually examine all animals during the lambing season and so the observer may be more reliant on group observation to identify signs of wool loss and pruritus. As this approach is performed from a distance, it may not be possible to assess the whole body of individual sheep that are managed indoors in highly stocked pens or those that express flocking behaviours in the field. In addition, on-farm experience suggested that group behaviour could be affected by the presence of the assessor or movement of the flock to fresh grazing. Therefore, it is possible that a group observation approach could miss small areas of wool loss or skin lesions (Napolitano *et al.*, 2009), for example, early lesions of sheep scab, which may be indicative of a serious risk to flock welfare.

Seasonal variation in the level of sheep lameness was detected using both group and individual methods of gait assessment. In this population, the odds of being lame increased during late summer and autumn and early winter. Previous studies (Chapter 5) found some correlation between the proportion of sheep affected by infectious foot lesions and the proportion of lameness, as identified by group observation. In the present study, a higher level of footrot appeared to follow the peaks in the proportion of lameness. In this study the higher level of footrot which occurred in late summer, autumn and early winter may have been associated with warmer and wetter environmental conditions during these periods (Whittington, 1995).

Variation in environmental conditions will also have affected the outcome of cleanliness assessments. High scores of belly dirtiness were indicative of poor environmental hygiene and/or the absence of a lie-back area for grazing animals. In this study, sheep were more likely to be assessed with a dirty or filthy belly score during winter, coinciding with wet weather and muddy field conditions. In contrast, sheep were more likely to be observed with dirty and filthy rear scores ('daginess') during spring and summer. Rear scores reflect the consistency of the faeces and were used to indicate a general risk for sheep health and welfare rather than the burden of gastrointestinal parasites *per se* (Pollott *et al.*, 2004). Faecal consistency may have been affected by nutritional changes, such as a move to lush grazing. As season and climate affect grass growth, the higher level of dirty rear scores observed during the springtime was likely to be the result of the increased grass growth and alterations in the composition of spring grass.

The outcome of nutritional management was also examined using body condition scoring using the Russel (1984) and Fit-Fat-Thin (Chapter 3) scoring scales. Sheep metabolise their body reserves to support a physiologically demanding process, therefore loss in body condition can be associated with lactation (Stubbings, 2008). Both scoring scales identified that sheep lost body condition between the spring and summer, which was likely to be due to the demands of peak lactation. Body condition was lowest following the weaning during July to August and improved from autumn onwards, coinciding with tupping. This improvement was likely due to the cessation of lactation and the added effect of specific management practices such as 'flushing' of ewes and the improved nutrition of rams, prior to their role within the breeding season. In common with previous findings (Chapter 7), the longitudinal study identified a very small number of sheep with extreme body condition scores (Russel, 1984). This may have been due to observer bias towards scores in the mid-range of the scale, or due to the farmer selection of samples animals. However, it was more likely that the low level of very thin or fat sheep was a true reflection of the body condition of sheep in the 12 study flocks.

Other welfare indicator scores, including dull demeanour, myiasis and thin body condition were observed at very low levels. Here the use of descriptive statistics provided a clear and simple means of demonstrating seasonal variation in conditions that were observed in a small proportion of the sample population. These methods were used to describe the study population and examine the responsiveness of the indicators and it was not the aim of this study to provide prevalence estimates of sheep welfare conditions. So, the data presented in this chapter should not be used to infer conclusions about the welfare of the wider sheep population.

8.5 Conclusion

The longitudinal study found that animal-based indicators were robust and responsive to identifying variation in the farm-level proportion of sheep affected by conditions associated with welfare. The low number of sheep affected by conditions, such as diffuse skin lesions, myiasis, and thin body condition, did not allow the modelling of seasonal variation. However, descriptive statistics clearly identified that there were seasonal differences in these conditions. This work provides further evidence of the validity of these measures and given these promising results, it would be useful to evaluate these welfare indicators on a sample population experiencing a greater level of sub-optimal welfare conditions.

Chapter 9

VALIDATING QUALITATIVE BEHAVIOUR ASSESSMENT AS AN ON-FARM INDICATOR OF SHEEP WELFARE

9.1 Introduction

The FAWC Five Freedoms have been used as the framework for the development of welfare indicators in this thesis. The Freedoms concepts stipulate the inputs required for good welfare (FAWC, 1994), and many of the welfare measures tested in previous chapters have focused on the physical outcomes of on-farm resources and management actions. For example, the presence of welfare conditions such as lameness and ectoparasitism has been measured using quantitative indicators of sheep behaviour such as changes in gait and posture or signs of pruritus. However, there is also a need to assess animal welfare beyond the physical appearance of the animal and this has led to demands for indicators that evaluate the quality of an animal's life (FAWC, 2009).

Both scientific research and European law (Treaty of Amsterdam, 1997) recognise that sheep are capable of experiencing emotions and feelings, such as pleasure or frustration (Greiveldinger *et al.*, 2007; Veissier *et al.*, 2009). Therefore, more recent research studies have been focused on the use of quantitative measures of animal behaviour such as fear tests (Forkman *et al.*, 2007), motivational preference tests (Greiveldinger *et al.*, 2007), and measures of emotional valence (Boissy *et al.*, 2011; Reefmann *et al.*, 2009) as a means of demonstrating the cognitive processing of emotions in sheep. However, the validity and reliability of some of these measures have been questioned (de Passillé and Rushen, 2005). In addition, many of the measures do not appear to be feasible for assessing large flocks of sheep produced under working farm conditions. Instead, an approach that takes an integrated and holistic approach to evaluating whether an animal has a 'life worth living' (Wemelsfelder *et al.*, 2001) could offer a valid, reliable and feasible means of assessing the on-farm welfare of sheep.

The method of Qualitative Behaviour Assessment (QBA), developed by Dr Françoise Wemelsfelder of the Scottish Agricultural College (SAC), examines the behavioural style and body language of the animal in order to assess welfare from the 'animal's point of view' (Wemelsfelder and Lawrence, 2001). This concept was informed by fundamental studies into behavioural expression, postures and quality of life scales (Goodall, 1990; Kessler and Turner, 1997; Morton and Griffiths, 1985; Roughan and Flecknell, 2003). A major advantage of QBA is that it considers the behaviour of the whole animal encompassing the mental and physical well-being and welfare rather than focusing on specific clinical signs or measures of physical health (Wemelsfelder and Lawrence, 2001).

The QBA methodology has been applied to assess the behaviour of both individual and groups of animals (Wemelsfelder *et al.*, 2001) and a further advantage is that the approach appears to be feasible measure that can be readily applied on farms. Many sheep in the UK are managed outdoors for large parts of the production cycle, so close inspection and assessment of individual sheep can require gathering and handling of the flock. This aspect of management may not only have effects on sheep behaviour but can also be time and labour consuming. Therefore, a method that does not involve major disturbance, requires few resources, and can be applied to groups of animals, offers clear benefits for measuring flock welfare.

Despite the obvious benefits of this approach, the method has received criticism for being subjective, with potential for anthropomorphism and the misinterpretation of animal behaviours by human observers. However, there is no direct measure or 'gold standard' for animal welfare (de Passillé and Rushden, 2005) and even measures, such as diagnostic blood tests, that are regarded to be 'quantitative' test require subjective interpretation. Whilst there may be concerns that humans cannot the identify feelings and emotions of other species, qualitative assessments of animal behaviour are intuitively and routinely used by stockpeople, veterinary surgeons and pet owners to assess animal health, welfare and well-being. Therefore, in the absence of a definitive test, human observations continue to form the foundation for animal welfare assessments (Wemelsfelder, 1997).

In this chapter the approach to investigating the validity of QBA was informed by previous research studies. The validity of QBA can be judged by different stakeholder groups to provide face, content and consensual validity (Wemelsfelder *et al.*, 2001). The test validity of the method can then be evaluated by studying the level of observer agreement. This approach identified good levels of inter-observer reliability when QBA was applied to pigs (Wemelsfelder *et al.*, 2000), poultry (Wemelsfelder *et al.*, 2009a) and cattle (Wemelsfelder *et al.*, 2009b). Research has also found construct and predictive validity between QBA and quantitative behavioural measures in dairy cattle (Rousing and Wemelsfelder, 2006) and physical measures of veal calf health (Brscic *et al.*, 2009).

The objective of the studies presented in this chapter was to investigate the validity of QBA as an on-farm measure of sheep welfare. Accordingly, a literature review and consensus of expert opinion were used in previous chapters (1 and 2) to judge the face, content and consensual validity of the method. The first objective of this chapter was to examine the level of inter-observer reliability of a range of assessors. Secondly, the on-farm feasibility of the method was examined. Finally, the association between QBA and physical measures of sheep welfare, and the ability of the method to identify between-farm and seasonal variation in sheep welfare was examined during a longitudinal study.

9.2 Materials and methods

9.2.1 QBA descriptive terms

The fixed-list terms used in this chapter were developed by a Quality Meat Scotland (QMS) project in which Inspectors from the Scottish Society for the Prevention of Cruelty to Animals (SSPCA) used a Free Choice Profiling (FCP) approach (Wemelsfelder *et al.*, 2001) to generate a fixed-list of 14 negative and positive descriptive terms that described sheep behaviour – relaxed, dejected, thriving, agitated, responsive, dull, demeanour, content, anxious, low demeanour, satisfied, bright, tense, vigorous, distressed (Wemelsfelder, 2009 personal communication).

9.2.2 Observer population

A pool of 17 assessors, including 10 observers previously recruited for other studies in this thesis (Chapter 3) and 7 farm assurance assessors (Soil Association) was used (Table 9.2). Observers were classified as experienced if they had applied sheep health and welfare assessments in the year prior to the study (Table 2). Observers who did not meet these criteria were classified as inexperienced. The occupation of each observer was recorded as either 'vet' (member of Royal College of Veterinary Surgeons), 'non-vet' (undergraduate veterinary or bio-veterinary science student) or 'farm assessor' (Soil Association farm assurance inspector). All observers were trained in the QBA methodology, which consisted of a 2 hour classroom based presentation given by Dr F Wemelsfelder.

Table 9.1 Descriptive terms of sheep behaviour

Term	Description and assessment of behavioural term
Relaxed	This reflects a positive mental state in which the sheep appears to be at ease with itself and with the environment. A relaxed sheep does not exhibit any muscle tension and this term does not imply a resting state – sheep may be lying down, standing or moving. A sheep that appears to be in pain or distressed in any way, showing muscle tension cannot be described as relaxed.
Dejected	Describes sheep which have 'given up' on life. This could be indicated by a low head carriage, listlessness or by behavioural separation of the individual sheep from the flock.
Thriving	A term often used as a clinical or stockperson descriptor, used to describe sheep that appear to be in good health, showing physical 'bloom' – implying that the sheep have been looked after well for some time prior to the assessment.
Agitated	Sheep appear to be 'on edge', becoming increasingly active and frustrated with a current event or situation. Occurs, for example, when ewes with lambs are approached by a human. Physical restlessness, twitchy, possibly foot stamping may be observed.
Responsive	The sheep respond to a stimulus, i.e. the presence of observers, dogs, or movement of other sheep. If this stimulus occurs spontaneously, for example the sheep 'notice' the presence of the assessors then this will suffice as a stimulus but if no obvious stimulus occurs the assessor should clap loudly, twice. Responsive sheep may stop whatever activity they are engaged in and raise their head towards the direction of the stimulus. A highly responsive flock may walk or trot away from the stimulus, or show a response to indicate the presence of the stimulus has been noticed.
Dull Demeanour	This is a subtle descriptor, applied to sheep that appear to be mentally dull and depressed. Sheep may exhibit a low head carriage, and individual sheep may show physical separation from the rest of the flock.
Content	A content sheep is one that is judged to be 'happy' with life; this could have been exhibited in numerous ways; for example lambs running and playing or sheep simply grazing. Observers should ask themselves – 'would you like to be a sheep at this farm?'
Anxious	A very general term that implies the sheep appear to be concerned and possibly nervous about an event that is actually happening or a potential event. Anxious, nervous sheep may appear 'twitchy', moving away from the subject or object or current situation.
Bright	A bright, alert sheep shows interest in its surroundings, and is mentally aware of any stimuli and current events. This term is considered to be the opposite term to 'dull demeanour'. 'Bright' does not apply to any physical description such as the colour or cleanliness of the fleece.
Tense	Tense refers to 'mental' and physical (muscular) tension, a sheep that does not appear 'at ease', and may have demonstrate obvious signs of physical tension such as a rigidly-held body posture.
Vigorous	Vigorous sheep are physically active at the time of assessment or have the potential to be physically active. Vigorous needs to be differentiated from thriving. For example, elephants in a zoo could in perfect health and can highly on the thriving term but, compared to wild elephants, they may be seen to have reduced lustre or 'vigour' in their behaviour.
Distressed	This is a very general term to describe the situation in which sheep appear to be unhappy or miserable and does not necessarily mean 'fearful' or 'stressed' as an animal can be distressed without being fearful.

Table 9.2 Observer population for QBA studies

Observer identity	Occupation	Experience Category
1	Vet	Experienced
2	Vet	Experienced
3	Vet	Experienced
4	Vet	Inexperienced
5	Student	Inexperienced
6	Student	Inexperienced
7	Student	Inexperienced
8	Student	Inexperienced
9	Student	Inexperienced
10	Student	Inexperienced
11	Farm certification assessor	Inexperienced
12	Farm certification assessor	Inexperienced
13	Farm certification assessor	Inexperienced
14	Farm certification assessor	Experienced
15	Farm certification assessor	Experienced
16	Farm certification assessor	Experienced
17	Farm certification assessor	Experienced

9.2.3 QBA methodology

The validity, reliability and feasibility of QBA were examined in 4 studies comprising video clips and on-farm visits. The protocol was selected according to the type of study and is described further below. In all studies, the observer performed independent assessments of sheep behaviour. Each descriptive term was scored along a VAS scale (125 mm long), labelled from 'minimum' to 'maximum' to represent the perceived level of behavioural expression. During Study 1, the sample animals were assessed using the 14 fixed-list terms developed by the QMS study. For the remaining studies, all sample animals were assessed using the 12 descriptive behavioural terms defined in Table 9.1.

Video clip assessments

The reliability of QBA was examined during video clip assessments (studies 1, 2 and 3), comprising 12 digital one-minute video clips (provided by Dr Wemelsfelder) of individuals or small groups of sheep (n < 20). Each clip was consecutively shown twice and at the end of this the observer scored each of the descriptive terms on the VAS scales. Study 1 was performed on 6th February 2009 by 8 assessors (observers 1 – 8) and informed the protocol used in subsequent studies. Prior to performing Study 2 on 30th June 2009, a group of 6 observers (including 4 assessors who participated in Study 1) spent 1 hour discussing the revised list of 12 fixed QBA terms in order to produce an agreed definition (Table 9.1). In Study 3, farm assurance assessors (n = 7), with a range of experience in performing sheep welfare assessments, participated in a video clip study held on 15th December 2009.

On-farm assessments

Study 4 was a prospective longitudinal on-farm study that was performed on 12 farms that had been previously recruited (Chapter 3) and were selected according to their location and farm type. The study was performed during May 2009 – April 2010 on 12 flocks, in which a sample of sheep from each farm was repeatedly assessed by the author at an interval of approximately 60 days. Sample sheep were selected for repeat assessments as previously described in Chapter 8. The sample group was firstly assessed using the QBA protocol described below and this was followed by the assessment of 8 animal-based indicators detailed in Chapter 3.

There were two parts to the assessment. Firstly an undisturbed observation in which the observer quietly approached the sample group and performed assessments from a distance – standing at the boundary of fields or several metres from penned animals with the aim of causing minimal disturbance to the group. QBA assessments commenced following a 5 minute period to allow sheep to settle and accustom to the presence of a group of assessors (n > 2). Thereafter, the number and location of observation points selected was according to the size of the field and the relative size of the sample group. The second part of the on-farm assessment method, required the observer to quietly approach the sample group to encourage any sheep that were resting or densely distributed to move and become more dispersed. Once the period of undisturbed observation was completed, the observer walked quietly to any additional observation points. Recording sheets were not examined during the on-farm periods of observation period and scorings were completed at the end of the farm visit.

9.2.4 Statistical analysis

VAS scales were visually examined and observer scores were recorded by measuring with a ruler the distance in millimetres between the minimum point and the mark on the line made on the scale. Data was analysed using Principal Component Analysis (PCA) in Minitab version 15.1 (Minitab, Inc, State College, PA). PCA was used to identify the least number of components that explained most of the variance in the data (Jolliffe, 2002).

As all QBA measurements in studies 1 to 3 were produced along the same scoring scale, PCA was performed using a covariance matrix (no rotation). Since QBA and physical measures were assessed along different scale, PCA for the longitudinal study was performed using a correlation matrix (no rotation) (Jolliffe, 2002). Linear orthogonal transformation of the original data produced a new set of uncorrelated variables called the 'principal components' (PC's). These new components contained the same information as the original data but they differed in the way that the information was distributed in each PC. More information was contained in the earlier components so that the first principal component (PC 1) explained the maximum level of variance and the last component explained the least amount of variance (Jolliffe, 2002). The amount of variance within each principal component was explained by a series of values known as eigenvalues. The distribution of these eigenvalues was visually examined by means of a scree plot (Jolliffe, 2002). PC 1 and PC 2 accounted for a cumulative variance ≥ 70 % for all observers, so only two components were retained in the subsequent analyses.

The correlation between each descriptive QBA term and PC 1 and PC 2 was contained in the series of values for each observer, known as the 'loadings'. These values were examined to verify that the positive terms included in PC 1 (relaxed, responsive, thriving, bright, content vigorous, content) were associated positive values and that the negative terms included in PC 2 (dejected, agitated, dull demeanour, tense, distressed) were attributed with negative values. Each observer's results were assessed case by case and where necessary, loadings were adjusted by multiplying all values in the PC by a factor of -1. The adjusted loadings were then plotted against the two-dimensional axes (PC 1 and PC 2) to explore the correlation between terms and components, and also to examine for any patterns in observer assessments (Rousing and Wemelsfelder, 2006). The PC 1 and PC 2 scores for each sample (farm visit or video clip) were adjusted in the same way.

The inter-observer reliability between observer PC scores was determined by Kendall's coefficient of concordance (W) (Kendall and Smith, 1939) and interpreted according to Martin and Bateson (2007). In addition, post-hoc visual examination of each VAS scales of each observer was performed to identify any scoring errors or difficulties with certain descriptive terms.

For Study 4, the association between PC 1 and PC 2 scores was examined using 3 loading plots: 1. a loading of the group welfare measures, 2. a loading plot of the QBA scores and 3. a combined plot of the QBA terms and group indicator scores. The loading values were examined to identify any association between the loadings of QBA terms and physical welfare indicators. In addition, the association between QBA terms and physical measures of welfare, such as lameness and cleanliness assessments were examined using Spearman's rank correlation coefficient (rho).

To identify whether QBA was capable of identifying variation in farm-level PC scores across the period of the longitudinal study, repeated-measures analysis of variance (ANOVA) were used in which farm identity was included as a repeated effect (Girden, 1992). In addition, linear regression models were fitted in which longitudinal study visit (1 – 6) was included as a fixed effect and farm identity was maintained as a random effect. The model outcomes were evaluated using coefficient (β), 95 % confidence interval (CI), and Wald p-value (Long and Freese, 2006) in which the baseline outcome for comparison was visit 1 (May to June 2009).

9.3 Results

Assessors in all studies were capable of distinguishing differences in the terms that contributed to PC 1 and PC 2. The first component (PC 1) distinguished the level of 'mood', ranging from 'bright', 'vigorous', 'content', to 'dull', 'distressed', 'dejected'. The second component (PC 2) distinguished the level of arousal or 'responsiveness', ranging from 'tense', 'anxious' and 'responsive' to 'relaxed'. As an example, the loading plots from Studies 1 (Figure 9.1) and 4 are shown (Figure 9.2). Overall, good levels of inter-observer reliability ($W > 0.80$) were achieved during most video clip assessments (Table 9.3).

Table 9.3 Inter-observer reliability of QBA studies determined by Kendall's W

Study visit	PC 1	p-value	PC 2	p-value
1	0.75	0.0001	0.71	0.0001
2	0.91	0.0001	0.77	0.0001
3	0.51	0.001	0.73	0.001

During May 2009 – April 2010, QBA and 8 physical indicators of sheep welfare assessed by group observation were tested on a total of 5740 adult sheep and growing lambs. Repeated sampling (n = 6) of sheep from 12 flocks produced a total of 72 on-farm assessments. Loading values identified that the fixed-list terms distinguished between PC1 (48.2% variation) – the level of 'mood,' which ranged from 'content/relaxed/thriving' to 'distressed/dull/dejected', and PC2 (19.8%) – the level of responsiveness, which ranged from 'anxious/agitated/responsive' to 'relaxed'.

Spearman's rank correlation coefficient (rho -0.56, p 0.001) identified a negative correlation between mood (PC 1) and the proportion of lame sheep. A correlation was also found between the proportion of lame sheep in the group and QBA terms of 'dull' (rho 0.50 p 0.001),' distressed' (rho 0.57, p 0.001) and 'dejected' (rho 0.57, p 0.001). By comparison, a correlation between QBA scores and the proportion of sheep with dirtiness of the rear was not identified. The interpretation of any association between QBA terms and indicators of dirty belly, coughing, pruritus, wool loss, excessive panting was precluded due to the low proportion of sheep observed with these indicator outcomes.

Differences in farm-level PC 2 scores were found for the combined PCA of group indicator assessment and QBA (Figure 9.3 and Tables 9.5 and 9.6). The position of farm-level scores, shown in Figure 9.3, are interpreted by examining the co-ordinates of the QBA terms and group indicator loadings presented in Figure 9.2. Repeated-measures ANOVA and linear regression models also concurred with these results and suggested that there was an effect of study visit on PC 2 ($p<0.001$), in which PC 2 scores were higher during visit 1 (May – June 2009) compared to study visits 2 – 6 (Table 9.6). By contrast, no effect of study visit was found on PC 1 ($p<0.31$).

Figure 9.1 Observer loading plots for study 2

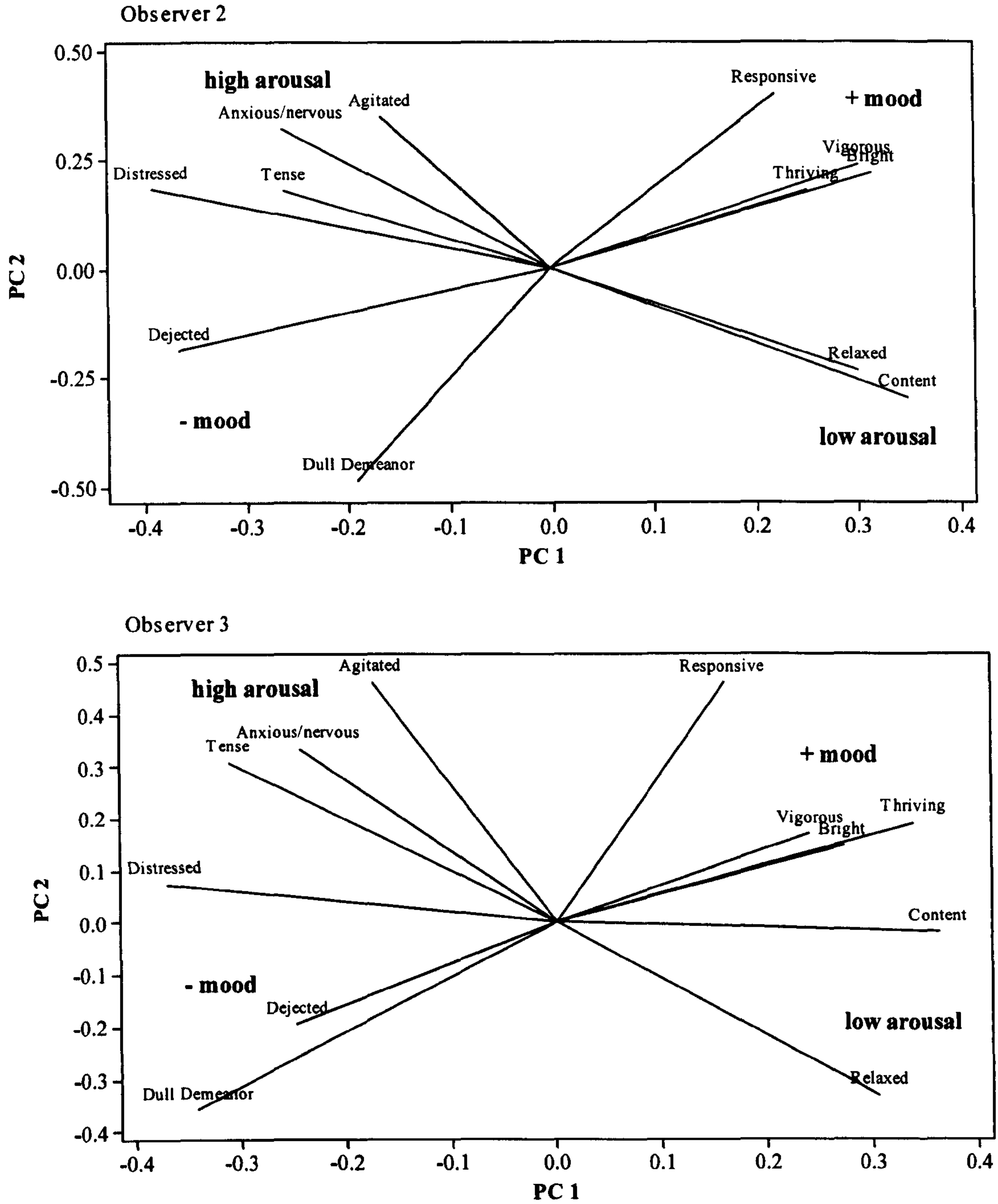


Figure 9.1 shows that in study 2 the first component distinguished between positive (upper right quadrant) and negative mood (lower left quadrant), and the second component distinguished between high (upper left quadrant) and low levels of arousal (lower right) for both observers.

Table 9.4 Repeated-measures ANOVA of the effect of study visit on PC scores

Principal component scores	F value	p-value
QBA – PC 1	1.24	0.305
QBA – PC 2	7.43	0.001
QBA and group indicators – PC 1	1.48	0.211
QBA and group indicators – PC 2	8.88	0.001

Table 9.5 Repeated-measures ANOVA of the effect of farm identity on PC scores

Principal component scores	F value	p-value
QBA – PC 1	20.18	0.001
QBA – PC 2	6.97	0.001
QBA and group indicators – PC 1	18.55	0.001
QBA and group indicators – PC 2	8.01	0.001

Figure 9.2 Loading plot of QBA terms and group indicators for Study 4

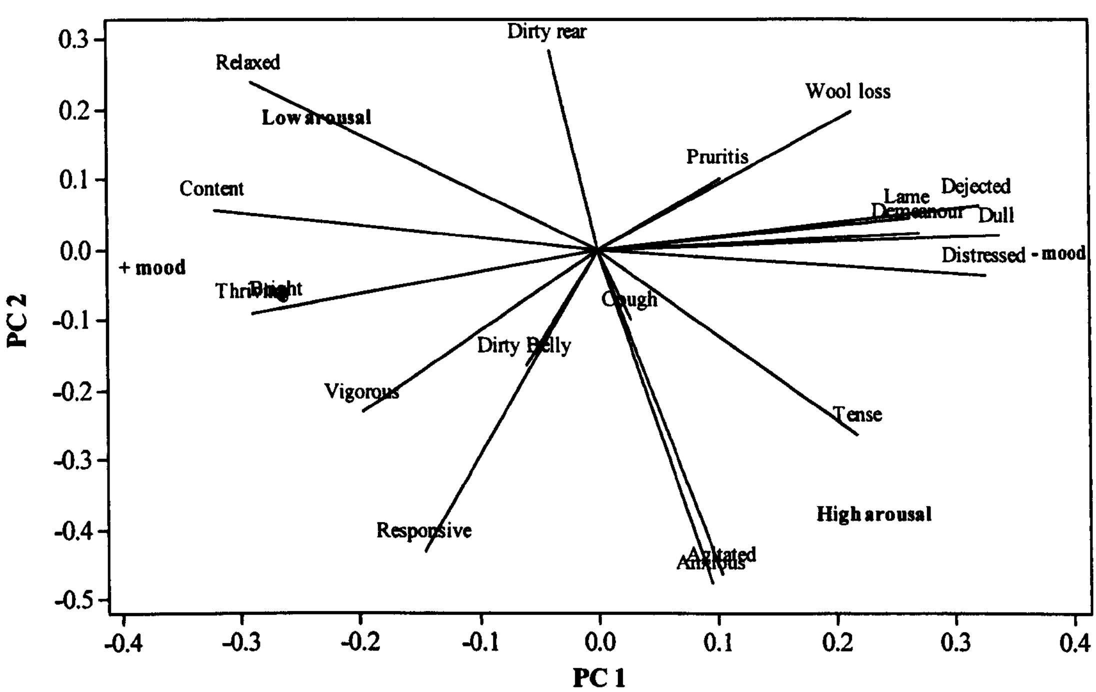
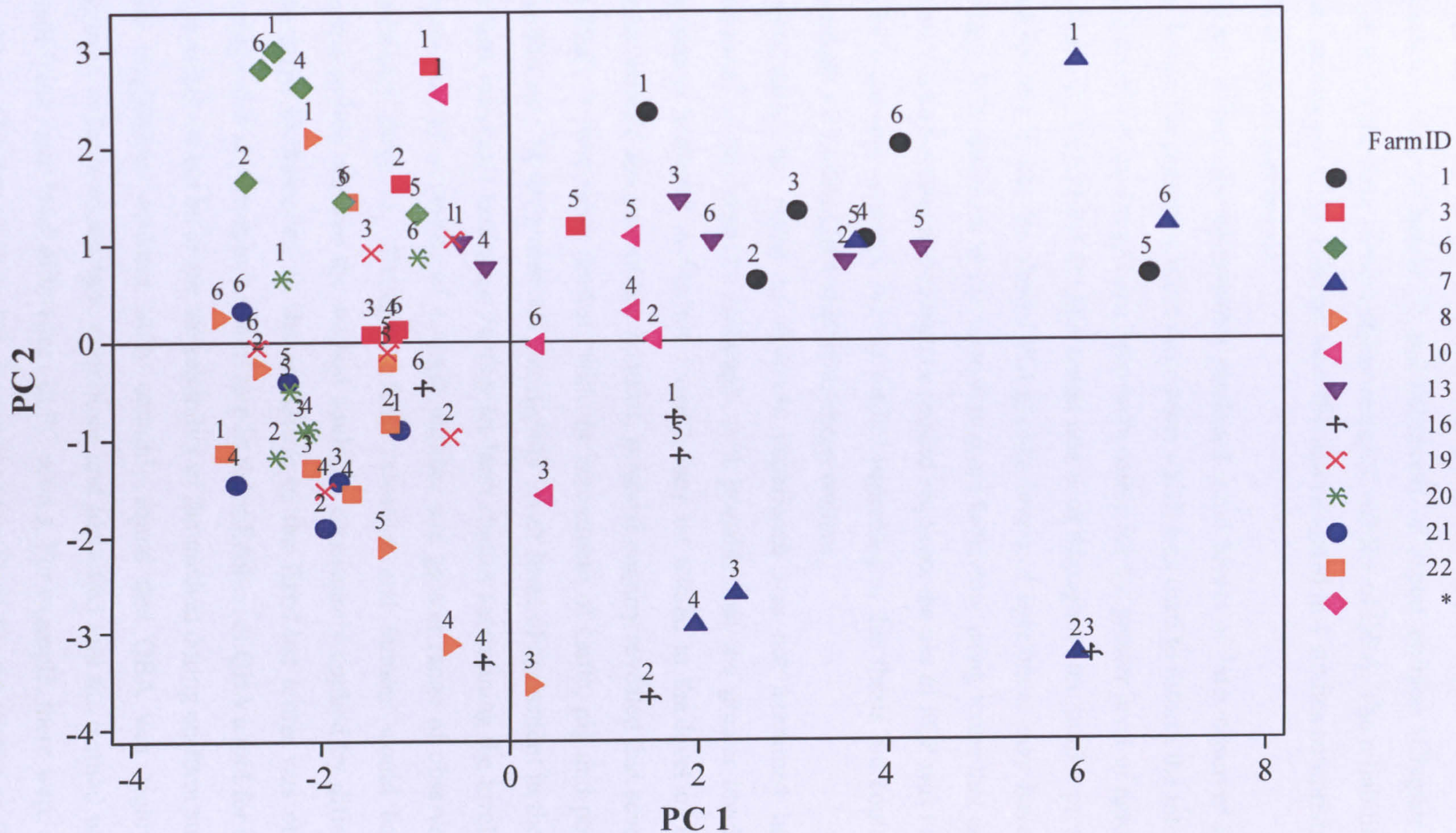


Figure 9.2 shows that terms associated with negative mood, such as 'dejected, dull and distressed' are loaded with physical measures of lameness and dull demeanour. Whilst physical measures of dirty belly, dirty rear and coughing do not appear to be loaded with any of the 12 fixed-list terms of sheep behaviour.

Table 9.6 Linear regression model of the effect of study visit on PC scores

Principal component	Study visit	β	95 % CI	p-value
QBA PC 1	1	0.21	-1.11 – 1.54	0.754
	2	-0.55	-1.47 – 0.36	0.234
	3	-0.54	-1.45 – 0.37	0.245
	4	0.11	-0.80 – 1.02	0.810
	5	-0.56	-1.47 – 0.35	0.229
	6	0.27	-0.64 – 1.19	0.558
QBA PC 2	1	1.25	0.46 – 2.04	0.002
	2	-1.82	-2.61 – -1.03	0.001
	3	-2.04	-2.83 – -1.25	0.001
	4	-1.93	-2.72 – -1.14	0.001
	5	-1.01	-1.80 – -0.21	0.013
	6	-0.73	-1.52 – 0.06	0.071
QBA and group measures PC 1	1	-0.41	-1.86 – 1.05	0.585
	2	0.58	-0.45 – 1.62	0.274
	3	0.66	-0.38 – 1.69	0.215
	4	-0.13	-1.16 – 0.91	0.810
	5	1.14	0.10 – 2.18	0.031
	6	0.18	-0.86 – 1.22	0.730
QBA and group measures PC 2	1	1.31	0.47 – 2.15	0.002
	2	-1.93	-2.73 – 1.12	0.001
	3	-2.10	-2.91 – -1.30	0.001
	4	-2.15	-2.95 – -1.34	0.001
	5	-1.17	-1.97 – -0.37	0.004
	6	-0.51	-1.31 – 0.29	0.213

Figure 9.3 Score plot of QBA terms and group indicators assessments of Study 4



In Figure 9.3, farm identity (ID) is shown by a colour coded symbol, interpreted using the legend (right of figure). The number above the symbol relates to the study visit: 1 (May-June), 2 (July-Aug), 3(Sept-Oct), 4 (Nov-Dec), 5 (Jan-Feb) and 6 (March-April). This shows both the seasonal and between-farm variation in principal component scores that occurred over the course of the longitudinal on-farm study.

9.4 Discussion

The objective of the studies presented in this chapter was to investigate the validity of using QBA as a measure of sheep welfare. In the absence of a reference standard, a literature review (Chapter 1) and consensus of expert opinion (Chapter 2) provided evidence of the face, content and consensual validity of QBA. The reliability, feasibility and consistency of the method were then investigated in 4 studies comprising video clip and on-farm assessments.

Overall, video clip assessments produced good levels of inter-observer reliability. In particular, the pre-assessment discussion which was used to inform the interpretation of each fixed-list term may have been responsible for the greater level of agreement found in Study 2. A fixed-list of QBA terms was used throughout the studies presented in this chapter and it may be argued that greater levels of agreement may have been found using a FCP approach as the observers score behaviour using terms that are familiar to them. Given the time and resources needed required, the use of FCP was not considered to be a feasible approach for the studies presented in this thesis but could be a useful approach for further QBA studies of sheep welfare.

Furthermore, the effect of observer experience was not examined in the studies presented in this chapter. Although, it is possible that the greater level of observer agreement achieved in Studies 1 and 2 may be related to the level of experience in sheep welfare assessment. By contrast, personal enquiry revealed that several assessors in Study 3 were more familiar with the assessment of cattle, pig and poultry welfare and this may be the reason for the slightly lower level of agreement in the PC 1 scores of farm assurance assessors. Further on-farm studies determining the level of inter- and intra-observer reliability of a larger number and greater range of observers, including veterinary surgeons, ethologists, farm assessors and farmers would be valuable in understanding whether the method could be consistently applied by different types of assessors. Furthermore, as the reliability of the fixed-list terms was only examined during video clip studies, a useful step in the validation of QBA a tool for sheep welfare assessment would be to test the reliability of the method during on-farm studies.

The longitudinal on-farm study certainly found that QBA was highly feasible to perform under working farm conditions and identified that the method was capable of identifying farm-level differences in PC scores. For example, there were differences in the levels of anxiety (Farm 16), responsiveness (Farm 8) and relaxation (Farm 6). This farm-level variation may be related to the breeds of sheep and familiarity with handling. The greater level of responsiveness of hill sheep (Farm 16) may reflect the reduced frequency of gathering and familiarity with handling (Boivin *et al.,* 2000). Although, sheep from lowland flocks (Farm 8) also showed very strong flocking instincts, which might have reflected the temperament of the breed or indicated the strong social stability of groups of sheep from certain study farms.

QBA was also capable of identifying variation in farm-level PCA scores over the course of the sheep production calendar. For example, there was variation in PC 2 scores with the highest scores recorded observed during May to June 2009 (study visit 1). This may indicate a possible 'relaxing' effect of the presence of young lambs on the sheep's responsiveness. Whilst it could be speculated that the altered PC 2 scores may be related to physiological changes and expression of maternal behaviour these factors were not examined by this thesis and so no inferences can be made here.

Repeated sampling also identified that certain farms had ongoing health and welfare problems. For example, sheep on Farms 7 and 13 were consistently scored with high PC 1 scores and higher levels of dull demeanour, distress and lameness throughout the one-year study period compared to other study farms. Across the study population, PC 1 scores did not significantly differ over the year and it appeared that the 'mood' of sheep in this study population remained stable throughout the longitudinal study. This was an interesting finding as farm scores on PC 1 ('general mood') are used as a QBA welfare indicator in Welfare Quality® on-farm welfare assessments (Wemelsfelder *et al.,* 2009a; Wemelsfelder *et al.,* 2009b). The relative stability of PC 1 farm scores over time suggest that there was consistency in the 'mood' of the flock and this could provide further support of the validity of QBA as an indicator of on-farm welfare.

Whilst the same farm could be consistently identified with poor welfare, the loading terms providing the overall PC scores could differ across the year. This suggests that not only was QBA stable, it was also sensitive to identifying deficiencies in welfare which may be due to different health and welfare issues during the course of the annual sheep production cycle.

The proportion of lame sheep was correlated with 'general mood' and particularly with terms such as 'dull', 'dejected' and distressed'. This meaningful association provides further evidence of the validity of QBA as a measure of sheep welfare. By contrast there was no clear relationship between rear cleanliness and any QBA term, suggesting that this physical measure was not considered to be an on-farm welfare issue in this study population. Variation in physical measures of welfare was also identified during the on-farm study. For example, during November – December, sheep from hill flocks (Farms 16 to 22) had a higher proportion of dirty bellies, which appeared to coincide with heavy rainfall and wet pasture conditions but did not appear to be linked to the PC scores. Other associations between QBA and physical measures of coughing, wool loss, skin irritation were not found and this may be related to the low proportion of these indicators in the study population.

It is also important to recognise that the same observer performed all assessments and the longitudinal study findings were not independent, as observations of poor standards of welfare during one study visit may have biased subsequent assessments. However, for the purposes of this study, it was useful to use the same assessor in order to maintain consistency in the application and interpretation of the descriptive terms and may also explain the stability in the assessment of the 'mood' (PC 1) of sample sheep.

The longitudinal findings were also limited by the small study farm population (n = 12) and the low proportion of sheep with sub-optimal welfare conditions may have been the result of farmer selection of the sample population. In addition, because of issues with the feasibility of selecting the same sheep for repeated exam there may have been variation in PC 2 because of the assessments of different groups of sheep from the same farm. Although, there was no evidence of deliberate bias in the selection of sample, it is possible that such factors may have affected the interpretation of the longitudinal study QBA scores. The position of each study farm on the score plot was anchored by the scores of other farms and the low proportion of sample sheep with conditions associated with poor welfare (Chapter 8) may have influenced the position of certain farm-level scores. Therefore, future studies need to examine the validity of QBA on a larger sample of sheep farms with a wider range of welfare problems to identify whether the findings of this chapter are applicable to the wider sheep population. This would provide further evidence of the value of QBA for on-farm assessments of sheep welfare.

The QBA on-farm protocol tested in this thesis also differed from the undisturbed assessment of group behaviour used to assess other farm animals, such as pigs (Wemelsfelder *et al.*, 2000). It was apparent that the dynamics of sheep behaviour meant that it was difficult to assess all animals within the group when they were grazing or resting in close proximity to one another. Therefore, the on-farm protocol tested in Study 4 was modified to include a period of disturbance to distinguish between immobility that was associated with calm, relaxed and restful states and recumbency that was due to an underlying physical or mental condition. As only a single on-farm protocol was tested it would be useful to examine the validity of other on-farm protocols. Further work could compare the effect of a single observation point to several points of observation, investigate the effect of disturbance on observer scores and develop alternative methods of observing sheep that flock closely together.

On-farm assessments may also have been influenced by the context and location of observations. Observers were blinded to information such as clinical or production records but were not blinded to factors such as weather conditions, farmer attitude or farm appearance, which could have influenced the impression of the assessor. Differences in the perception of extensive and intensive management systems may also have affected on-farm studies. The effect of environment on QBA score of pig welfare has been investigated (Wemelsfelder *et al.*, 2009c), so, further studies need to address the effect on the environment on the outcome of sheep welfare assessments.

9.5 Conclusion

The work presented in this chapter identified that QBA was a feasible, robust and responsive means of assessing a range of sheep from different farms during a one-year period. QBA assesses welfare beyond the physical appearance and has the potential to address the quality of the on-farm experience of sheep. As video clips assessments demonstrated good levels of inter-observer reliability, the next step would be to evaluate the inter-observer reliability of a broader range of assessors and test the reliability on the method during on-farm studies. As video clips were easy to conduct, they offered a practical means for observer training and interpretation of fixed-list terms. Since the QBA terms need to be relevant and understood by the observers who use them, further studies could generate a new set of terms using FCP. The relative stability of farm-level PC 1 scores of QBA over time and their meaningful association with lameness, support the reliability and validity of QBA as an indicator of the on-farm welfare of sheep.

Chapter 10

IDENTIFYING CUT-OFF POINTS FOR ANIMAL-BASED INDICATORS OF SHEEP WELFARE

10.1 Introduction

A number of valid, reliable and feasible animal-based indicators, akin to diagnostic tests, have been developed in this thesis to assess the on-farm welfare of sheep and lambs. Each test is applied to assess the welfare of the individual animal but the outcome of an assessment can be interpreted in terms of the proportion of the flock affected by a particular welfare condition. Therefore, a threshold value or 'cut-off point' can be used to distinguish between 'acceptable' and 'unacceptable' standards of animal welfare (Mendl, 1991). In this way, cut-off points could be used by farm assurance schemes or statutory animal welfare inspection to identify sheep farms with good standards of sheep welfare as well as identifying those with unacceptable standards which require further investigation or improvements in specific aspects of flock management.

Currently there are no scientifically-validated thresholds or 'cut-off points' that define acceptable levels for conditions that impact on sheep welfare, such as thin body condition, pruritus or lameness. Cut-off values for diagnostic tests can be selected on the basis of validation studies so that a threshold value with good sensitivity (Se) and specificity (Sp) can be applied (Greiner and Gardner, 2000). However, it is not possible to select cut-off points for animal welfare indicators solely on the results of their diagnostic performance, as any welfare standards also need to be ethically acceptable to a wide sector of society (Sørensen and Fraser, 2010).

Cut-off points that address ethical concerns for animal welfare have been developed for indicators of cattle, pig and poultry welfare in the Welfare Quality® project using the opinion of a group of experts (Botreau *et al.*, 2009). Therefore, the objective of this chapter was to use an informed consensus of expert opinion to ascertain preliminary cut-off points for acceptable and unacceptable levels of welfare conditions for sheep.

10.2 Materials and Methods

10.2.1 Expert panel meeting

The selection of expert members was described in Chapter 2. As a result, all members (n = 30) were invited to attend a final meeting of the expert panel in order to ascertain the consensus opinion regarding the potential application of the welfare indicators and identify preliminary cut-off points for each animal-based indicator.

A final meeting of the expert panel, held on 15th June 2010, was attended by 23 members including sheep farmers (n = 6), specialist sheep veterinary surgeons (n = 3), government policy advisors on farm animal welfare (n = 2), welfare inspectors (n = 2), farm assurance assessors (n = 2), veterinary surgeons from general practice (n = 2), agricultural sheep consultants (n = 3), welfare research (n = 2), and a representative from an animal welfare organisation (n = 1).

The meeting was conducted according to the principles of the National Institute of Health (NIH, 1990) method. The panel was maintained as a whole group throughout the course of the meeting. Following an initial update on the project, the chair of the meeting (the author) gave a presentation to clarify indicator scoring systems and on-farm assessment methods. Thereafter, experts were asked to provide their individual and independent opinion on the final set of animal-based indicators for sheep welfare, which were described in the final set of Standard Operating Procedures (SOP's) (Appendix C).

Each expert was provided with 2 workbooks, labelled 'A', and 'B', which were numerically coded so that paired workbooks could be later identified during data analysis. Workbooks were maintained anonymous and experts only recorded their primary area of occupation (farmer, vet in practice, sheep vet specialist, sheep industry, welfare assessment, welfare research, governmental agency or non-governmental organisation). The importance of independent and individual expert opinion was outlined and experts were requested not to confer or disclose their scores with other panel members at this stage.

A synopsis of expert comments and feedback from previous meetings was summarised and presented in workbook A. To ascertain whether this was the consensus statement of the panel, experts were asked to provide their agreement by marking a tick-box labelled 'agree' or 'disagree'. A majority-rule approach to consensus of opinion (Scott and Black, 1991) was taken, which was defined in this study as > 90 % expert agreement i.e. 21 out of 23 experts.

Experts also used workbook A to provide a cut-off point for each animal-based indicator. The cut-off point was defined as the percentage (%) of the flock affected by each welfare condition that was considered to be acceptable for welfare. Levels above this value distinguished unacceptable levels of flock welfare. The cut-off points were recorded using a horizontal, visual analogue scoring (VAS) scale (120 millimetres long) which was labelled at one end 0 % and at the other end 100 %, to represent the proportion of the flock that was affected by each welfare condition. Experts were asked to provide an 'acceptable' cut-off by indicating on the VAS the maximum % of sheep affected by each welfare outcome that they considered to be acceptable and the minimum % of sheep affected that they considered to be unacceptable for welfare. Following previous discussions with the panel it was evident that a single mark on the VAS could be used to distinguish between the acceptable and unacceptable cut-off points but the panel were given the option of providing two scores to acknowledge that there may be an area of 'potential risk' for sheep welfare.

As well as providing cut-off points for the young lamb welfare indicators, experts were asked whether they agreed or disagreed with the use of an overall welfare index for lambs in which the presence of one or more young lamb welfare condition would produce a 'non-thriving' score. Again, a majority approach to consensus was taken and following completion of this exercise, all experts handed in workbook A so as to remain blinded to the scores provided.

To investigate whether the cut-off points altered when experts were provided with on-farm data, the chair presented a series of graphs which showed the outcome of welfare indicator assessments taken during a one-year longitudinal study, as described in Chapter 8. Experts were informed of the non-random nature of farm recruitment and potential bias toward farms with veterinary contact and those consenting to participate in a welfare research study. Experts were then asked to complete workbook B by providing cut-off points for each animal-based indicator. Following completion of workbook B, a chaired, open discussion was conducted to allow experts to provide comments and feedback on the meeting. These comments were recorded by the chair of the meeting and an additional note-taker.

10.2.2 Statistical analysis

According to the area of occupation recorded by each member of the panel, experts were categorised as either 'sheep farmer', 'vet (veterinary surgeon in general practice or specialist sheep veterinary surgeon) or 'other' (welfare researcher, farm assurance assessor, industry consultant, animal welfare charity, government policy advisor).

Each workbook was electronically scanned and the cut-off points provided on each VAS scale were electronically measured (in millimetres) using ImageJ (Abramoff *et al.*, 2004). Data was analysed using Stata version 10 (StataCorp LP, Texas) and the cut-off values were log-transformed to improve the distributional assumptions of normality and homeoscedasticity (Kirkwood and Sterne, 2003). Differences between the cut-off points provided in workbooks A and B were examined graphically and using paired t-tests. Differences between the cut-off points provided by different categories of expert (vet, farmer, other) and acceptable and unacceptable values were analysed using a two-way repeated-measures analysis of variance (ANOVA).

10.3 Results

Workbooks A and B were completed by 23 experts who were categorised as sheep farmers ($n = 7$), veterinary surgeons ($n = 6$) and other occupations ($n = 10$). Data analysis found that there was no statistical difference ($p > 0.05$) between the cut-off value provided in workbook A and B for several animal-based indicators. As the cut-offs determined by workbook B were informed by on-farm data, the results of workbook B were presented in this chapter. The acceptable and unacceptable cut-off points were expressed in terms of the geometric mean and 95 % confidence interval (CI) (Tables 10.1 and 10.2). In cases where the effect of expert category was not statistically significant ($p > 0.05$), the geometric mean of the whole panel was provided. Where there was a significant effect ($p \leq 0.05$), the cut-off determined by each expert category was presented in Tables 10.1 and 10.2.

Graphical representation of the data and results of the two-way repeated-measures ANOVA found that, for most indicators, there were no significant differences in the cut-offs suggested in workbook B by different categories of expert. However, there were differences between experts in the cut-off levels for lameness, wool loss, dirty rears and short tail length in adult sheep and growing lambs (Table 10.2). For example, compared to 'other' experts, sheep farmers and veterinary surgeons considered that higher levels of wool loss were acceptable for flock welfare (Figure 10.1).

All experts agreed that the provisional cut-off points provided may be used to indicate the on-farm sheep welfare standards, but exceeding the cut-off points does not necessarily indicate non-compliance with legal standards. Instead the panel agreed that the cut-off points should be taken to indicate that further investigations are required. The expert panel also agreed that welfare indicator assessments should be performed by experienced and trained assessors who are familiar with the on-farm assessment of sheep and trained to the scoring systems used. The panel also agreed that the determination of cut-off points for the indicators should be an active and on-going process in which the scoring systems and cut-off points should be refined and modified according to on-farm experience and scientific advances. All but one expert (95.7 %) agreed that the young lamb indicators could be applied as an overall 'thriving lamb' score.

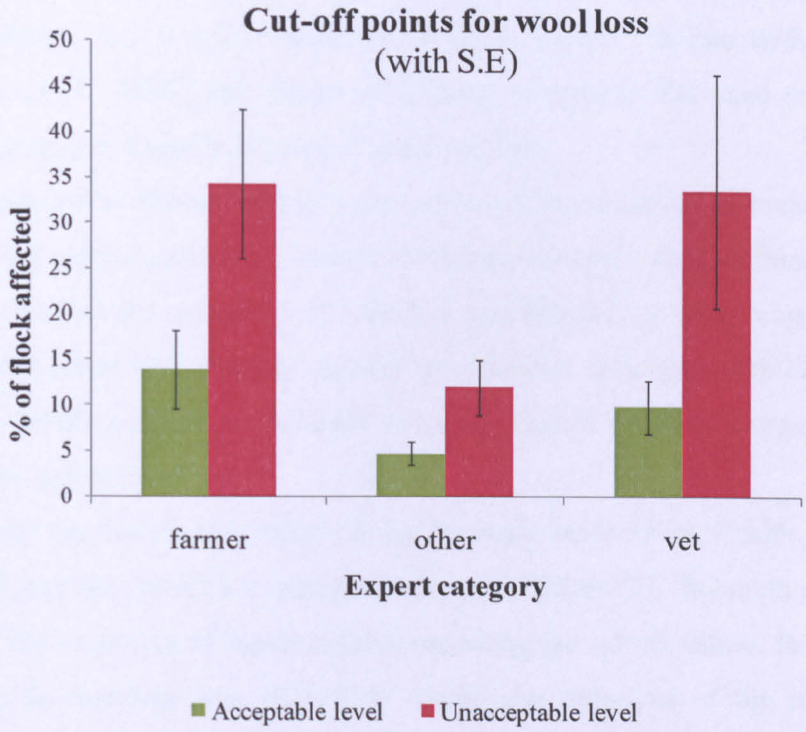
Table 10.1 Cut-off points for indicators of adult sheep and growing lamb welfare

Welfare indicator	Acceptable % cut-off point (95 % CI)	Unacceptable % cut-off point (95 % CI)
Dull demeanour	**2.0** (1.1 – 3.6)	**6.1** (4.2 – 9.0)
Skin irritation	**1.2** (0.8 – 2.0)	**3.8** (2.7 – 5.4)
Lameness	**3.8** (2.3 – 6.4)	**12.6** (8.9 – 17.9)
Wool loss	**5.8** (3.8 – 8.8) **10.7** (5.5 – 21.04) farmer, **3.3** (1.7 – 6.2) other, **7.1** (2.5 – 20.1) vet	**16.4** (11.0 – 24.5) **27.0** (12.6 – 58.1) farmer, **9.4** (5.6 – 15.7) other, **23.0** (8.3 – 64.0) vet
Dirty rear (group)	11.7 (8.2 – 16.8) **21.8** (11.8 – 40.3) farmer, **9.3** (5.1 – 17.2) other, **8.3** (4.4 – 15.8) vet	**27.4** (20.6 – 36.4) **46.4** (26.2 – 82.4) farmer, **23.4** (15.1 – 36.2) other, **19.2** (12.3 – 30.1) vet
Dirty belly (group)	**9.9** (5.3 – 18.4)	**24.4** (14.7 – 40.5)
Thin condition (lowland)	**4.1** (2.3 – 7.2)	**9.5** (5.3 – 16.9)
Thin condition (hill)	**4.8** (2.5 – 9.3)	**11.8** (6.6 – 20.9)
Fat condition (lowland)	**8.8** (4.2 – 18.6)	**18.7** (11.0 – 31.7)
Fat condition (hill)	**8.3** (4.2 – 16.5)	**17.5** (9.0 – 34.0)
Dirty belly (individual)	**10.5** (6.7 – 16.3)	**28.8**(18.7 – 44.4)
Filthy belly (individual)	**3.2** (1.5 – 6.9)	**10.9** (5.8 – 20.6)
Incisor loss	**12.3** (6.1 – 23.5)	**29.8** (17.6 – 50.2)
Molar disease	**3.2** (1.9 – 5.3)	**7.2** (4.5 – 11.6)
Eye abnormality	**1.4** (0.7 – 2.9)	**4.6** (2.7 – 8.0)
Nasal discharge	**2.0** (1.1 – 3.5)	**5.8** (3.9 – 8.5)
Small skin lesion	**5.6** (3.1 – 10.1)	**11.7** (6.9 – 20.1)
Hand-sized skin lesion	**2.7** (1.4 – 5.0)	**7.1** (4.2 – 12.0)
Diffuse skin lesion	**0.6** (0.3 – 1.4)	**2.0** (0.9 – 4.3)
Mastitis	**4.6** (2.6 – 8.0)	**12.3** (7.9 – 19.4)
Short tail length	**0.7** (0.3 – 1.3) **1.6** (0.2 – 10.7) farmer, **0.3** (0.12 – 0.6) other, **1.1** (0.3 – 3.9) vet	**1.1** (0.5 – 2.4) **2.8** (0.3 – 24) farmer, **0.4** (0.2 – 1.0) other, **1.9** (1.4 – 3.2) vet
Myiasis	**1.0** (0.5 – 2.2)	**2.1** (1.0 – 4.5)

Table 10.2 Cut-off points for indicators of young lamb welfare

Welfare indicator	Acceptable % (95 % CI)	Unacceptable % (95% CI)
Dull demeanour	**3.0** (1.8 – 4.9)	**6.4** (4.4 – 9.4)
Unable to stand	**1.3** (0.8 – 2.3)	**3.0** (2.0 – 4.6)
Weak on standing	**2.1** (1.3 – 3.3)	**4.4** (3.1 – 6.3)
Bloated abdomen	**1.6** (0.9 – 3.0)	**3.9** (2.5 – 6.3)
Hollow abdomen	**1.7** (0.8 – 3.5)	**4.6** (2.8 – 7.5)
Hunched posture	**2.1** (1.1 – 3.9)	**4.6** (2.9 – 7.4)
Thin body condition	**1.6** (0.9 – 3.5)	**3.8** (2.1 – 6.5)
Eye abnormality	**2.8** (1.4 – 5.4)	**6.7** (3.8 – 11.9)
Lameness	**2.5** (1.3 – 4.9) **9.5** (5.2 – 17.6) farmer, **1.3** (0.4 – 3.9) other, **1.7** (0.3 – 8.1) vet	**6.3** (3.6 – 11.0) **20.0** (9.3 – 43) farmer, **3.4** (1.2 – 8.9) other, **4.7** (2.8 – 7.9) vet
Salivation	**1.1** (0.6 – 2.0)	**2.1** (1.2 – 3.7)
Shivering	**1.4** (0.7 – 2.9)	**2.8** (1.4 – 5.5)

Figure 10.1 Effect of expert category on cut-off points for wool loss and tail length



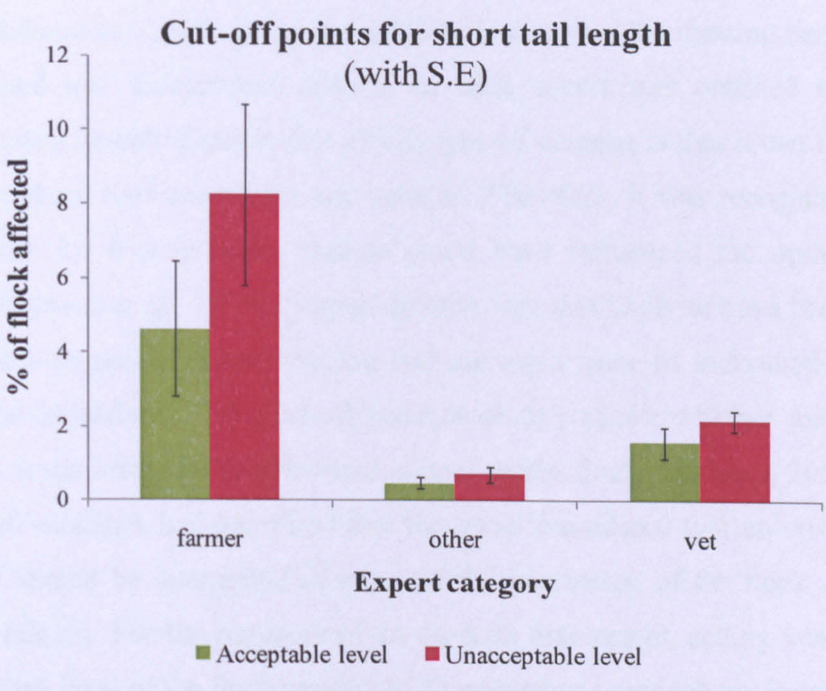


Figure 10.1 shows the differences between each expert category in terms of the geometric mean cut-off point along with the standard error (S.E) of these values.

10.4 Discussion

The objective of the study presented in this chapter was to identify cut-off points for valid, reliable and feasible indicators of sheep welfare. In line with previous studies (Botreau *et al.*, 2009), the opinion of a panel of experts was used to ascertain cut-off points for animal-based indicators of sheep welfare.

As the concept of animal welfare varies across different sectors of society (Phillips *et al.*, 2009), the expert panel was selected to include members with different views of sheep welfare, so that the cut-offs were ethically acceptable to a wide range of stakeholders. The expert panel was therefore selected to comprise members from different sectors of society including farmers, veterinary surgeons, animal protection organisations, welfare inspectors and researchers.

Given the familiarity and value of the National Institute of Health (NIH) consensus methodology for previous expert consultations (Chapter 2), this method was selected to identify the consensus of expert opinion regarding the cut-off values. In line with the NIH method, the meeting was chaired to clarify the objective of the meeting, scientific information was used to inform the decisions of the panel and finally a consensus output statement was achieved.

Experts remained in a single group throughout the course of the meeting and the need for the individual and independent opinion of each expert was outlined to the panel. However, one potential disadvantage of this type of meeting is that it can be difficult to prevent members conferring with one another. Therefore, it was recognised that there was potential for bias as some experts could have influenced the opinion of other members (Delbecq *et al.*, 1975). Expert opinion was also likely to have been influenced by their view regarding animal welfare and the importance of individual versus flock welfare. The identification of a cut-off point needs to consider whether sheep welfare is interpreted at the level of the individual animal or the flock (Goddard, 2008). Previous expert panel meetings had identified that the panel considered that an on-farm welfare assessment should be interpreted in terms of the percentage of the flock affected by a specific condition. For the purposes of an on-farm assessment, setting cut-off points at the cumulative level of the flock appears to be a practical approach but it does mean that, depending on the application of the cut-off values, a few individuals that are affected by very serious welfare conditions might not receive the appropriate attention and action.

As animal welfare can be viewed on a continuous scale from very good to very poor standards of welfare (Webster, 1994), the cut-off points were recorded on a continuous VAS scale. The selection of this scoring scale was also informed by the findings of previous expert panel meetings, which suggested that categorical scoring scales provided difficulty for many expert panel members. Although the continuous scale provided an easier means of providing the cut-off points, it was apparent that there were still difficulties with this approach. This is because it was evident that a clear cut view between 'acceptable' and 'unacceptable' standards of animal welfare was not held by all members of the panel. Whilst many experts provided a single mark along the VAS scale to indicate the distinction between acceptable and unacceptable welfare, other members provided two lines on the VAS scale, leaving a gap between the acceptable and unacceptable cut-off points. This indicates that there may be a level at which welfare may be 'at risk' but not necessarily 'unacceptable' that expert opinion alone may not be able to define. Experience from this study may inform further expert consultation processes, and it may be beneficial to ask experts to only provide a single cut-off point. Although the problem in identifying the line between the two welfare states may still arise because of the range of views that are held regarding animal welfare assessment.

Despite this, it was clear that for the vast majority of indicators, no differences between the cut-off points provided in workbooks A or B were found. The results of workbook B were therefore selected in this chapter because they were informed by scientific data and the panel had more time to reflect on the cut-off points. As experts completed workbook A, they were likely to have become more familiar with the use of VAS scales and so it is suggested that fewer scoring errors may have occurred during the completion of workbook B.

As there were few differences between the cut-off provided in workbooks A and B, it could be suggested that experts were able to consistently provide threshold values for sheep welfare assessment. However, the cut-off levels varied between workbooks A and B for a small number of measures including thin and fat body condition scores and skin lesions. One shortcoming of the study was that the reliability of the VAS scoring scale was not examined and it was possible that experts could have provided different cut-off points in the second exercise regardless of whether the farm data was shown.

It is also likely that the lower cut-off points provided in workbook B were influenced by the presentation of data from the longitudinal study (Chapter 8). The 12 flocks participating in the study were non-randomly recruited and may not have been representative of the welfare standards of all flocks in England and Wales. Farms participating in a welfare study may have been biased towards those with higher welfare standards. Furthermore, the indicators were only applied to a small sample of the flock so it is not known whether the longitudinal data reflected the true welfare status of the 12 study farms. However, a major issue for interpreting many welfare conditions of sheep is that accurate and up-to-date prevalence estimates do not exist. So, in the absence of any other validation studies, the longitudinal study data was used as a means of informing the expert panel with some farm data. Few sheep in the study population were observed with any of the welfare conditions and this may have implications for findings reported in this chapter. This is because it is possible that the cut-off points may be set at levels which identify many farms with unacceptable standards of welfare. As a result, it is important that the cut-offs presented in this thesis are tested on a wide range of farms before they can be used in any formal on-farm inspection processes.

Evidently, the welfare standards need to be set at levels which have good levels of sensitivity and specificity. For example, if the focus of an on-farm welfare assessment is to find farms with poor standards of welfare, then using the acceptable cut-off point might be more appropriate. Alternatively, if the aim of an assessment is to identify farms with good standards of welfare, setting the cut-off point at a higher level, such as the unacceptable value, might be more appropriate. Concerns over the potential use of the cut-offs, such as those used in statutory schemes with possible financial penalties, might be the reason why farmers provided higher cut-off points for wool loss, rear cleanliness and short tail length in adult sheep and growing lambs and lameness in young lambs. Alternatively, the higher cut-offs might reflect different perspectives on the impact of tail length, breech dirtiness or wool loss on flock welfare. Furthermore, it is likely that the expert panel was keen to involve experienced and trained assessors to avoid misinterpretation of certain indicators. For example, the panel suggested that rear cleanliness or 'dagginess' can be affected by the breed of sheep and characteristics of the fleece. This may be important if there are serious implications arising from an on-farm welfare assessment.

Results also demonstrated that the expert panel provided different acceptable levels of welfare conditions for sheep managed on lowland and hill farming systems. A higher percentage of thin sheep was more acceptable for animals managed under hill conditions compared to those on lowland farms. Therefore, the interpretation of the cut-off points was influenced by specific farming practices and management decisions.

Given the feasibility of a one-day meeting, the expert panel was not asked to provide guidance on the assessment of resource- and management-based indicators. However, the provision and availability resources, such as clean water or provision of a lie-back, could be interpreted as acceptable (present) and unacceptable (absent). The study did not identify acceptable levels of sheep and lamb mortality so further studies, involving the collection of mortality surveillance data and expert consultation may be needed to develop cut-off points for these measures.

Currently, each animal-based welfare indicator can be views as a separate screening test used to assess particular flock welfare issues. The panel agreed that exceeding any cut-off point did not necessarily indicate non-compliance with legal standards but instead should be used to indicate that further investigation is required. This may involve identifying the reasons behind the welfare problem or determining whether appropriate action is being performed. Therefore, these cut-off points need to be evaluated on a large and diverse sheep population before any conclusions or valid interpretation of this data can be made.

Experts also suggested that an overall 'thrift score' for young lambs could be derived by weighting indicators, such as demeanour, posture, eye condition, body condition and lameness. However, the development of an overall welfare assessment system was not within the scope of this thesis, so any overall welfare indices or assessment systems need to be addressed by further research studies.

10.5 Conclusion

A consensus of expert opinion has identified a set of preliminary cut-off points for animal-based indicators of sheep welfare developed in this thesis. It is clear that the cut-off values presented in this chapter could have been biased by the composition of the expert panel and the low level of welfare conditions observed during the longitudinal study. Therefore, the cut-off points should be viewed as preliminary levels that need to be evaluated on a range of sheep farms before they can be applied in any on-farm welfare assessment system.

Chapter 11
CONCLUDING DISCUSSION

11.1 Introduction

The objective of the work presented in this thesis was to develop valid, reliable and feasible indicators for the on-farm assessment of sheep welfare. The approach undertaken has applied the methods used in animal welfare science, veterinary epidemiology, and the behavioural and social sciences to develop scientific measures of sheep welfare. In common with work by Main *et al.*, (2001) and Whay *et al.*, (2003), the welfare indicators were developed using the framework of the Five Freedoms (FAWC, 1994). A review of the scientific literature (Chapter 1) and a consensus of expert opinion (Chapter 2) used the Freedom's framework to develop indicators that were sensitive to the current on-farm welfare issues for sheep. Whilst previous researchers have elicited expert opinion in using questionnaire-based methods conducted by postal (Whay *et al.*, 2003) or face-to-face meetings (Phillips *et al.*, 2009), this was the first study to conduct expert consultations in animal welfare research using the consensus method of National Institute of Health (NIH, 1990). The NIH approach was found to be a useful means of identifying the opinions of individuals from diverse backgrounds and occupations, feasible to use and proved a productive approach, identifying 193 welfare issues and 26 animal-, 13 resource- and 22 management-indicators of sheep welfare.

The Five Freedoms concentrate on the inputs and resources required to provide good animal welfare and this may have been the reason why many indicators suggested by the panel were focused on physical measures of health and welfare. It may also have reflected the awareness or bias of certain members of the expert panel towards measures of biological-functioning. However, members of the panel also recognised the increasing move towards welfare assessments which focus on the quality of an animal's life as per the Qualitative Behaviour Assessment (QBA) approach (Wemelsfelder *et al.*, 2000). Therefore, following the development of an on-farm study protocol (Chapter 3), these animal-based indicators were tested on a total of 10588 from 50 flocks comprised of a range of breeds, ages and types of sheep (Chapter 3).

Following previous research, the animal-based indicators were examined in terms of their reliability (Harkins, 2005; Napolitano *et al.*, 2009) and in the absence of a gold standard, the approach of Burn *et al.*, (2009) and Mullan *et al.*, (2009), was applied in which the assessments of an experienced and trained assessor (the author) were used as the reference standard for comparison (Chapters 4, 5 and 6). Results suggested that this was a suitable observer to provide training and use as a reference standard as LCA suggested that this observer achieved higher levels of Se and Sp than other study observers. Overall, few scoring disagreements in indicator assessments occurred and the vast majority of indicators including demeanour, lameness, body condition, mastitis were deemed to be 'reliable' and showed greater levels of observer agreement in comparison to previous reliability studies in sheep (Harkins, 2005).

In contrast to the Bristol Welfare Assessment Program (BWAP) (Whay *et al.*, 2003) and the WelfareQuality® project (Knerium and Winckler, 2009), this thesis took a new and more extensive approach to the validation of animal welfare indicators by evaluating the diagnostic sensitivity (Se) and specificity (Sp) of each indicator test (Chapters 5 and 6). This approach identified that many of the animal-based indicators had high Sp, suggesting that the indicator tests could work well if they were used to identify farms with good standards of animal welfare. A further novel aspect of the research presented in this thesis was the field testing of the robustness and responsiveness of animal-based outcomes, such as lameness, demeanour, body condition and mastitis, to seasonal and management variation.

The vast majority of previous research into the development of animal welfare indicators has concentrated of measures for cattle, pigs, poultry (Knerium and Winckler, 2009; Main *et al.*, 2003). By comparison, few studies have examined the reliability and feasibility of sheep welfare indicators (Harkins, 2005; Napoliano *et al.*, 2009). However, it is difficult to make cross-study comparisons because of differences in analysis used e.g. Spearman's rank rather than use of kappa (Napolitano *et al.*, 2009) and differences in the study population prevalence of specific welfare conditions (Fernstein and Cichetti, 1991). In spite of these issues, similarities were found with reliability studies performed in dairy cattle (Kristensen *et al.*, 2006), in which a range in the degree of observer reliability was found for assessments of body condition. In addition, much higher levels of observer reliability were found for body condition, lameness and mastitis assessments of sheep by the studies presented in this thesis compared to those of Harkins (2005). Furthermore, the indicators presented were examined on a considerably larger sheep population and by a larger observer population than previously reported (Harkins, 2005; Napolitano *et al.*, 2009).

11.2 Recommendations for use of sheep welfare indicators

Following evaluation of the reliability, robustness and responsiveness of each welfare indicator (Appendix C), a list of indicators recommended for current use in on-farm assessments of sheep welfare has been established (Table 11.1), which is discussed in further detail below.

11.2.1 Animal-based indicators

Whilst there are different scoring systems and assessment criteria for adult and growing sheep and young lamb measures the key animal-based indicators that worked well for all stages of production were: 1. demeanour, 2. lameness, and 3. body condition. These were measures that were reliable, robust and responsive tests could be used as key 'iceberg' indicators that could be applied to a range of different breeds and types of sheep and to both extensive and intensive flocks.

It was suggested that, for adult and growing sheep, demeanour and lameness were more reliable and feasible to assess on the basis of group observation rather than individual examination (Chapter 7). An advantage of the group observational method of assessment is that it does not require handling or gathering of individual sheep and so appears to be a feasible and useful method of assessing groups of sheep in grazing and housed environments. The indicators were tested on groups of 24 – 120 sheep, although a sample size of 70 sheep appeared to be a feasible number for most farms and was selected on the basis of pilot studies examining the feasibility of performing the study protocol during a single day. Therefore, it is not known whether indicators assessed by group observation could be feasibly applied to the larger group sizes suggested by standard sample size calculations (Appendix C).

On-farm experience suggested that a group observation of wool loss, skin irritation and cleanliness of the belly and rear may be affected by the number of sheep, as well as other factors such as size of the field, terrain and weather conditions. For example, whilst cleanliness of the belly was reliably assessed by group observation (Chapter 4), it may be more difficult to accurately count numbers when a large proportion of the group is affected and it would be useful to assess the indicators on a population with a greater proportion of affected animals. One possible solution would be to engage the farmer and shepherding dogs to divide large groups of sheep into smaller groups in order to facilitate indicator assessments. Alternatively, use of an all-terrain vehicle (ATV) might be helpful when sheep are widely dispersed over rough terrains. Similarly, whilst wool loss and skin irritation are clearly good indicators of the presence of pruritic skin lesions, such as sheep scab (Beriatua *et al.*, 2001), the failure to observe these signs during a group observation should not rule out the presence of pruritic skin lesions, because it is known that small lesions may be missed by observers (Napolitano et al., 2009) or because sheep may not be observed to show signs of pruritus during the period of observation. Few sheep in the study population were observed with wool loss or skin irritation and given that wool loss was not found to be reliable by group observation (Chapter 4), it is suggested that individual examination of animals may be required to assess both wool loss and skin irritation.

Individual animal assessment was required to assess a key indicator of sheep welfare – body condition. Evaluation of the diagnostic performance of all study observers identified that, overall, body condition could be reliably and feasibly assessed using the Russel (1984) or 'fit-fat-thin' scoring system. These two scoring approaches are both based on the 6-point scoring scale suggested by Russel (1984) but the fit-fat-thin method provides an interpretation of body condition scores. The fit-fat-thin system was more reliable than the 6-point Russel (1984) scale, although the fit-fat-thin scale had less precision and so may be more appropriate for the purposes of a flock welfare assessment rather than as a routine clinical or management tool (Lovatt, 2010).

Other individual animal indicators which were recommended for inclusion in on-farm assessments were assessment of tail length with legal compliance, mastitis, eye condition as these indicators were found to be reliable and feasible measures to perform on all study farms (Chapters 5 and 7). Assessment of the gait of individual animals for signs of lameness may be used in place of or in addition to a group observation of lameness. However, during assessments of adult sheep and growing lamb indicators, on-farm experience suggested that factors including the location of individual examination and quality of lighting and flooring as well as pen size and shape, affected the ability to perform individual gait assessments on a small number of study farms (Chapter 5). The familiarity with handling, breed and stage of production may also have influenced the ability to assess gait. In spite of these potential issues, cross-sectional study results showed that individual lameness assessment provided good levels of reliability, suggesting that farm conditions may not have caused considerable difficulties for the vast majority of observer assessments (Chapter 5). Therefore, the ability of these animal-based welfare indicators to work under varying test conditions provided further evidence of their validity and robustness.

Compared to the adult sheep and growing lamb indicators, the young lamb measures stood out as having much higher levels of observer Se, Sp and reliability (Chapter 6). This may have reflected the fact that all assessors were trained and a smaller number of well-defined tests were applied. Particular indicators of young lamb welfare which were deemed to be both reliable and feasible were demeanour, lameness and posture. As close observation was needed to assess body condition, eye condition, and abdominal fill, these indicators could not be consistently applied to all lambs. However, as entropion was identified as particular welfare issue in the study population (Chapter 7), it seems sensible to retain this measure where possible. Similarly, body condition and abdominal fill are considered to be key indicators of young lamb welfare and therefore are recommended for use where feasible to assess. This highlights one of the issues of the thesis which involved conducting the research on working farms. As such it was not possible to standardise the conditions for observation and assessment but this is part of the 'real life' scenario for on-farm assessor. Given that young lambs managed under different systems can be assessed by a different set of indicators, one possible solution would be to develop a weighted overall young lamb 'thriving score', based on on-farm research and expert opinion, to ensure that lambs managed under a variety of rearing and farming systems are capable of attaining the same overall welfare score.

Some indicators were not recommended for current use in on-farm assessments because they were not feasible to assess, for example assessment of coughing at both group and individual examination. Whilst foot lesions, which produced high levels of reliability, Se and Sp were not recommended for use (Table 11.1) because they were considered to form the further diagnostic step in the investigation of a flock lameness issue. Similarly, the assessment of tooth condition and particularly molar abnormalities (Chapter 10), was considered to be part of the next step in an on-farm assessment following assessment of body condition. Furthermore, as all forms of ear tagging are suggested to produce some degree of ear lesions (Edwards *et al.*, 2001), there was difficulty in interpreting the significance of ripped and chronic lesions, therefore, following guidance of the expert panel (Chapter 10), this indicator was not included in the list of indicators recommended for use in on-farm assessments of sheep welfare (Table 11.1).

Table 11.1 Recommendations for on-farm indicators of sheep and lamb welfare

Recommendation	Group assessment of adult sheep and growing lambs	Individual examination of adult sheep and growing lambs	Individual assessment of young lambs	Resources and management
Recommend use as welfare indicator	Demeanour Skin irritation Lameness Dirty rear Dirty belly	Demeanour Eye condition Nasal discharge In-growing horns Dirty rear Dirty belly Tail length Mastitis Wool loss Skin irritation Skin lesion Body condition score Fit-fat-thin Lameness Myiasis	Demeanour Stimulation Standing ability Body condition Posture Abdominal fill Lameness Eye condition	Provision of shelter Provision of lie-back area Provision of clean and accessible water supply Stocking rates Comfort and hygiene of bedding Adequate lighting for inspection Assessment of maintenance of equipment and facilities Presence of carcases amongst live animals Medicine records Mortality records Tail-docking policy Castration policy
Not recommended	Excessive panting Coughing	Tooth condition Ear lesions Coughing Dirty legs Crystals Foot lesion 'Significant' foot lesion White line lesion Scald Foot rot CODD Other foot lesions Joint swelling	Salivation Shivering	Sward height assessment Assessment of boundaries and fences Freedom to move in field or housing Flooring type Provision of bedding Moisture content of bedding Presence of food in troughs Accessibility of troughs Forage assessment Presence of isolation pen Assessment of handling Presence of foot bath Presence of flock health plan Mixing of horned and unhorned sheep
Further work needed	Wool loss QBA	Injury and wounds	Play behaviour - QBA	

Other indicators, such as excessive panting, crystals (urolithiasis), shivering and salivation in young lambs, were not included in the final list of recommendations (Table 11.1) because there was insufficient data from the study to allow a full interpretation of their diagnostic performance. The results of the group observation of wool loss were difficult to interpret because, with the exception of one study visit in which the measure was found to be unreliable, there were insufficient observations to deduce any meaningful estimation of the level of inter- and intra-observe reliability. As wool loss can clearly alert to serious welfare issues for sheep (Beriatua *et al.*, 2001), it is recommended that further work needs to be conducted on sheep with a greater degree of wool loss to confirm whether the measure can be consistently applied by different assessors. The evaluation of the test performance of other indicators was also affected by the low proportion of sheep with conditions associated with sub-optimal welfare. For example, the full range of categorical scores of injuries and wounds could not be fully tested and this was associated with poor levels of inter-observer agreement in this thesis. This indicator needs to be examined on a population experiencing a higher prevalence of wounds and injuries since these measures have been suggested to be valid, reliable and feasible measures in other species, such as pigs (Leeb *et al.*, 2001). The issue surrounding the prevalence of welfare conditions in a study population is not specific to this thesis and has also been identified to cause issues for test evaluation when animal welfare indicators are found at a high prevalence (Burn *et al.*, 2009). Some authors have suggested that a study population with 50% prevalence of the condition of interest should be tested (Hoelher, 1990) but this was not feasible here given the number of welfare conditions under study and the criteria for farm recruitment.

11.2.2 Qualitative Behaviour Assessment

The work presented in this thesis also provided further evidence in support of the validity of QBA as an on-farm measure of sheep welfare and was the first study to test the reliability of QBA for sheep welfare assessment using a diverse group of assessors (Chapter 9). High levels of reliability achieved for video clip assessments (Chapter 9), comparable to those found during pig and poultry on-farm assessment (Wemelsfelder *et al.*, 2009a,b) were found but it was not within the scope of this thesis to examine the on-farm reliability of a pool of observers. However, the method was found to be easy to apply and capable of detecting variation in farm-level scores and across the seasons of the sheep calendar. QBA appears to offer a means of capturing both the physical and mental experience of an individual or flock of sheep and it is possible that important on-farm welfare issues, such as dystocia and vaginal prolapses, which were not evaluated by animal-based outcomes in this thesis, are identified by the approach.

Currently, QBA farm scores on PC 1 ('general mood') are used as a welfare indicator in Welfare Quality® protocols for cattle, pig and poultry welfare assessments and have been incorporated into Scottish farm assurance protocols (Wemelsfelder, personal communication, 2011). By demonstrating the relative stability of PC 1 scores over time, and their meaningful association with sheep lameness, the work presented in this thesis supports the application of QBA as a valid, reliable and feasible tool in the on-farm welfare assessment of sheep. However, the on-farm reliability of the method needs to be examined on a wide range of sheep flocks before any final recommendations on its use as an on-farm indicator of sheep welfare can be made in this thesis. Therefore, further research addressing the on-farm reliability, development of new QBA terms using Free-Choice Profiling, testing a range of on-farm protocols and application of QBA for assessments of young lamb behaviour is highly recommended (Table 11.1).

11.2.3 Resource- and management-based indicators

The main focus of this thesis has been the validation of direct observations of animal-based measures. A major confounding factor that could influence the outcome of sheep welfare was the skill, attitude and behaviour of the stockperson. Therefore, in addition to seasonal variation, the outcome of animal-based indicators of sheep welfare may have been influenced by changes in flock nutrition, genetic composition or breeding practices. It was not feasible to measure these qualities and as a result, an assessment of stockperson skills and interactions with the flock were not tested in this thesis. Instead, the feasibility of resource- and management-based indicators, such as assessment of water sources and grazing and mortality records, was tested (Chapter 7). Given the nature of this study, farmers did not play a large role during the assessment of the indicators and it was not always feasible to access farm records or conduct a brief, farmer interview.

The difficulty in accessing records or conducting interviews may not be as apparent for welfare assessments performed for certification or regulatory purposes, which may require a greater level of farmer involvement. However, management-based indicators, such as castration and tail-docking policies and mortality rates, are still reliant on farmer recall and the quality and accuracy of records. Furthermore, these measures do not assess the quality of management procedures, such as castration and tail-docking as the observer needed to be present on the farm at the time these mutilations were being performed.

11.3 Limitations of thesis

A major limitation of work presented in this thesis was the recruitment of the study population. Ideally, a randomised selection of farms and stratified sampling process would have been taken but this was not possible due to the need for farmer consent for participation and the feasibility of travelling and conducting farm visits within the limits of a single day. Instead, a non-random and convenience sampling approach was taken in which flocks representative of the British Sheep Stratification system (Pollott and Stone, 2006) were selected so that the indicators were thoroughly tested under a range of farm management conditions and a variety of breeds of sheep (Chapter 7). It is recognised that this recruitment method may have introduced selection bias for participation of farms with higher standards of health and welfare. Consequently, this may have been the reason for the low numbers of sheep affected with signs of pruritus, myiasis and emaciation. Trading standards inspectors and Animal Health agencies were approached to assist with the recruitment of farms experiencing a higher level of sub-optimal conditions but this proved to be a disappointing approach as no farms were offered. Furthermore, given the vast areas and varied locations of flocks of sheep, farmers were asked to select the sample animals and it is recognised that this may have led to the low level of sheep observed with many welfare conditions (Chapters 7 and 8). On-farm experience suggested that rams were not always accessible and therefore the location of rams needs to be considered when conducting future on-farm assessments to ensure that the welfare status of all production stages within the flock is examined.

Given the non-random approach to farm recruitment and sampling it is important to highlight that the proportion data presented in this thesis is not generalisable to the wider sheep population in England and Wales and this was not the objective of the thesis. The approach was deemed to be appropriate for the evaluation of the diagnostic performance and seasonal responsiveness of the animal-based indicators. However, the low proportion of affected animals in the study population did present issues for the interpretation of certain indicators, such as emaciation (BCS <2), pruritis, myiasis, injuries and wounds, and skin lesions. As few animals were observed with these conditions, no information can be deduced on the performance of these tests in flocks with a higher proportion of affected animals.

The low proportion of animals with specific welfare conditions also affected the analysis of observer agreement. For example, a trained and experienced observer attained a κ of 0 for body condition scoring, despite 99% agreement with the test standard, because the homogeneous nature of the sample population affected the cross-tabulation of κ scores (Chapter 5). Similarly, for LCA, the evaluation of Se is driven by the number of affected animals in the study population and the lower Se found for some individual animal indicators, such as pruritus, myiasis and emaciation, was likely to be due to the low level of study sheep affected by these conditions. The assessment of pruritus in the individual animal by means of the 'nibble test' was easy to perform and clearly identified animals with skin irritation, as evidenced by the good level of observer reliability. Therefore, it is recommended to include pruritus as an indicator of sheep welfare (Table 11.1) as study results should be interpreted in the light of proportion data (Chapter 7).

Since many of the indicators were capable of identifying welfare conditions in a population with a low proportion of affected animals (Chapters 4, 5 and 6), the skills of the observer to identify the presence of particular outcomes was tested to a high degree. As on-farm experience suggested that it was easier to identify the presence of a condition than confirm the absence of an outcome, it is suggested that the indicators could perform even better on farms with a greater level of sub-optimal welfare conditions. However, a limitation of the work presented in this thesis is that it is not known whether the indicators do work well when tested under conditions of higher prevalence and further evaluation of all the animal-based indicators on farms experiencing a higher degree of health and welfare issues is recommended.

Another outstanding issue for the application of the animal-based indicators lies in the interpretation of indicator assessments. The expert panel provided cut-off points for 28 welfare outcomes, in terms of the flock prevalence (%) each outcome that was considered to be 'acceptable' and 'unacceptable' for animal welfare (Chapter 10). However there are issues with the cut-off points as there was potential for bias by the composition of the expert panel, and for other indicators, the low numbers of sheep observed with signs of pruritus and myiasis during the longitudinal study may have meant that the cut-off points were set at levels which diagnose almost all sheep farms as having unacceptable levels of these conditions. This could risk losing the confidence of consumers and producers in the sheep farming industry and the measures developed in this thesis. There are also ethical considerations with the use of cut-off points which are interpreted on the basis of flock prevalence. Setting the cut-off points at too high a level means that it is possible for individual sheep to experience very poor standards of welfare but this may not affect the outcome of a flock assessment. For example, the cut-offs may suggest a farm has an 'acceptable' level of lameness, but this does not mean that lameness levels under the acceptable threshold are justified as any lame sheep should be receiving appropriate care and treatment. Therefore, the assessor needs to look for evidence of appropriate action and treatment (Appendix C). Similarly, if the cut-off points identify farms with 'unacceptable' levels of lameness this does not automatically mean that the farmer should not be penalised since farms with good standards of sheep welfare and those with preventive flock health plans in place can still experience unexpected outbreaks of disease. Again, evidence of appropriate action, could be used to discriminate whether the welfare standards of the farm are 'acceptable'. Preliminary cut-off points were only determined for animal-based indicators, and for resource-based measures, the absence of clean water or a lie-back area, could be interpreted as 'unacceptable'. By contrast, the interpretation of on-farm records is less clear and further work evaluating mortality surveillance data and further expert consultations could be used to inform the acceptable and unacceptable threshold levels.

This is the first time that cut-off points of sheep welfare have been produced and further research is needed to test whether they are set at the appropriate level for detecting both good and poor standards of sheep on a large and diverse sample population before being applied in any formal means of animal welfare inspection. This could inform current policies of voluntary assurance schemes and enforcement agencies as well as improving the advice offered to farmers regarding the assessment. As well as identifying farms in which improvements and interventions in welfare are needed, it is important that farmers are rewarded and encouraged to maintain high standards of animal welfare. A positive slant that rewards those who go beyond legal compliance with minimal standards could be a more useful approach to striving for the highest standards of farm animal welfare.

11.4 Potential applications for sheep welfare indicators

Currently, the intended application of the welfare indicators developed in this thesis is not known. However, the indicators developed in this thesis should be viewed as separate measures or 'screening tests' that are employed to diagnose the presence or absence of a particular condition. Whilst the measures cannot be used as an overall indicator of farm compliance with welfare legislation, the indicators can be used to inform further investigations and assessments to identify the reasons behind an on-farm welfare problem. The indicators may be included as part of a one-off or 'spot check' flock assessment, although this approach would not identify welfare problems that are specific to certain periods within the production cycle (Chapter 8). An advantage of repeated sampling was that it revealed that high levels of lameness (> 10 %) were consistently found on certain study farms and it identified farms with welfare issues that only arose at specific points in the production cycle, for example at post-weaning, pre-tupping, or at lambing. Therefore, it is suggested that the indicators are used repeatedly over the course of the production cycle to identify farms at risk of specific welfare problems at key periods of production – pre-tupping, mid-pregnancy, lambing and post-weaning. For assurance and statutory purposes the repeat assessment outcomes could be used to inform the number and type of future assessments as part of a risk-based approach.

There is great potential for the recommended indicators (Table 11.1) to be applied within farm assurance and certification schemes, industry benchmarking, veterinary flock health and welfare planning, routine stockpeople assessments as well as being incorporated into veterinary and agricultural training and education. The measures could also complement welfare assessments of sheep at other stages in the production chain, for example during transportation, at markets and slaughter houses. It is important that the indicators developed and tested in this thesis are transparent and consistently used within different applications and by different assessors. Therefore, the Standard Operating Procedures (SOP's), which describe the assessment of each indicator and scoring system with photographic examples (Appendix C), were produced for recommended indicators.

It is also important to ensure that the indicators are applied by calibrated assessors. On-farm results suggested that experienced and trained observers achieved higher levels of test performance, but being experienced alone did not predict good levels of diagnostic ability (Chapter 5). Although sheep farmers routinely apply measures such as gait assessment in the course of their daily management tasks they may require training to become standardised to the methods of assessment and to highlight certain welfare issues, such as entropion in young lambs. Therefore, it is recommended that experienced and trained assessors should apply these indicators in the future. This would not only help maintain the consistency of cross-farm assessments but would give greater social validity to the indicators and confidence for the end-users of these measures (Sørensen and Fraser, 2010). Reliability exercises could be used to select suitable on-farm assessors and re-calibration exercises and re-testing the reliability of different groups of assessor is also recommended to ensure that trained assessors applying the indicators during cross-farm inspections remain consistent and calibrated to the method of assessment.

11.5 Conclusion

This thesis has contributed to the application of new methodologies and techniques to inform the development and testing of a range of measures that are considered to be sensitive to the current on-farm welfare issues for sheep. On the basis of the studies outlined in this thesis, a set of 15 animal-based indicators for adult and growing lambs, assessed by a combination of group and individual animal assessment, 8 animal-based indicators for young lambs and 12 resource- and management-based indicators have been recommended for use in future on-farm welfare assessments.

Key indicators such as demeanour, lameness and body condition, which can be used on a wide variety of farms and all production stages, could now be incorporated into voluntary farm assurance and certification body assessment schemes, including the BWAP (Main *et al.*, 2003) and WelfareQuality® protocols (Knerium and Winckler, 2009), which do not currently include valid, reliable, feasible and responsive indicators of sheep welfare. Before being used as part of any statutory assessment protocol, the indicators and cut-off points need to be validated on farms with a higher level of sub-optimal welfare standards to check that they remain reliable, robust and responsive on these types of flocks – an aspect of research which clearly requires the full support and engagement of enforcement agencies such as Animal Health and Trading Standards as well as the involvement of key societies and associations within the sheep industry.

Whilst the indicators do not encompass all the potential welfare issues that can arise during the on-farm period of sheep production, such as obstetrical care, or the issues that arise for specific management system, such as dairy flocks, the methodology outlined here could be used to inform the development, testing and evolution of welfare indicators for sheep managed under different systems and geographical areas as well as those for other farm animal species. In this way the measures developed and tested in this thesis can evolve and improve in line with advances in scientific knowledge, our understanding of animal welfare and evidence from on-farm applications and field research findings.

References

Abramoff, M.D., Magelhaes, P.J., Ram, S.J., 2004. Image processing with ImageJ. Biophotonics International 11, 36-42.

Abramson, Z.H., Abramson, J.H., 2008. Research methods in community medicine: surveys, epidemiological research, programme evaluation, clinical trials. John Wiley and Sons Ltd.

Alban, L., Ersb, A.K., Bennedsgaard, T.W., Johnsen, P.F., 2001. Validation of Welfare Assessment Methods at Herd Level: An Example. Acta Agriculturae Scandinavica, Section A - Animal Sciences 51, 99 - 102.

Altman, D.G., 1994. Practical statistics for medical research. Chapman and Hall, London.

Angus, K., 1991. Arthritis in lambs and sheep. In Practice 13, 2-4-207.

Athanasiadou, S., Houdijk, J., Kyriazakis, I., 2008. Exploiting synergisms and interactions in the nutritional approaches to parasite control in sheep production systems. Small Ruminant Research 76, 2-11.

Baldock, N.M., Sibly, R.M., 1990. Effects of handling and transportation on the heart-rate and behavior of sheep. Applied Animal Behaviour Science 28, 15-39.

Bartussek, H., 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. Livestock Production Science 61, 179-192.

Benoit, M., Laignel, G., 2002. Constraints under organic farming on French sheepmeat production: a legal and economic point of view with an emphasis on farming systems and veterinary aspects. Veterinary Research 33, 613-624.

Berriatua, E., French, N.P., Broster, C.E., Morgan, K.L., Wall, R., 2001. Effect of infestation with *Psoroptes ovis* on the nocturnal rubbing and lying behaviour of housed sheep. Applied Animal Behaviour Science 71, 43-55.

Bertrand, P., Benichou, J., Grenier, P., Chastang, C., 2005. Hui and Walter's latent-class reference-free approach may be more useful in assessing agreement than diagnostic performance. Journal of Clinical Epidemiology 58, 688-700.

Binns, S.H., Cox, I.J., Rizvi, S., Green, L.E., 2002. Risk factors for lamb mortality on UK sheep farms. Preventive Veterinary Medicine 52, 287-303.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307-310.

Boissy, A., Aubert, A., Désiré, L., Greiveldinger, L., Delval, E., Veissier, I., 2011. Cognitive sciences to relate ear postures to emotions in sheep. Animal Welfare 20, 47-56.

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., 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. Physiology & Behavior 92, 375.

Boivin, X., Lensink, J., Tallet, C., Veissier, I., 2003. Stockmanship and farm animal welfare. Animal Welfare 12, 479-492.

Boivin, X., Nowak, R., Garcia, A.T., 2001. The presence of the dam affects the efficiency of gentling and feeding on the early establishment of the stockperson-lamb relationship. Applied Animal Behaviour Science 72, 89-103.

Boivin, X., Tournadre, H., Le Neindre, P., 2000. Hand-feeding and gentling influence early-weaned lambs' attachment responses to their stockperson. Journal Of Animal Science 78, 879-884.

Bonde, M., Toft, N., Thomsen, P.T., Sorensen, J.T., 2010. Evaluation of sensitivity and specificity of routine meat inspection of Danish slaughter pigs using Latent Class Analysis. Preventive Veterinary Medicine 94, 165-169.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H.C.E., Grp, S.S., 2003a. Towards complete and,accurate reporting of studies of diagnostic accuracy: the STARD initiative. British Medical Journal 326, 41-44.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Moher, D., Rennie, D., de Vet, H.C.W., Lijmer, J.G., 2003b. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Clinical Chemistry 49, 7-18.

Botreau, R., Veissier, I., Perny, P., 2009. Overall assessment of animal welfare: strategy adopted in Welfare Quality (R). Animal Welfare 18, 363-370.

Bracke, M., Hopster, H., 2006. Assessing the importance of natural behavior for animal welfare. Journal of Agricultural and Environmental Ethics 19, 77.

Bracke, M.B.M., Edwards, S.A., Engel, B., Buist, W.G., Algers, B., 2008. Expert opinion as 'validation' of risk assessment applied to calf welfare. Acta Veterinaria Scandinavica 50, 29.

Bracke, M.B.M., Hulsegge, B., Keeling, L., Blokhuis, H.J., 2004. Decision support system with semantic model to assess the risk of tail biting in pigs 1. Modelling. Applied Animal Behaviour Science 87, 31-44.

Brambell, F.W.R., 1965. Report of the Technical Committee to enquire into the welfare of animals kept under intensive livestock husbandry systems. Her Majesty's Stationery Office (HMSO), London.

Broom, D.M., 1986. Indicators of poor welfare. British Veterinary Journal 142, 524-526.

Broom, D.M., Johnson, K.G., 1993. Stress and animal welfare. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Broster, J.C., Dehaan, R.L., Swain, D.L., Friend, M.A., 2010. Ewe and lamb contact at lambing is influenced by both shelter type and birth number. Animal 4, 796-803.

Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardol, F., Cozzi, G., Van Reenen, C.G., 2009. Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. Italian Journal Of Animal Science 8, 601-603.

Burkholder, W.J., 2000. Use of body condition scores in clinical assessment of the provision of optimal nutrition. Journal Of The American Veterinary Medical Association 217, 650-654.

Burn, C.C., Pritchard, J.C., Whay, H.R., 2009. Observer reliability for working equine welfare assessment: problems with high prevalences of certain results. Animal Welfare 18, 177-187.

Butterworth, A., Knowles, T.G., Whittington, P., Matthews, L., Rogers, A., Bagshaw, C.S., 2007. Validation of broiler chicken gait scoring training in Thailand, Brazil and New Zealand. Animal Welfare 16, 177-179.

Calavas, D., Sulpice, P., Lepetitcolin, E., Bugnard, F., 1998. Appreciation de la fidelite de la pratique d'une methode de notation de l'etat corporel des brebis dans un cadre professionnel. Veterinary Research 29, 129-138.

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

Capdeville, J., Veissier, I., 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Agriculturae Scandinavica Section A-Animal Science, 62-68.

Caroprese, M., 2008. Sheep housing and welfare. Small Ruminant Research 76, 21-25.

Caroprese, M., Albenzio, M., Muscio, A., Sevi, A., 2006. Relationship between welfare and udder health indicators in dairy ewes. Veterinary Research Communications 30, 83-94.

Caroprese, M., Casamassima, D., Rassu, S.P.G., Napolitano, F., Sevi, A., 2009. Monitoring the on-farm welfare of sheep and goats. Italian Journal of Animal Science.

Christley, R.M., Morgan, K.L., Parkin, T.D.H., French, N.P., 2003. Factors related to the risk of neonatal mortality, birth-weight and serum immunoglobulin concentration in lambs in the UK. Preventive Veterinary Medicine 57, 209.

Clements, A.C.A., Mellor, D.J., Fitzpatrick, J.L., 2002. Reporting of sheep lameness conditions to veterinarians in the Scottish Borders. Veterinary Record 150, 815.

Cockram, M., 2005. Welfare issues associated with routine husbandry procedures. In: Proceeding of the Sheep Veterinary Society, Pitlochry, Scotland, pp. 23-25.

Cockram, M.S., 2004. A review of behavioural and physiological responses of sheep to stressors to identify potential behavioural signs of distress. Animal Welfare 13, 283-291.

Cockram, M.S., Imlah, P., Goddard, P.J., Harkiss, G.D., Waran, N.K., 1993. The behavioral, endocrine and leukocyte response of ewes to repeated removal of lambs before the age of natural weaning. Applied Animal Behaviour Science 38, 127-142.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37-46.

Conington, J., Hosie, B., Nieuwhof, G., Bishop, S., Bünger, L., 2008. Breeding for resistance to footrot – the use of hoof lesion scoring to quantify footrot in sheep. Veterinary Research Communications 32, 583-589.

Coop, R.L., Sykes, A.R., Angus, K.W., 1982. The effect of three levels of intake of *Ostertagia-Circumcincta* larvae on growth-rate, food-intake and body-composition of growing lambs. Journal of Agricultural Science 98, 247-255.

Corke, M.J., Broom, D.M., 1999. The behaviour of sheep with sheep scab, *Psoroptes ovis* infestation. Veterinary Parasitology 83, 291-300.

Cronbach, L.J., 1947. Test "reliability": its meaning and determination. Psychometrika 12, 1-16.

Cronbach, L.J., Gleser, G.C., Harinder Nanda, A.N., Rajaratnam, N., 1972. The dependability of behavioural measurements: theory of generalizability for scores and profiles. John Wiley & Sons, Ltd.

Cronin, G.M., Barnett, J.L., Edge, M.K., Hemsworth, P.H., 2002. Identifying animal welfare issues for sheep in Australia. Wool Technology And Sheep Breeding 50, 534-540.

Cross, P., Edwards-Jones, G., Omed, H., Williams, A.P., 2010. Use of a Randomized Response Technique to obtain sensitive information on animal disease prevalence. Preventive Veterinary Medicine 96, 252-262

D'Angelo, A., Maurella, C., Bona, C., Borrelli, A., Caramelli, M., Careddu, M.E., Jaggy, A., Ru, G., 2007. Assessment of clinical criteria to diagnose scrapie in Italy. Veterinary Journal 174, 106-112.

D'Eath, R.B., Conington, J., Lawrence, A.B., Olsson, I.A.S., Sandøe, P., 2010. Breeding for behavioural change in farm animals: practical, economic and ethical considerations. Animal Welfare 19, 17-27.

Dawkins, M.S., 1990. From an animals point of view - motivation, fitness, and animal-welfare. Behavioral and Brain Sciences 13, 1-&.

de Passillé, A.M., Rushen, J., 2005. Can we measure human-animal interactions in on-farm animal welfare assessment? Some unresolved issues. Applied Animal Behaviour Science 92, 193-209.

Defra, 2003. Code of recommendations for the welfare of livestock - sheep. Department for the Environment, Food and Rural Affairs (Defra) publication, London.

Delbecq, A.L., Van de Ven, A.H., Gustafson, D.H. (Eds.), 1975. Group techniques for program planning. Scott, Foresman and Company, Illinois.

Dobson, H., 1990. The assessment of stress in sheep. In: Proceedings of the Sheep Veterinary Society, pp. 46-47.

Dolan, S., Field, L.C., Nolan, A.M., 2000. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation. Pain 86, 311-320.

Done-Currie, J.R., Hecker, J.F., Wodzicka-Tomaszewska, M., 1984. Behavior of Sheep Transferred from Pasture to an Animal House. Applied Animal Behaviour Science 12, 121-130.

Doney, J.M., Smith, W.F., 1961. The fleece of the Scottish Blackface sheep I. Seasonal changes in wool production and fleece structure. The Journal of Agricultural Science 56, 365-374.

Ducker, M.J., Fraser, J., 1973. Effect of Level of husbandry at lambing on lamb viability and subsequent performance. Animal Production 16, 91-94.

Duncan, I.J.H., 1996. Animal welfare defined in terms of feelings. Acta Agriculturae Scandinavica Section A-Animal Science, 29-35.

Dwyer, C.M., 2004. How has the risk of predation shaped the behavioural responses of sheep to fear and distress? Animal Welfare 13, 269-281.

Dwyer, C.M., 2007. Genetic and physiological determinants of maternal behaviour and lamb survival: implications for low input sheep management. Journal of Animal Science 86, 246-258.

Dwyer, C.M., 2008a. Environment and the sheep. Breed adaptations and welfare implications, In: The welfare of sheep. First Ed. Springer, pp. 41-79

Dwyer, C.M., 2008b. The welfare of the neonatal lamb. Small Ruminant Research 76, 31-41.

Dwyer, C.M., 2009. Welfare of sheep: providing for welfare in an extensive environment. Small Ruminant Research 86, 14-21.

Dwyer, C.M., Bornett, H.L.I., 2004. Chronic stress in sheep: assessment tools and their use in different management conditions. Animal Welfare 13, 293-304.

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., Brown, H.E., Simm, G., 1996. Effect of ewe and lamb genotype on gestation length, lambing ease and neonatal behaviour of lambs. Reproduction Fertility and Development 8, 1123-1129.

Edwards, D.S., Johnston, A.M., Pfeiffer, D.U., 2001. A comparison of commonly used ear tags on the ear damage of sheep. Animal Welfare 10, 141-151.

Edwards, G.T., 2005. Veterinary Laboratories Agency welfare surveillance. In: Proceedings of the Sheep Veterinary Society, Pitlochry, Scotland, pp. 27-20.

EFSA, 2007. Basic information for the development of the animal welfare risk assessment guidelines. EFSA (European Food Standards Agency) publication. Palma, Italy. pp. 1-36. http://www.efsa.europa.eu/en/scdocs/doc/130ar.pdf (accessed 1 November 2007).

Egerton, J.R.R., D. S., 1971. Vaccination against ovine footrot. Journal of Comparative Pathology 81, 179-185.

El Balaa, R., Marie, M., 2006. Animal welfare considerations in small ruminant breeding specifications. Journal of Agricultural & Environmental Ethics 19, 91-102.

Evans, D.G., 1978. Interpretation and analysis of subjective body condition scores. Animal Production 26, 119-125.

Evans, S., Scott, P.R., 1999. The future for veterinary services on sheep farms. In: Proceedings of the Sheep Veterinary Society, Dumfries, Scotland, pp. 129-133.

Everitt, G.C., 1962. On the Assessment of body compostition in live sheep and cattle. In: Proceedings of the Australian Society of Animal Production, pp. 79-89.

Farkas, R., Hall, M.J.R., Kelemen, F., 1997. Wound myiasis of sheep in Hungary. Veterinary Parasitology 69, 133-144.

FAWC, 1994. Report on the welfare of sheep. PB 1755. Farm Animal Welfare Council (FAWC) publication, London.

FAWC, 2008. Report on the implications of castration and tail docking for the welfare of lambs. Farm Animal Welfare Council (FAWC) publication, London. http://www.fawc.org.uk/pdf/report-080630.pdf (accessed 28 January 2009)

FAWC, 2009. Report on Farm Animal Welfare in Great Britain: Past, Present and Future. Farm Animal Welfare Council (FAWC) publication, London. http://www.fawc.org.uk/pdf/ppf-report091012.pdf (accessed 19 November 2010).

Feinstein, A.R., Cicchetti, D.V., 1990. High agreement but low kappa .1. The problems of 2 paradoxes. Journal of Clinical Epidemiology 43, 543-549.

Fernie, A.C., Tribe, A., Murray, P.J., Phillips, C.J.C., 2008. The creation and implementation of a Great Ape Welfare Index. In: Proceedings of the International Primatological Society XXII Congress, Edinburgh, Scotland, pp. 261.

Fink, A., Kosecoff, J., Chassin, M., Brook, R.H., 1984. Consensus methods - characteristics and guidelines for use. American Journal of Public Health 74., 979-983.

Fisher, M.W., 2004. A review of the welfare implications of out-of-season extensive lamb production systems in New Zealand. Livestock Production Science 85, 165-172.

Fisher, M.W., Mellor, D.J., 2002. The welfare implications of shepherding during lambing in extensive New Zealand farming systems. Animal Welfare 11, 157-170.

Fisher, M.W., Scobie, D.R., 2003. How do sheep rank different welfare compromises? In: Proceedings of workshop on improving sheep welfare on extensively managed flocks: economics, husbandry and welfare. Scottish Agricultural College, Aberdeen, Scotland, pp. 23-26.

Fitzpatrick, J., Scott, M., Nolan, A., 2006. Assessment of pain and welfare in sheep. Small Ruminant Research 62, 55.

Fitzpatrick, R., Boulton, M., 1994. Qualitative methods for assessing health care. Quality in Health Care 3, 107 - 113.

Fleiss, L.L., 1981. Statistical methods for rates and proportions. John Wiley and Sons Ltd.

Flower, F.C., Weary, D.M., 2009. Gait assessment in dairy cattle. Animal 3, 87-95.

Forkman, B., Boissy, A., Meunier-Salaun, M.C., Canali, E., Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiology & Behavior 92, 340.

Fox, W.M., 1989. The improved Nominal Group Technique. Journal of Management Development 8., 20 - 27.

Fraser, D., Weary, D.M., Pajor, E.A., Milligan, B.N., 1997. A scientific conception of animal welfare that reflects ethical concerns. Animal Welfare 6, 187-205.

French, N., Wall, R., Cripps, P.J., Morgan, K.L., 1994a. Blowfly strike in England and Wales - the relationship between prevalence and farm and management factors. Medical and veterinary entomology 8, 51-56.

French, N.P., Morgan, K.L., 1992. Neuromas in docked lambs tails. Research in Veterinary Science 52, 389-390.

French, N.P., Morgan, K.L., 1996. A model of ovine cutaneous myosis using the predicted abundance of *Lucilia sericata* and a pattern of sheep susceptibility. Preventive Veterinary Medicine 26, 143-155.

French, N.P., Parkin, T.D., Morgan, K.L., 1996. A case control study of blowfly strike in lambs. Veterinary Record 139, 384-388.

French, N.P., Wall, R., Morgan, K.L., 1994b. Lamb tail docking: a controlled field study of the effects of tail amputation on health and productivity. Veterinary Record 134, 463-467.

FSA, 2007. Clean sheep for slaughter. A guide for producers. Food Standards Agency (FSA) publication, London.

Fthenakis, G.C., 2000. Field evaluation of flunixin meglumine in the supportive treatment of ovine mastitis. Journal of Veterinary Pharmacology and Therapeutics 23, 405-407.

Garabed, R.B., Perez, A.M., Johnson, W.O., Thurmond, M.C., 2009. Use of expert opinion for animal disease decisions: An example of foot-and-mouth disease status designation. Preventive Veterinary Medicine 92, 20-30.

Girden, E.R., 1992. Repeated measures ANOVA. Sage Press, California.

Glaser, E.M., 1980. Using behavioural science strategies for defining the state-of-the-art. Journal of Applied and Behavioural Science 16, 79-92.

Goddard, P., 2008. The management of sheep. In: The welfare of sheep. First Ed. Springer, pp. 291-323.

Goddard, P., Waterhouse, T., Dwyer, C., Stott, A., 2006. The perception of the welfare of sheep in extensive systems. Small Ruminant Research 62, 215-225.

Goodall, J., 1990. Through a window: thirty years with the chimpanzees of Gombe. Phoenix Giant, London.

Gordon, I.J., Illius, A.W., Milne, J.D., 1996. Sources of variation in the foraging efficiency of grazing ruminants. Functional Ecology 10, 219-226.

Gougoulis, D.A., Kyriazakis, I., Fthenakis, G.C., 2010. Diagnostic significance of behaviour changes of sheep: A selected review. Small Ruminant Research 92, 52-56.

Graham, M.J., Kent, J.E., Molony, V., 1997. Effects of four analgesic treatments on the behavioural and cortisol responses of 3-week-old lambs to tail docking. The Veterinary Journal 153, 87-97.

Gray, D., 2002. Can parasitic gastro-enteritis be used as an indicator of welfare in organic sheep? In: UK Organic Research Aberystwyth, pp. 347-350.

Green, L.E., Morgan, K.L., 1994. Risk factors associated with postpartum deaths in early born, housed lambs in southwest England. Preventive Veterinary Medicine 21, 19-27.

Greiner, M., Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. Preventive Veterinary Medicine 45, 3-22.

Greiveldinger, L., Veissier, I., Boissy, A., 2007. Emotional experience in sheep: Predictability of a sudden event lowers subsequent emotional responses. Physiology & Behavior 92, 675-683.

Grogono-Thomas, R., Cook, A. J. & Johnston, A. M, 1998. Lame Excuses? In: Proceeding of the Sheep Veterinary Society, Scarborough, England, pp. 77-82.

Hall, M., Wall, R., 1995. Myiasis of humans and domestic animals. Advances in Parasitology 35, 257-334.

Hargreaves, A.L., Hutson, G.D., 1990. Some effects of repeated handling on stress responses in sheep. Applied Animal Behaviour Science 26, 253-265.

Harkins, L.S., 2005. Development of a prototype welfare tool for use in sheep. Thesis, Master of Science, University of Glasgow.

Harrison, R., 1964. Animal machines. Vincent Stuart Ltd, London.

Hemsworth, P.H., 2007. Ethical stockmanship. Australian Veterinary Journal 85, 194-200.

Herva, T., Peltoniemi, O., Virtala, A.-M., 2009. Validation of an Animal Needs Index for cattle using test theory. Animal Welfare 18, 417-425.

Hewetson, M., Christley, R.M., Hunt, I.D., Voute, L.C., 2006. Investigations of the reliability of observational gait analysis for the assessment of lameness in horses. Veterinary Record 158, 852-858.

Hoehler, F.K., 2000. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. Journal of Clinical Epidemiology 53, 499-503.

Hosie, B.D., Carruthers, J., Sheppard, B.W., 1996. Bloodless castration of lambs: results of a questionnaire. British Veterinary Journal 152, 47-55.

Hovi, M., Sundrum, A., Thamsborg, S.M., 2003. Animal health and welfare in organic livestock production in Europe: current state and future challenges. Livestock Production Science 80, 41-53.

Hsu, C., Sanford, B.A., 2007 The Delphi technique: making sense of consensus. Practical assessment, research and evaluation, 12. http://www.joe.org/joe/2010february/rb2.php (accessed 12 April 2010).

Hughes, B.O., 1976. Behaviour as index of welfare. In: Proceedings of the 5th European Poultry Conference, Malta, pp. 1005-1012.

Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36, 167-171.

Hurnik, J.F., 1988. Welfare of farm animals. Applied Animal Behaviour Science 20, 105-117.

Huxley, J.N., Burke, J., Roderick, S., Main, D.C.J., Whay, H.R., 2004. Animal welfare assessment benchmarking as a tool for health and welfare planning in organic dairy herds. Veterinary Record 155, 237-239.

Jackson, F., Coop, R.L., 2000. The development of anthelmintic resistance in sheep nematodes. Parasitology 120 S95-107.

Jefferies, B.C., 1961. Body condition scoring and its use in management. Tasmanian Journal of Agriculture 32, 19-21.

Jerram, P.J.S., 1987. Animal welfare considerations of laparoscopic A.I. - a reply. New Zealand Veterinary Journal 35, 157.

Jolliffe, I.T., 2002. Principal component analysis. Springer, New York.

Kaler, J., Green, L.E., 2008a. Naming and recognition of six foot lesions of sheep using written and pictorial information: a study of 809 English sheep farmers. Preventive Veterinary Medicine 83, 52-64.

Kaler, J., Green, L.E., 2008b. Recognition of lameness and decisions to catch for inspection among sheep farmers and specialists in GB. BMC Veterinary Research 4, 41-.

Kaler, J., Green, L.E., 2009a. Farmers' practices and factors associated with the prevalence of all lameness and lameness attributed to interdigital dermatitis and footrot in sheep flocks in England in 2004. Preventive Veterinary Medicine 92, 52-59.

Kaler, J., Wassink, G.J., Green, L.E., 2009b. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. The Veterinary Journal 180, 189-194.

Kaplan, R.M., Burke, J.M., Terrill, T.H., Miller, J.E., Getz, W.R., Mobini, S., Valencia, E., Williams, M.J., Williamson, L.H., Larsen, M., Vatta, A.F., 2004. Validation of the FAMACHA eye color chart for detecting clinical anemia in sheep and goats on farms in the southern United States. Veterinary Parasitology 123, 105-120.

Kendall, M.G., Smith, B.B., 1939. The problem of m rankings. The Annals of Mathematical Statistics 10, 275-287.

Kent, J.E., Molony, V., Graham, M.J., 1998. Comparison of methods for the reduction of acute pain produced by rubber ring castration or tail docking of week-old lambs. Veterinary Journal 155, 39-51.

Kent, J.E., Thrusfield, M.V., Molony, V., Hosie, B.D., Sheppard, B.W., 2004. Randomised, controlled field trial of two new techniques for the castration and tail docking of lambs less than two days of age. Veterinary Record 154, 193.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Animal Welfare 6, 243-254.

Kirkwood, B.R., Sterne, J.A.C., 2003. Essential medical statistics. Blackwell Science, Oxford.

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. Animal Welfare 18, 451-458.

Kristensen, E., Dueholm, L., Vink, D., Andersen, J.E., Jakobsen, E.B., Illum-Nielsen, S., Petersen, F.A., Enevoldsen, C., 2006. Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. Journal of Dairy Science 89, 3721-3728.

Krueger, R.A., Casey, M.A., 2009. Focus groups: a practical guide for applied research. SAGE Publishing, Los Angeles.

Kynn, M., 2008. The 'heuristics and biases' bias in expert elicitation. Journal of the Royal Statistical Society Series A- Statistics in Society. 171, 239-264.

Le Neindre, P., Boivin, X., Boissy, A., 1996. Handling of extensively kept animals. Applied Animal Behaviour Science 49, 73-81.

Leeb, C., Troxler, J., Schuh, M., 2001. Skin lesions and callosities in group-housed pregnant sows: animal-related welfare indicators. Acta Agriculturae Scandinavica, Section A - Animal Sciences 51, 82 - 87.

Lees, A.D., Milne, A., 1951. The seasonal and diurnal activities of individual sheep ticks (*Ixodes-Ricinus*). Parasitology 41, 189-208.

Ley, S.J., Waterman, A.E., Livingston, A., 1995. A field study of the effect of lameness on mechanical nociceptive thresholds in sheep. Veterinary Record 137, 85-87.

Long, J.S., Freese, L., 2006. Regression models for categorical dependent variables Using Stata. Stata Press Publications, Texas.

Lovatt, F.M., 2005. Investigation of Pasteurellosis outbreaks-the reality of getting vets on farms and the consequent welfare implications. In: Proceedings of the Sheep Veterinary Society Pitlochry, Scotland, pp. 37-40.

Lovatt, F.M., 2010. Clinical examination of sheep. Small Ruminant Research 92, 72-77.

Lucas, N.P., Macaskill, P., Irwig, L., Bogduk, N., 2010. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). Journal of Clinical Epidemiology 63, 854-861.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: evolution, critique and future directions. Statistics in Medicine 28, 3049-3067.

Lynch, J.J., Hinch, G.N., Adams, D.B., 1992. The behaviour of sheep: biological principles and implications for production. CAB International, Wallingford.

Macfarlane, J.M., Matheson, S.M., Dwyer, C.M., 2010. Genetic parameters for birth difficulty, lamb vigour and lamb sucking ability in Suffolk sheep. Animal Welfare 19, 99-105.

Macleod, J., 1992. Blowfly species composition in sheep myiasis in Britain. Medical and veterinary entomology 6, 177-178.

Macnab, A., 1998a. Welfare and productivity of hill sheep farming in the Highlands & Islands: Part One. Government Veterinary Journal 8, 11-14.

Macnab, A., 1998b. Welfare and productivity of hill sheep farming in the Highlands & Islands: Part Two Government Veterinary Journal 8, 15-17.

Main, D.C.J., Webster, F., Green, L.E., 2001. Animal welfare assessment in farm assurance schemes. Acta Agriculturae Scandinavica Section A-Animal Science, 108-113.

Main, D.C.J., Whay, H.R., Green, L.E., Webster, A.J.F., 2003. Preliminary investigation into the use of expert opinion to compare the overall welfare of dairy cattle farms in different farm assurance schemes. Animal Welfare 12, 565-569.

Main, D.C.J., Whay, H.R., Lee, C., Webster, A.J.F., 2007. Formal animal-based welfare assessment in UK certification schemes. Animal Welfare 16, 233-236.

Martin, P.R., Bateson, P.P.G., 2007. Measuring behaviour: an introductory guide. Third Ed. Cambridge University Press, Cambridge.

Mason, G.J., Latham, N.R., 2004. Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? Animal Welfare 13, S57-S69.

May, R., van Dijk, J., Forland, J.M., Andersen, R., Landa, A., 2008. Behavioural patterns in ewe-lamb pairs and vulnerability to predation by wolverines. Applied Animal Behaviour Science 112, 58-67.

McCoy, M.A., Edgar, H.W., Kenny, J., Gordon, A.W., Dawson, L.E., Carson, A.F., 2005. Evaluation of on-farm faecal worm egg counting in sheep. Veterinary Record 156, 21-23.

McDermott, J.J., Schukken, Y.H., Shoukri, M.M., 1994. Study design and analytic methods for data collected from clusters of animals. Preventive Veterinary Medicine 18, 175-191.

McGreevy, P.D., George, S., Thomson, P.C., 2007a. A note on the effect of changes in flooring on the behaviour of housed rams. Applied Animal Behaviour Science 107, 355-360.

Melches, S., Mellema, S.C., Doherr, M.G., Wechsler, B., Steiner, A., 2007. Castration of lambs: A welfare comparison of different castration techniques in lambs over 10 weeks of age. The Veterinary Journal 173, 554-563.

Mellor, D., Stafford, K., 1999. Assessing and minimising the distress caused by painful husbandry procedures in ruminants. In Practice 21, 436-446.

Mellor, D.J., Stafford, K.J., 2003. Assessing the welfare status of newborn farm animals. Animal Welfare 12, 695-698.

Mellor, D.J., Stafford, K.J., 2004. Animal welfare implications of neonatal mortality and morbidity in farm animals. The Veterinary Journal 168, 118-133.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. Applied Animal Behaviour Science 31, 136-146.

Milligan, K.E., Broadbent, J.S., 1974. Condition scoring of sheep. In: Proceedings of the New Zealand Society of Animal Production, Dunedin, New Zealand, p. 114.

Mollenhorst, H., Rodenburg, T.B., Bokkers, E.A.M., Koene, P., de Boer, I.J.M., 2005. On-farm assessment of laying hen welfare: a comparison of one environment-based and two animal-based methods. Applied Animal Behaviour Science 90, 277-291.

Molony, V., Kent, J.E., 2007. Sheep welfare: castration and tail docking. In: Diseases of Sheep. Fourth Ed. Blackwell Publishing, Oxford, pp. 27-31.

Molony, V., Kent, J.E., McKendrick, I.J., 2002. Validation of a method for assessment of an acute pain in lambs. Applied Animal Behaviour Science 76, 215.

Molony, V.K., J. E., 1997. Assessment of acute pain in farm animals using behavioural and physiological measurements. Journal of Animal Science 75, 266-272.

Morgan-Davies, C., Waterhouse, A., Milne, C.E., Stott, A.W., 2006. Farmers' opinions on welfare, health and production practices in extensive hill sheep flocks in Great Britain. Livestock Science 104, 268-277.

Morgan-Davies, C., Waterhouse, A., Pollock, M., L. and Milner, J. M., 2008. Body condition score as an indicator of ewe survival under extensive conditions. Animal Welfare 17, 71-77.

Mork, T., Waage, S., Tollersrud, T., Kvitle, B., Sviland, S., 2007. Clinical mastitis in ewes: bacteriology, epidemiology and clinical features. Acta Veterinaria Scandinavica 49, 23.

Morton, D.B., Griffiths, P.H.M., 1985. Guidelines on the Recognition of Pain, Distress and Discomfort in Experimental-Animals and an Hypothesis for Assessment. Veterinary Record 116, 431-436.

Mullan, S., Edwards, S.A., Butterworth, A., Whay, H.R., Main, D.C.J., 2009. Interdependence of welfare outcome measures and potential confounding factors on finishing pig farms. Applied Animal Behaviour Science 121, 25-31.

Muñoz, C., Carson, A. F., McCoy, M. A., Dawson, L. E. R., O'Connell, N. E. and Gordon, A. W., 2008. Nutritional status of adult ewes during early and mid-pregnancy. 1. Effects of plane of nutriton on ewe reproduction and offspring performance to weaning. animal 2, 52-63.

Munsterhjelm, C., Valros, A., Heinonen, M., Halli, O., Peltoniemi, O.A.T., 2006. Welfare index and reproductive performance in the sow. Reproduction in Domestic Animals 41, 494-500.

Napolitano, F., Cifuni, G.F., Pacelli, C., Riviezzi, A.M., Girolami, A., 2002. Effect of artificial rearing on lamb welfare and meat quality. Meat Science 60, 307-315.

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 Ruminant Research 83, 49-57.

Napolitano, F., Grasso, F., Bordi, A., Tripaldi, C., Saltalamacchia, F., Pacelli, C., De Rosa, G., 2005. On-farm welfare assessment in dairy cattle and buffaloes: evaluation of some animal-based parameters. Italian Journal Of Animal Science 4, 223-231.

Nielsen, L.R., Toft, N., Ersboll, A.K., 2004. Evaluation of an indirect serum ELISA and a bacteriological faecal culture test for diagnosis of Salmonella serotype Dublin in cattle using latent class models. Journal of Applied Microbiology 96, 311-319.

Nieuwhof, G.J., Conington, J., Bünger, L., Haresign, W., Bishop, S.C., 2008. Genetic and phenotypic aspects of foot lesion scores in sheep of different breeds and ages. animal 2, 1289-1296.

NIH, 1990. Consensus development at the NIH: improving the programme. Report of a study by a committee of the Institute of Medicine council on health care technology. National Academy Press, Washington D.C. Retrieved 4 January 2008, from http://www.nap.edu/catalog.php?record_id51563

Nowak, R., 1996. Neonatal survival: contributions from behavioural studies in sheep. Applied Animal Behaviour Science 49, 61.

Nunnally, J.C., .Bernstein., I. H., 1994. Psychometric theory. McGraw-Hill Inc, New York.

O'Connor, C.E., Jay, N.P., Nicol, A.M., Beatson, P.R., 1985. Ewe maternal behaviour score and lamb survival. In: Proceedings of the New Zealand Society of Animal Production 45, pp. 159-162.

Orgeur, P., Mavric, N., 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. Applied Animal Behaviour Science 58, 87.

Peddie, S., Goddard, P., Stott, A., 2003. Footrot control and sheep welfare: applying adaptive conjoint analysis (ACA) to elicit stakeholder opinion. In: Goddard, P. (Ed.), Improving sheep welfare on extensively managed flocks: economics, husbandry and welfare. Scottish Agricultural College, City, pp. 57-64.

Petersen, H.H., Enoe, C., Nielsen, E.O., 2004. Observer agreement on pen level prevalence of clinical signs in finishing pigs. Preventive Veterinary Medicine 64, 147-156.

Phillips, C.J.C., Phillips, A.P., 2010. Attitudes of Australian sheep farmers to animal welfare. Journal of International Farm Management 5, 1-26.

Phillips, C.J.C., Wojciechowska, J., Meng, J., Cross, N., 2009. Perceptions of the importance of different welfare issues in livestock production. Animal 3, 1152-1166.

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. Animal Welfare 16, 489.

Pitmann, E.J.G., 1939. A note on normal correlation. Biometrika 31, 9-12.

Plant, J.W., 2006. Sheep ectoparasite control and animal welfare. Small Ruminant Research 62, 109-112.

Pollott, G.E., Karlsson, L.J.E., Eady, S., Greeff, J.C., 2004. Genetic parameters for indicators of host resistance to parasites from weaning to hogget age in Merino sheep. Journal of Animal Science. 82, 2852-2864.

Pollott, G.E., Stone, D.G., 2006. The breeding structure of the British sheep industry 2003. Defra publications, London.

Prettejohn, M.W.H., 1990. Sheep welfare - the ministry view. In: Proceedings of the Sheep Veterinary Society, pp.

Pritchard, D., 2008. The role of government in imporiving the welfare of sheep - what will be different in 2012. In: Proceedings of a workshop on the welfare goals from the persepective of extensively managed sheep, Scottish Agricultural College, Aberdeen, Scotland, pp. 79-88

Pritchard, J.C., Barr, A.R.S., Whay, H.R., 2007. Repeatability of a skin tent test for dehydration in working horses and donkeys. Animal Welfare 16, 181-183.

Reefmann, N., Kaszas, F.B., Wechsler, B., Gygax, L., 2009. Ear and tail postures as indicators of emotional valence in sheep. Applied Animal Behaviour Science 118, 199-207.

Ridler, A.L., West, D.M., 2010. Examination of teeth in sheep health management. Small Ruminant Research 92, 92-95.

Roger, P.A., 2008. The impact of disease and disease prevention on sheep welfare. Small Ruminant Research 76, 104-111.

Rollin, B.E., 1991. Animal welfare, science, and value. Journal of Agricultural and Environmental Ethics 6, Supplement 2, 44-50

Rothwell, P.M., 2006. Factors that can affect the external validity of randomised controlled trials. PLoS Clinical Trials 1, 9.

Roughan, J.V., Flecknell, P.A., 2003. Pain assessment and control in laboratory animals. Laboratory Animals 37, 172-172.

Rousing, T., Jakobsen, I.A., Hindhede, J., Klaas, I.C., Bonde, M., rensen, J.T., 2007. Evaluation of a welfare indicator protocol for assessing animal welfare in AMS herds: researcher, production advisor and veterinary practitioner opinion. Animal Welfare 16, 213.

Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Applied Animal Behaviour Science 101, 40-53.

Russel, A., 1984. Body condition scoring of sheep. In Practice 6, 91-93.

Russel, A.J.F., Doney, J. M., Gunn, R. G., 1969. Subjective assessment of body fat in live sheep. Journal of Agricultural Science 72, 451-454.

Rutherford, K.M.D., 2002. Assessing pain in animals. Animal Welfare 11, 31-53.

Sanson, D.W., West, T.R., Tatman, W. R., Riley, M. L., Judkins, M. B. and Moss, G. E., 1993. Relationship of body composition of mature ewes with condition score and body weight. Journal Of Animal Science 71, 1112-1116.

Sargison, N., 2000. Welfare issues associated with 'easy care'. In: Proceedings of the Sheep Veterinary Society meeting, pp. 35-40.

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

Scott, E., Black, N., 1991. When does consensus exist in expert Panels? Journal of Public Health Medicine 13, 344-344.

Scott, E.M., Nolan, A.M., Fitzpatrick, J.L., 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. Acta Agriculturae Scandinavica, Section A - Animal Sciences 51, 5 - 10.

Scott, P.R., 2003. A questionnaire survey of ovine dystocia management in the United Kingdom. Animal Welfare 12, 119.

Scott, P.R., 2005a. The management and welfare of some common ovine obstetrical problems in the United Kingdom. The Veterinary Journal 170, 33-40.

Scott, P.R., 2005b. The potential for improving disease control and welfare standards in UK flocks using veterinary flock health plans. In: Proceedings of the Sheep Veterinary Society meeting, pp. 19-22.

Scott, P.R., Sargison, N.D., Penny, C.D., Strachan, W.D., 1995. A questionnaire survey of the correction methods for vaginal prolapse in ewes. Animal Welfare 4, 251.

Scott, P.R., Sargison, N.D., Wilson, D.J., 2007. The potential for improving welfare standards and productivity in United Kingdom sheep flocks using veterinary flock health plans. The Veterinary Journal 173, 522-531.

Sevi, A., 2007. Ewe welfare and ovine milk and cheese quality. Italian Journal Of Animal Science 6, Supplement 1, 521-526.

Sevi, A., Napolitano, F., Casamassima, D., Dell'Aquila, S., 2001a. Effect of milk source on welfare and weight gain of lambs. Animal Welfare 10, 163.

Sevi, A., Taibi, L., Albenzio, M., Annicchiarico, G., Muscio, A., 2001b. Airspace effects on the yield and quality of ewe milk. Journal of Dairy Science 84, 2632-2640.

Shands, C.G., McLeod, B., Lollback, M.L., Duddy, G., Hatcher, S., Halloran, W.J., 2009. Comparison of manual assessments of ewe fat reserves for on-farm use. Animal Production Science 49, 630-636

Silanikove, N., 2000. Effects of heat stress on the welfare of extensively managed domestic ruminants. Livestock Production Science 67, 1.

Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical Therapy 85, 257-268.

Smidt, D., 1983. Indicators relevant to farm animal welfare. In: Proceedings of the EEC programme of co-ordination of research on animal welfare. Martinus Nijoff, Boston.

Smith, R.F., Dobson, H., 2002. Hormonal interactions within the hypothalamus and pituitary with respect to stress and reproduction in sheep. Domestic Animal Endocrinology 23, 75-85.

Sørensen, J.T., Fraser, D., 2010. On-farm welfare assessment for regulatory purposes: Issues and possible solutions. Livestock Science 131, 1-7.

Sørensen, J.T., Rousing, T., Moller, S.H., Bonde, M., Hegelund, L., 2007. On-farm welfare assessment systems: what are the recording costs? Animal Welfare 16, 237-239.

Spoolder, H., De Rosa, G., Horning, B., Waiblinger, S., Wemelsfelder, F., 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529-534.

Stafford, K.J., 1995. Electroejaculation: a welfare issue. Surveillance 22, 15-17.

Stafford, K.J., Spoorenberg, J., West, D.M., Vermunt, J.J., Petrie, N., Lawoko, C.R.O., 1996. The effect of electro-ejaculation on aversive behaviour and plasma cortisol concentration in rams. New Zealand Veterinary Journal 44, 95.

Stewart, K.E.J., Bourn, N.A.D., Thomas. J.A. 2001. An evaluation of three quick methods commonly used to assess sward height in ecology. Journal of Applied Ecology 38, 1148–1154.

Stolwijk, A.M., Straatman, H., Zielhuis, G.A., 1999. Studying seasonality by using sine and cosine functions in regression analysis. Journal of Epidemiology and Community Health 53, 235-238.

Stubbings, L.A., 2008. Ewe management for reproduction. In: Diseases of sheep. Fourth Ed. Blackwell Science, Oxford, pp. 53-60.

Stubsjøen, S.M., Bohlin, J., Skjerve, E., Valle, P.S., Zanella, A.J., 2010. Applying fractal analysis to heart rate time series of sheep experiencing pain. Physiology and Behavior 101.

Stubsjøen, S.M., Flø, A.S., Moe, R.O., Janczak, A.M., Skjerve, E., Valle, P.S., Zanella, A.J., 2009. Exploring non-invasive methods to assess pain in sheep. Physiology & Behavior 98, 640-648.

Tadich, N.A., Carey, A., Porter, R., Ridley, J., Green, M.J. 1998. Case control study of risk factors for toxic mastitis in 26 dairy herds. Veterinary Record 143, 362-365.

Taylor, M.A., Coop, R.L., Wall, R.L., 2007. Veterinary Parasitology. Blackwell publishing, Oxford.

Thornton, P.D., Waterman-Pearson, A.E., 1999. Quantification of the pain and distress responses to castration in young lambs. Research in Veterinary Science 66, 107-118.

Toft, N., Åkerstedt, J., Tharaldsen, J., Hopp, P., 2007a. Evaluation of three serological tests for diagnosis of Maedi-Visna virus infection using latent class analysis. Veterinary Microbiology 120, 77-86.

Toft, N., Innocent, G.T., Gettinby, G., Reid, S.W.J., 2007b. Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. Preventive Veterinary Medicine 79, 244-256.

Tribe, D.E., Seebeck, R.M., 1962. Effect of live-weight and live-weight change on the lambing performance of ewes. The Journal of Agricultural Science 59, 105-110.

Turner, S.P., Dwyer, C.M., 2007. Welfare assessment in extensive animal production systems: challenges and opportunities. Animal Welfare 16, 189.

Tuyttens, F.A.M., Sprenger, M., Van Nuffel, A., Maertens, W., Van Dongen, S., 2009. Reliability of categorical versus continuous scoring of welfare indicators: lameness in cows as a case study. Animal Welfare 18, 399-405.

van den Broek, A.H., Huntley, J.F., 2003. Sheep scab: the disease, pathogenesis and control. Journal of Comparative Pathology 128, 79-91.

Vang, J., 1986. The consensus development conference and the European experience. International Journal of Technology Assessment in Health Care 2, 65-76.

Veissier, I., Boissy, A., Desire, L., Greiveldinger, L., 2009. Animals' emotions: studies in sheep using appraisal theories. Animal Welfare 18, 347-354.

Villalba, J.J., Provenza, F.D., Manteca, X., 2010. Links between ruminants' food preference and their welfare. Animal 4, 1240-1247.

von Borell, E., Langbein, J., Despres, G., Hansen, S., Leterrier, C., Marchant-Forde, J., Marchant-Forde, R., Minero, M., Mohr, E., Prunier, A., Valance, D., Veissier, I., 2007. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - a review. Physiology and Behavior 92, 293.

von Keyserlingk, M.A.G., Rushen, J., de Passillé, A.M., Weary, D.M., 2009. Invited review: The welfare of dairy cattle--Key concepts and the role of science. Journal of Dairy Science 92, 4101-4111.

Waiblinger, S., Knierim, U., Winckler, C., 2001. The development of an epidemiologically based on-farm welfare assessment system for use with dairy cows. Acta Agriculturae Scandinavica Section A-Animal Science, 73-77.

Waterhouse, A., 1996. Animal welfare and sustainability of production under extensive conditions - A European perspective. Applied Animal Behaviour Science 49, 29-40.

Waterhouse, T., P., C., Vipond, J., Goddard, P.E., 2003. Identification of key issues - a summary of the final group session. . In: Proceedings of a workshop on improving sheep welfare on extensively managed flocks: economics, husbandry and welfare. Scottish Agricultural College, Aberdeen, Scotland, pp. 73-77.

Webster, J., 1994. Animal welfare - a cool eye towards Eden: a constructive approach to the problem of man's dominion over the animals. Blackwell Science, Oxford.

Welsh, E.M., Gettinby, G., Nolan, A.M., 1993. Comparison of a visual analogue scale and a numerical rating scale for assessment of lameness, using sheep as a model. American Journal of Veterinary Research 54, 976-983.

Wemelsfelder, F., 1997. The scientific validity of subjective concepts in models of animal welfare. Applied Animal Behaviour Science 53, 75-88.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Animal Welfare 16, 25.

Wemelsfelder, F., Farish, M., 2004. Qualitative categories for the interpretation of sheep welfare: a review. Animal Welfare 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. Applied Animal Behaviour Science 67, 193-215.

Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the 'whole animal': a free choice profiling approach. Animal Behaviour 62, 209.

Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M., Sandilands, V., 2009a. Qualitative behaviour assessment. In: Assessment of animal welfare measures for layers and broilers. Welfare Quality reports No. 9, Cardiff, pp. 113-119.

Wemelsfelder, F., Lawrence, A.B., 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Acta Agriculturae Scandinavica Section A-Animal Science, 21-25.

Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009b. Qualitative behaviour assessment. In: Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves. Welfare Quality reports No. 11, Cardiff, pp. 215-224.

Wemelsfelder, F., Nevison, I., Lawrence, A.B., 2009c. The effect of perceived environmental background on qualitative assessments of pig behaviour. Animal Behaviour 78, 477-484.

Whay, H.R., Main, D.C.J., Green, L.E., Webster, A.J.F., 2003a. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. Animal Welfare 12, 205-217.

Whay, H.R., Main, D.C.J., Green, L.E., Webster, A.J.F., 2003b. An animal-based welfare assessment of group-housed calves on UK dairy farms. Animal Welfare 12, 611-617.

Whay, H.R., Main, D.C.J., Green, L.E., Webster, A.J.F., 2003c. Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records. Veterinary Record 153, 197-202.

Whittington, R.J., 1995. Observations on the indirect transmission of virulent ovine footrot in sheep yards and its spread in sheep on unimproved pasture. Australian Veterinary Journal 72, 132-134.

Winter, A., 2001. Mastitis in ewes. In Practice, 160-163.

Winter, A., 2004a. Lameness in sheep 2. Treatment and control. In Practice 26, 130-136.

Winter, A., 2004b. Lameness in sheep: 1. Diagnosis. In Practice 26, 58.

Winter, A.C., 1997. Welfare concerns in sheep farming in the United Kingdom. In: Proceedings of the 4th International Congress for Sheep veterinarians, Armidale, Australia, pp. 11-15.

Winter, A.C., 2008. Lameness in sheep. Small Ruminant Research 76, 149-153.

Winter, A.C., Fitzpatrick, J. L, 2007. Sheep welfare: standards and practices. In: Diseases of Sheep. Fourth Ed. Blackwell Publishing, Oxford, pp. 155-21.

Wood, G., Molony, V., 1992. Welfare aspects of castration and tail docking of lambs. In Practice, 14, 2-7.

Yates, W.J., Gleeson, A.R., 1975. Relationships between condition score and carcass composition of pregnant Merino sheep. Australian Journal of Experimental Agriculture 15, 467-470.

Appendix A

SUPPLEMENTARY INFORMATION RELATING TO CHAPTER 1

On-farm welfare issues for sheep identified through a pre-meeting worksheet

Ewes

1. Footrot
2. Scab
3. Metabolic disorders
4. Under and over-nutrition
5. Abortion
6. Reproductive disorders and prolapse
7. Management
8. Lameness
9. Vaccination
10. Anthelmintics and endoparasite control
11. Live fluke
12. Blowfly strike
13. Body condition
14. Mastitis and udder health
15. Poor dental health, tooth loss
16. Value as cull animal
17. Subsidies
18. Wool value
19. Transport of lame sheep & legislation
20. Lack of shelter and shade
21. Poor environmental conditions
22. Failure to treat disease
23. Dystocia, lack of supervision at lambing time
24. Inappropriate and rough lambing invention
25. Lack of veterinary input
26. Failure to investigate ill thrift
27. Scab
28. Respiratory disease
29. Enteric disease
30. Marketing & transport
31. Transport of pregnant ewes
32. Quality of handling: facilities and stress
33. Poor culling policy
34. Caesarian section
35. Conformation
36. Perineum
37. Con-specifics
38. Mature weight
39. High prolificacy
40. Artificial insemination
41. Cast sheep
42. Winter housing
43. Wool length
44. Clostridial control
45. Timing of weaning
46. Chronic disease
47. Tagging losses
48. Turnout after winter shearing
49. Contagious conjunctivitis
50. Cross-fostering
51. Early lambing
52. Breeding elderly ewes
53. Ecto-parasitism

Rams

1. Footrot
2. Scab
3. Under or over-nutrition
4. General management
5. Lameness
6. Vaccination
7. Inappropriate use of anthelmintics
8. Ecto-parasitism
9. Rearing as young animal compared with final environment
10. Neglect out-of breeding season
11. Low cull value
12. Subsidies
13. Legislation on movement
14. Lack of shelter and shade
15. Poor environmental and pasture conditions
16. Failure to treat disease
17. Expected to serve too many ewes
18. General foot health and care
19. Respiratory disease
20. Scrapie
21. Lack of housing
22. Chorioptic mange
23. Conformation and selection for physical attributes
24. Ease of lambing
25. Use of raddling harness
26. Bullying
27. Use of electro-ejaculation
28. Body condition score
29. Hormonally active
30. Physically able
31. Cleanliness of brisket
32. Weight correct for ewe
33. Fighting injuries
34. Tooth loss
35. Wool length
36. Maggot control
37. Group size
38. Handling facilities
39. Post-tupping recovery
40. Losses due to tagging
41. Stress associated with management
42. Vasectomy
43. Blindness caused by occlusion of horns
44. Endo-parasitism
45. Handling stress

Young lambs

1. Predation
2. Blowfly strike
3. Joint Ill
4. Inter-digital dermatitis
5. Control of sheep scab
6. Dystocia
7. Ease of parturition
8. Tail docking
9. Castration
10. Starvation
11. Hypothermia
12. Coccidiosis
13. Anthelmintics and endoparasites
14. Lack of housing and shelter
15. Presence of shepherd: management
16. Hygiene
17. Colostrum quality and supply
18. Birth weight
19. Ewe body condition
20. Pen size
21. Movement restrictions
22. Poor hygiene and poor bio-security,
23. Isolation or overcrowding
24. Poor ewe nutrition
25. Poor outdoor environmental conditions
26. Failure to treat disease
27. Lack of lambing supervision
28. Poor rearing pet lambs
29. Failure to foster and/ or artificially feed
30. Late docking & castrating
31. Mis-mothering and poor maternal bond
32. Respiratory disease
33. Neglect early in life
34. Ewe mastitis
35. Lameness
36. Orf
37. Handling
38. Ear tagging and notching
39. Transport
40. Weaning
41. High mortality: pre- and post-partum
42. Lack of veterinary involvement
43. Slowness to stand and suck
44. Acidosis
45. Survival
46. Growth rates
47. Temporary ewe-lamb separation
48. Watery mouth
49. Injury/physical damage
50. Stocking rate
51. Multiple births
52. Failure to correct entropion,
53. Deaths due to failure to vaccinate

Growing and fattening lambs

1. Blowfly strike, maggot control
2. Parasitic gastro-enteritis
3. Anthelmintic resistance
4. Lameness
5. Scab
6. Management of ewe and entire ram lambs
7. Feed availability
8. Management
9. Transport
10. Lack of shelter and shade
11. Interdigital dermatitis ('scald)
12. Footrot
13. Weaning method
14. Subsidies
15. Respiratory disease
16. Poor environmental conditions
17. Failure to treat disease
18. Inappropriate health plans
19. Lack of veterinary intervention
20. Failure to investigate ill thrift
21. Joint disease
22. Poor nutrition and nutrient supply
23. Orf
24. Liver fluke
25. Routine handling: handling stress and facilities
26. Routine husbandry treatments
27. Acidosis
28. Urinary calculi
29. Dog worrying and dog control
30. Infectious diseases
31. Nutrient supply
32. Foot care, foot health
33. Group size and selection
34. Intensive indoor finishing
35. Provision of grazing or fodder
36. Transport and loading facilities
37. Misadventure
38. Failure to vaccinate
39. Mortality
40. Overcrowding indoors
41. Water deprivation
42. Ecto-parasitism

Table A.1 Potential indicators of sheep welfare identified by a consensus of expert opinion

Five Freedoms	Young lambs	Growing lambs	Ewes	Rams
Freedom from hunger and thirst	Ewe nutrition and body condition	Provision of appropriate and adequate diet and water	Provision of appropriate and adequate diet	Nutrition & body condition
	Ewe and lamb behaviour	Management of dietary change	Reproductive and general health status	General health status
	Lamb nutritional management	Appropriate stocking density		Reproductive management
Freedom from discomfort	Indoor and outdoor environmental conditions	Indoor and outdoor environmental conditions	Indoor and outdoor environmental conditions	Indoor and outdoor environmental conditions
			Presence of lie-back area	Absence of housing/shelter
Freedom from pain, injury or disease	Presence of disease (parasitic, foot and limb, ocular and infectious disease)	Presence of disease (skin, nutritional, foot and limb, dental, respiratory, endo-parasitic, infectious and zoonotic)	Presence of disease (skin, foot, metabolic, endo-parasitic, ocular, reproductive and udder disease)	Presence of disease (skin, foot, dental, endo-parasitic, inheritable, reproductive disease)
	Failure to take appropriate control/treatment of disease	Failure to take appropriate treatment/control of disease	Failure to take appropriate action to treat/control disease	
			Transport of pregnant and lame sheep	Management practices associated with pain (electro-ejaculation, raddling, vasectomy, fighting)
	Management practices associated with pain (mutilations, dystocia, predation)		Presence/absence of shepherd (inspection and lambing)	
Freedom to express normal behaviour	Quality of housing and management	Management of housed and grazing lambs	Reproductive management practices	Mating management and control
Freedom from fear and distress	Stockmanship	Stockmanship	Stockmanship	Stockmanship
	Stocking density	Use and presence of dogs	Quality of handling (skills and facilities)	Quality of handling (skills and facilities)

Appendix B

SUPPLEMENTARY INFORMATION RELATING TO CHAPTERS 4 – 9

Recording sheets used for welfare indicator assessments

Indicators assessed by group observation (Table B.1)

Indicators assessed by individual examination during cross-sectional (Table B.2) and longitudinal studies (Table B. 3)

Young lamb indicators (Table B.4)

Resource -based indicators (Table B.5)

Management-based indicators (Table B.6)

Assessment of on-farm records (Table B.7)

Data capture form

Table B.1 Indicators assessed by group observation

Cross-sectional study			
Date: / / Time:	Assessor Name:	Total number of sheep in sample:	**FARM ID:**
Weather during assessment: WET/DRY	Location of assessment:	Sheep type: ram/ewe/young lamb/growing-fat lamb	
ANIMAL-BASED INDICATORS ASSESSED GROUP OBSERVATION: RAMS, EWES & GROWING LAMBS			
WELFARE INDICATOR	**TALLY**	**Total number observed with condition**	**Comments**
Dull demeanour			
Excessive Panting			
Coughing			
Skin irritation			
Wool loss			
Lame sheep			
Dirty rear			
Dirty belly			

Table B.2 Indicators assessed by individual examination (Chapters 4 and 7)

Date: / / Time:	Assessor Name:	Total number of sheep in sample:	**FARM ID:**
Weather at time of assessment:	Location of assessment:	Sheep type: ram/ewe/young lamb/growing-fat lamb	

ANIMAL-BASED INDICATORS ASSESSED BY INDIVIDUAL EXAMINATION: RAMS, EWES & GROWING LAMBS

SHEEP ID	Demeanour (0=bright/alert, sick/dull/depressed)	Eye Abnormality (0= absent 1= present)	Nasal Discharge (0= absent 1= present)	Tooth Disease (0=no, 1=broken mouthed, 2= molar, 3 = incisor & molar	Age (Years/months/production stage)	Coughing (0= absent 1= present)	Damaged deformed ears (0= absent 1= present)	In-growing horns (0= absent 1= present)	Belly (0=clean, 1=dirty mud, 2=dirty faecal 3=filthy)	Legs (0=clean, 1=dirty mud, 2=dirty faecal, 3=filthy)	Rear (0=clean, 1=dirty mud, 2=dirty faecal, 3=filthy)	EWES: (0=no mastitis, 1=one gland, 2=both)	RAMS: (0=no crystals, 1=crystals prepuce/penis)	Tail length (0=OK , 1 =too short)	Wool Loss (0= absent 1= present)	Skin Irritation (0= absent 1= present)	50p size, 3= single hand-size, 4=multiple hand-sized , 5=diffuse,6=small <5p size areas < 5 areas)	Injury/wounds (0=no, 1= ≤ 5 scratches ≤5cm, 2 = ≥ 5 scratches ≥5cm, 3=healing wounds, 4=single open wound, 5=multiple open wounds)	Original BCS System: 0-5 scale Full Score	Very Thin (0=no, 1=yes)	Fit for purpose (0=no, 1=yes)	Very Fat (0=no, 1=yes)	Lameness (0= absent 1= present)	Any foot lesion (0= absent 1= present)	Lesion significant to cause lameness (0=no, 1=yes)	White Line (0= absent 1= present)	Scald (0= absent 1= present)	Footrot (0= absent 1= present)	CODD (0= absent 1= present)	Other (1=Toe granuloma, 2=Interdigital growth, 3=Pedal abscess, 4=unknown lesion)	Joint Swelling (0= absent 1= present)	Myiasis (0=none, 1= <50p size 2= multiple 50p size, 3=single hand size, 4=multiple hand size, 5=diffuse)

Table B.3 Indicators assessed by individual examination (Chapter 8)

LONGITUDINAL STUDY			
Date: / Time:	Assessor Name:	Total number of sheep in sample:	FARM /ID:
Weather at time of assessment:	Location of assessment:	Sheep type: ram/ewe/young lamb/growing-fat lamb	

INDIVIDUAL SHEEP OBSERVATIONS & EXAMINATION: RAMS, EWES, GROWING LAMBS, YOUNG LAMBS

SHEEP ID	Sheep Type: Ram (R) Ewe (E) Growing lamb	Demeanour (0=bright/alert,	Eye Abnormality (0 =absent 1= present)	Nasal Discharge (0 =absent 1= present)	*Tooth Disease (0=no, 1=incisor 2= molar, 3 =*	Coughing (0 =absent 1= present)	Damaged ears (0=no, 1 =active, 2=inactive	In-growing horns 0 =absent 1= present)	Belly – MUD (0=clean, 1= dirty, 2 =filthy)	Rear - MUD (0=clean,1= dirty, 2 =filthy)	Rear - FAECES (0=clean,1=dirty, 2 = filthy)	*Ewes: Udder (0=no disease, 1=one side affected 2=both sides affected)*	Tail length (0= covers anus or vulva, 1 = too short)	Wool Loss (0 =absent 1= present)	Skin Irritation (0=no, 1=yes)	Skin Lesion (0=no, 1=single area 50p size, 2=multiple areas 50p size, 3= single hand-sized area, 4=multiple hand -sized areas, 5=diffuse, 6=small <5p size areas < 5 areas)	Injury/wounds (0=no, 1= scratches ≤5cm/5 in no. 2=scratches ≥5cm/6 in no. 3=healing wounds, 4=single open wound, 5=multiple open wounds)	Original BCS System: 0-5 scale	Fit = 0, Fat = 1, Thin = 2	Lame (0 =absent 1= present)	*White Line* 0 =absent 1= present)	*Scald* 0 =absent 1= present)	*Footrot* (0 =absent 1= present)	*CODD* 0 =absent 1= present)	*Other (1=Toe granuloma, 2=Interdigital growth, 3=Pedal abscess, 4=unknown lesion)*	*Joint Swelling* 0 =absent 1= present)	Myiasis (0 = absent, 1 = present)
1																											
2																											
3																											
4																											
5																											
6																											
7																											
8																											
9																											
10																											

Table B.4 Young lamb indicators (Chapter 6)

ANIMAL-BASED INDICATORS FOR YOUNG LAMBS: from birth to 6 weeks old

te: / /	Location of assessment:	Assessor Name:	Total number of lambs in sample:	**FARM ID:**

LAMB ID	**AGE:** 0 to ≤3 days = 1, 3days to ≤7 days = 2, ≥ 1 week to ≤ 6 wks =3	**MANAGMENT:** Housed individual pen = P, in groups = G, Outside = O	**REARING: 0 =** With Ewe (loose), 1 = With Tethered Ewe, 2= orphan lamb	**DEMEANOUR:** (0= bright, alert, 1 = dull, depressed)	**STIMULATION TEST:** (0=Bright, alert and responsive when	**STANDING ABILITY:** 0=stands easily, 1=weak on legs, 2 = cannot stand, recumbent)	**BODY CONDITION:** (0= no ***prominent*** hip bone, good cover over backbone and body 1= ***prominent*** hip bone, poor cover over backbone - thin body condition)	**POSTURE :** (0= Normal, relaxed posture, 1= Hunched tucked up posture)	**ABDOMEN:** (0 = normal abdominal fill, 1 = abnormally distended, 2 = hollow abdomen)	**LAMENESS:** (0=sound, no obvious lameness/stiffness/ joint ill, 1 = lame, obvious signs of lameness/stiffness/joint ill)	**SALIVATION:** (0=no evidence of salivation, 1=presence of salivation)	**SHIVERING:** (0=no shivering , 1= lamb shivering)	**EYE ABNORMALITY:** (0=no eye disease, 1=presence of eye disease/entropion)	**PLAY BEHAVIOUR:** (min = no play behaviour with objects or other lambs or sheep, max = could not be more playful with other lambs, other sheep or objects) min max
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														

Table B.5 Resource-based indicators (Chapter 7)

Date: / / Time:	Assessor Name:		FARM ID:
RESOURCE-BASED INDICATORS OF SHEEP WELFARE			
ASSESSMENT OF GRAZING AREAS	Yes/No	Circle If Applicable	Comments
Presence of facilities to deal with weather			
Grazing provided		Pasture Type: Grass / Soil / Rough Hill / Stubble Turnip /...............	
Lie back area provided			
Water trough/supply provided in field			
Water supply accessible			
Water Quality	N/A	Clear & fresh / murky / filthy with debris & faecal contamination	
Good boundaries and fences			
Freedom to move in field		Lots of room / adequate / over-crowded	
Number of sheep in field/observation area			
Size of field (hectacre)			
Length of time on field being measured	.... Days/ Weeks/ Months		
Sward Height (cm)			
ASSESSMENT OF HOUSING	Yes/No	Circle If Applicable	Comments
Number of sheep per housed/pen area			
Width of pen/housing (metres)			
Length of pen/housing (metres)			
Freedom to move inside housing		Lots of room / adequate / over-crowded	
Adequate lighting in housing for inspection			
Flooring type	N/A	Flooring type: Concrete / wire mesh / slats /	
Bedding provided		Bedding type: Straw / Shavings /.....................................	
Bedding of sufficient quality for comfort & hygiene			
Squelch Test (*If applicable)	N/A	*Squelch test: Dry / Squelchy / Obvious wet & pooling of water	
Water trough/supply provided in housed area			
Water supply accessible			
Water Quality	N/A	Clear & fresh / murky / filthy with debris & faecal contamination	
Presence of food in troughs/racks			
Feed trough/rack accessible			
Hay/Straw/Silage of sufficient quality		Forage type: Hay/Straw/Silage Restricted/ Ad lib	

Table B.6 Management-based indicators (Chapter 7)

Date: / / Time:	Assessor Name:	FARM ID:
MANAGMENT-BASED INDICATORS OF SHEEP WELFARE		
	Yes/No	**Circle If Applicable / Further comments**
Presence of foot bathing facilities		
Presence of isolation pen/area		
Facilities to deal with weather/shelter		Trees / Straw bales / Stone wall / Housing / Polytunnel
Evidence of appropriate health control & action		No disease/ Disease-Appropriate Action /-Disease Inappropriate Action
Number of dead bodies amongst live sheep		.
Presence of sheep handling equipment		
Equipment well maintained and in good order		
Evidence of rough or inappropriate handling		
Mixing of horned and unhorned sheep		

Table B.7 Assessment of on-farm records

Date: / / Time:	Assessor Name:			FARM ID:	
ON-FARM RECORDS					
RECORD TYPE	Records MAINTAINNED? Yes/No	Records USEFUL? Yes/No	Comments	INFORMATION PROVIDED	(Yes/No)
Mortality records present?				Reason for culling /death	
				Age at culling/death	
				Time period/dates	
				Number of sheep that died	
Medicine records present?				Use of local anaesthetic	
				Treatments used & reasons	
Disease records present?				Lameness	
				Mastitis	
				Dystocia/reproductive	
				Scab/skin disease	
				Gut worms/fluke	
Flock Health plan present? (* *if applicable*)					
*Veterinary Involvement with flock health plan?					
Are the following specific areas covered by plan:	Yes/No	Comments			
Lameness prevention					
Lameness treatment					
Scab prevention					
Scab treatment					
Obstetric advice and treatment					
Worm and Fluke control					
Tail docking policy					
Castration policy					
Vaccination policy					
Planning for emergencies, movement restrictions					
Policy for hypothermia management: lambing time					

Data capture form

Date: **Farm ID:**

Farm type and purpose
Farm type: Lowland / Upland/ Hill / Mountain
Pedigree / Commercial / Hobby
Replacements: bought-in / home-bred

Management system: Housed/ Outdoors / ....
Weather on day of assessment:...
Weather conditions in previous 1-2 weeks: ....
Farm Assured: No/Yes : Name of scheme ....

Flock details

Production Stage	Number	Breeds
Ewes		
Rams		
Young lambs (at foot/orphaned)		
Growing/fattening lambs		
Replacement females		

Performance
Age/Weight when sent for slaughter:....................
% sent for slaughter:............ % stores .... % retained for breeding ....
Total number of ewes put to the tup:
Scanning performed: yes/no Scanning % ....
Lambing %: Method lambing % calculated ....
Lambs weaned per ewe: Number of lambs sold:
Replacement rate: Ewes:.................Tups.........................
Culling %: When are sheep culled:...

Reproductive management
Breeding policy: Ewe: Ram ratio ....
Reproductive management: Use of Synchronisation: yes/no
Use of Sponges: yes/no
Use of teaser rams: yes/no
Use of melatonin: yes/no
Use of AI: yes/no
Use of ET: yes/no

Sire selection: How are rams selected: auction ram sales / private vendor / breed own
Bought –in sheep: quarantined: yes/no:
If quarantined: Length of quarantine .... days/weeks

Quarantine treatments given: foot bath/ footrot vaccine/injectable avermectin/ dipped/pour on/ worm drench/fluke treatment
Number of people who are regularly employed to work with sheep ....
Additional employment needed at lambing yes/no number of staff at lambing: ....
How regularly do you inspect sheep: .... times per day/week /month / year
How are casualty animals are dealt with?
....

If animals are housed: how often is food replenished: ....
Feeding regime for growing/finishing lambs ....
Policy specific for winter ration: yes/no and details ....
Forage: what is being fed: silage/ hay / straw / at grass
Are concentrates offered to ewes pre-lambing yes/no
quantity fed ....
how often are concentrates given .... times day/week
bought in / homemade mix and details..

Tail docking performed: yes/no
Method: rubber ring/budizzo/surgical/other ....
Local anaesthetic used: yes/no
When performed: age .... days / weeks
Castration performed: yes/no
Method: rubber ring/bloodless castrator/knife/ combined method
Local anaesthetic used: yes/no
When performed: age .... days / weeks

Lambing
Time of year: .... Indoor/outdoor lambing / both
Does the flock have many problems with dystocia/need for assistance: yes/no
Caesareans: approx number .... in previous lambing season
Prolapses: Do you observe this condition in the flock: yes/no
How are prolapses treated...
What is the on-farm policy for hypothermia prevention/treatment? ....
Naval dipping performed: yes/no What is applied ....

Flock health and welfare
What do you think of your flock welfare (subjective assessment by farmer e.g. good/above average/could be improved): ....

Do you have a flock health plan: yes/no
If yes, was there veterinary involvement with the health plan: yes/no

Do you have any particular problems with the following sheep diseases and if so what actions have been taken?

Lameness: yes/no ....
Scab: yes/no ....
Gut worms: yes/no ....
Fluke: yes/ no ....

Consider the health and welfare of the whole flock on the day of assessment:
Farmer estimate % lame in flock:...................
Farmer estimate % flock with skin condition:..................

In your opinion, what, if any, are the main flock health and welfare problems:
....
....

Routine foot care details: none / foot bathing / foot trimming / combination
Frequency: .... times per year / month

Routine scab and fly control: Dipping / Showering / Pour-on / Injectable
Frequency .... times per year/ month

Routine dagging/crutching performed: yes/no
When performed (time of year)...................... Frequency .... times per year

Recently sheared? yes/no Periods/dates of shearing: ....

Vaccinations routinely administered:

	Tick if given
Clostridial diseases	
Pasteurella	
Foot rot	
Orf	
Toxoplasma abortion	
Enzootic abortion	
Others:	

Which sheep receive vaccinations:
....
....

How often has vet been on-farm for sheep in past 12 months: approx .... times
Is vet attention sought for individual animals: yes/no
How frequently is veterinary advice requested: frequently /occasionally/ never

Appendix C

SUPPLEMENTARY INFORMATION RELATING TO CHAPTER 10

Standard operating procedures for on-farm indicators of sheep welfare

Selecting indicators recommended for application in on-farm assessment of sheep and lamb welfare based on expert opinion and results of on-farm validation studies.

Table C.1 Selecting indicators for use in on-farm assessments of sheep and lamb welfare

INDICATORS ASSESSED BY GROUP OBSERVATION							
Welfare Indicator	Legislation		Interpretation of reliability	Responsiveness to		Comments	Recommend for use
	W of FA	SMR's		Seasonal variation	Farm-level variation		
Demeanour	1 & 5 & 6	A1, A4, A5	Excellent [a]	Yes	Yes	Indicator of general health	Yes
Excessive panting	5 & 6, 13 &17	A1, A4, A5, A 11, A13	[b]	No	No	May indicate heat stress, over exertion or gathering in hot weather	No [b]
Skin irritation	1 & 5 & 6	A1, A4, A5	Excellent [a]	Yes	No	Undisturbed assessment is crucial to identifying this indicator and there can be a high risk of false negative even if not observed during assessment. Therefore, individual examination using the nibble test on sample of individual sheep is recommended.. If observed may indicate the presence of an external skin disease or in the individual animal a neurological condition (scrapie)	Yes, but if not seen does not rule out skin irritation
Wool loss	1 & 5 & 6	A1, A4, A5	Poor [a]	Yes	Yes	Need to be able perform close observe of entire sheep. Flocking may hamper observation so wool loss could be assessed using the individual examination method. If observed may indicate the presence of an external skin disease, general ill-health or in the individual animal a neurological condition (scrapie). Wool loss may also be a normal seasonal spring phenomenon in sheep.	No, considered to be a useful[a] but poor reliability
Lameness	1 & 5 & 6	A1, A4, A5	Excellent	Yes	Yes	Sufficient assessment space required. Ease of assessment affected by flocking instinct and 'flightiness' of sheep and quality of assessment surface. Sheep need to be walked away from observer in relatively quiet manner, when walking away from observer.	Yes
Coughing	1 & 5 & 6	A1, A4, A5	Poor [a]	Yes	No	Not possible to accurately count the number of affected animals.	No [b]
Dirty rear	1 & 5 & 6, 11	A1, A4, A5, A9, B2	Excellent	Yes	Yes	Relatively close-up assessment needed, observation affected by flocking instinct of sheep. Cleanliness may be affected by endoparasitism, bacterial enteritis and dietary changes and breed variation. Difficult to assess in sheep with dark coloured fleece	Yes
Dirty belly	1 & 5 & 6, 17	A1, A4, A5, A13	Excellent [a]	Yes	Yes	Used to indicate cleanliness and dryness of environment.	Yes
QBA	1 & 5 & 6		Good – video clip	Yes	Yes	Great potential but further work needed to examine on-farm observer reliability testing before can recommend for current use.	No [b] see comment

ANIMAL-BASED INDICATORS ASSESSED BY INDIVIDUAL EXAMINATION

Welfare Indicator	Legislation: W of FA	Legislation: SMR's	Interpretation of reliability	Responsiveness to		Comments	Recommend for use
Demeanour	1 & 5 & 6	A1, A4, A5	Excellent	Yes	Yes	Indicator of general Health	Yes
Eye condition	1 & 5 & 6, 29	A1, A4, A5, B8	Fair to Good	Yes	No	Used to indicate eye disease	Yes
Nasal discharge	1 & 5 & 6	A1, A4, A5	Fair to Good [a]	No	No	Used to identify respiratory disease	Yes
Tooth condition	1 & 5 & 6, 22-27	A1, A4, A5, A16	Fair to Good	No	No	Used to assess the ability of sheep to masticate, however the degree that tooth disease affects this depends on diet, quality of care and would be reflected in the animal's body condition,.	No, use body condition
Coughing	1 & 5 & 6	A1, A4, A5	Fair to Good [a]	No	No		No [b]
Ear lesions	1 & 5 & 6, MPP	A1, A4, A5, B6	Fair to Good	No	No	Could be used to indicate poor tagging technique, however the significance of certain lesions is unknown	No
In-growing horns	1 & 5 & 6	A1, A4, A5	[b]	No	No	Not observed in sample population. Important indicator to detect physical penetration of the eye or skull.	Yes [b]
Dirty rear	1 & 5 & 6, 11	A1, A4, A5, A9, B2	Poor	Yes	No	Used to indicate faecal soiling which may be due to gastro-intestinal disorders or dietary changes, risk factor for myiasis. Can be assessed by group observation	Yes, poor reliability. suggest simplified scoring
Dirty belly	1 & 5 & 6, 17	A1, A4, A5, A13	Fair to Good	Yes	Yes	Used to indicate environmental conditions; dirty or unhygienic environments.	Yes
Dirty legs	1 & 5 & 6, 17	A1, A4, A5, A13	Fair to Good	n/a	n/a	Used to indicate environmental conditions; dirty or unhygienic environments but recommend use of ventral abdomen instead	No
Mastitis	1 & 5 & 6	A1, A4, A5	Fair to Good	Yes	No	Used to identify chronic and acute mastitis in ewes. Difficult to identify the presence of mastitis during the post-weaning period	Yes
Crystals	1 & 5 & 6, 22-27	A1, A4, A5, A16	No data	No	No	Not seen in sample population but used to identify urolithiasis. Demeanour may identify clinical cases.	No [b]
Tail Length	MPP	A1, B6	Excellent	No	No	Used to assess compliance with current legislation.	Yes

Fleece Loss	1 & 5 & 6	A1, A4, A5	Excellent [a]	Yes	yes	Indicates the presence of an external skin disease, general ill-health or a neurological condition (scrapie)	Yes
Skin irritation	1 & 5 & 6	A1, A4, A5	Fair to Good [a]	No	No	Indicates the presence of an pruritic external skin disease or a neurological condition (scrapie)	Yes
Skin lesion	1 & 5 & 6	A1, A4, A5	Fair to Good [a]	Yes	No	Indicates presence of pruritic and non-pruritic skin conditions. Degree of fleece cover can hamper detection of small lesions.	Yes
Injury and wounds	1 & 5 & 6, 11, 12, 17	A1, A4, A5, A9, B2, A10, A13	Poor [a]	No	No	Indicates the presence of wounds and injuries. May indicate inadequate housing design, poor management practices, inappropriate use of dogs and poor handling systems. Fleece cover can hamper detection of skin lesions.	No, poor reliability
BCS	1, 22-27, 2, 5	A1, A16, A2, A4	Fair to Good [a]	Yes	Yes	Used to indicate the adequacy of nutrition in the previous months. Seasonal changes in body condition are expected, particularly reproductive ewes. Not appropriate for assessing growing lambs.	Yes
Fit-fat-thin	1, 22-27, 2, 5	A1, A16, A2, A4	Fair to Good [a] of thin and fat	Yes	Yes	Used to indicate the adequacy of previous nutritional intakes	Yes
Lameness	1 & 5 & 6	A1, A4, A5	Fair to Good	Yes	Yes	Conditions for assessment of lameness and the 'flightiness' of the individual sheep affect the ability to identify this indicator.	Yes
Foot lesion	1 & 5 & 6	A1, A4, A5	Fair to Good	No	No	Significance of foot lesion requires veterinary clinical knowledge of sheep health and medicine	No, assess lameness[d]
'Significant 'Foot Lesion	1 & 5 & 6	A1, A4, A5	Fair to Good	No	No	Diagnosis and assessment of specific lesions requires veterinary clinical knowledge.	No, assess lameness[d]
White Line Lesion	1 & 5 & 6	A1, A4, A5	Fair to Good			Lameness assessment should be sufficient	
Scald	1 & 5 & 6	A1, A4, A5	Fair to Good				
Foot Rot	1 & 5 & 6	A1, A4, A5	Fair to Good	n/a	n/a		
CODD	1 & 5 & 6	A1, A4, A5	Excellent				
Other Foot Lesions	1 & 5 & 6	A1, A4, A5	Fair to Good				
Joint swelling	1 & 5 & 6	A1, A4, A5	Excellent	Yes	Yes	May be useful measure but should be detected by lameness assessment	No, assess lameness[d]
Myiasis	1 & 5 & 6	A1, A4, A5	Excellent [a]	Yes	No	Important indicator for identifying myiasis .Requires close and meticulous inspection of individual sheep. Seasonal effect on presence of maggots.	Yes

ANIMAL-BASED INDICATORS OF YOUNG LAMB WELFARE							
Welfare Indicator	Legislation		Interpretation of reliability	Responsiveness to		Comments	Recommend for use
	W of FA	SMR's		Farm-level variation	Seasonal variation		
Demeanour	1 & 5 & 6	A1, A4, A5	Good to Fair	Yes	c	Indicator of general health. Inclement weather can affect lamb demeanour. Observer needs to ensure lamb is awake and not sleeping.	Yes
Stimulation	1 & 5 & 6	A1, A4, A5	Good to Fair	Yes	c	Indicators of general health. Requires close observation and physical stimulation	Yes
Standing	1 & 5 & 6	A1, A4, A5	Good to Fair	Yes	c	Indicator to identify general ill-health, muscular or neurological conditions	Yes
Body condition	1, 22-27, 2, 5	A1, A16, A2, A4	Good to Fair	Yes	c	Used to indicate inadequate nutrition of the ewe or lamb. Requires close observation,	Yes
Posture	1 & 5 & 6	A1, A4, A5	Good to Fair	Yes	c	Indicator to identify general ill-health, muscular or neurological conditions. Posture may be affected by inclimate weather.	Yes
Abdomen	1, 22-27, 2, 5	A1, A16, A2, A4	Good to Fair	Yes	c	Used to indicate inadequate/inappropriate nutrition of lamb. Requires close observation, possibly palpation by observer	Yes
Lameness	1 & 5 & 6	A1, A4, A5	Good to Fair	No	c	Include lameness due to joint or other causes.	Yes
Salivation	1 & 5 & 6	A1, A4, A5	Good to Fair	No	c	Used to indicate watery mouth, requires close observation [b]	Yes *
Shivering	1 & 5 & 6 & 17	A1, A4, A5, A13	Good to Fair	No	c	Very low proportion of affected sheep in in study population , and could be considered a normal response to cold.	No [b]
Eye condition	1 & 5 & 6	A1, A4, A5	Good to Fair	Yes	c	Requires close observation	Yes
Play behaviour	1 & 5 , 22-27, 5 & 6	A1, A2, A4, A4, A5, A16	Poor	No	c	Positive sign of welfare when present, but not significant if not observed in brief observation period	No

Welfare Indicator	Legislation		Reliability	Responsiveness		Comments	Recommend For use
	W of FA	SMR's					
RESOURCE- & MANAGMENT-BASED INDICATORS							
Provision of shelter	17	A13	c	c	c	Requirements may vary, may depend on sheep breed, management system	Yes
Grazing provided	22-27	A16	c	c	c	Grazing or alternative feed in field	No, assess body condition
Sward height		A16	c	c	c	Assessment of 40 measures in each grazed field not practical and influenced by length of grazing period. Recommend use of body condition assessment instead.	No, assess body condition
Lie back provided	17	A13,	c	c	c	Used to assess dry lying area available, for example when grazing stubble root crops	Yes, could assess sheep cleanliness
Water supply provided	22-27	A17, A18	c	c	c		Yes
Water supply accessible	22-27	A17, A18	c	c	c		Yes
Water quality	22-27	A17, A18	c	c	c		Yes
Good boundaries and fences	17	A13	c	c	c	Used to assess action taken to prevent hazards/injuries and biosecurity. Not applicable in extensive systems and not feasible to perform full assessment of farm perimeter	No
Freedom to move in field	9	B1	c	c	c	Very subjective measure. Recommend sufficient grazing/alternative feed in field and presence of lie-back area indicators	No
Stocking density	9	B1	c	c	c	Difficult to assess field size of across entire farm acreage. Recommend assess sufficiency of grazing and lie-back area .	No
Size of field/area	9	B1 A16	c	c	c	Difficult to assess field size of all farm acreage. Recommend sufficient grazing/alternative feed in field and presence of lie back area	No
Stocking rate of housed sheep		B1	c	c	c	Number of sheep/pen measurements based on Defra Sheep Welfare Codes	Yes
Adequate lighting for inspection	3, 14-16	A3, B3	c	c	c		Yes
Freedom to move inside housing	9	A8 B1	c	c	c	Subjective assessment. Recommend assess stocking rate	No
Flooring type	4	A9 A5A10 A13	c	c	c	Difficult to be prescriptive, should be based on quality, hygiene, comfort to animals	no
Bedding provided	4	A5A9 A11	c	c	c		No, assessed by below
Bedding sufficient quality for comfort and hygiene	4	A5 A9 A11	c	c	c	Similar to squelch test used for cattle assessments (Taldich *et al.*, 2006).	Yes

Moisture content of bedding	4	A5 A11	c	c	c		No assessed by above
Presence of food in troughs	22-27	A16,	c	c	c		No, assess body condition
Feed trough/rack accessible	22-27	A16	c	c	c		No, assess body condition
Forage of sufficient quality	22-27	A16 A18	c	c	c		No , assess body condition
Presence of isolation pen	6	A5	c	c	c	Isolation facilities may be created as needed so absence not necessarily negative sign	No
Rough handling	1	A1	c	c	c	Use animal based measure skin injuries and wounds	No
Presence of sheep handling equipment	2, 5	A2 A10	c	c	c		No
Maintenance of equipment	11, 12	A9, B2, A10	c	c	c		Yes
Presence of foot-bathing facilities	2, 5	A4,	c	c	c	Foot bathing is not the only way of treating lameness and not assessment of quality of foot bath or bathing practices	No
Dead bodies amongst live sheep	2, 5	A2 A4	c	c	c		Yes
Mixing horned & unhorned sheep			c	c	c	Assess injuries /wounds instead	No
Presence of medicines records	7, 8	A6, A7	c	c	c		Yes
Presence of mortality records	7, 8	A6, A7	c	c	c		Yes
Presence of disease records	1, 2, 5	A1, A2, A4,	c	c	c	Positive sign if present	Not a legal requirement
Flock health plan present	1, 2, 5	A1, A2, A4	c	c	c	Positive sign if present	Not a legal requirement
Tail-docking policy	MPP	B6	c	c	c	Information from farmer/ stock person regarding age/method/use of local anaesthetic	Yes
Castration policy	MPP	B6	c	c	c		Yes

W of FA = Welfare of Farmed Animals Regulations (England) 2007 (numbers signify relevant paragraph of the legislation)
SMR's = Statutory Management Regulations, MPP = Mutilations (Permitted Procedures) (England) Regulations 2007
a = low proportion of sheep affected in study population,
b = insufficient data collected,
c = not examined during study,
d = suggested use as part of further investigation following initial indicator assessment

Standard Operating Procedures (SOP's)

Welfare indicators for the on-farm assessment of sheep and lambs

University of Liverpool
Sheep Welfare Project

PhD research output
February 2011

Standard Operating Procedures (SOP's) for indicators of sheep welfare

Recommendations and guidance on SOP's

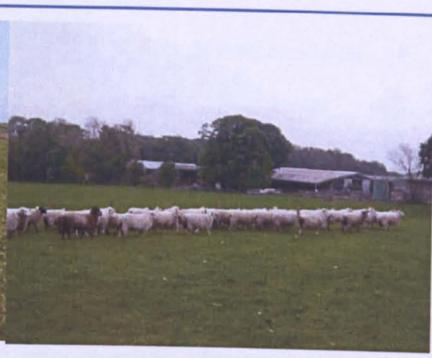
The animal-, resource- and management-based indicators for sheep and lambs detailed in the SOP's have been recommended for use in on-farm welfare assessments on the basis of their diagnostic performance in the field, expert opinions and feasibility for assessment.

Specific indicators may be selected according to welfare assessment scheme requirement. However, key animal-based indicators recommended by the author for adult sheep, growing lambs and young lamb welfare assessments are: lameness, demeanour and body condition – please note there are different methods and scoring systems for assessing these different production stages.

The welfare indicator assessments should be performed by experienced and trained assessors, familiar with the on-farm assessment of sheep and standardised to the scoring systems detailed in the SOP's.

The provisional cut-off points determined by the expert panel may be used to indicate the on-farm sheep welfare standards. However, exceeding cut-off points does not necessarily indicate non-compliance with legal standards but should indicate that further investigations by the assessor are required. This may involve investigations into the reasons behind the problem, or determining whether appropriate current action is being taken. The determination of cut-off points for the indicators should be an active and on-going process in which the indicator scoring systems and cut-offs should be refined and modified according to on-farm experience and scientific advances.

Animal-based indicators of sheep welfare

Welfare indicators of adult sheep and growing lambs assessed by group observation	
Protocol	The group observational method is used to assess the indicators of adult sheep (aged > 1 year) and growing lambs (aged > 6 weeks to < 1 year) detailed below. The method starts with and undisturbed assessment of group behavior. The observer stands at the barrier of the field or paddock or housing to observe the group of sheep for a period of 5 minutes. *After this* initial observation the observer enters the group assessment area and quietly walks around the sheep to assess *individual sheep* within the group. The observer then walks the group around the assessment area to assess the gait of individual sheep and assess each indicator. For the group assessment of 70 sheep, the time taken is approximately 20- 30 minutes. The assessment area should be a relatively flat surface, well drained, of approximately 1 to 2 acres. Note, the 'flightiness' of sheep may affect the ability of the observer to perform indicator assessments based on a group observation.
Example	  Observing group behaviour Assessing lameness

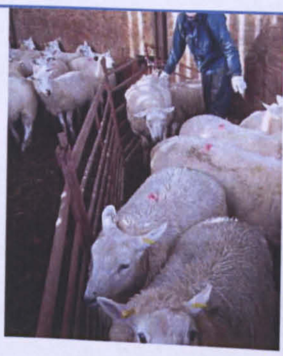
Indicator	Demeanour
Type	Animal Based, Group Observation
Welfare Issue	General Health Indicator: Freedom from Pain Injury and Disease
Method	The number of individuals animals in the group showing any or all of the following signs: Separation from the rest of the group; appearing dull; lowered head carriage; reduced responsive to the approach of observers. Demeanour can be affected by adverse weather conditions.
Score	Number of animals in group showing dull demeanour (percentage of the group).
Example	 Sheep showing appearance of dull demeanour

Indicator	Lameness
Type	Animal Based, Group Observation,
Welfare Issue	Freedom from Pain Injury and Disease: Lameness.
Method	The assessment of lame sheep is performed when sheep are walking and not running within the observation area. Ease of assessment is affected by flocking and flightiness of sheep. A 'sound' sheep is defined as one that bears weight evenly on all four feet. Lameness is defined as the observation of any or a combination of the following clinical signs: visible nodding of head in time with short stride, grazing on knees, uneven gait, arching of the back during locomotion, non-weight bearing on affected limb when standing. One or more limb may be affected by any one, or a combination of these signs.
Score	Number of individual animals in group identified as lame (percentage of the group).
Example	 Sheep classed as a lame

Indicator	Skin irritation
Type	Animal Based, Group Observation,
Welfare Issue	Freedom from Pain, Injury and Disease: Diseases associated with pruritus.
Method	Any one or a combination of the following signs: rubbing and scratching along walls, posts, fences or other objects, scratching, biting and nibbling at areas of the body. Undisturbed assessment is often important as pruritic sheep may cease the behaviour when disturbed, giving a risk of a false negative result.
Score	Number of animals in group showing skin irritation (percentage of the group).

Indicator	Rear cleanliness ('Dirty rear')
Type	Animal Based, Group Observation
Welfare Issue	Freedom from Pain, Injury and Disease: Indicator of gastrointestinal disturbance
Method	The rear area covers the perineum, around and under the tail, over the superficial aspect of the gluteal region (hindquarters) and the caudal aspect of the hind limb(s) to the region of the hock. Dirtiness is defined as contamination of the rear area with faecal matter, mud or soil in the form of discrete or solid plaques. Reasonably close observation is required as flocking behavior can impede observation. There may be breed variation in the level of dirtiness.
Score	Numbers of animals in group showing dirtiness of rear (percentage of the group).
Example	 Sheep with dirty rear.

Indicator	Belly cleanliness ('Dirty belly')
Type	Animal Based, Group Observation,
Welfare Issue	Freedom from Discomfort: Cleanliness of the indoor or outdoor environment
Method	Dirtiness of the ventral abdomen is defined as faecal, mud or soil contamination of this area, in the form of discrete or solid plaques of faecal matter or mud. This can be difficult to assess at group level in field situations due to ability to inspect ventral abdomen
Score	Numbers of animals in group with dirtiness of the ventral abdomen (percentage of group).
Example	 Solid plaques of dirt

Welfare indicators of adult sheep and growing lambs assessed by individual examination	
Protocol	This protocol is used to assess indicators of adult sheep (aged > 1 year) and growing lambs (aged > 6 weeks to < 1 year) detailed below. For assessment of individual sheep, animals are gathered into a holding pen. Two sheep at a time are then moved into an examination pen for assessment. The pen should have a level walking surface, be of sufficient size to walk the sheep around for lameness assessment, and be adequately lit to allow thorough examination. Lameness and demeanour are indicator tests that are assessed as the animal enters and moves round the examination pen. To assess other indicators, the individual sheep is restrained and held quietly by placing a hand under the jaw to allow the observer to facilitate the indicator assessment (eye condition, nasal discharge, tooth disease, coughing, ear lesions, body condition, fit fat thin, skin irritation, wool loss, skin lesions, injuries and wounds, tail length and cleanliness of rear).To assess foot conditions, cleanliness of ventral abdomen, and mastitis the sheep should be turned over so that it is sitting on its hind quarters and body weight is supported by the assessor. Alternatively a turning crate can be used.
Example	 Assessing lameness and demeanour  Body condition scoring  Assessing indicators in a turned sheep

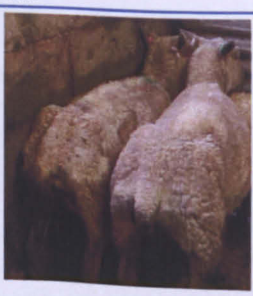
Indicator	Demeanour
Type	Animal Based
Welfare Issue	Freedom from Pain, Injury and Disease: Indicator of general health and welfare
Method	The general demeanour of individual sheep is assessed before and whilst the sheep is moved into the examination area. Individual demeanour is assessed using the following signs: a sheep that is not alert or responsive to the approach of observers, low head carriage, appearing dull or depressed, isolation from rest of flock.
Score	Bright, alert, responsive 1 = Dull, reduced responsiveness
Example	

Indicator	Lameness
Type	Animal Based,
Welfare Issue	Freedom from Pain, Injury and Disease: Lameness
Method	As the sheep enters the assessment area the gait is evaluated to identify any clinical signs of lameness. The observer walks the sheep around the pen to examine the gait in both directions. Lameness is defined as the observation of a single or any combination of the following clinical signs: visible nodding of head in time with short stride, uneven gait, arching of the back during locomotion, non-weight bearing on affected limb when standing. One or more limbs may be affected.
Score	0 = Sound 1 = Lame
Example	 Sheep classified as a lame Non-weight bearing right hind limb.

Indicator	Eye condition
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: ocular disease and abnormalities
Method	The sheep is examined for the presence of any eye abnormality, including, the following signs eye held partially or fully closed (blepharospasm), corneal opacity or ulceration, presence of an ocular discharge (muco-purulent, watery or haemorrhagic) and inversion of the lower eyelid (entropion)
Score	No eye abnormality 1 = Presence of eye abnormality
Example	 Ocular discharge  Entropion

Indicator	Nasal Discharge
Welfare Issue	Freedom from Pain Injury and Disease: Respiratory Disease
Method	The presence of a mucoid, purulent or bloody nasal discharge is recorded.
Score	0= no discharge 1= presence of discharge
Example	

Indicator	Body Condition Score
Type	Animal Based, Individual Level
Welfare Issue	Freedom From Hunger and Thirst: Indicator of nutritional status, chronic disease or dental disease
Method	The method of Russel (1984) is applied in which the lumbar vertebrae and transverse processes are manually palpated to assess the sharpness and prominence of the spinal process, the coverage of the loin (*longissimus dorsi* muscle) and degree of fat cover. Note that the body condition score of sheep is expected to change through the sheep production season.
Score	0 = Extreme emaciation, impossible to detect any muscle or fat between the skin and bone 1 = Spinous processes are prominent and sharp. Transverse processes are also sharp; fingers pass easily under the ends and it is possible to feel between each process. Eye muscles are shallow with no fat cover 2 = Spinous processes still feel prominent but smooth, individual process can be felt only as fine corrugations. Transverse processes are smooth and rounded and it is possible to pass the fingers under the ends with a little pressure. Eye muscle areas are of moderate depth but have little fat cover 3 = Spinous processes are detected only as small elevations; they are smooth and rounded and individual bones can be felt only with pressure. Transverse processes are smooth and well covered and firm pressure is required to feel over the ends. Eye muscle areas are of moderate depth with some fat cover 4 = Spinous processes can just be detected, with pressure as a hard line between the fat covered areas. Transverse processes cannot be felt. Eye muscles are full and have a thick covering of fat. 5 = Spinous processes cannot be detected even with firm pressure and there is depression between the layers of fat in the position where the spinous processes would normally be felt. Transverse processes cannot be detected. Eye muscle areas are very full with very thick fat cover. There may be large deposits of fat over the rump and tail.

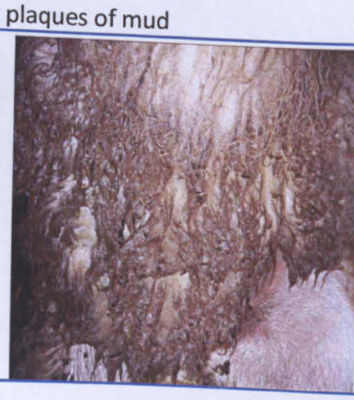
Indicator	Fit –Fat-Thin
Type	Animal based
Welfare Issue	Freedom From Hunger and Thirst Indicator of nutritional status, chronic disease or dental disease
Method	The assessment of body condition using the three-point fit – fat – thin system is based on the Russell (1984) body condition scoring method. The term 'fit; is equivalent to body condition scores between 2 and 4. 'Thin' is categorised as body condition below a score of 2, and 'Fat' is categorised as body condition above a score of 4. The lumbar vertebrae and transverse processes are manually palpated to assess the sharpness and prominence of the spinal process, the coverage of the loin (*longissimus dorsi* muscle) and degree of fat cover.
Score	Thin = Spinous processes can be felt as prominent and sharp. Transverse processes are also sharp; the fingers pass easily under the ends and it is possible to feel between each process. Eye muscle areas may be shallow with no fat cover Fit =Spinous processes are smooth: they may be detected as small elevations that are smooth and rounded, felt only with pressure so that individual process can be felt only as fine corrugations or they may be more prominent but still smooth and rounded on palpation. Transverse processes are also smooth and rounded; they may be well covered so that pressure is required to feel over the ends or it may be possible to pass the fingers under the ends with a little pressure. The eye muscle areas are of moderate depth but have little fat cover. Fat = Spinous processes may not be detected even with firm pressure. There can be a depression between the layers of fat in the position where the spinous processes would normally be felt. The ends of the transverse processes cannot be felt. Eye muscles are full and have a thick covering of fat. There may be large deposits of fat over the rump and tail
Example	'   Thin' equivalent to body condition score of 1 'Fat' sheep equivalent to body condition score of 5 

Indicator	Wool loss
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease Indicator of skin disease
Method	Assessment of wool cover is made in the shorn and unshorn sheep. Any size and area of wool loss is recorded. The interpretation of wool loss should account for natural shedding of wool, which can be seen in some breeds and at certain times of year (e.g.: spring). Wool loss may also result following rough handling.
Score	0 = No wool loss observed 1 = Area of wool loss observed

Indicator	Skin Irritation (Pruritus)
Type	Animal based
Welfare Issue	Freedom from Pain Injury and Disease: pruritic skin condition
Method	Assessed using the 'nibble test' which is performed by rubbing the fingertips on the skin of the sheep along the lumbar, flank and shoulder regions. A positive response is seen if the animal shows any of the following head and neck extension, nibbling and chewing movements after manual stimulation and is indicative of pruritus.
Score	0 = No response to nibble test 1 = Positive response to nibble test

Indicator	Skin Lesions
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease Parasitic and non parasitic skin disease
Method	An assessment of the integrity of the skin covering the body and head is made in the standing and turned sheep. Discoloration of the fleece, exudation, and presence of any odours can be used as an alert to the presence of a skin lesion. In the fully fleeced sheep the hands are run through the wool and areas of the wool are parted in order to examine the integrity of skin.
Score	0 = No skin lesions observed 1 = Presence of a small single lesion: area size of a 50 pence coin or less 2 = Presence of a single area: approx hand-sized (10x5cm) 3 = Presence of diffuse or generalized skin lesions (larger
Example	 Score 1  Score 2  Score 3

Indicator	Tail Length
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease Docking of tail too short
Method	With the individual sheep remaining in a standing position, the length of the tail is assessed according to the U.K legislation in which a short tail is defined as one that does not cover the anus in males or the vulva in females.
Score	0 = Appropriate tail length: tail covers anus in males or vulva in females 1 = Inappropriate tail length: tail does not cover anus in males or vulva in females
Example	 Score 1: short tail length Tail does not cover vulva

Indicator	Belly Cleanliness ('Dirty belly')
Type	Animal Based
Welfare Issue	Freedom from discomfort: Cleanliness of indoor or outdoor environment
Method	Whilst restrained in an upended position the cleanliness of the ventral abdomen of is assessed. Dirtiness is defined as dried on or freshly contaminated soiling of the area by mud
Score	Clean, splashing or occasional solid plaque of mud Dirty multiple solid or confluent plaques of mud
Example	 

Indicator	Mastitis
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: Mastitis
Method	In the ewe both mammary glands are palpated for areas of thickening and hard masses with or without signs of active inflammation: heat, swelling discharge, discomfort and engorgement of the glands and teats.
Score	0 = No evidence of mastitis in any gland 1 = One or both glands affected by mastitis
Example	 Acute gangrenous mastitis

Indicator	Myiasis
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: maggot Infestation (myiasis)
Method	The head, body and feet of are visually examined for the presence of maggots on any area and of any size of the sheep
Score	0= No maggots 1= Presence of maggots
Example	 Maggot infestation of a foot lesion

Welfare indicators for the assessment of young lambs	
Protocol	Each indicator test is assessed on individual young lambs - lambs aged 6 weeks and under. Depending on the management system, each individual lamb is assessed either by standing outside of a pen or by walking around field areas. Young lambs managed in individual lambing pens may need to be lifted out of the pen to assess some indicators. Minimal disturbance to ewe-lamb bonding is necessary and lambs managed outdoors may need to be assessed from a distance. In addition, these indicators are not performed on lambs less than 12 hours old. As some indicators require close observation it may be difficult to assess the lambs for those in all management situations. Approximately 5 minutes is required to assess an individual lamb using all the welfare indicators detailed below.

Indicator	Response to Stimulation
Type	Animal based
Welfare Issue	Freedom from Pain Injury and Disease and Hunger and Thirst: Indicator of General Health
Method	Stimulation was assessed as the responsiveness of the lamb to observer presence; in group situations or resting lambs this may require observer to whistle or wave or gently pat to assess awareness and response to stimulation. Ensure lambs are fully awake when performing this assessment
Score	0: Responsive to stimuli either observer presence or other sheep 1: Unresponsive to stimuli either observer presence or other sheep

Indicator	Demeanour	
Type	Animal Based	
Welfare Issue	Freedom from Pain Injury and Disease: Indicator of General Health	
Method	Demeanour was a subjective assessment of the general appearance and behaviour of the lamb. Note that inclement weather can affect lamb demeanour. Assessor needs to ensure that lamb is fully awake.	
Score	0: Bright, alert demeanour 1: Dull, depressed demeanour	
Example		Young lamb showing dull demeanour

Indicator	Standing Ability
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: Indicator of General Health and/or Lameness
Method	The standing ability of the lamb was scored by observing the movement of lambs around the pen area or field. A lamb that stands easily was observed to bear weight equally on all four legs without collapsing or swaying.
Score	0: Stands easily and steady on four legs, without difficulty 1: Weak, and unstable when standing or recumbent
Example	 Weak and unstable collapses when attempts to stand

Indicator	Lameness
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: Lameness
Method	Lameness included the observation of any clinical signs of lameness including a three-legged gait, holding a foot off the ground, an obvious head nod in time with movement of a limb or clinical signs of joint ill including large, swollen hot joints of any limb and/or a stiff, stilted gait.
Score	0: Sound; no lameness, stiffness or joint swelling 1: Lame; signs of lameness, stiffness or joint swelling
Example	 Lame lamb non-weight bearing on front limb.

Indicator	Posture
Type	Animal Based,
Welfare Issue	Freedom from Pain Injury and Disease: General Health and Lameness
Method	The posture of lambs was assessed as 'hunched' if lambs showed arching of the backbone with a tucked up abdomen. Note the posture of lambs can be affected by inclement weather conditions
Score	0: No evidence of a hunched or tucked up, or back arching posture 1: Hunched or tucked up, back arching posture
Example	 Tucked up abdomen

Indicator	Body Condition
Type	Animal Based
Welfare Issue	Freedom from Hunger and Thirst: Adequate nutrition
Method	Body condition is assessed by examining the cover of fat and muscle over the hip bone (ilial crest) and backbone (spinal vertebrae) of the lamb. In lambs with woolly fleeces or wrinkled skins, it was necessary to palpate the backbone and hip to assess the degree of cover. The hip bone) may be distinguished as a raised area but was covered should not be sharp or prominent in a lamb with appropriate body condition. Lambs with a thin body condition have a hip bone and backbone which appears sharp and prominent and there is little or no fat cover over the skeleton. This indicator requires close observation and possible palpation.
Score	0: Appropriate body condition. Skeleton is not prominent through the skin. 1: Thin body condition. The skeleton appears sharp and prominent, with little or no fat covering the spinal vertebrae

Indicator	Abdominal Fill
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease, Freedom from Hunger and Thirst: Inadequate nutrition, or abdominal distention
Method	The degree of abdominal fill was assessed visually and, where necessary, through gentle palpation of the abdomen. This is performed to determine the presence of a bloated or hollow abdomen. *Requires close observation, possibly palpation*
Score	0: Normal abdominal fill 1: Abdomen distended (ballooning of abdomen) or hollow appearance

Indicator	Salivation
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease: septicaemia or enteritis
Method	Lambs were examined for the presence of excess salivation around the lips of lambs ('watery mouth'). Slight frothing of milk around the lips, which may occur in lambs following sucking, was not scored as salivation. Close observation required
Score	0: No salivation 1: Presence of salivation around mouth

Indicator	Eye Condition
Type	Animal Based, Individual Lamb < 12 weeks
Welfare Issue	Freedom from Pain Injury and Disease: Ocular Disease
Method	Lambs were assessed for signs of eye disease. The presence of an eye condition was assessed if any one or a combination of the following signs was observed: blepharospasm (holding the eye lids tightly closed), presence of an ocular discharge (purulent, mucoid or haemorrhagic), corneal opacity, and inversion of the lower eyelid (entropion). This indicator requires close observation
Score	0: No signs of eye disease 1: Presence of eye disease
Example	 Young lamb with entropion and ocular discharge

Resource-based indicators of sheep welfare

Indicator	Provision of Clean Accessible Water
Type	Resource
Welfare Issue	Freedom from Hunger and Thirst: Provision of clean and accessible water
Method	Assess the provision of clean and accessible source of drinking water. A sample of water is collected in a clear container and visually examined. The water source should be clean – free from contamination by decayed food material, faeces or excessive mud and at a height that is accessible for the production stage. The provision of water for housed animals could include water troughs or buckets or bowls, in fields this maybe in troughs or a natural source, such as a stream, river or pond.
Score	0: Provision of clean, accessible water 1: No provision of water, or provision of water that is unclean or inaccessible
Example	   Clean · Clean (straw in drinker) · Dirty (contamination)

Indicator	Presence of Lie Back Area
Type	Resource
Welfare Issue	Freedom from Discomfort: Physical comfort for grazing sheep
Method	Assess provision of a well drained lying area. Applicable to animals that are housed and those in fields and particularly those grazing root crops.
Score	0: Well drained lying area 1: No well drained lying area

Indicator	Provision of shelter and shade
Type	Resource
Welfare Issue	Freedom from Discomfort: Protection from inclement weather conditions
Method	Assess provision of shelter for animals in fields. This could be a hedge, wall or trees that animals are able to use to protect themselves from adverse weather. Requirement for shelter vary depending on breed, age of animal and management system
Score	0: Shelter available 1: No shelter available
Example	 Shelter provided by hedgerows and trees

Indicator	Stocking Rate of Housed Sheep
Type	Resource
Welfare Issue	Freedom from discomfort and Freedom to express normal behaviour
Method	Measure pen area and count number sheep in pen to calculate the stocking rate. The required rates depend on the age and stage of production and are stated in the Defra codes of recommendations for the welfare of sheep: Lowland ewes (60-90kg) pregnant : 1.2-1.4m^2 Lowland ewes after lambing with lambs at foot up to 6 weeks age: 2.0-2.2m^2 Hill ewes (45-65kg): 1.0-1.2m^2 Hill ewes after lambing with lambs at foot up to 6 weeks age : 1.8-2.0m^2 Lambs up to 12 weeks old: 0.5-0.6m^2 Lambs and sheep 12 weeks to 12 months old : 0.75-0.9m^2 Rams: 1.5-2.0m^2 These space allowance can be reduced by 10% for winter shorn sheep
Score	0: stocking rate acceptable 1: housing over stocked

Indicator	Adequate Lighting
Type	Resource
Welfare Issue	Freedom from Pain Injury Disease
Method	This indicator only applicable to housed sheep when a subjective assessment of the adequacy of lighting required to allow inspection of individual sheep is made.
Score	0: lighting adequate to inspect sheep 1: lighting inadequate to inspect sheep

Indicator	Maintenance of Equipment and Facilities
Type	Resource
Welfare Issue	Freedom from Pain Injury and Disease
Method	Walk round farm buildings, yards and fields assess whether equipment facilities are maintained in such a way as to avoid injury to sheep. For example no sharp edges or protrusions
Score	0= no evidence that facilities or equipment could injure animals 1= evidence that facilities or equipment could injure animals

Indicator	Bedding quality and quantity
Type	Resource
Welfare Issue	Freedom from Discomfort
Method	When sheep are housed a subjective assessment is made of the quantity and hygiene of the bedding
Score	0= bedding adequate and clean and dry 1= bedding inadequate, dirty or wet
Example	 Clean, dry bedding  Dirty, wet bedding

Indicator	Presence of Sheep Carcasses Amongst Live Sheep
Type	Animal Based
Welfare Issue	Freedom from Pain Injury and Disease
Method	Walk round farm buildings, yards and fields, record presence of sheep carcasses amongst live sheep
Score	0= no carcasses present 1= 1 or more carcasses present

Management-based indicators of sheep welfare

Indicator	Presence of Medicines Records
Type	Management
Welfare Issue	Freedom from Pain Injury and Disease
Method	Inspect medicines records. Ensure correct information recorded i.e., animals treated, date, drug, dose, batch numbers, meat or milk withdrawal times.
Score	0= Medicines records present and correctly recorded 1= medicines records absent or incorrectly recorded

Indicator	Presence of Mortality Records
Type	Management
Welfare Issue	Freedom from pain injury disease, or freedom from hunger and thirst
Method	Inspect mortality records to ensure number of deaths is recorded
Score	0= mortality records present and correct 1= mortality records not present or incorrect

Indicator	Tail docking and castration policy
Type	Management
Welfare Issue	Freedom from Pain Injury Disease: Practices used to perform mutilations
Method	Identify management procedures for tail docking and castration from discussion with farmer, examination of medicines records.
Score	0= castration and tail docking carried out within legal requirements 1= castration and/ or tail docking not carried out within legal requirements

Suggested sample size for animal-based indicator assessments

The sample size can be based on the estimated prevalence of welfare condition. Where no estimate can be made, the 50 % prevalence value can be used to provide an accuracy of 0.05.

Estimated Prevalence (%)	Flock size					
	50	100	200	500	1000	2000
	Estimated sample size					
10	37	58	82	108	121	129
20	42	71	110	165	197	219
30	43	76	123	196	244	278
40	44	79	130	212	269	311
50	44	79	132	217	278	322
60	44	79	130	212	269	311
70	43	76	123	196	244	278
80	42	71	110	165	197	219
90	37	58	82	108	121	129