UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# WELFARE ASSESSMENT IN PORTUGUESE DAIRY GOAT FARMS: ON-FARM OVERALL FEASIBILITY OF AN INTERNATIONAL PROTOTYPE

EDNA MARIA DE ANDRADE VIEIRA CAN

CONSTITUIÇÃO DO JÚRI
Doutor Miguel Luís Mendes Saraiva Lima
Doutora Ilda Maria Neto Gomes Rosa
Doutor George Thomas Stilwell

ORIENTADOR
Doutor George Thomas Stilwell
CO-ORIENTADOR
Dra. Ana Catarina Lopes Vieira Godinho de Matos

2015

LISBOA

UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# WELFARE ASSESSMENT IN PORTUGUESE DAIRY GOAT FARMS: ON-FARM OVERALL FEASIBILITY OF AN INTERNATIONAL PROTOTYPE

EDNA MARIA DE ANDRADE VIEIRA CAN

DISSERTAÇÃO DE MESTRADO INTEGRADO EM MEDICINA VETERINÁRIA

CONSTITUIÇÃO DO JÚRI

Doutor Miguel Luís Mendes Saraiva Lima

Doutora Ilda Maria Neto Gomes Rosa

Doutor George Thomas Stilwell

ORIENTADOR

Doutor George Thomas Stilwell

CO-ORIENTADOR

Dra. Ana Catarina Lopes Vieira Godinho de Matos

2015

LISBOA

À minha Mãe

**Acknowledgments**

I would like to express my acknowledgements to the AWIN (Animal Welfare Indicators) integrated 7FP project, funded by the European Commission, as carrying out the present study was only possible with the project support.

I would like to thank my supervisor Professor George Stilwell for providing me the opportunity, scientific advice and guidance throughout the present study.

I would like to express my deepest gratitude to my co-supervisor Dr. Ana Vieira for her excellent guidance, permanent help, caring and patience, and for providing me with an excellent atmosphere for doing this study.

Great thanks go to Dr. Telmo Pina Nunes for having the time and interest to review this study, providing advices, precious tips and encouragement at the final stage of this work.

Special thanks go to Dr. Inês Ajuda and Dr. Inês Barão for all the companionship, encouragement and support through all stages of this research.

My gratitude goes to Mónica Battini for helping me so willingly with the Qualitative Behaviour Assessment (QBA) and for all the suggestions made.

Many thanks go to *Direcção-Geral de Alimentação e Veterinária* (DGAV), for providing the national data.

I would like to thank and recognize the time and open doors of the farmers who participated in the field study, as without them this work would not have been possible.

Special thanks go to my friends for the moral support and unconditional belief in me. Thank you all for your affection, energy and motivation.

Finally, I would like to thank my Mother and all my family. They were always there cheering me up, inspiring me to follow my dreams and helping me through the good times and bad, being vital for finishing my academic degree.

I am indebted to all of you who have discussed this study with me at one time or another. Thank you.

**Welfare assessment in Portuguese dairy goat farms: on-farm overall feasibility of an international prototype**

Edna Can

**Abstract**

This study describes and assesses the application of the on-farm welfare assessment prototype for dairy goats (*Capra hircus*) developed by the AWIN project. Thirty Portuguese dairy goat farms were assessed from January to March 2014. Pen-level observations were carried out on 2715 animals and detailed individual observations were performed on 1172 of these animals. The main areas of concern were associated with claw overgrowth, queuing at feeding, overweight animals, poor hair coat condition and improper disbudding. The results obtained show that these welfare issues are related to farm sizes, with larger farms heading higher concerns. Furthermore, the reliability and feasibility of the animal-based indicators were tested. Overall, moderate to high levels of agreement between observers were identified, with the exception of Qualitative Behaviour Assessment (QBA). From all stages of the prototype 'Queuing' and 'Clinical scoring' were the most time consuming, with the mean time required to apply the prototype being longer in large farms. In conclusion, the protocol has shown the potential not only for legislative and regulatory purposes, but also as a certification, advisory/management and research tool, probably following a two-step approach.

**Key words:** animal welfare, AWIN, dairy goats, animal-based indicators

Faculdade de Medicina Veterinária - Universidade de Lisboa
Avenida da Universidade Técnica
1300-477 Lisboa
Portugal

E-mail: ednacan@gmail.com

**Avaliação de Bem-Estar Animal em Explorações Portuguesas de Caprinos de Leite: exequibilidade de um protótipo internacional de avaliação de bem-estar**

Edna Can

**Resumo**

Este estudo tem como objectivo descrever e avaliar a aplicação do protótipo de avaliação de bem-estar animal, desenvolvido pelo projecto AWIN em explorações de caprinos de leite em regime intensivo. Trinta explorações portuguesas foram avaliadas de Janeiro a Março de 2014, tendo sido efectuadas avaliações no parque de 2715 caprinos de leite e observações individuais a 1172 desses animais. Os principais problemas identificados nas explorações encontram-se associados a um crescimento excessivo das unhas, filas na manjedoura, animais com condição corporal elevada, com má condição do pêlo e alvo de uma má descorna. Os resultados obtidos indicam que estes problemas de bem-estar animal encontram-se relacionados com a dimensão das explorações, em que as de maior dimensão demonstram prevalências mais elevadas. A repetibilidade e exequibilidade dos indicadores que compõem o protótipo foram também testadas. De uma forma global, os níveis de repetibilidade entre os observadores, são moderados a elevados, com excepção da Avaliação Qualitativa do Comportamento (AQC). Das várias etapas que constituem o protótipo, as que envolvem mais tempo são a 'Presença de filas' e a 'Avaliação clínica', com o tempo médio necessário para a sua aplicação atingindo valores mais elevados em explorações maiores. A realização deste estudo permite concluir que o protocolo final de avaliação de bem-estar animal em caprinos de leite terá potencial não só como uma base legislativa e regulamentar, mas igualmente como uma ferramenta de certificação, consultoria/gestão e de pesquisa, seguindo uma estratégia baseada em dois níveis de avaliação.

**Palavras-chave:** bem-estar animal, AWIN, caprinos de leite, indicadores baseados no animal

Faculdade de Medicina Veterinária - Universidade de Lisboa
Avenida da Universidade Técnica
1300-477 Lisboa
Portugal
E-mail: ednacan@gmail.com

**Table of contents**

**Figures list**

**Tables list**

**Acronyms**

| | |
|---|---|
| ACTH | Adrenocorticotropic hormone |
| AD | Avoidance distance |
| ANI | Animal Needs Index |
| AWIN | Animal Welfare Indicators |
| BA | Body abscesses |
| BC | Body condition |
| BCS | Body Condition Score |
| BLAS | Body lesions and swellings |
| CI | Confidence interval |
| CAE | Caprine arthritis and encephalitis |
| CLA | Caseous lymphadenitis |
| CO | Claw overgrowth |
| COT | Consistency of indicators over time |
| D | Diarrhoea |
| DGAV | Direcção Geral de Alimentação e Veterinária |
| DGSANCO | General Directorate for the Health and Consumer Protection |
| DL | Decreto-lei |
| EC | European Commission |
| EFSA | European Food Safety Authority |
| EU | European Union |
| FAO | Food and Agriculture Organization of the United Nations |
| FAOSTAT | Food and Agriculture Organization of the United Nations Statistics Division |
| FAWC | Farm Animal Welfare Council |
| HA | Head abscesses |
| HAR | Human-animal relationship |
| HC | Hair coat condition |
| HLAS | Head lesions and swellings |
| HQA | Hindquarters abscesses |
| HQC | Cleanliness - hindquarters |
| HQLAS | Hindquarters lesions and swellings |
| ICC | Intra-class correlation |
| ID | Improper disbudding |
| INE | Instituto Nacional de Estatística |
| INRA | Institut National de la Recherche Agronomique |

| | |
|---|---|
| IOR | Inter-observer reliability |
| KF | Kneeling at feeding |
| KP | Kneeling in pen |
| KLASS | Knee lesions and swellings score |
| LAS | Lesions and swellings |
| LLC | Cleanliness - lower legs |
| LLLAS | Lower legs lesions and swellings |
| NA | Neck abscesses |
| ND | Nasal discharge |
| NLAS | Neck lesions and swellings |
| NS | Non–significant |
| O | Oblivious |
| OD | Ocular discharge |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PS | Panting score |
| QBA | Qualitative Behaviour Assessment |
| QF | Queuing at feeding |
| QD | Queuing at drinking |
| RSPCA | Royal Society for the Prevention of Cruelty to Animals |
| SD | Standard deviation |
| SL | Severe lameness |
| SNIRA | Sistema Nacional de Informação e Registo Animal |
| SPSS | Statistical Package for the Social Sciences |
| SS | Shivering score |
| UA | Udder abscesses |
| UC | Cleanliness - udder |
| UAS | Udder asymmetry |
| VAS | Visual analogue scale |
| VD | Vulvar discharge |
| VT | Very thin |
| VF | Very fat |
| WP | Work Package |

**Internship report**

The present study was performed during the 6th year of Integrated Master in Veterinary Medicine (*Faculdade Medicina Veterinária* – University of Lisbon), assigned to curricular practical training, with guidance of Professor George Stilwell. The official curricular internship had a total duration of 3 months, beginning on February 2014 and lasting until the end of May, 2014, during which I engaged in Professor George Stilwell's Large Animal Clinics practical classes and clinical cases discussions (5th year). Clinical activities included, among others, general physical examinations, special examination of the respiratory and gastro-intestinal tracts, rectal palpations, diagnosis of udder disorders, disbudding of calves, functional and curative hoof trimming, treatment of downer cows, surgical resolution of left abomasal displacement, vaccinations, injections (subcutaneous, intramuscular, intravascular, epidural).

Along with the practical classes, the field study for the present work was conducted. Hence, visits to 30 Portuguese intensive dairy goat farms were completed in order to test an on-farm welfare assessment prototype for dairy goats developed by the AWIN project (*Faculdade Medicina Veterinária* – University of Lisbon and *Università degli Studi di Milano*, Italy), during the period of January to March 2014. I participated in the research project by joining the application and testing of the prototype in these 30 farms, and by inputting and statistically analysing the collected data. A second assessment was conducted in July 2014 during which 10 farms were revisited.

In addition, during this period data were gathered in a dairy farm in order to perform a follow-up study on the reproductive and productive performance, and behavioural characteristics of purebred Holstein-Friesian, compared with their crossbreds with Montbeliarde and Swedish Red.

## Introduction

The public awareness of what happens to farm animals in intensive animal production has grown (Appleby, 1999; Webster, 2005; Miele, Veissier, Evans & Botreau, 2011). Consumers now expect animal-based products, in particular food, to be produced with greater consideration for the welfare of the animals (Blokhuis, Jones, Geers, Miele & Veissier, 2003; Rushen, Butterworth & Swanson, 2011), resulting in an increasing requirement for scientifically valid and feasible welfare assessment systems (Waiblinger, Knierim & Winckler, 2001; Ofner, Amon, Amon, Lins & Boxberger, 2002; Main, 2009). In response to this demand, the scientific community encouraged the development of welfare indicators in order to produce more accurate outcomes, with several on-farm welfare assessment protocols being established in Europe (Johnsen, Johannesson, & Sandøe, 2001; Bracke, Spruijt, Metz & Schouten, 2002) and elsewhere (e.g., United States, New Zealand).

Welfare assessment requires a multidimensional approach (Mason & Mendl, 1993; Fraser, 1995), implying that all the component dimensions are most adequately assessed by particular criteria (Fraser, 2003; Botreau, Capdeville, Perny & Veissier, 2008).

In 2008, the EU Welfare Quality® project expanded the 'Five Freedoms' (Brambell Committee, 1965) framework on animal welfare definition and assessment, establishing a holistic concept that covers the different domains of animal welfare (Blokhuis et al., 2013). Four welfare principles divided into twelve criteria (Blokhuis, Veissier, Miele & Jones, 2010; Rushen et al., 2011) were described, with each being formulated to communicate an important welfare issue and branched into different criteria (Welfare Quality, 2009).

Two categories of indicators can be used to assess animal welfare at the farm level: animal and resource-based (Johnsen et al., 2001). Traditionally, on-farm welfare assessment focused on the evaluation of resources provided to the animal (e.g. Bartussek, 1999, Bracke et al., 2002). However, providing good management and environmental resources does not automatically reflect higher standards of welfare. Capdeville and Veissier (2001) stress that an animal-based strategy seems more appropriate for measuring the actual welfare state of the animals, which represents a considerable change in perspective. The Welfare Quality® project followed this line of thought (Blokhuis et al., 2010), assessing welfare from the animals' "point of view", hence reflecting a more direct assessment of their welfare (Whay, Main, Green & Webster, 2003a). These animal-based indicators are then combined in protocols to provide an assessment of the welfare of the animals (European Food Safety Authority [EFSA]; 2012), and must fulfil the requirements of validity, reliability and feasibility to be used at farm level (Waiblinger et al., 2001; Winckler, 2006). In spite of how challenging it is to select and establish valid, reliable and simultaneously feasible indicators for on-farm welfare assessment systems (Winckler, 2006), in all dimensions of animal welfare the accurate assessment of its main aspects is essential (Spoolder, De Rosa, Hörning, Waiblinger & Wemelsfelder, 2003; Meagher, 2009).

A schematic presentation of the thesis structure (Figure 1), illustrating the sequential workflow of this study, can help connect all the research parts.

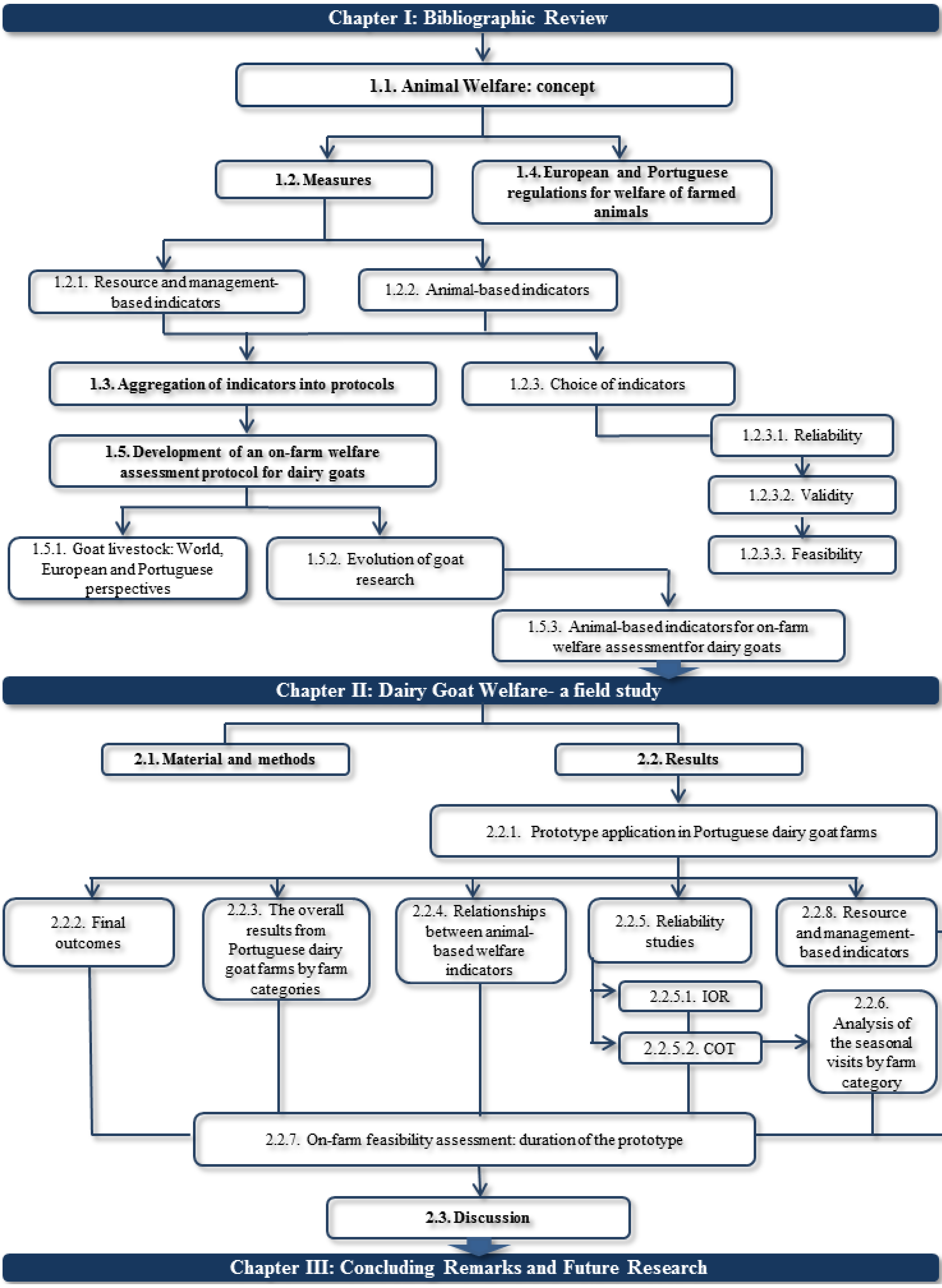


Figure 1- Thesis structure: objectives and organization.

The aims of this research are linked to specific objectives associated with the thesis structure. The main aim is to describe and assess the application of the on-farm welfare assessment prototype for dairy goats developed by the AWIN project, in Portugal. This prototype includes animal-based indicators established following the Welfare Quality® strategy, which provides a concise but complete framework for assessing the welfare of animals regarding its state of nutrition, comfort, health, and behaviour. Another aim of the present study is to have a first insight into some of the main problems impairing the welfare of intensively kept dairy goats in Portugal, hence gathering information on the potential animal welfare issues a future European protocol may find in Portuguese farms.

# CHAPTER I – Bibliographic review

## 1.1. Animal welfare: concept

Since the 1960s there has been an increasing concern on the effects of intensive production on animal welfare, which led to a progressive advance on scientific research in the area (Millman, Duncan, Stauffacher & Stookey, 2004; Lassen, Sandøe & Forkman, 2006; Carenzi & Verga, 2009). With the publication of *Animal Machines: The New Farming Industry* by Ruth Harrison (1964) and the Brambell Committee Report (1965) on the welfare of farm animals, issued by the British government, animal welfare science as a recognised discipline emerged (Duncan, 2006; Blokhuis, Miele, Veissier, & Jones, 2013; Mellor & Webster, 2014).

As highlighted by Broom (2007), an agreement should be reached on the concept of animal welfare for use in accurate scientific measurements, in legal documents and in public statements or discussions.

Three different approaches on animal welfare have developed: one centred on the affective state, one on biological function and one that relies on natural living as the fundamental measure (Fraser, 2003). Although these three views comprise quite different areas of importance in assessing animal welfare, they constitute three complementary starting points for identifying and solving animal welfare problems, often leading to similar conclusions (Fraser, Weary, Pajor & Milligan, 1997; Fraser, 2003). Nevertheless, there is a scientific debate on how to define animal welfare. Whilst numerous definitions have been suggested, a few can be highlighted as being widely cited and generally approved. One of the most broadly accepted definitions of animal welfare are the 'Five Freedoms', delineated by the Brambell Committee in the United Kingdom in 1965, and which form the basic philosophy of the Farm Animal Welfare Council (FAWC). They consist of (Farm Animal Welfare Council [FAWC], 2009):

1) Freedom from hunger and thirst, by ready access to water and a diet to maintain health and vigour;
2) Freedom from discomfort, by providing an appropriate environment;
3) Freedom from pain, injury and disease, by prevention or rapid diagnosis and treatment;
4) Freedom to express normal behaviour, by providing sufficient space, proper facilities and appropriate company of the animals' own kind;
5) Freedom from fear and distress, by ensuring conditions and treatment that avoid mental suffering.

The ‘Five Freedoms’ establish the aspects that define the animals’ own perception of their welfare state and express the necessary requirements to support that state, taking into account both physical fitness and mental suffering (Webster, 2001).

As reported by Botreau et al. (2008) and Blokhuis et al. (2013) the ‘Five Freedoms’ (FAWC, 2009) define conceptual states to aim for, overlapping each other, and to be useful in practice these definitions need to be converted into more operational descriptions with assessable indicators.

A list of mutually exclusive dimensions that could be evaluated, starting from the concept of the animals’ ‘Five Freedoms’ (FAWC, 2009), was constructed by the multidisciplinary EU funded research project Welfare Quality®, which defined four welfare principles, associated to twelve criteria (Blokhuis, Veissier, Miele, & Jones, 2010; Rushen et al., 2011). Each principle is phrased in order to report a key welfare question and divided into different criteria, with each welfare criterion symbolizing a specific area of welfare and specifying an area of concern. Accordingly, criteria are independent of each other and form ‘an exhaustive, but minimal list’ (Botreau et al., 2007c; Welfare Quality, 2009).

### 1.2. Measures of animal welfare

Resulting from these considerations, it is apparent that animal welfare is a multidimensional concept covering physical, physiological and psychological components, and an overall welfare assessment must address all these components (Mason & Mendl, 1993; Fraser, 1995; Botreau, Veissier, Butterworth, Bracke & Keeling, 2007a; Miele et al., 2011; Rushen et al., 2011; Blokhuis et al., 2013) corresponding to a multicriteria evaluation problem (Botreau et al. 2007c; Carenzi & Verga, 2009). As animal welfare is a multidimensional concept its assessment includes environmental-based (e.g. space allowance, type of floor, climate control systems, etc.) and animal-based indicators (e.g. injuries, fear, lameness, etc.; Johnsen et al., 2001; Smulders, Verbeke, Mormède & Geers, 2006). It is usually recognized that both categories – environmental and animal-based – are significant aspects of animal welfare, and that the most valid assessment of it is achieved when these are used in association (Johnsen et al., 2001; Waiblinger et al., 2001; Botreau et al., 2007a).

#### 1.2.1. Resource and management-based indicators

In the first phases of the development of animal welfare assessment systems, many resource-based indicators were used to estimate the welfare status of animals at farm level (Amon, Amon, Ofner & Boxberger, 2001), as they are less subjective, often easier to audit (requiring relatively little training of the assessor), very convenient (one short visit of a farm is usually sufficient for assessing all indicators) and highly repeatable, frequently having high inter and intra-observer repeatability (Capdeville & Vessier, 2001; Johnsen et al., 2001; Whay, 2007; Blokhuis et al., 2013). Resource-based observations focus on what has been given to the animal, such as shelter, length of stalls, comfort, space allowance, access to pasture, nutrition, feeding and drinking facilities and companionship (Hörning, 2001; Johnsen et al., 2001). Whereas the resources delivered define the physical situation for the animal, management-based indicators are also very significant. Main management choices concerning the animal's life consist on how and when they are fed, moved and mixed with other animals and in routine practices like beak trimming, tail docking or dehorning (Blokhuis et al., 2013). This consists in a more indirect aspect of animal welfare, centred on the idea that if we deliver the right environment and care for the animal then it will have a high standard of welfare. They are also usually the type of indicators used for legislation, and when well-chosen resource-based indicators should prevent welfare problems from taking place, allowing the determination the risk factors or hazards that can affect animal welfare (Rushen et al., 2011). However, one should bear in mind that a farm's environment and management routines do not necessarily define animal welfare, and that farms with equivalent production systems may present an enormous variation in animal welfare (Sandøe, Munksgaard, Bådsgård & Jensen, 1997).

### 1.2.2. Animal-based indicators

Assessments centred on resource-based indicators can fail to completely answer questions about the actual state of welfare of the animals on a particular farm (Rushen et al., 2011). Consequently, in order to more precisely reproduce the animal's own welfare perception, regardless of how they are housed or managed, it was suggested that animal-based indicators of welfare assessment were more suitable (Bartussek, 1999; Main, Whay, Green & Webster, 2003a; Whay et al., 2003a; Rushen et al., 2011; Webster, 2009; Rushen et al., 2011) and were established for several species (Whay et al., 2003a; Whay et al., 2003b; Anzuino, Bell, Bazeley & Nicol, 2010). As measurements of animal-based indicators assess the animals' responses to particular environments, it appears to be a consensus in the literature on the notion that animal welfare should be then recorded by an association of indicators, that fall within the categories of behaviour, health, performance and physiology (Fraser, 1995; Johnsen et al., 2001; Smulders, et al., 2006; Whay, 2007; Ellegaard et al., 2010; EFSA, 2012). As a result, several welfare assessment protocols containing animal and resource-based assessment indicators have been created for many farm species (Whay et al., 2003a; Knierim & Winckler, 2009; Anzuino et al., 2010).

Behaviour varies in reaction to several environmental difficulties (Broom, 1991). Using behaviour in the assessment of animal welfare has major advantages, as it is non-invasive and in many cases it is also non-intrusive (Dawkins, 2004), with changes in behaviour due to disease being often used by veterinarians on diagnosis (Broom, 1987; Fraser & Broom, 1990; reviewed by Broom, 2006). Abnormal behaviours such as stereotypies, self-mutilation, tail-biting in pigs, feather-pecking in hens or extremely aggressive behaviour can also show that the animal's welfare is poor (Broom, 2007). Viñuela-Fernández, Jones, Welsh and Fleetwood-Walker (2007) stated that behaviour is the most frequently used measure to record pain at farm level.

Animal health is the basis of all good welfare (Fraser & Broom, 1990; Appleby & Hughes, 1997; Dawkins, 2001; reviewed by Dawkins, 2004). According to Broom (2006) health concerns the state of the body and brain regarding the responses to pathogens, parasites, tissue damage or physiological disorder. All of this range of responses implicate pathology, thus the health of an animal is its state regarding its attempt to deal with disease. Considering the welfare of an individual as its state with regard to its efforts to cope with the environment (Broom, 1986) and that disease is one of the effects of environment, then health constitutes a part of welfare.

However, poor welfare does not always implicate poor health (Broom, 2007). Coping involves having control of mental and bodily stability and prolonged inefficacy to cope originates failure to grow, reproduce or death which implies fitness reduction and stress (Broom & Johnson, 1993; Broom, 2007), consequently affecting the animal and its performance. While this last one against goals is not always a reflexion of welfare, it can, nevertheless, provide an overview of welfare management of the considered farm when in association with other indicators. Thus, variation in production and reproductive performances can be suggestive of potential welfare problems (Colditz, Ferguson, Collins, Matthews & Hemsworth, 2014). Nonetheless, Dawkins (1980) defends that production indicators, such as growth, morbidity or mortality, are not relevant to animal welfare, as they are measured at farm level and the production of a farm can be satisfactory even if some animals are in poor condition (reviewed by Botreau et al., 2007c). The Scientific Veterinary Committee (1997) considered that health problems remain some of the main welfare issues for farm animals. Several health problems, such as lameness in dairy cattle, also have major economic impact. Hence, a larger use of health indicators will help to enhance the ability to record animal welfare at herd level (Rushen, 2003).

Some physiological measurements, as increased heart-rate, adrenal activity, adrenal activity following adrenocorticotropic hormone (ACTH) challenge, or reduced immunological response following a challenge, can demonstrate that welfare of these animals is poorer than that of animals which do not suffer such changes, acting as warning signs of poor welfare (Broom, 2007). Still, these results must be carefully interpreted (Moberg, 1985; reviewed by Broom, 2007), since corticosteroids levels can increase in situations other than stress (Toates, 1995; reviewed by Dawkins, 2004). However, physiological indicators record at farm level is frequently limited for feasibility reasons, due to their cost and the need for the animals to be handled (Bartussek, 2001; Capdeville & Vessier, 2001; Johnsen et al., 2001; Smulders et al., 2006; Winckler, 2006).

Animal-based indicators are considered more direct indicators of welfare than their environmental complements, (Johnsen et al., 2001) allowing the comparison of the welfare of animals kept in different types of farming systems (Blokhuis et al., 2013). They have the benefit of being practical and focused, where important data for measuring animal health indicators are frequently available on databases of health records based on registrations completed by, for instance, the assigned veterinarian (Johnsen et al., 2001). Nevertheless, there is a great level of subjectivity involved in this type of assessment and the analysis of the impact of the outcomes to the animals themselves constitutes a constant challenge (Whay, 2007). Although there are advantages in increasing the use of animal-based indicators on welfare assessment, their practical use within existing assurance schemes, as mentioned, is problematic in numerous ways (Waiblinger et al., 2001; Rushen & De Passillé, 2009; Rushen et al., 2011), and there is always pressure to minimize the costs by limiting the duration of the visit, the frequency of visits to farms, or the number of animals observed (Rushen et al., 2011). Nonetheless, once a valid animal-based system is established, it can be applied to determine on-farm risk factors regarding the provision of resources, management, stockmanship and other farm aspects (Smulders et al., 2006).

### 1.2.3. Choice of indicators

When considering a multidimensional evaluation model a first step consists in defining a direct set of criteria, i.e., indicators with an established objective that can be used to assess the welfare of an animal (Botreau et al., 2007c). Waiblinger et al. (2001) specified that for an assessment tool to be effectively used on different farms it must include the following characteristics: contain indicators that are reliable and valid, be easily employed by trained people, need limited time so that repeated assessments on several holdings are conceivable, expose the reasons of reduced welfare and thus, possible improvements of the husbandry and management system. In all facets of animal welfare the accurate and valid measurement of the key aspects is crucial, since welfare includes several indicators for many of which the measurement is subjective or made at an ordinal level, it is not integrated over species or management practices, and there is a lack of a ‘golden standard’ (Scott, Nolan & Fitzpatrick, 2001; Spoolder et al., 2003; Meagher, 2009).

#### 1.2.3.1. Reliability

Reliability, which refers to the degree to which measures are free from random errors (Martin & Bateson, 2007), is an essential requisite of scientific measurement and is determined by precision, sensitivity, resolution and consistency. Reliability (consistency) is usually tested before testing validity, since an instrument cannot be used if it is unreliable (Meagher, 2009). Reliability of an indicator shows the relative similarity of measurements accomplished on one animal in several occasions, referring to the repeatability of the indicator (Amon et al., 2001; De Passillé & Rushen, 2005). The essential aspects of reliability are inter-observer and intra-observer reliability, and also test–retest reliability. Inter-observer reliability measures the agreement between different observers, intra-observer reliability (or observer consistency) evaluates the agreement between the same observer on several occasions and test-retest reliability refers to the agreement between observations performed on the same individual on at least two different occasions (Scott et al., 2001), being especially important when it is to be used for certification purposes in terms of welfare labelling (Blokhuis et al., 2013). Reliable assessments are necessary to show the objectivity and robustness of indicators applied by different observers, to assess the welfare of animals kept under different farming systems (Phythian et al., 2012). In on-farm welfare evaluation, each farm will preferably be visited only once, and approximately the same number of observers should be able to be used (De Passillé & Rushen, 2005). Thus, inter-observer reliability testing may be used to investigate either the suggested indicators or possible observers for their appropriateness (Mullan, Edwards, Butterworth, Whay & Main, 2011). Additionally, high levels of observer reliability offer confirmation of the validity of the aspects measured (Hewetson, Christley, Hunt & Voute, 2006; Meagher, 2009) and ensure the objectivity of a welfare outcome assessment, being important that reliability evaluations are directed with representative samples and observers, under proper research settings and adopting precise assessment systems. If the repeatability is low, the variable has to be evaluated many times in order to achieve a reliable result, therefore being inappropriate for animal welfare monitoring at farm level (De Rosa, Grasso, Pacelli, Napolitano & Winckler, 2009).

#### 1.2.3.2. Validity

After the reliability of a test has been determined, the next phase involves evaluating its validity. This one deals with the relation between a variable, as a measure of behaviour, and what it is supposed to measure or predict and is determined by accuracy, specificity and scientific validity (Martin & Bateson, 2007). As reliability, validity can be divided into three categories: content, construct and criterion (Cronbach & Meehl, 1955).

In content validity the test items are a sample of a universe that the investigator is studying and it is supposed to be established deductively, by determining a universe of items and sampling repeatedly within this universe in order to create the test for the phenomenon to be measured (Cronbach & Meehl, 1955). There are no direct statistical tests of content validity, however it can be evaluated by having experts rate the suitability of the items involved and testing their agreement (Lawshe, 1975; reviewed by Meagher, 2009). Face validity can be considered a weak form of content validity (Scott & Mazhindu, 2014), consisting in a subjective judgment that a measure is believed to be valid, as judged by one or more experts (Scott et al., 2001). Accordingly, content and face validity are subjectively evaluated from expert judgment (Scott, Fitzpatrick, Nolan, Reed & Wiseman, 2003).

Construct validity is assessed by constructing hypotheses based on the relationship between welfare and other variables (Scott et al., 2001). The process of construct validation implicates defining how well a measure evaluates a construct as this one has been conceptualized (Jones & Gosling, 2005). This conceptualization involves stipulating the aspects to which the construct should be related or unrelated. These two components are identified as convergent- degree to which a measure is correlated with others to which it is theoretically predicted to be related with (Meagher, 2009) - and discriminant validity– when a measure is unrelated to other measures that are conceptually unrelated (Campbell & Fiske, 1959).

Criterion validity can be defined as the accuracy in predicting a score on, if available, a 'gold standard' criterion measure (Meagher, 2009) by picking criteria or standards to assess a scale, such as a predictive or a concurrent measure (Acock, 2008; reviewed by Costa, Murray, Dai, Canali & Minero). Thus, criterion validity can be divided in predictive and concurrent validity (Cronbach & Meehl, 1955): concurrent validity is the accuracy in predicting results of a criterion measure determined at essentially the same time as the measure being validated, whereas predictive validity is considered as the accuracy in predicting results of a criterion measurement obtained sometime after (Meagher, 2009).

Criterion validity is based on the association of an indicator to another welfare relevant indicator, while construct validity is based on the experimental proof that the welfare state is connected to the indicator in question (Blokhuis et al., 2013). Therefore, in animal welfare, which cannot be measured directly and for which there are several definitions, validity must be studied accurately before any method can be established for common use (Scott et al., 2003), since an overall assessment system is only as valid as the indicators used to establish it (Dalmau et al., 2010).

#### 1.2.3.3. Feasibility

Feasibility refers to the degree to which the suggested measurement procedure is possible, practicable and worthwhile (Martin & Bateson, 2007). Indicators taken should also justify the cost, effort involved and the disturbance to the subjects.

The feasibility can be particularly limiting when recording welfare on-farm: the frequently limited amount of time available for data collection and the conditions under which data have to be gathered exclude the use of most physiological indicators and various behavioural indicators (Spoolder et al., 2003).

Sørensen, Rousing, Møller, Bonde and Hegelund (2007) state that costs in a welfare assessment system may be attenuated by reducing the number of indicators of a given protocol, by decreasing the frequency of recording and the sample size without losing validity (e.g. Waiblinger & Menke, 2003), by cooperating with data recording systems used for other instances and by adopting a change from external to internal recordings, where the farmer himself perform some of the recording.

In practice, a welfare assessment tool must be centred on simple observations and records related to aspects of management, resources and welfare so that an observer can collect this information during a single visit (Webster et al., 2004). Even if simple, such protocols uniting the several aspects recorded, should provide a complete and valid picture of the welfare state of commercially kept animals (Smulders et al., 2006).

## 1.3. Aggregation of indicators into protocols

### 1.3.1. Overall assessment system

The several indicators used to assess animal welfare need to be combined to determine an overall level of animal welfare on farms (Botreau et al., 2008; De Vries et al., 2013; Blokhuis et al., 2013), since approaches centred only on behaviour, emotional state, physical appearance, or performance records, can never give a welfare overview (Webster et al., 2004).

### 1.3.2. Applications of a welfare assessment system

In general, there are four categories of applications for animal welfare assessment systems: research, legislative requirements (non-voluntary), certification systems (voluntary) and advisory/management tools (Johnsen et al., 2001; Main, et al., 2003; reviewed by Main, 2009). These applications may have several objectives, as advising farmers on animal welfare improvement (Sørensen, Sandøe & Halberg, 2001), checking compliance with legislative requirements, implementing welfare certification schemes (Main et al., 2001) and comparing systems to improve legislation (Bracke et al., 2002; Main et al., 2003). Different applications will then require other sets of indicators, based either on the farm or housing environment, or on the animals themselves (Botreau et al., 2007c; Main, 2009).

### 1.3.3. Existing methods for welfare assessment

During the last few decades several animal welfare monitoring systems have been developed in Europe, such as ethical account in Denmark (Sørensen et al., 2001), systems based on minimum requirements, as Freedom Food farm assurance and food labelling scheme (Royal Society for the Prevention of Cruelty to Animals [RSPCA], 1994), index systems of welfare assessment, as 'Animal Needs Index (ANI) 35 L' (Bartussek, 1999) and 'ANI 200' (Sundrum et al., 1994), and an operational decision support system (relational database) established to assess the welfare state of pregnant sows (Bracke et al., 2002). More recently, there has been a growing interest in measuring how the resources truly affect the animal, with several protocols being developed, aiming to deliver a more holistic animal welfare evaluation at farm level (Wemelsfelder & Mullan, 2014). Following this perspective, the on-farm assessment method of dairy cattle welfare, that considers the fulfilment of animal needs as the fundamental principle for the evaluation of animal welfare status, suggested by Capdeville and Veissier (2001), and the animal-based welfare evaluation protocols for dairy cattle, pigs and laying hens proposed by the University of Bristol, in an investigation study evaluating the welfare impact of Freedom Food Scheme (RSPCA, 1994; Whay et al., 2003b) were created. These systems are based on the logic of the 'Five Freedoms' (FAWC, 2009), providing a comprehensive framework for assessing animal welfare and being mostly focused on animal-based indicators.

In 2008, a scientifically sound and feasible overall assessment system of the welfare of cattle, pigs and poultry was proposed, with its origin based on the fact that the evaluation of welfare varies among the existing schemes and on the absence of a standard for animal welfare appraisal at farm level, and for product information related to animal welfare intended for consumers (Botreau et al., 2008). The European Welfare Quality® (2004–2009) project was a pioneer in developing animal-based on-farm and slaughter welfare assessment systems to address the main extents of feeding, housing, health, disease, and behaviour that engage numerous welfare criteria (Rushen et al, 2011).

The Welfare Quality® project constructed a multicriteria evaluation model for welfare assessment at unit level (farms, slaughterhouses), where 12 key animal welfare criteria were identified (Botreau et al., 2007c): 'Absence of prolonged hunger'; 'Absence of prolonged thirst'; 'Comfort around resting'; 'Thermal comfort'; 'Ease of movement'; 'Absence of injuries'; 'Absence of disease'; 'Absence of pain induced by management procedures';

‘Expression of social behaviour’; ‘Expression of other behaviours’; ‘Good human-animal relationship’ and ‘Absence of general fear’. Data recorded on an animal unit are used to check unit compliance with the 12 welfare criteria. Then, the scores achieved at criterion level are assembled to evaluate unit compliance with four main welfare principles: ‘Good feeding’, ‘Good housing’, ‘Good health’ and ‘Appropriate behaviour’. Each of the four principles is independent from the others and coincides with a particular management aspect on a farm, having implications for farmers' routines and strategic choices (Botreau et al., 2008). At last, these principle scores are used to determine an overall evaluation, hence following a hierarchical aggregation process (Botreau et al., 2007b; Botreau et al., 2008; Blokhuis et al., 2013). The four scores are aggregated to create the overall assessment and four welfare categories were defined: ‘Excellent’, ‘Enhanced’, ‘Acceptable’, and ‘Not classified’ (Botreau, Veissier & Perny, 2009).

Although these systems are very different (Figure 2), considering they may differ both in their perception of the definition of ‘animal welfare’ and their eventual goals, they all state to assess animal welfare (Johnsen et al., 2001; Rousing, Bonde & Sørensen, 2001).

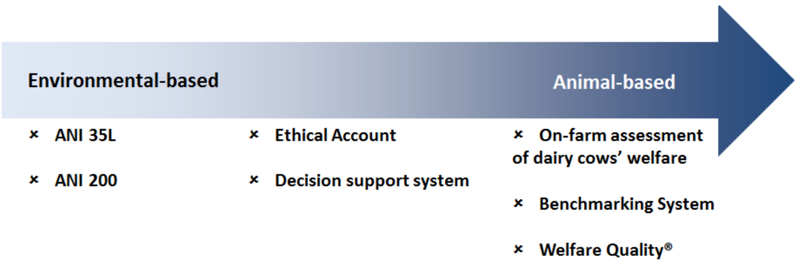



Figure 2 - Combination of welfare indicators in systems to assess animal welfare at farm level (adapted from Johnsen et al., 2001).

## 1.4. European and Portuguese regulations for the welfare of farmed animals

Animal welfare is not a new topic for regulation in most developed countries, due to a refined consumer foundation and a larger exposure to animal welfare issues (Vapnek & Chapman, 2010), collecting more legislative attention in Europe than in other regions (Van Horne & Achterbosch, 2008). In this continent, animal protection laws are delivered and drawn by national governments. Nevertheless, specific initiatives are formulated by supra-national institutions, as the European Council (the first supranational organization proposing measures to guarantee animal welfare) and the European Union (EU), which specify minimum requirements that need to be implemented by all member states (Appleby, 2003; Veissier et al., 2008). The General Directorate for the Health and Consumer Protection (DGSANCO), a department of the European Commission, is responsible for animal protection. When a decision is made about elaborating a new part of legislation to protect animals, DGSANCO consults a scientific committee (the scientific committee on Animal Health and Animal Welfare, part of EFSA) that reviews the scientific evidence on any aspect that may impair animal welfare, and delivers recommendations on how to protect animals. Subsequently, DGSANCO may agree to draft a directive that is submitted to the Council of Ministers of the EU and turns into a Council Directive after receiving their approval (Veissier, Beaumont & Lévy, 2007; Veissier et al., 2008; Le Neindre, 2009). Generally, the work of the EU is guided by the 'Five Freedoms' (FAWC, 2009; Vapnek & Chapman, 2010) and the common trends of EU Directives for the breeding of farmed animals consist in: 1) increasing space allowance per animal; 2) allowing interactions between animals, and therefore supporting group housing; 3) giving more freedom of movement; 4) offering animals an enriched environment; 5) feeding animals a diet based on their physiological and behavioural needs; 6) limiting painful interventions (Veissier et al., 2008).

The EU is a signatory to the European Convention for the protection of animals kept for farming purposes, which was translated into an EU directive (Directive 98/58/EC) that covers the minimum animal welfare standards for farmed animals, also applying to goats. EU directives are translated into national regulations (e.g. decrees) in order to be implemented in each country. For instance, the Directive 98/58/EC have been transposed to Portuguese regulation by means of the decree number 64/2000 from April 22 (*Decreto-Lei (DL) n. º 64/2000, de 22 de Abril*), in order to be applied in Portuguese farms. To ensure that the owners or keepers of animals fulfil this decree, the competent national authorities, *Direcção-Geral de Alimentação e Veterinária* (DGAV), carry out an annual inspection of the registered farms.

On January 2012, as part of the EU Strategy for the Protection and Welfare of Animals 2012-2015 "a simplified EU legislative framework for animal welfare" has been suggested (European Commission [EC], 2012), in order to provide more uniform and appropriate implementation (Dalla Villa, Matthews, Alessandrini, Messori & Migliorati, 2014). To tackle the main common drivers that compromised animal welfare in the EU specific objectives were set (EC, 2012): objective 1) to improve enforcement of the EU legislation in a consistent approach across the Member States; objective 2) to provide for open and fair competition for EU business operators that implement or go beyond EU requirements; objective 3) to improve knowledge and awareness of EU business operators regarding animal welfare; objective 4) to improve the coherence of animal welfare across animal species. To accomplish these objectives four main options were identified, reflecting the problems and its drivers (EC, 2012): 1) strengthening Member States' compliance; 2) benchmarking voluntary schemes; 3) establishing a European network of reference centres; 4) streamlining requirements for competence and using animal welfare indicators (and investigate the possibility of extending the scope of this option). According to the European Commission (2012), this framework considers the use of science-based animal welfare indicators as a potential way to simplify the legal framework and allows flexibility to encourage competitiveness of livestock producers. Furthermore, it also considers an increase on the transparency and adequacy of information to consumers on animal welfare for their purchase choice, which will improve competitiveness in the EU food industry, by adding economic value to animal welfare (Dalla Villa et al., 2014). Although Europe has an extensive body of legislation intended to ensure the welfare of farm animals (Blandford, 1999), progressively animal welfare measures are being endorsed by non-state actors (Maciel & Bock, 2013). Numerous European supermarkets, non-government organizations (NGOs) and industries are implicated in joint actions to preserve the sake of animals (Blandford, 1999; Veissier et al., 2008). Thus, nowadays the quest relies in increasing food animal production, while consecutively ensuring animal welfare and protecting food security (Vapnek & Chapman, 2010). As referred by Lundmark, Berg, Schmid, Behdadi and Röcklinsberg (2014), the following step in the development of legislation of animals kept by humans appears to reside in a shift from the protection of animals from unnecessary suffering to giving them a quality of life that is worth living (FAWC, 2009), as a result of the increased focus on the positive welfare (Boissy et al., 2007; Yeates & Main, 2008).

## 1.5. Development of an on-farm welfare assessment protocol for dairy goats

### 1.5.1. Goat livestock: World, European and Portuguese perspectives

The goat (*Capra hircus*) is one of the earliest small ruminant species to be domesticated and was used for numerous purposes (milk, meat, fibre, skin or work) in different conditions, at least since 2500 B.C. in the Middle East (Dubeuf, Morand-Fehr & Rubino, 2004; Dubeuf & Boyazoglu, 2009).

The development of goat husbandry is fairly more common under the extreme settings of very intensive and very extensive systems of animal husbandry due to the formidable resilience, high adaptability to very different environments (Aviles, 2002) and nutritional regimes, high productivity and low maintenance cost (Morand-Fehr et al., 2004; Boyazoglu, Hatziminaoglou & Morand-Fehr, 2005; Aziz, 2010; Popescu, 2013) of this species.

There are variations among the different parts of the world considering the number of goats, with a larger of number of herds being in Asia (59.4%), followed by Africa (35%), Americas (3.6%), Europe (1.6%) and Oceania (0.4%), proving that the highest portion is in the developing countries where goat milk is a basic food, particularly for rural population

(Boyazoglu et al., 2005; Devendra, 2013; Popescu, 2013; Food and Agriculture Organization of the United Nations Statistics Division [FAOSTAT], 2015). According to Food and Agriculture Organization of the United Nations (FAO; 2015) the largest number of goats in the world is in China, followed by India, Pakistan and Nigeria. Considering the dairy goats, the larger number of these animals is in Bangladesh, India and Mali (FAO, 2015).

In the past 20 years a new and growing interest in goat milk and goat milk products took place all over the world, with several varieties of goat milk cheeses being produced, determined by the diversity in locality, milk composition and manufacturing practices used (Yangilar, 2013). However, the most organized programs for selection, processing and commercialization of goat milk are situated in the developed European countries (Dubeuf, 2010), largely used for cheese production (Le Jaouen & Toussaint, 1993; Castel et al. 2010). The specialized dairy goat breeds used in developed countries present a higher genetic potential for milk production than breeds found in the developing world, the most common dairy goat breeds being Saanen, Anglo-Nubian, Toggenburg, Alpine and West African Dwarf (FAO, 2015). According to Haenlein (2004), there are three aspects for demand of goat milk: home consumption, *connoisseur* interest in goat milk products, and a medical purpose, for people with cow milk allergies and other gastrointestinal conditions.

In 2012, in Europe, the largest amount of goat milk was produced in France (0.6 million tons), Spain (0.44 million tons) and Greece (0.4 million tons; FAOSTAT, 2015), being the only continent in which goat milk presents substantial economic importance and organization (Dubeuf, 2010). Dairy goat farming has a paramount importance to the economies of the Mediterranean countries (Boyazoglu & Morand-Fehr, 2001; Pirisi, Lauret & Dubeuf, 2007; Escareño et al., 2012), where goat milk is a typical product traditionally consumed directly or as handmade cheese (Dubeuf, Morand-Fehr & Rubino, 2004; Boyazoglu et al., 2005; Dubeuf, 2010; Yangilar, 2013).

In some Portuguese regions, the breeding of small ruminants has (and has had across time) a pronounced socioeconomic value, not only because of meat and milk production, but also related to the fact that farmers are dealing with mountainous areas or regions where other economic activities are rare, with agriculture being the main activity (Silva, Fitas da Cruz & Barbosa, 2007). Goat distribution is quite irregular, being the caprine livestock more frequent in the driest inland areas (Barbosa, 1993), as in Alentejo, Centre and North inland regions[1]. In the national perspective these areas are often considered less-favoured and depopulated (Santos, Fitas da Cruz & Barbosa, 2007). As mentioned by Bruno-de-Sousa et al. (2011), the intensification of agriculture that happened mostly during the second half of the 20th century

[1] Information regarding goat livestock distribution in Portugal is presented in Table 13 (Annex 1).

resulted in the predominant use of a reduced number of exotic breeds under intensive production systems (e.g. Murciano-Granadina, Saanen and Alpine), with the native breeds of goats (e.g. Serrana, Bravia and Charnequeira) being mostly bred in small farms in marginal and forest areas of the country. According to *Instituto Nacional de Estatística* (INE; 2011) and Tibério and Diniz (2014) the Portuguese goat production is more meat than dairy oriented, where the majority of goat farms are small with an average size of 13 animals per herd, and farms with more than 100 animals (3% of the total number of goat farms) represent 41% of the goat livestock. The dairy herd, approximately 0.15 million head, represents about 35% of the goat population, distributed over approximately 12000 farms (36%), with about 80% of farms having less than 9 animals (INE, 2011). Portugal, with a goat population of approximately 0.4 million head in 2013 and a yield of nearly 30 million litres of milk in 2012 (INE, 2015; FAOSTAT, 2015), shows the aptitude for goat's milk production as well, not only for a national market but mostly for the European market, where there is a major demand for a variety of products (Barbosa, 1993).

### 1.5.2. Evolution of goat research

Although there has been less research on goats than on other production species (Barbosa, 1993; Dubeuf et al., 2004), the number of this throughout the world is increasing (Sahlu & Goetsch, 2005) and the overall recognition of this long-underestimated species has developed, improving its importance in the livestock sector (Boyazoglu et al., 2005).

While Welfare Quality® project recently designed methods for the overall welfare assessment of cattle, pig, and poultry welfare, on-farm and at slaughter, only occasional research has been carried out on the welfare of goats, and in particular of dairy ones. With the increasing interest in these animals as production ones, approaches to assess their welfare are mandatory. Nevertheless, looking at scientific literature there is little information available on which to base welfare assessment protocols for dairy goat (Battini et al., 2014), maybe due to the assumption these animals have several adaptation mechanisms to harsh environments and are still mostly raised in extensive production systems (Caroprese, Casamassima, Rassu, Napolitano & Sevi, 2009). Recently, several studies into the health and welfare of dairy goats emerged, focusing on specific aspects that have the potential to affect the welfare of this species, which will be mentioned in the next sub-section along with the discussion and presentation of the animal-based welfare indicators.

More recently, the AWIN (Animal Welfare Indicators) project funded by the European Commission, in the Seventh Framework Programme (FP7-KBBE-2010-4), emphasized the need for a science-based approach to assess and improve animal welfare by developing, integrating and disseminating practical on-farm welfare assessment protocols for commercially important husbandry species, often forgotten in those evaluations (AWIN, n.d.). The AWIN project, involving 10 Institutions in nine countries, approached animal welfare indicators in four separate but complementary Work Packages (WP), developing and testing these indicators, including those of pain, for sheep, goats, horses, donkeys and turkeys (WP1), studying the relationship between diseases and animal welfare (WP2), examining the influence of prenatal and early-postnatal environments and handling methods on welfare and health of pregnant females and their offspring (WP3), and effectively communicating with stakeholders and interested parties on animal welfare investigation, teaching, and outreach opportunities, by creating a global hub (Animal Welfare Science Hub) for research and education in animal welfare (WP4; AWIN, 2014a). The Animal Welfare Science Hub (www.animalwefarehub.com) incorporates research and learning materials, in part derived from deliverables of the WP1, WP2 and WP3 research.

In Portugal, the AWIN Portuguese research team from *Faculdade Medicina Veterinária* – University of Lisbon, conducted its studies on dairy goats, through the work of WP1 and WP2 that were established for this species. The WP1, in close collaboration with the AWIN Italian research team from the *Università degli Studi di Milano*, assessed the psychometric proprieties (validity, reliability, and feasibility) of different indicators to be incorporated in a welfare assessment protocol for dairy goats. Focused on the findings of this assessment AWIN scientists established a research action plan (Figure 3) to approach the lack of knowledge concerning the above mentioned proprieties, with the resulting list of indicators, including pain, and environmental-based indicators, being tested on-farm by trained assessors.

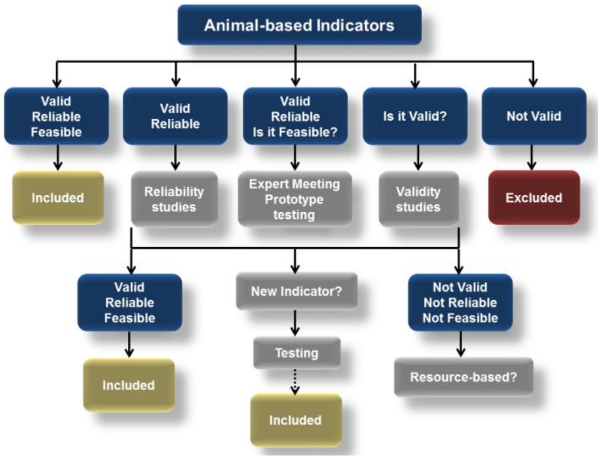



Figure 3 – Operational methodology supporting the selection of potential indicators to be included in the prototype of the welfare assessment protocol (adapted from Mattiello et al., 2014).

WP1 researchers also encouraged a participatory approach in this project through the involvement and the collaboration with stakeholders, in order to assure the effectiveness and sustainability of the final assessment protocols. The AWIN work in Portugal was performed entirely in commercial farms, which showed great advantages concerning the validation of indicators and the feasibility of the protocol (AWIN, 2014a).

### 1.5.3. Animal-based indicators for on-farm welfare assessment for dairy goats

The AWIN prototype for dairy goats is based on the use of the four principles – 'Good feeding', 'Good housing', 'Good health' and 'Appropriate behaviour' – which are linked to 12 criteria, developed by Welfare Quality® (Table 1). Some indicators, such as 'Body Condition Score', 'Hair coat condition' or 'Oblivious', provide information related to more than one welfare criteria.

| Principles | Welfare Criteria | Animal-based indicators |
|---|---|---|
| **Good Feeding** | Absence of prolonged hunger | BCS<br>Queuing at feeding<br>Hair coat condition |
| | Absence of prolonged thirst | Queuing at drinking |
| **Good Housing** | Comfort around resting | Cleanliness |
| | Thermal comfort | Shivering and Panting score |
| | Ease of movement | Kneeling at feeding rack |
| **Good Health** | Absence of injuries | Lameness, claw overgrowth, kneeling in pen, lesions and swellings and udder asymmetry |
| | Absence of disease | BCS, abscesses, hair coat condition, discharges, udder cleanliness, oblivious and diarrhoea |
| | Absence of pain induced by management procedures | Improper disbudding |
| **Appropriate Behaviour** | Expression of social behaviours | Queuing at feeding and drinking |
| | Expression of other behaviours | Oblivious |
| | Good human-animal relationship | Latency to first contact and Avoidance distance tests |
| | Positive emotional state | QBA |

Table 1 - Welfare principles, criteria and indicators for on-farm welfare assessment for dairy goats.

#### 1.5.3.1. Good Feeding

**a) Absence of prolonged hunger**

Nutrition plays an essential and singular role in goat farming, since it is the production factor that goat farmers can act on the most easily and rapidly, it has the most noticeable effect on production costs and due to the fact that feeding directly influences other aspects of this production, such as pathological conditions and the reproductive performance of the herd (Morand-Fehr, 2005).

Absence of prolonged hunger is defined by Welfare Quality® (2009) as "animals should not suffer from prolonged hunger, i.e., they should have a suitable and appropriate diet".

### Body Condition Score (BCS)

Morand-Fehr et al. (2005) states that body condition, which considers body reserves, mostly lipid reserves, is a useful indicator in assessing nutritional *status* under several conditions. The Body Condition Score (BCS) is a commonly subjective method used in livestock to evaluate the animal's lipid and protein reserves that are used at different stages of production. This evaluation is useful for monitoring the adequacy of the feeding program both under intensive and extensive conditions (Smith & Sherman, 2009) and contributes to the decisions concerning different management practices during production cycle (Short, Grings, MacNeill, Heitschimidt & Haferkamp, 1996).

In the adult goat, appropriate body weight is related to breed, frame size, and stage of gestation or lactation. Frame size of an individual adult is constant, but deposition of fat and muscle alters with nutritional and physiologic state (Smith & Sherman, 2009). According to Chilliard et al. (1981), in goats there is a different distribution of fat deposits, most of the dairy goat's body fat is deposited in the omentum and perirenal tissues (reviewed by Smith & Sherman, 2009), therefore goat body condition is the result of fat mass (the stock of energy reserves) and muscle mass (protein accumulation; Gaias, 2012). Critical phases for evaluating goats might comprehend dry-off, the last two weeks of gestation, six weeks into lactation, the turn-out onto pasture, the beginning of the dry season, and the beginning of the breeding season (Morand-Fehr et al., 1989; reviewed by Smith & Sherman, 2009), existing an optimum BCS for each animal and stage of their production cycle (Koyuncu & Altınçekiç, 2013). For instance, the *Institut National de la Recherche Agronomique* (INRA; 1988) defined the BCS recommended in different physiological stages of the animals: 2.25-3.5 during dry period, 2.75-3.5 at parturition, and higher than 2 at the lactation peak.

This parameter is considered a valid welfare indicator for several species as cattle (Winckler et al., 2003; Welfare Quality, 2009), buffaloes (Winckler et al., 2003; De Rosa et al., 2009), sheep (Russel, Doney & Gunn, 1969; Caldeira, Belo, Santos, Vazques & Portugal, 2007; Phythian et al., 2012), and goats (Santucci et al., 1991; McGregor & Butler, 2008; Anzuino et al., 2010). Several BCS systems have been created and used for investigation and practical monitoring on commercial farms, the assessment might rely on a visual method, palpation of specific parts of the body of small ruminants or in an association of both methods (Battini et al., 2014). In goats, the most common method was established by Hervieu and Morand-Fehr (1999), which is a six-point scale with intermediate scores (0.25) involving palpation of the sternum and the lumbar vertebrae and requiring an individual assessment, that is not always feasible. The concurrent validity of the scale was evaluated by Santucci et al. (1991), as also the inter-observer reliability, with results of 0.6. However, the inclusion of BCS in on-farm welfare assessment schemes has the purpose of identifying the endpoints of a scale (animals that are too thin or too fat), thus the scoring system does not need to be exceptionally thorough (e.g. Welfare Quality, 2009; Anzuino et al., 2010).

**Queuing at feeding**

Intensively kept dairy goats are typically fed in early and late hours of the daytime. This system may lead goats to go through inadequate nutrient consumption, poor performance and behavioural alterations (Görgülü et al., 2008; Battini et al., 2014). Feed can be a restricted resource either because the quantity is limited or because the feeding space is not available for all animals. That may not only decrease the average feeding time, but as the competition increases the difference between low and high ranking animals is expected to also rise (e.g. Tregenza, 1994; Andersen et al., 2008; Jørgensen, Andersen & Bøe , 2007). Age, body size and horns seem to be related to dominance, at least in wild and feral goats (Barroso, Alados & Boza, 2000; Shinde, Verma & Singh, 2004). In order to keep a regular level of consumption as the competition develops animals may increase feed intake or eat at different times of the day, and low ranking ones may, for instance, feed when the others are resting (Olofsson, 1999; Shinde et al., 2004; Jørgensen et al., 2007). Thus, recording the number of queuing goats may be used not only to identify animals suffering from hunger, due to insufficient number of feed places or improper feed distribution along the feeding rack, but also the expression of social behaviours.

**Hair coat condition**

Several authors propose that hair coat condition can be established as a first sign of goats' nutritional and health status (Veit, McCarthy, Friedericks, Cashin & Angert, 1993; Smith & Sherman, 2009; Sarkar et al., 2010; Lengarite, Mbugua, Gachuiri & Kabuage, 2012). Goats with poor hair coat condition might have extremely low BCS and suffer from mineral deficiencies and imbalances, thus exerting a significant effect on health and productivity of livestock (Dar et al., 2014). This notion achieved consensus among farmers and recent studies (e.g. Battini et al., 2015) appear to support that this indicator can be valid and feasible for on-farm welfare assessment, since goats with rough or scurfy hair are easily recognized in the herd.

**b) Absence of prolonged thirst**

**Queuing at drinking**

Although several reports have documented the ability of goats tolerating dehydration (Silanikove, 1994; Silanikove, 2000; Alamer, 2006; Alamer, 2009), animals should have a sufficient and accessible water supply (Welfare Quality, 2009), being their welfare compromised if they cannot drink whenever they need to, either because fresh water is not available or because of competition (Battini et al., 2014).

According to Ehrlenbruch, Pollen, Andersen and Bøe (2010), goats drink mostly during feeding (Rossi & Scharrer, 1992) and this behaviour is usually socially enabled (Forkman, 1996) and synchronized (Rook & Penning, 1991). Reduced opportunity to simultaneous feeding and drinking can result in decreased feeding/drinking time and therefore, lower the feed/water intake, with the low ranking individuals particularly suffering from this (Milinski & Parker, 1991; reviewed by Murray, Eberly & Pusey, 2006; Zupan, Bojkovski, Štuhec & Kompan, 2010).

**1.5.3.2. Good housing**

**a) Comfort around resting**

According to Welfare Quality® (2009), "animals should have comfort when they are resting". Goats are sensitive to the influences of their surroundings, thus adequate house holdings should be important to the breeder (Toussaint, 1997).

**Cleanliness**

The assessment of body cleanliness may provide information on animal comfort as well as stockpeople's attitudes and care for animals (De Rosa et al., 2009). Animal cleanliness is used as a welfare indicator in pigs (Scott et al., 2007), poultry and cattle (Hughes, 2001; Whay et al., 2003b; Andreasen & Forkman, 2012). Anzuino et al. (2010) identified dirt in several areas of the goats' body (limbs, body, head, udder and teats), using cleanliness as a potential welfare indicator. Comparing to dairy cattle, goats generally have a cleaner environment being housed on straw bedding all year, as they present much drier faecal matter than cows. Although manure managing is much easier in goats than in cattle, the way the first ones are moved and handled (e.g. for the milking parlour), as well as the cleanliness and dryness of pathways, may considerably influence the cleanliness of the herd (Anzuino et al., 2010). In large farms this assessment may be time consuming, thus the development of a representative sampling strategy may be necessary.

**b) Thermal comfort**

**Shivering and panting score**

As defined by Welfare Quality® (2009), "animals should have thermal comfort, i.e., they should neither be too hot nor too cold".

Ruminants have wide comfort zones and a high degree of thermal tolerance (Sejian & Srivastava, 2010). In spite of presenting well developed mechanisms of thermoregulation, ruminants do not preserve homeothermy when suffering from heat stress (Lu, 1989; Silanikove, 2000). The common homeostatic responses to thermal stress in mammals consist of raised respiration (Yousef, 1985), panting, drooling, reduced heart rates and abundant sweating (Blazquez et al., 1994), decreased feed intake (Silanikove, 1992) and reduced milk production (Albright & Alliston, 1972; Lu, 1989; reviewed by Silanikove, 2000). Although limited information in this field is available, principally for goat breeds usually raised under Mediterranean and subtropical conditions, goats are known to be adapted to harsh environments (Silanikove, 2000). However, their productivity is affected by extreme climatic conditions (Sejian & Srivastava, 2010). According to Toussaint (1997), temperature is one of the factors of production resulting from the animal's metabolism, climatic parameters and from a change of surface characteristics. Goats try to maintain a near-constant body temperature of 38.5 ºC that must be preserved by thermal changes, such as contribution or loss. Therefore, Toussaint (1997) recommends that temperatures for goats kept indoor range from a minimum of 6°C to a maximum of 27°C (with an optimum from 10° to 18°C), with relative humidity from 60 to 80%, and with 0.5 m/s maximum air speed.

High temperatures, high direct and indirect solar radiation and humidity are stressing factors that disturb the animals. In domestic ruminants a rise of body temperature leads to physiological (sweating, panting), hormonal (cortisol, thyroid gland activity), and behavioural thermoregulatory responses (Silanikove, 2000). Reproduction, milk production and growth become compromised under heat stress due to the severe changes in biological functions caused by the stress (Habeeb et al., 1992; Silanikove, 2000; reviewed by Hristov et al., 2012), being these responses warning signs of poor welfare (Broom & Johnson, 1993). Evaluation of the respiratory rate provides reliable and practical information for assessing the degree of heat stress in farm animals, considering that the respiration rate follows evaporative water loss (Silanikove, 2000). A panting score has already been applied to cattle (Gaughan, 2003) and it revealed a predictive validity in goats (Darcan, Cedden & Cankaya, 2008; Fioni, 2014).

Low temperatures, wind and rain will increase the animals' heat loss with animals responding with physical and behavioural mechanisms of thermoregulation (e.g. huddling), in order to reduce heat loss (Curtis, 1981; reviewed by Bøe & Ehrlenbruch, 2013). Investigation supports the idea that goats can suffer when dealing with low air temperatures (Holmes & Moore, 1981; McGregor, 2002; Bøe & Ehrlenbruch, 2013), shivering once exposed to inclement weather (Fioni, 2014).

For instance, simple animal-based indicators as shivering or panting can be adopted to identify cold or heat stress, respectively, in many species (Blokhuis et al., 2013), including goats.

**c) Ease of movement**

**Kneeling in pen and at feeding rack**

This criterion is centred on the assumption that "animals should have enough space to be able to move around freely" (Welfare Quality, 2009) without possibility of injury and at a proper stock density. Anzuino et al. (2010) proposed kneeling in the pen and in the trough area as potential welfare indicators, which involve goats standing or walking on their front knees. This behaviour origin is still unknown and may be related to discomfort, due to inadequate farm-household structure.

The prevalences of severely lame goats and goats kneeling in the pen area assessed by Anzuino et al. (2010) were significantly correlated, which suggest that kneeling behaviour may be associated to painful limb injuries. For example, kneeling is a clinical sign of Caprine arthritis and encephalitis (CAE) and often exhibited in infected farms (Smith & Sherman, 2009). The high prevalence of British farms with kneeling goats (87.5%) recorded by Anzuino et al. (2010) supports further investigation, either as an indicator of ease of movement or as absence of disease or injury. The on-farm feasibility for kneeling is high, since this behaviour is easy to detect, albeit requesting a sampling strategy (Battini et al., 2014).

### 1.5.3.3. Good health

The health of animals is an important part of their welfare (Broom & Corke, 2002; O'Callaghan, Cripps, Downham & Murray, 2003; Whay et al., 2003b; Weary, Niel, Flower & Fraser, 2006), and diseases of animals, particularly those associated with pain, lead to welfare impairment (Whay, Waterman & Webster, 1997; Weary, Huzzey & von Keyserlingk, 2009).

**a) Absence of injuries**

As determined by Welfare Quality® (2009) "animals should be free of injuries, e.g. skin damage and locomotory disorders".

**Lameness**

Lameness, a serious disease with a multifactorial aetiology consisting in the clinical sign of compromised locomotion or abnormal gait, is one of the most severe and common welfare problems in many species of livestock and poultry (Smith & Sherman, 2009; Nonga, Makungub, Bittegekob & Mpandujib, 2009; Grandin, 2010). Pain is a serious characteristic of lameness which can origin distress and limit the animal performance, having a negative effect in production and reproductive potentials of the affected animals (Whay et al., 1997; O'Callaghan et al., 2003, KilBride, Gillman, Ossent & Green, 2009a; KilBride, Gillman & Green 2009b). Hence, lameness may affect all of the 'Five Freedoms' (Anil, Anil & Deen, 2009; FAWC, 2009) presenting economic repercussions that are difficult to quantify (Mohammed, Badau & Kene, 1996). This condition may affect productivity in dairy goats by reducing milk yield and weight (Hill et al., 1997; Christodoulopoulos, 2009), influence fertility (resulting in longer kidding interval and decreased number of kids; Hill et al., 1997; Eze, 2002; Christodoulopoulos, 2009), conduce to pregnancy toxaemia (Eze, 2002; Lima, Pascoal, Stilwell & Hjerpe, 2012b) and neonatal diseases (Eze, 2002), leading to early culling (Hill et al., 1997). The predisposing factors for lameness in goats include poor nutrition, unhygienic ground conditions, hard and rough environment terrain, wetness, poor digital conformation, claw overgrowth, penetrating injuries, trauma, fracture, inflammation of anatomical structures and presence of infectious agents that affect the limb joints, as CAE and caprine contagious agalactia (Mohamed et al., 1996; Bergonier, Berthelot & Poumarat, 1997; Hill et al., 1997; Bokko & Chaudhari, 2001; Smith & Sherman, 2009; Winter, 2011).

Lameness scores have been validated in some species such as in cattle (Welfare Quality, 2009) and sheep (Winter, 2008). However, in goats there are no established gait scoring systems and lameness is assessed applying different scales (Hill et al., 1997; Mazurek, Marie & Desor, 2007; Anzuino et al., 2010) or by classifying the animals as lame or not lame (Christodoulopoulos, 2009).

**Claw overgrowth**

A major predisposing factor for lameness in commercial dairy goat farms may be overgrown claws, which can be scored from moderate to severe, presenting a significant correlation with lameness (Hill et al., 1997; Christodoulopoulos, 2009; Anzuino et al., 2010; Winter, 2011), and is a major problem in intensively kept animals due to lack of claw wear (Anzuino et al., 2010; Muri et al., 2013).

**Lesions and swellings**

Anzuino et al. (2010) observed that some goats on all the assessed farms presented visible skin lesions, with the location and type of the lesions varying between farms and between goats on the same one. The presence of lesions (including skin damages, swellings and hair losses) may not be painful but can be important indicators of welfare, since they may reflect the impact of the surrounding environment on the animal's body (De Rosa et al., 2009). These alterations can be caused by contact with hard floors, presence of physical obstructions to normal behaviours (e.g. bars restricting access to feed), or may result from trauma or ectoparasites (Smith & Sherman, 2009).

**Udder asymmetry**

Udder and teat lesions can affect welfare and production in dairy goats (Contreras et al., 2007; Leitner, Silanikove & Merin, 2008), thus the relevance of udder health indicators in a dairy goat's welfare assessment. Udder asymmetry consists in a chronic alteration of the udder that perseveres even after an udder has recovered from infection (Klaas, Enevoldsen, Vaarst & Houe, 2004) or injury. This alteration has been associated with chronic intramammary infection as CAE and caprine contagious agalactia, causing fibrosis and atrophy of one half (Alawa, Ngelea & Ogwu, 2000; Smith & Sherman, 2009; Paterna et al., 2014). This indicator can be detected by visual assessment at a short distance (Anzuino et al. 2010; Muri et al., 2013; Battini et al., 2014).

**b) Absence of disease**

Welfare Quality® (2009) states "animals should be free from disease, i.e., animal unit managers should maintain high standards of hygiene and care".

**BCS**

Low BCS is related to an excessive mobilization of body fat reserve (and nitrogen reserve at a lesser extent), due to reduced energy intake and/or increased energy output, which mainly occur under high heat load situations (Sevi et al., 2001) and during suckling, and early stage of lactation (Sevi et al., 2001; Albenzio et al., 2003). On the contrary, high BCS can be a sign of overfeeding or extreme confinement of animals (Caroprese et al., 2009). This parameter allows identifying chronically ill goats, being reduced for example in chronic contagious diseases, as caseous lymphadenitis (CLA), or CAE, gastro-intestinal parasitism, painful illnesses, as arthritis and laminitis (Smith & Sherman, 2009; Battini et al., 2014). Laporte-Broux et al. (2011) state that BCS can be useful in determining which periparturient animals are at risk from pregnancy toxaemia. Very thin or obese pregnant dairy goats present a greater risk of developing this condition (Brozos, Mavrogianni & Fthenakis, 2011; Lima, Pascoal & Stilwell, 2012a). Therefore, the health status of the herd can be evaluated by scoring the body condition, with BCS being also a valid and feasible indicator to assess 'Absence of disease' criterion.

**Abscesses**

Enlargement of one or more lymph nodes in a goat typifies caseous lymphadenitis, which is a chronic contagious disease where enlargement and abscessation of one or more peripheral lymph nodes occurs (mostly parotid, mandibular, prescapular, prefemoral, and supramammary lymph nodes; Smith & Sherman, 2009). However, internal abscesses, particularly in the lungs, may appear if *Corynebacterium pseudotuberculosis* reaches the thoracic lymph duct or if it is inhaled. Other aetiologies may also cause hyperplasia or abscessation of the regional node, such as other bacterial infections (Gezon et al., 1991; reviewed by Smith & Sherman, 2009), CAE, skin diseases (e.g. sarcoptic mange and contagious ecthyma) and occasionally lymphosarcoma. Mantova (2012) observed a reduced feeding time and low BCS in goats with external abscesses, indicating a general poor condition of the animals, which shows the validity and feasibility of this animal-based indicator for on-farm welfare assessment (reviewed by Battini et al., 2014).

**Hair coat condition**

According to Berg, Jolly, Rambeloarivony, Andrianome and Rasamimanana (2009) hair coat condition can be regarded also as a potential indicator of animal health, reflecting internal or systemic diseases in addition to the expected factors, as parasites and skin infections. Endocrine dysfunctions and nutritional deficiencies, organic disorders, fever, immune-mediated diseases or general severe illness may affect the growth of hair and thus the hair coat condition (Kahn et al., 2006; reviewed by Berg et al., 2009), supporting the assumption that this one may reveal hidden pathologies.

**Discharges**

Welfare Quality® (2009) has already included ocular, nasal and vulvar discharges in cattle welfare assessment protocols, and their feasibility is accepted. Ocular and nasal discharges are important signs of upper and lower respiratory tract infections, as sinusitis caused by the larvae of *Oestrus ovis*, nasal foreign bodies, nasal tumours and pneumonias. In Norwegian farms, Muri et al. (2013) found that the majority of goats with ocular discharge only had mild symptoms being unclear, for the authors, how important this is as a welfare problem. Nevertheless, Muri et al. (2013) state that this may indicate the presence of a primary pathogenic challenge or of a sub-optimal environment. Mild conjunctivitis may result from an upper respiratory tract infection, normally involving *Mycoplasma spp.* and dust. However, foreign bodies and entropion can be considered as non-infectious causes of ocular discharge (Harwood, 2006; Smith & Sherman 2009). Regarding vulvar discharge, it is an important sign of reproductive tract infection, as metritis (Smith & Sherman, 2009), that may impair animal welfare.

**Udder cleanliness**

In dairy cattle, udder cleanliness has been used as an indicator to evaluate the risk of mastitis (Hughes, 2001; Schreiner & Ruegg, 2003; Reneau et al., 2005), thus being considered related to the health status in this species. Dairy cattle with dirtier udders, teats and hind limbs present a higher prevalence of intramammary infection (Schreiner & Ruegg, 2003, Reneau et al., 2005), and this may also occur in dairy goats.

**Oblivious**

Farmers describe that sick goats try to keep themselves away from the herd, usually standing immobile, sometimes facing the wall or other parts of the household (Battini et al., 2014). Goats are very gregarious, preferring to stay together, with individuals hardly ever seen apart from the group (Ross & Berg, 1956; reviewed by Miranda-de la Lama & Mattiello, 2010), thus the presence of obviously dull/sick goats (Anzuino et al., 2010) may allow the detection of early modifications in natural behaviours, that may reveal poor welfare (Miranda-de la Lama & Mattiello, 2010), being an early indicator of health problems.

**Diarrhoea**

Diarrhoea is described as an increase in the frequency, fluidity, or volume of bowel movements (Smith & Magdesian, 2008), being an important sign of disease or nutrition errors that can result from different origins. There are several known infectious, parasitic and non-infectious causes of diarrhoea with the frequency of occurrence varying to a large extent with the age of the animals affected (Smith & Sherman, 2009). In goats the most frequent causes of diarrhoea are enteritis, mal-absorption, gastrointestinal structural lesions, endotoxic conditions and nutritional problems leading to ruminal acidosis. Diarrhoea can also depress the immune system, becoming the animal more vulnerable to other conditions, being an important origin of economic losses due, for example, to decreased milk production.

**c) Absence of pain induced by management procedures**

This welfare criterion takes into account that "animals should not suffer pain induced by inappropriate management, handling, slaughter, or surgical procedures (e.g. castration, dehorning)" (Welfare Quality, 2009).

**Improper disbudding**

Disbudding and dehorning are procedures that may affect goat welfare. The presence of horns or scurs when kids are inefficiently disbudded may be a welfare problem in adult animals, since horned goats often adopt agonistic behaviour, which is intended to dissuade, injure, cause pain, or reduce the freedom of dehorned ones (Sisto, 2004; reviewed by Miranda-de la Lama & Mattiello, 2010). Also, horn regrowth may lead these animals to be caught in places of the housing system (as fences or pen partitions), or the scurs can press against the head or eye producing lesions and pain (Smith & Sherman, 2009). Disbudding of young kids is much more common in goat dairy farms, but dehorning of adult animals is occasionally performed. This procedure can be very painful and stressful, being associated with tetanus, heat meningitis, sinusitis, brain abscesses, ketosis, scurs and loss of social status (Smith &

Sherman, 2009; Battini et al., 2014). The identification of pain in animals and the development of management plans, to deal with this pain, are crucial to animal productivity and welfare (Plummer & Schleining, 2013). Adults that have been dehorned or present traces of horns due to improper disbudding are easily assessed on-farm, particularly when goats are at the feed trough.

#### 1.5.3.4. Appropriate behaviour

This principle refers to the ability of animals to express social and species-specific behaviours and the promotion of good human- animal relationships and positive emotions, as security or contentment (Welfare Quality, 2009).

Behaviour is one of the most significant early indicators of the welfare and the adjustment to the surrounding environment, echoing the interaction between the animal and its environment (Metz & Wierenga, 1997; reviewed by Altinçekiç & Koyuncu, 2012).

**a) Expression of social behaviours**

Social behaviour can differ significantly in reaction to different environmental factors (Mattiello, 2001). In domesticated goats, particularly in intensive production systems, routine management practices can restrict the expression of social behaviours. Reduced space, modifications in feeding practices, regrouping, and animal manipulations during periods as weaning and gestation, can inhibit the expression of animals' natural behaviour, often promoting stress-related responses and competition for resources, becoming the social structure of the group unstable (Miranda-de la Lama & Mattiello, 2010). Limiting access to important resources will generally increase the aggression level (Bøe, Berg & Andersen, 2006).

**Queuing at feeding and drinking**

As stated before, in order to assess social interactions 'Queuing at feeding rack' may be recorded. According to Jørgensen et al. (2007), goats are not well synchronized in their feeding behaviour, particularly in a competitive situation where a dominant animal can control the space in front of the feed barrier, demonstrating that the cost of increased competition is much higher for subordinate animals than dominants, with the latter having a high-priority access to the food (Barroso et al., 2000). Although the validity of 'Queuing at drinking' as an indicator of social behaviour has not yet been established (Ehrlenbruch et al., 2010), it is expected that results similar to 'Queuing at the feeding' may appear when the number of goats per nipple drinkers increases.

**b) Expression of other behaviours**

This criterion refers to the possibility of animals to express species-specific natural behaviours, as foraging (Welfare Quality, 2009). As a result, the expression of behaviours that are not species-specific in natural conditions, or that are performed with low frequency, can be considered as a warning sign of poor welfare (Battini et al., 2014), such as the presence of obviously dull/sick goats (Anzuino et al., 2010).

**c) Good human-animal relationship (HAR)**

**Latency to first contact and avoidance distance tests**

This criterion is defined by Welfare Quality® (2009) as “animals should be handled well in all situations, i.e., handlers should promote good human-animal relationships”.

Human-animal interaction is a regular aspect of modern intensive farming systems, and research has revealed that the interactions between stock people and their animals can limit their productivity and welfare (Lensink, Boissy & Veissier, 2000; Waiblinger, Menke & Coleman, 2002; Hemsworth, 2003). The HAR is a dynamic process that can be defined as the degree of affiliation or distance between the animal and the human, i.e., the reciprocal awareness, which develops and expresses itself in their mutual behaviour, with the catalogue of previous interactions between the animal and humans establishing the nature and perception of future interactions (Waiblinger et al., 2006). The differences in the HAR found among farms can be related to the variation in number, duration and nature of daily interactions between stock people and the animals. The nature/quality of human–animal interactions can vary from frequent, calm and friendly to infrequent and mainly negative ones (Waiblinger et al., 2002), with the prolonged evocation of fear impairing the welfare, productivity, product quality and profitability of farm animals (Waiblinger et al., 2006). In opposition, the development of a positive HAR, with low levels of fear or high levels of confidence in people, can be favourable.

Measuring animals’ reactions to humans allows us to gain insight on how they perceive specific human beings or people in general, with the several tests that have been applied to evaluate the HAR falling into three categories (reactions to a stationary human, to a moving human and responses to handling/restraint), according to the degree of human involvement (De Passillé & Rushen, 2005; Waiblinger et al., 2006). According to Battini et al. (2014), in goats the most promising tests identified so far belong to the first two main categories. Jackson and Hackett (2007) applied the latency to approach a stationary man test to examine the positive effect of a gentle handling treatment in goats, resulting in subject goats approaching the observer more quickly than control ones, and habituating to their presence faster. These results advocate that human gentling has a positive impact on the stress perceived by dairy goats, having repercussions on the welfare and productivity of these animals. This outcome is shared by other authors in relation to cattle, as latency to approach is a common measurement in literature (Jago, Krohn & Matthews, 1999; Hemsworth, Coleman, Barnett & Borg, 2000; Lensink et al., 2000).

The avoidance distance (AD) test to a moving man developed for cattle (Welfare Quality, 2009) evaluates the distance at which an animal retreats from an approaching human. Its validity as a welfare indicator has been tested in dairy cows, by correlating the flight distances to stockmen behaviour and to other human–animal tests, and it was validated in goats by Mattiello et al. (2010). According to these authors the AD test appears to be feasible in goats and allowed to identify differences depending on farm size and, accordingly, on management practices. Frequent manipulation of the animals during daily activities was an important aspect in reducing fear reactions towards humans. Behavioural tests for measuring human-animal relationship seem to be valid, feasible and reliable in numerous species, as in sheep (Napolitano, De Rosa, Girolami, Scavone & Braghieri, 2011), beef heifers (Mazurek et al., 2011), buffaloes (De Rosa et al., 2009) and dairy cows (Rousing & Waiblinger, 2004). However, time and training are essential in order to carry out the tests properly, and probably these assessments induce stress to the animals. Moreover, Muri et al. (2013) suggested that the feasibility of behavioural tests may be influenced by different breeds, or production systems.

d) **Positive emotional state**

**Qualitative Behaviour Assessment (QBA)**

The last criterion is based on the assumption that “negative emotions such as fear, distress, frustration or apathy should be avoided, whereas positive emotions such as security or contentment should be promoted” (Welfare Quality, 2009). At the moment it is commonly acknowledged that good welfare is more than the absence of negative experiences, residing primarily on the presence of positive experiences, such as pleasure (Boissy et al., 2007). If positive welfare concerns are included, it is possible to model welfare as a continuum. Nevertheless, there is still no agreement on how to assess these positive experiences, although they are believed to be a core aspect of good welfare (Fraser, 1995; Duncan, 2005). Battini et al. (2014) refer that currently, the only promising approach is the Qualitative Behaviour Assessment (QBA), which is an integrative methodology that reliably combine the detailed information collected, into a whole-animal approach of welfare (Wemelsfelder, 2007; Wemelsfelder & Mullan, 2014). The qualitative assessment of behaviour integrates and summarizes the several aspects of an animal’s dynamic interaction with its surroundings, where an observer applies descriptors such as calm, anxious, timid or confident in its assessment (Wemelsfelder & Lawrence, 2001). These descriptors have an expressive connotation reflecting on an animal's experience of a circumstance, and consequently are particularly appropriate for evaluation of its welfare (Wemelsfelder, 1997). Different research teams have applied this approach with several species as pigs, cattle, poultry, sheep, buffaloes and horses and have found good agreement between observers' assessments, even when these observers had different backgrounds and levels of experience (e.g. Minero, Tosi, Canali & Wemelsfelder, 2009; Wemelsfelder, Millard, De Rosa & Napolitano, 2009; Rutherford, Donald, Lawrence & Wemelsfelder, 2012).

# CHAPTER II – Dairy Goat Welfare- a field study

## Aims

The primary aim of this study is to describe and evaluate the application of the on-farm welfare assessment prototype for dairy goats developed by the AWIN project in Portugal. Moreover, this will allow to identify the main welfare problems affecting intensively kept dairy goats in Portugal, therefore gaining insight into the potential welfare problems that future European protocols may find in Portuguese farms. Finally, the methods and results of the present study contribute to advances in the assessment of the welfare status in adult lactating goats' scientific research field, particularly in its feasibility.

## 2.1. Material and methods

### 2.1.1. Ethics Statement

All procedures were approved by the ethics committee of University of Lisbon - *Faculdade de Medicina Veterinária* and were in agreement with the recommendations of the *DL 113/2013, 7 de Agosto*. Efforts were taken to reduce unnecessary disturbances to animals used in the prototype testing.

### 2.1.2. Farm recruitment and sampling

The study population was dairy goat farms under intensive production system. Information regarding the study population was requested from *Direcção-Geral de Alimentação e Veterinária* (DGAV), and records from a national database, *Sistema Nacional de Informação e Registo Animal* (SNIRA) for small ruminants, were obtained in the beginning of January 2014, regarding farm codes, number of animals per farm, farm locations, animals' age, animals' gender and ear tags. Subsequently, these data were entered on to a spread sheet ("WP1_farms2014" data set; Microsoft Office Excel 2013) and information management was conducted. The first step consisted on the analysis of the distribution of the number of farms according to their herd size. Not only with the purpose of obtaining a representative sample of the population, but also to conduct a stratified random one, different farm size categories were created. Thus, prior to sampling, the population was divided into mutually exclusive strata based on farm size categories. As in Portugal very small intensive dairy farms do not exist, only those with a total number of adult dairy goats above 50 were taken into consideration. Therefore following size distribution, three categories were created: [50-99], [100-499] and [>500]. The number of Portuguese sampling units to test (dairy goat farms) was pre-determined by AWIN project, resulting in 10 farms from each category, comprising a total of 30 farms. Then, within each category, a simple random sample, where every element in the study population had an equal probability of being included (Dohoo, Martin & Stryhn, 2003), was chosen.

According to DGAV (personal communication), there are a total of 3058 Portuguese dairy goat farms, with 269 of these farms being under intensive production system ([50-99], n=92; [100-499], n=161; and [>500], n=16). Ten farms from each category were selected using the "WP1_farms2014" data set, through a simple random sampling performed in Microsoft Office Excel 2013. Therefore, the target population consisted in the total national population of intensively kept adult dairy goats, in which breeds such as Murciano-Granadina, Saanen and Serrana are predominant (DGAV, personal communication).

#### 2.1.3. Preliminary considerations

Farm managers were contacted before the farm visits to discuss the visit's objectives, timetable and methods, and to specify how and for how long the farm manager and/or the stockperson would be involved. In addition, some information was collected on the particular farm, as goats' numbers and breed(s), feed sources and feeding time, type of housing and presence of outdoor grazing or exterior pen. It was also certified that the day of the visit was a regular one, to avoid for example, veterinary visits, that would disrupt the normal functioning of the routine. Lastly, security and biosecurity issues were discussed to be assured that all farm rules were followed.

#### 2.1.4. Farm visits

There were a total of six assessors with varying degree of experience working with dairy goats: two were veterinarians, two were veterinary students, one was a zootechnical engineering student, and one a biologist. Before the farm visits were conducted, to minimize differences and achieve high repeatability, the assessors were given equal training. This one consisted on classroom presentations and exercises, and then practical field assessments. The training was fostered in the beginning of January 2014.

The welfare assessment prototype was tested in 30 Portuguese intensive dairy goat farms, during the period of January to March, 2014. Each farm was visited by two trained assessors, while one conducted the welfare assessment, the other performed the general evaluation questionnaire and provided eventual help. On each farm, all data were collected on the same day and by the same person. On arrival, boots and clean (or disposal) overalls were put and boots were washed with water and disinfectant (e.g. common bleach or aldehydes' based solution), for biosecurity reasons. Subsequently, the farm safety rules were discussed with the manager (or the stockperson responsible for the animals) including areas to be aware of from a hazard point of view, as electric fences, vehicle paths, chemical storage areas, dangerous machinery, etc. During this conversation the welfare prototype was presented, including the objectives, the assessment duration, the assessors' schedules and activities and the indicators' collection order, so the farmer could acknowledge where the assessors would be at any time. Evaluations did not take place at the same time of the day at every farm, but always started at the end of feed distribution. The equipment used on field consisted on a stopwatch, checklists (Annex 3) and an animal marker.

### 2.1.5. Flow of data collection: on-farm assessments

The prototype comprehended six stages of assessment of animal-based indicators plus two final stages, when a questionnaire, consisting of two sections, was made to the farmer, so data related to management and resource-based indicators could be collected. Observations began right after feeding distribution and followed the order illustrated in Figure 4. The welfare indicators' order of assessment was rigid and was designed to ensure a continuous flow of collection, reducing the disturbance for both animals and farmers, and to guarantee that the results of the behavioural observations were not influenced by animal handling or other sources of disruption. Accordingly, all the behavioural data regarding the four initial stages were collected from outside the pen and by group observation that did not require individual animal handling, but involved counting the number of goats. Afterwards, the assessor entered the pen to evaluate human-animal relationship (stage five) and other animal-based indicators associated with the welfare principles of good feeding, housing and health (stage six).

On each farm the pen-level observations were made on a single pen of adult dairy goats, and each pen was evaluated as a whole. Assessments were always carried out without males being inside the pen, since their presence may influence the results. A checklist (Annex 3) was used to ensure that all the observations were completed in a standard order, and the time needed to collect each indicator was recorded (on-farm feasibility assessment).

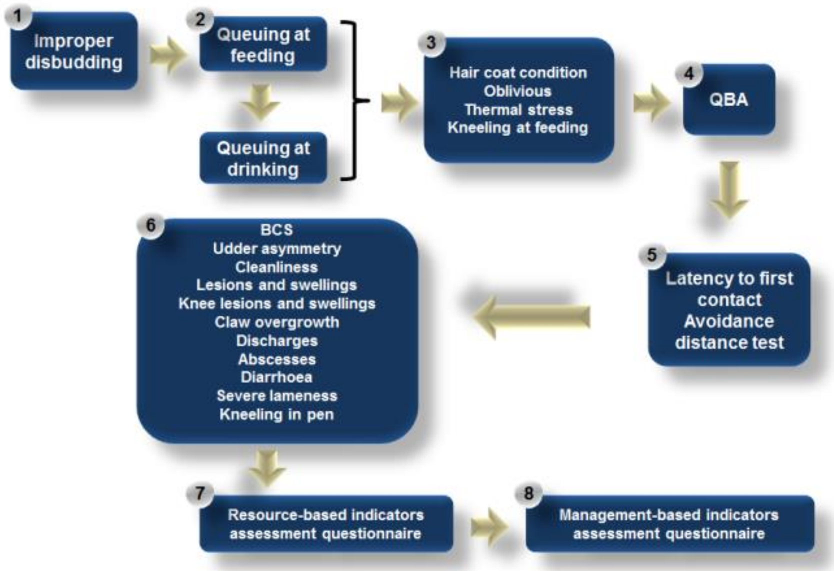



Figure 4 - Flow of data collection of the AWIN welfare assessment prototype for dairy goats.

### 2.1.6. Animal-based indicators: a description

In this sub-section the animal-based indicators scoring is presented in the order they appeared to the assessor.

**Stage 1 – Improper disbudding**

The first indicator (Figure 5) was measured from outside the pen during feeding time, with the assessor walking along the feeding rack, keeping a distance so goats would not react negatively to the human presence, and registering the number of animals with presence of some kind of horn tissue (Figure 32). Animals were considered as "improperly disbudded" (score 1) if they had been disbudded but any kind of horn tissue was visible even without head manipulation.

| | | |
|---|---|---|
| **Score 0** |  | **0 - Properly disbudded** (No evidence of horn tissue on the goat's head) |
| **Score 1** |  | **1 -Improperly disbudded** (Presence of smaller, weaker and deformed horn(s) on the goat´s head) |

Figure 5 - Assessment of 'Improper disbudding': score 0 – Properly disbudded; score 1 – Improperly disbudded (AWIN, 2014b).

**Stage 2 – Queuing at feeding and drinking**

The goats were assessed from outside the pen and the observation period started two minutes after feeding distribution, with all the feeding rack (or other feed places) being assessed simultaneously. The number of queuing goats was recorded by instantaneous and scan sampling method during 15 min/observation (1 min/scan; Figure 32). A goat was considered as queuing if it was standing within 0.5 m behind another goat that was feeding, with her head oriented towards the feed barrier. Attention was given to differentiate goats that were queuing from those that were transiting in the feeding alley.

From the moment the first goat started drinking, the number of queuing goats at water place was counted as well, for the same 15 minutes/observation (1 min/scan). All the functioning water places were recorded at the same time.

**Stage 3 - Hair coat condition, oblivious, thermal stress and kneeling at feeding rack assessment**

In stage three, moving slowly outside the pen, the assessor visually recorded the number of goats with poor hair coat condition, the number of oblivious individuals, those exhibiting symptoms of thermal stress (panting or shivering animals) and the number of goats in kneeling position at feeding rack (Figure 32).

**a) Hair coat condition**

The assessor started to locate the goats with poor hair coat condition, such as matted, rough, scurfy, uneven, shaggy hair coat and longer than normal, and recorded their number (Figure 6). The hair coat condition was assessed considering the whole body, with the exception of head and legs below the joints (knees and elbows). In case of doubt, the hair coat was compared with other goats. Factors as moulting season and the breed under assessment were taken into consideration.

**Normal hair coat**

The hair coat is shiny and sheen
The hair coat is homogeneous and adherent to the body

**Poor hair coat**

The hair coat is matted on the whole body
The hair coat is longer especially on hind quarters

Figure 6 - Assessment of 'Hair coat condition': normal hair coat and poor hair coat (AWIN, 2014b).

**b) Oblivious**

While observing the pen, the assessor identified the goats that seemed isolated, regarding not only physical, but also mental isolation, and recorded their number (Figure 7). The posture, behaviour and localization of the oblivious animals were compared to the rest of the group.

**Oblivious**

.

Figure 7 - Oblivious' assessment: a goat is isolated from the rest of the herd (mental isolation), standing immobile with its face towards the wall (AWIN, 2014b).

**c) Thermal stress**

**Shivering score**

The number of goats with (or without) signs of cold stress was recorded, focusing on hair coat on the back, postures and movement of the body. Shivering score was made in three levels: 'Score 0 – hair coat is flat on the back, no signs of cold stress are visible', 'Score 1 – the hair is bristling on the back; the goat has a thick coat' and 'Score 2 – the goat is shivering and may take a posture with arched back and head down'. Special attention was given to not include animals during agonistic interactions in score 1, as they frequently raise the hair on the back**.**

**Panting score**

The number of goats with (or without) signs of heat stress was assessed, concentrating on the goat's respiration, as goats suffering from heat stress normally have an accelerated respiration rate with open-mouth and excessive salivation. Panting score was ranked in three levels: 'Score 0 – normal respiration: the mouth is closed, the flank moves regularly (slightly visible)', 'Score 1 – elevated respiration: from slightly to moderate panting with mouth closed' and 'Score 2 – panting: from heavy to severe open-mouthed panting'. The assessor was attentive to not include animals with abnormal respiration sounds**.**

**d) Kneeling at the feeding rack**

The number of goats in kneeling position at feeding rack was recorded (Figure 8).

Figure 8 - Kneeling at the feeding rack' assessment: a goat is kneeling at the feeding rack to reach the feed (AWIN, 2014b).

**Stage 4 – Qualitative Behaviour Assessment (QBA)**

In stage 4, the goats were observed from outside the pen and the assessment was conducted on the whole pen, at least 30 min after the feed distribution, by selecting the suitable observation points and, consequently, the timing of the observations. The selection of these points was made regarding the different structures of the housing environment (e.g. deep straw barn, outside field, pens of different sizes in different areas or corners of the farm). The number of observation points depended on the complexity of the housing environment and the group size, with observation sessions lasting from 10 to 20 minutes and with the time spent at each observation point ranging from 2.5 (8 points) to 10 minutes (1 or 2 points). This assessment took place during activity periods of goats, where different behavioural expression might be exhibited. At the end of the observation period, the assessor rated a list of 13 descriptors (as 'aggressive', 'alert', 'content', 'relaxed', among other) using the visual analogue scale (VAS) according to the overall general behaviour of the pen (Figure 33).

**Stage 5 - Group assessment on human-goat relationship**

At this stage, the assessor entered the pen to perform tests to evaluate the human-animal relationship. These assessments consisted in measuring the latency in seconds of the first goat that entered in contact with the test person (latency to first contact), and the number of goats that could be contacted or that accepted to be gently stroke (avoidance distance test; Figure 34).

**a) Latency to first contact test**

The assessor approached the pen and waited at the gate for 30 seconds before entering it. Inside the pen, walked to a pre-decided place located in the middle of the pen, and the test started as soon as this place was reached, with the stopwatch being launched. The assessor stood motionless with his back to the wall, looking around the pen or at the ground but not at the goats (Figure 9). The test finished when the first goat entered in contact with any part of the body of the assessor. If no goats entered in contact, the test was capped at 300 seconds.

Figure 9 – Latency to first contact test: the test person stands motionless waiting to the first goat to enter in contact with any part of his body (AWIN, 2014b).

**b) Avoidance distance test**

Inside the pen, the assessor stood in front of a single animal at a distance of 200 cm, establishing a reciprocal visual contact. The goat should be on the opposite side of the assessor or at most, 45° twisted. After establishing a visual contact the test began: the assessor moved slowly towards the animal at a speed of one step/sec, 60 cm/step, with the arm lifted with an inclination of 45° and the hand palm directed downwards, without looking into the animal's eyes, but looking at the muzzle (Figure10). With the purpose of approaching as many goats as possible, the assessor walked from one side of the pen to the other without stopping for more than 10 seconds. When reaching the opposite side, the test finished. The number of goats that could be contacted or that accepted to be gently stroke was recorded.

Figure 10 – Avoidance distance test: the test person tries to approach as many goats as possible by walking from one side to the other of the pen. The test includes two levels: 1) contact; 2) acceptance (AWIN, 2014b).

**Stage 6 – Clinical Scoring**

Detailed individual observations were then carried out using a sample of goats. Welfare indicators as 'Body condition score', 'Udder asymmetry', 'Cleanliness', 'Lesions and swellings (LAS)', 'Abscesses', 'Claw overgrowth', 'Discharges' and 'Diarrhoea' were recorded (Figure 35).

With the purpose of reducing the execution time within an acceptable timeframe and minimize disturbance, it was adopted a sampling strategy similar to Welfare Quality® for dairy cows (Welfare Quality, 2009). This strategy involved the inspection of a number of animals proportional to the pen size, with percentages ranging from 100% of the subjects (in pens with fewer than 30) to a minimum of 25% of the subjects, in the pens with more than 150. The same animals were used to collect all the indicators included in this stage, with both sides (left and right) being considered. Each welfare indicator, except 'BCS' and knee lesions, was scored using a binary assessment system (present or absent). After being restrained and individually observed each goat was physically marked on the tail base, and the group was assessed for severe lameness and kneeling position in pen. The descriptive criteria used to assess each indicator is presented below.

## a) Body condition score

Body Condition assessment was carried out on an individual goat, from behind, observing the rump region and by applying a validated visual body condition scale developed by AWIN (2014). An adult dairy goat could be scored according to three levels: '-1 - Very thin', '0 - Regular' or '1- Very fat' in the pictorial body condition scale (Figure 11).

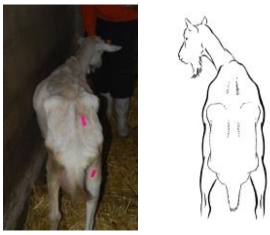
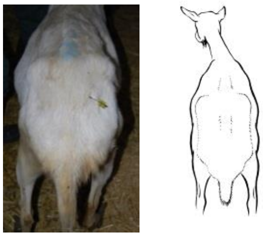
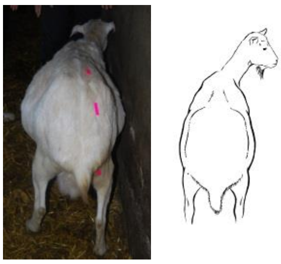
| | | |
|---|---|---|
| **Very thin** |  | General condition: Raw or slightly-raw boned goat, with backbone and some ribs visible.<br>Rump region: Hip and pin bones are prominent. The line that connects the hip bone and the thurl assumes a markedly concave shape. There is little muscle and/or fat between the skin and bone structures. |
| **Normal** |  | General condition: Backbone not prominent but still visible and ribs difficult to assess visually.<br>Rump region: Hip and pin bones still visible, but not prominent. The line that connects the hip bone the thurl assumes a slightly concave or straight shape. It is possible to realize some muscle and/or fat between the skin and bone structures. |
| **Very fat** |  | General condition: Backbone and ribs not visible. Goat has a rounded appearance, sometimes with abdominal fat deposits visible.<br>Rump region: Hip and pin bones are difficult to identify. The line that connects the hip bone the thurl assumes a slightly or markedly convex shape. The entire rump region is coated by muscle and fat, contributing to the rounded appearance of the goat. |

Figure 11 - Three point pictorial scale to assess BCS in dairy goats, developed by AWIN (2014b).

## b) Udder asymmetry

The goats were individually assessed from behind, making sure there was a good visualization of the udder. According to Figure 12, the animals were assessed considering two levels, with all animals in which one half was at least 25% longer that the other (excluding the teats) being considered as "with udder asymmetry" (score 1).

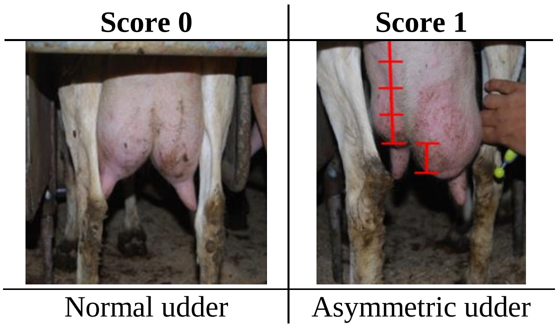



Figure 12 – ‘Udder asymmetry’ assessment: normal udder (score 0) and asymmetric udder (score 1; AWIN, 2014b).

**c) Cleanliness**

While assessing cleanliness on an individual animal, three regions were evaluated: hindquarters, lower legs (front and hind legs – above the coronary band) and udder. All regions were assessed on both sides of each goat. On each side, the assessment started from the rear to the front, in order to minimize stress to the animal. In all the three regions, the assessment was performed in two levels: ‘Score 0 - No dirt or minor splashing’ or ‘Score 1 - Wet hair, separate or continuous plaques of dirt’.

**d) Lesions and swellings**

Lesions and swellings assessment was done on an individual animal, regarding both sides, and considering the following regions: hindquarters, lower legs, body (area other than the head, neck, lower legs, udders and knees), neck (shoulders to base of head), head (including periorbital and nasal areas, and ear tear). The assessment was carried out in two levels: ‘Score 0 - Absence of lesions/swellings (skin damage with/without hair loss)’ or ‘Score 1- Presence of lesions/swellings (skin damage with/without hair loss)’.

**e) Knee lesions and swellings**

While assessing this indicator on an single animal, as shown in Figure 13, the knee region was graded in three levels: ‘Score 0 - No lesions, hair loss or skin thickening’, ‘Score 1 - Skin damage with/without hair loss and reddened skin, but no enlargement of any joint’ or ‘Score 2 - Skin damage with hair loss and enlargement of at least one joint; a thick callus is present over one or both knees’.

| Score 0 | Score 1 | Score 2 |
|---|---|---|
| No lesions, hair loss or skin thickening**.** | Skin damage with/without hair loss and reddened skin, but no enlargement of any joint. | Skin damage with hair loss, and enlargement of at least one joint, showing a thick callus. |

Figure 13 – ‘Knee lesions and swellings’ assessment: the knee region was ranked in three levels - score 0, score 1 and score 2 (AWIN, 2014b).

**f) Abscesses**

Abscesses assessment was done regarding both sides of each animal, and considering the following regions: hindquarters, udder, body (area other than the head, neck, hind quarter, lower legs, udder and knees), neck (shoulders to base of head) and head. Each goat had to be carefully observed, paying special attention to the main superficial lymph nodes and bearing in mind that abscesses could appear in three different stages: not ruptured, ruptured and scar. The assessment was performed in two levels: ‘Score 0 – Absence of abscesses’ or ‘Score 1 – Presence of abscesses’ and no distinction was made between the different stages.

**g) Claw overgrowth**

The four claws of the animal were assessed without lifting the limbs of the ground. The animals were recorded as ‘Score 0 - without severe overgrown claws’ or ‘Score 1 - with severe overgrown claws’. An animal was considered as ‘Score 1 - with severe overgrown claws’ if it presented at least one rear claw with severe overgrowth (Figure 14).

| Score 0 | Score 1 |
|---|---|
|  |  |
| Normal claw - Perfectly healthy claw. | Claw with severe overgrowth. |

Figure 14 – ‘Claw overgrowth’ assessment: ‘Score 0 - without severe overgrown claws’ or ‘Score 1 – with severe overgrown claws’ (AWIN, 2014b).

**h) Discharges**

**Ocular discharge**

The animals were assessed from the front. Considering ocular discharge can be defined as any discharge from the eye, an individual dairy goat was scored in two levels: ‘Score 0 - Absence of ocular discharge’ or ‘Score 1 – Presence of ocular discharge’.

**Nasal discharge**

Any unilateral or bilateral mucous or purulent discharge observed inside the nostrils or hanging from the nose was registered. The scoring of the animals was performed in two levels: ‘Score 0 - Absence of nasal discharge or a watery discharge’ or ‘Score 1 - Presence of mucous or purulent nasal discharge’.

**Vulvar discharge**

The animals were assessed from behind, making sure there was a good visualization of the vulva, perineum, tail area and back legs. An individual animal was recorded in to two levels: ‘Score 0 – Absence of vulvar discharge’ or ‘Score 1 – Presence of vulvar discharge’.

**i) Diarrhoea**

Considering diarrhoea as loose watery manure around and below the tail head, the goats were evaluated in two levels: ‘Score 0 – Absence of diarrhoea’ or ‘Score 1- Presence of diarrhoea’. The animals were assessed from behind, in order to obtain a good observation point of the area under and on both sides of the tail.

**j) Severe lameness**

To assess the number of severely lame animals, the assessor walked slowly in the pen, visualizing all the animals and counting the number of animals that presented an extremely irregular gait in time and space, severe head nodding and an accentuated arched rump. All the goats lying down were forced to stand and walk at least a few steps. The animals were assessed considering two levels: ’Score 0 - Non-lame, slightly or moderate lame goats’ or ‘Score 1 -Severely lame goats’.

**k) Kneeling in the pen**

The number of goats in kneeling position in the pen was recorded.

As a final part of the visit, a questionnaire (consisting of two sections) was made to the farmer to gather data concerning management and resource-based indicators (Figure 36). General information on management procedures, farm structures and routines, prevention of disease, feeding, hygiene, production and mortality records was obtained.

**Stage 7 – Resource-based indicators assessment questionnaire**

This section of the questionnaire was related to general characteristics of facilities, number of workers on the farm, breeds present at farm, replacement goats per year, among other.

**Stage 8 - Management-based indicators assessment questionnaire**

Lastly, data related to management procedures (e.g. pen grouping strategy, disbudding/dehorning routines, frequency of claw trimming), production and reproduction data (e.g. average age within milking goats, average number of days in milking , average annual charge of somatic cells) were collected.

Once all the data had been gathered on the dairy farms, the prevalence of each animal-based indicator was calculated at farm and farm category level, to evaluate their suitability for an on-farm welfare assessment prototype following a stepwise approach. Thus, a threshold was set at a prevalence of 5%, and indicators were categorised as most or less prevalent, accordingly.

### 2.1.7. Reliability studies

#### 2.1.7.1. Inter-observer reliability (IOR)

With the purpose of testing inter-observer reliability (IOR), in 10 of these farms the assessments were performed simultaneously by two assessors.

#### 2.1.7.2. Consistency of the indicators over time (COT)

In the beginning of July, to investigate consistency over time (COT) of the animal-based indicators, 10 of the 30 farms were revisited by the same assessor who executed the initial assessment. All the farms visits followed the same prototype methods.

### 2.1.8. Data management and statistical analysis

**General outline**

Data were entered, compiled and statistically analysed using SPSS v22 (IBM® SPSS® Statistics, NY, USA). To perform an initial data analysis, the prevalence of each indicator was determined at farm level and categorised according to the 5% prevalence threshold established. Subsequently, farms were clustered in the three categories considered: small farms (>50 and <100 adult dairy goats, n= 10); medium farms (>100 and <500 adult dairy goats, n = 10); and large farms (>500 adult dairy goats, n = 10), and indicators prevalences were determined at farm category level. Prevalences of animal-based indicators of feeding, housing, health and avoidance distance test were expressed as the proportion of animals/farms affected on the total of animals/farms assessed. Measures of central tendency (mean and median) and dispersion (minimum, maximum, percentiles and standard deviation) were determined for each indicator and presented in tables (Annex 1) and figures to summarize and describe the collected information.

For the purpose of QBA data analysis, for each of the descriptors, the distance from minimum to where the assessor ticked the VAS scale was measured in mm. Data were submitted to statistical analysis with ‘farm size category’ as experimental unit. In order to summarize the 13 QBA descriptors, a Principal Component Analysis (PCA) was carried out using a correlation matrix with no rotation. PCA forms the basis for multivariate data analysis (Jackson, 1991), and its objective is to extract the important information from the data set and to express this information as a set of new orthogonal variables called principal components (Abdi & Williams, 2010). These principal components are ordered so that the first few retain most of the variation present in all of the original variables. PCA provides an alternative set of coordinate axes given by the principal components representing the original data set (Coleman, 2010).

### Relationships between animal-based welfare indicators

Spearman's rank correlations ($r_s$) were used to determine relationships between animal-based welfare indicators. Only indicators that were significant at the 0.05-level were considered and only associations that were relevant, biologically conclusive or of special importance regarding animal welfare were taken into account. According to Martin and Bateson (2007), an $r_s$ value of 0.4–0.7 points to a moderate correlation and values above 0.7 indicate a high correlation (Table 2).

| Value of correlation coefficient ($r_s$) | Informal Interpretation |
|---|---|
| **< 0.2** | Slight - almost negligible relationship |
| **0.2-0.4** | Low correlation - definite but small relationship |
| **0.4-0.7** | Moderate correlation - substantial relationship |
| **0.7-0.9** | High correlation - marked relationship |
| **0.9-1.0** | Very high correlation - very dependable relationship |

Table 2 - Informal phrases used to interpret correlation coefficients of different sizes (from Sprinthall, 2003; reviewed by Martin & Bateson, 2007).

### Inter-observer reliability

Inter-observer reliability is defined as the agreement between different assessors separately rating the same individual, and refers to the relative measurement error, i.e., the variation between individuals as measured by different observers in relation to the total variance of the measures (Streiner & Norman, 2008). Agreement provides insight into the capability of a measure to produce the same value on multiple occurrences and indicates absolute measurement error (De Vet et al., 2006).

For categorical data, the inter-observer reliability between two observers was tested using kappa (κ) and weighted kappa ($\kappa_w$) coefficients (Cohen, 1968). Kappa (κ) consists in a measure of "true" agreement that reflects the proportion of agreement fully chance corrected. Weighted kappa ($\kappa_w$) penalizes disagreements in terms of their seriousness, whereas unweighted kappa (κ) handles all disagreements equally not taking order of categories into account, thus, being inappropriate for ordinal scales (Cohen, 1968). The quadratic weighting scheme, where disagreement weights are proportional to the square of the deviation of individual ratings (Brenner & Kliebsch, 1996), was used.

For continuous data, intra-class correlations (ICCs) were calculated with a two-way mixed effects model (Shrout & Fleiss, 1979), i.e., the subjects in the study were considered to be random but the observers (raters) were not random effects, with absolute agreement and consistency being estimated.

The ICC is a commonly used statistics for assessing inter-observer reliability for ordinal, interval, and ratio variables. There are several ICC variants[2] that must be selected based on the nature of the study and the type of agreement the researcher wishes to estimate (Hallgren, 2012).

The lowest limit of 0.4 for κ values (Fleiss et al., 2003) and Landis and Koch threshold values for $\kappa_w$ (Table 3) were assumed (Landis & Koch, 1977). Estimates for ICC were interpreted using Shrout (1998) guidelines (Table 4). All analyses were conducted using the R statistical language (R Development Core Team, 2013) with base, "irr" (Gamer et al., 2012) and "psych" (Condon & Revelle, 2014) packages.

| **Fleiss threshold values for κ** | **Agreement** |
|---|---|
| **0 - 0.40** | Poor |
| **0.41–0.75** | Fair to good |
| **0.76–1** | Excellent |
| **Landis and Koch threshold values for $\kappa_w$** | **Agreement** |
| **< 0** | Poor |
| **0.00–0.20** | Slight |
| **0.21–0.40** | Fair |
| **0.41–0.60** | Moderate |
| **0.61–0.80** | Substantial |
| **0.81–1** | Almost perfect |

Table 3 - Interpretation of agreement beyond chance by Kappa (Fleiss et al., 2003) and by weighted Kappa (Landis & Koch 1977).

| **Interpretation of ICC estimates** | **Agreement** |
|---|---|
| **0–10%** | Virtually none |
| **11–40%** | Slight |
| **41–60%** | Fair |
| **61–80%** | Moderate |
| **81–100%** | Substantial agreement |

Table 4 - Interpretation of ICC estimates using Shrout (1998) guidelines.

Pearson's correlation coefficient (r) was used to determine inter-observer reliability for QBA's dimensions. For descriptors, Spearman's rank correlations ($r_s$) were applied assuming, for both, Martin and Bateson (2007) thresholds (Table 2).

**Consistency over time of animal-based indicators**

A Wilcoxon signed rank test was used to test whether the prevalences obtained during the two visits were significantly different at the 0.05-level. Investigations on the correlation of animal-based welfare indicators between two consecutive farm visits were done. For each animal-based indicator, Spearman's rank correlations ($r_s$) between the two visits at the same 10 farms

[2] For a comprehensive analysis on the ICC variants see, e.g. Shrout and Fleiss (1979).

were calculated. Consistency of the animal-based indicators was found acceptable if correlation coefficients were equal or exceeded the threshold of 0.7 (Martin & Bateson, 2007).

**Analysis of the seasonal visits by farm category**

An analysis, concerning farm categories, of the most prevalent indicators of the two consecutive visits at the same ten farms was performed.

**On-farm feasibility assessment: duration of the prototype**

The mean time necessary to perform each stage of the prototype was recorded in 26 farms.

## 2.2. Results

### 2.2.1. Prototype application in Portuguese dairy goat farms

From January to March 2014, the welfare assessment prototype was tested in adult dairy goats kept under intensive conditions in 30 Portuguese farms. Nine of these farms were situated in the Centre region, nine were located in Lisboa e Vale do Tejo, seven in the Alentejo and five in the North region. As shown in Figure 15, most of the large and medium size farms visited were placed in Centre and Lisboa e Vale do Tejo regions, and half of the small farms visited were located in the North region.

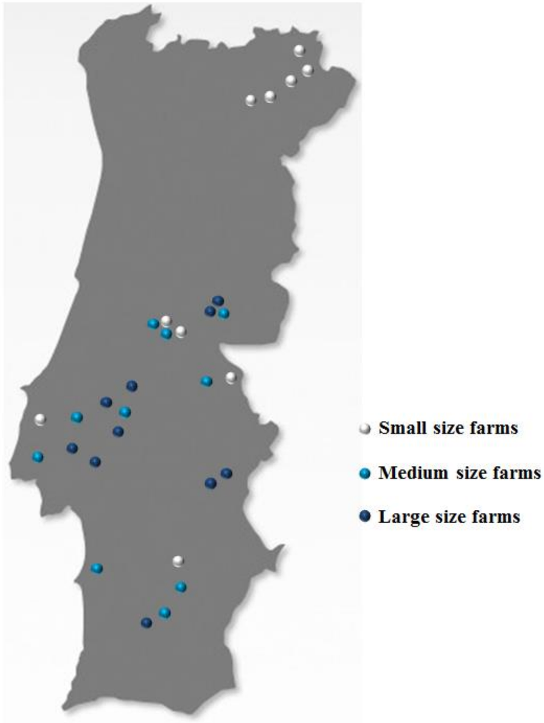



Figure 15 - Location of the farms where the prototype was tested, organized by farm category.

All farms had an indoor production system on concrete floor, soil or grit with straw as bedding material. In 23 farms there was the presence of outdoor grazing or exterior pen, where the goats had the opportunity to exercise. Only in three farms the milking processed manually.

The number of adult dairy goats on each farm ranged from 50 to 2000 animals, with a mean (±SD) of 292 (±410) goats. With regard to farm categories, the average herd size in small, medium and large farms, was 79 (±17), 309 (±74) and 834 (±451) adult dairy goats, respectively (Table 5).

| Farm size category | Number of adult dairy goats | | | |
|---|---|---|---|---|
| | Mean | Std. Deviation | Min | Max |
| **Small farms** | 78.9 | 17.3 | 50 | 100 |
| **Medium farms** | 309.1 | 73.7 | 200 | 440 |
| **Large farms** | 833.5 | 451.4 | 500 | 2000 |

Table 5 - Average number of adult dairy goats in the 30 Portuguese farms visited during AWIN prototype trial period.

On these farms, detailed individual observations were carried out on 1172 adult dairy goats and pen-level observations on 2715 animals. The average number of animals in the assessed pen was 113.1 ± 83.9 adult dairy goats and the animal sample ranged from 30 to 55 animals. During the days of assessment, environment temperatures ranged from 7ºC to 25ºC and relative humidity from 43% to 93%.

### 2.2.2. Final outcomes

As stated previously, initially, data from the 30 visited farms were analysed in order to determine the prevalence of each indicator, and categorise it according to the pre-determined threshold (5%). A preliminary analysis of the collected data showed that there was no evidence of statistical difference in the amount of lesions, abscesses and dirtiness scored on the right and left side of the animal (Table 14). Therefore, the presented prevalences of the indicators related to 'Abscesses', 'Lesions and swellings (LAS)' and 'Cleanliness' in this study are regarding the left side of the animals.

**a) Most prevalent indicators**

As graphically illustrated in Figure 16, starting from pen-level observations, the most prevalent (>5%) animal-based indicators were 'Queuing at feeding' (QF), 'Queuing at drinking' (QD), 'Hair coat condition' (HC) and 'Improper disbudding' (ID), showing prevalences between 5-27%. Regarding individual assessment, the most prevalent indicators varied from 5% ('BCS – very thin'; VT) to 83% ('Knee lesions and swellings (LAS) – score 1'; KLASS1). 'Claw overgrowth' (CO) and 'Head LAS' (HLAS) prevalence values were around 34% and 26%, respectively. 'Cleanliness – hindquarters' (HQC), 'BCS – very fat' (VF), 'Cleanliness - lower legs' (LLC),'Body LAS' (BLAS), 'Neck LAS' (NLAS), 'Knee LAS – score 2' (KLASS2),'Body abscesses' (BA) and 'Hindquarters LAS' (HQLAS) presented prevalences between 10-20%. 'Lower legs LAS' (LLLAS), 'Head abscesses' (HA), 'Udder asymmetry' (UAS),'Ocular discharge' (OD) and ' Nasal discharge' (ND) prevalence values were below 10%. Further details are given on Table 15 (Annex 1).

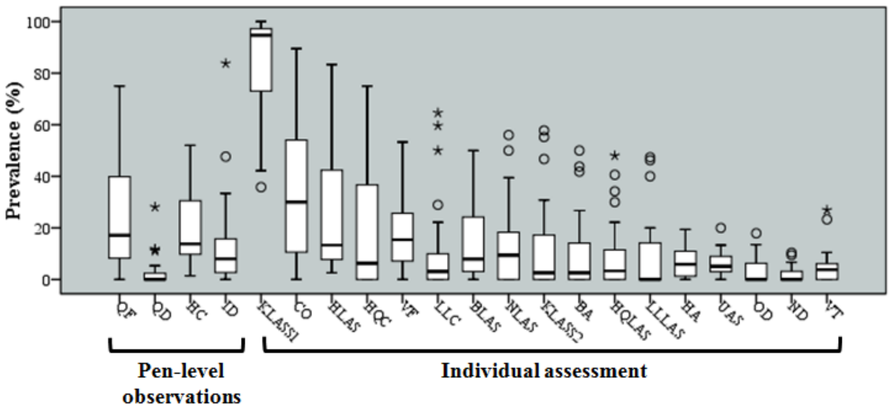



Figure 16 - Variation in farm level prevalences on the 30 Portuguese visited farms: most prevalent animal-based indicators of the AWIN prototype for dairy goats. Points (o) and asterisks (*; extreme values) represent outliers. Key to acronyms: QF, queuing at feeding, QD, queuing at drinking, HC, hair coat condition, ID, improper disbudding, KLASS1, knee lesions and swellings - score 1, CO, claw overgrowth, HLAS, head lesions and swellings, HQC, hindquarters cleanliness, VF, very fat animals, LLC, lower legs cleanliness, BLAS, body lesions and swellings, NLAS, neck lesions and swellings, KLASS2, knee lesions and swellings - score 2, BA, body abscesses, HQLAS, hindquarters lesions and swellings, LLLAS, lower legs lesions and swellings, HA, head abscesses, UAS, udder asymmetry, OD, ocular discharge, ND, nasal discharge and VT, very thin animals.

**b) Less Prevalent indicators**

The less prevalent indicators (<5%) in the 30 farms, as depicted from Figure 17, regarding pen-level observations, were ‘Shivering score 1’ (SS1), ‘Shivering score 2’ (SS2), ‘Panting score 1’ (PS1), ‘Panting score 2’ (PS2), ‘Avoidance distance (AD) – acceptance’ (AD- A), ‘AD – contact’ (AD– C), ‘Oblivious’ (O), ‘Kneeling at feeding’ (KF), ‘Kneeling in pen’ (KP) and ‘Severe lameness’ (SL), with prevalences between 0-3%. At individual level, indicators as ‘Neck abscesses’ (NA), ‘Cleanliness – udder’ (UC), ‘Diarrhoea’ (D) and ‘Udder abscesses’ (UA) showed prevalence values of around 3-4%. Hindquarters abscesses’ (HQA) and ‘Vulvar discharge’ (VD) presented prevalences below 1%. Additional information on the considered indicators is shown in Table 16 (Annex 1).

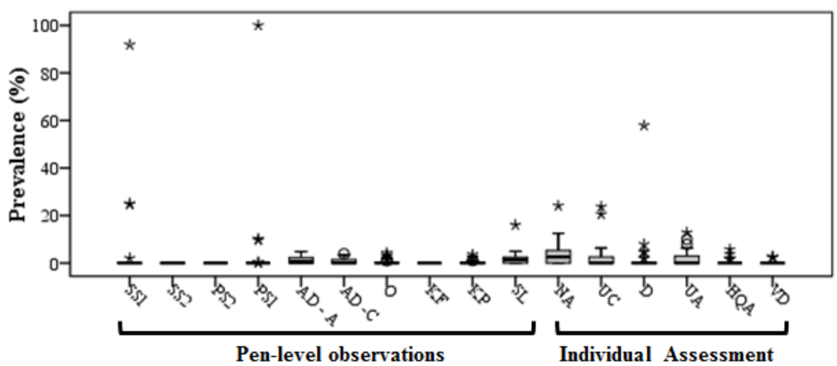



Figure 17 - Variation in farm level prevalences from the 30 recorded farms, where the AWIN prototype for dairy goats was applied: less prevalent animal-based indicators. Points (o) and asterisks (*; extreme values) represent outliers. Key to acronyms: SS1, shivering score 1, SS2, shivering score 2, PS1, panting score 1, PS2, panting score 2, AD – A, avoidance distance – acceptance, AD – C, avoidance distance - contact, O, oblivious animals, KF, kneeling at feeding, KP, kneeling in pen, SL, severe lameness, NA, neck abscesses, UC, udder cleanliness, D, diarrhoea, UA, udder abscesses, HQA, hindquarters abscesses and VD, vulvar discharge.

**c) Latency to first contact**

Considering ‘Latency to first contact’, the mean time to contact was 125.9 ± 129.2s (2.1 ± 2.2min) ranging from 0 to 300s (5min; Table 17).

### 2.2.3. The overall results from Portuguese dairy goat farms by farm categories

Subsequently, data were submitted to statistical analysis to examine how the prevalences of these 24 indicators varied along the three farm categories.

**a) Most Prevalent Indicators**

The prevalences of the indicators differed across farm categories, showing highest values in large farms, in general. Considering the most prevalent indicators, 'Queuing at feeding' oscillated from 13% in medium farms to 40% in large farms. 'Queuing at drinking' had zero prevalence in small farms, and presented values around 6% in medium and large farms. 'Improper disbudding' showed prevalence values around 12-13% in small and large farms, and around 23% in medium farms (Figure 28). 'Poor hair coat condition' had similar values along the three farm categories, as shown in Table 18 (Annex 1). Regarding individual assessments, 'Knee LAS – score 1' prevalence was about 94% in small farms, and revealed values around 79% in medium and large farms, 'Claw overgrowth' presented a lower prevalence value in small farms (about 12%), with medium and large farms having values around 38-42% (Figure 29). Moreover, that was also verified for 'Hindquarters LAS', 'Head LAS', 'Neck LAS' and 'Ocular discharge' prevalences. 'BCS – very fat' showed a higher prevalence in large farms, about 25%, with small and medium farms presenting values around 13 -17% (Figure 18). Whilst, BCS – very thin' presented prevalences near 5% in the three farm categories. Furthermore, 'Cleanliness – lower legs', 'Body LAS' and 'Udder asymmetry' also presented identical prevalences along the three farm categories.

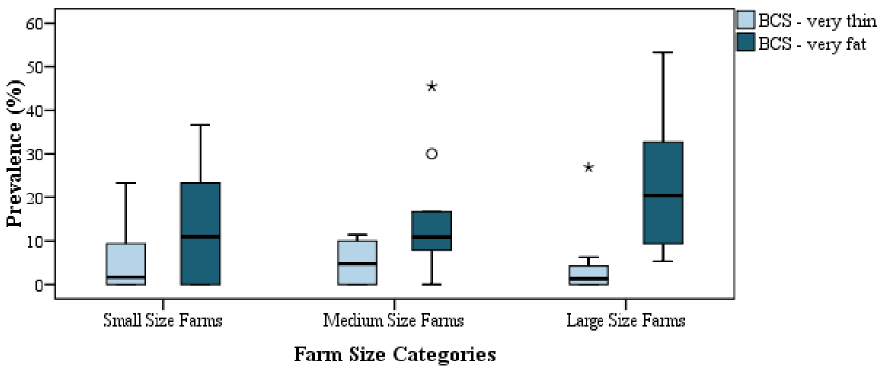



Figure 18 - Variation of 'BCS – very thin' and 'BCS – very fat' prevalences in the 30 Portuguese assessed farms, where the AWIN prototype for dairy goats was tested, organized by farm size categories. Points (o) and asterisks (*; extreme values) represent outliers.

'Cleanliness – hindquarters' prevalence varied from 9% (medium farms) to 30% (large farms), 'Knee LAS – score 2' from 5% (small farms) to 17% (large farms) and 'Lower legs LAS' from 0.7% (small farms) to 16% (large farms). 'Body abscesses' and 'Head abscesses' showed prevalences around 3-9% in small and medium farms, and around 13-14% in large farms (Figure 19). 'Nasal discharge' presented a prevalence value of about 8% in medium farms, while small and large farms showed prevalences below 5% (Figure 30). Detailed information concerning the variation of these indicators among farm categories is presented on Table 18 (Annex 1).

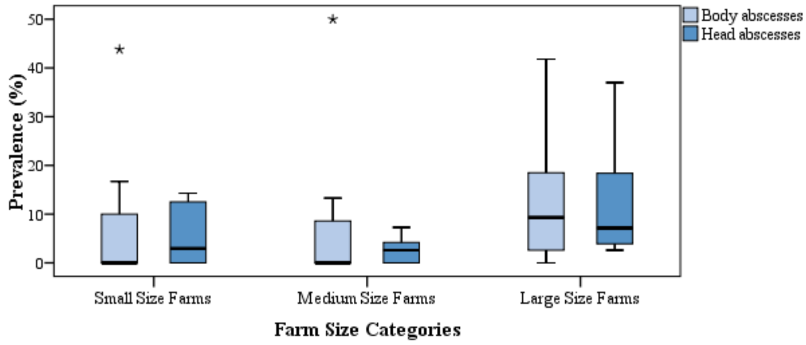



Figure 19 – Variation of 'Body abscesses' and 'Head abscesses' prevalences in the 30 Portuguese assessed farms, where the AWIN prototype for dairy goats was applied, clustered by farm size categories. Asterisks (*; extreme values) represent outliers.

### b) Less Prevalent Indicators

Each indicator categorised as less prevalent presented identical values across the three farm categories considered, except for 'Panting score 1', 'Shivering score 1', and 'Diarrhoea'. 'Panting score 1' and 'Shivering score 1' revealed higher prevalences in small farms (11 to 14%), and similar values in medium and large farms. Diarrhoea prevalence varied from 0.3%, in small farms, to 6%, in medium farms (Figure 31). 'Severe lameness' prevalence was below 2%, in small and medium farms, and reached a value of around 3% in large farms (Figure 20). The prevalence of animals that accepted to be gently stroke (acceptance; range 0% to 4.8%) or that could be contacted (contact; range 0% to 3.9%) was higher among large size farms, 2.1% and 1.9%, correspondingly. The prevalence of these indicators is given on Table 19 (Annex 1), according to farm category.

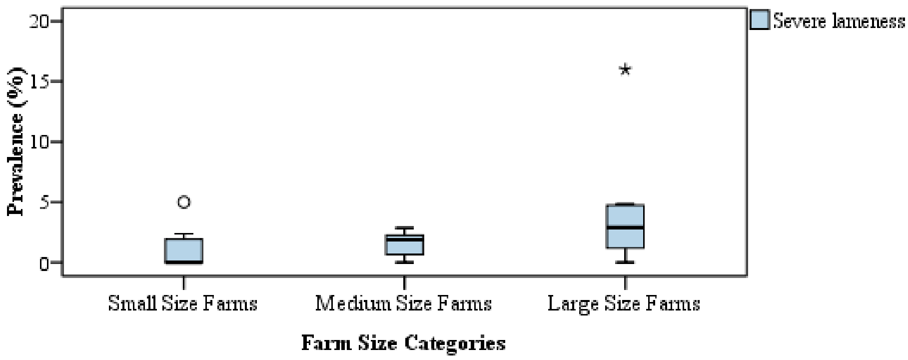



Figure 20 - Variation of 'Severe lameness' prevalence in the 30 Portuguese visited farms, where the AWIN prototype for dairy goats was tested, organized by farm size categories. Points (o) and asterisks (*; extreme values) represent outliers.

#### c) Latency to first contact

As shown in Figure 21, the mean time needed to perform 'Latency to first contact' test was inferior in large farms (81.1 ± 117s; 1.4 ± 2 min) ranging from 0 to 300s. In small farms the mean time was 139.7 ± 139s (2.3 ± 2.3min; range 10 to 300s) and in medium farms 156.9 ±131.4s (2.6 ± 2.2min; range 7 to 300s; Table 17).

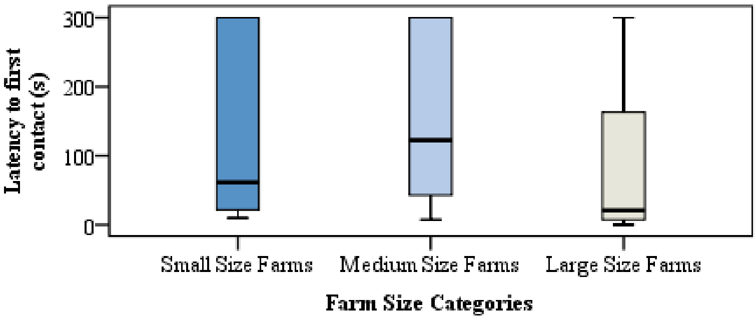



Figure 21 – Variation of 'Latency to first contact' test in the 30 Portuguese farms where the AWIN prototype for dairy goats was performed, organized by farm size categories.

#### d) QBA

Principal components analysis with extraction of two components was performed. The two factors explained 27.6% and 17.7% of the variance. To comply with the standardized way of analysing QBA data, no rotation was performed. A loading plot showing the relationship among the dairy goat QBA descriptors is given in Figure 22.

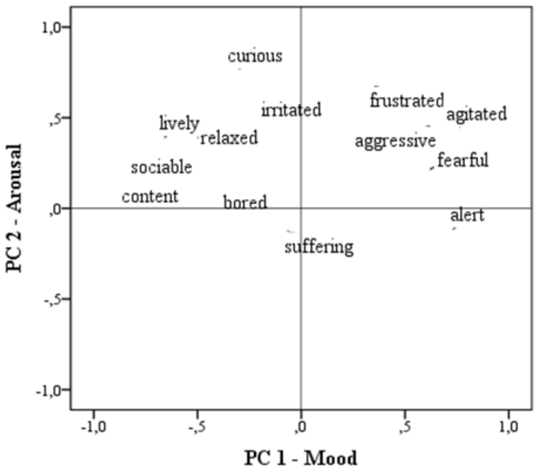



Figure 22 - Word chart of the QBA assessed in the 30 dairy goat farms. This 2-dimensional loading plot shows the relationship among the 13 QBA descriptors representing dairy goat behaviour on the two principal PCA dimensions.

Descriptors as agitated (0.76), alert (0.72), aggressive (0.63), fearful (0.61), as well as lively (-0.67), sociable (-0.65), content (-0.58) and relaxed (-0.52), presented the highest loadings on the first axis. The second axis was characterised with descriptors as curious (0.75), frustrated (0.68), irritated (0.63), as well as alert (-0.13), suffering (-0.12) and bored (-0.05).

The position of each farm category on the basis of the QBA was plotted in Figure 23, showing a homogenous overall distribution of farms throughout the two axes and presenting little dispersion. However, there is a small farm presenting a higher score in the first axis and a medium and a large farm showing higher scores in the second axis.

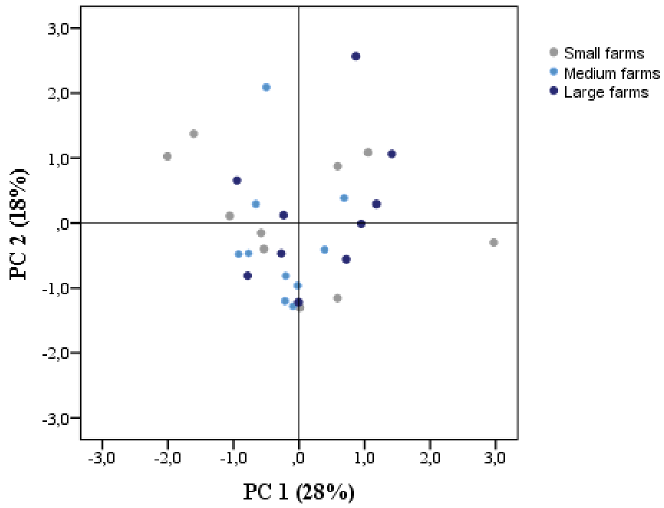



Figure 23 - QBA analysis of dairy goats' behaviour among the three farm size categories considered in the present study.

### 2.2.4. Relationships between animal-based welfare indicators

Spearman's rank correlations ($r_s$) were used to evaluate the relationship between the collected indicators, associating individual and pen-level observations.

There is a positive correlation between dirty hindquarters and dirty lower legs ($r_s$=0.52, P=0.004), similar to 'Knee LAS - score 2' and 'Cleanliness – lower legs' ($r_s$=0.38, P=0.04), and to 'Knee LAS - score 2' and 'BCS - very thin' ($r_s$=0.48, P=0.008). Furthermore, 'Claw overgrowth' is significantly correlated with 'Severe lameness' ($r_s$=0.51, P=0.004) and 'Knee LAS - score 2' ($r_s$=0.58, P=0.001).

Hindquarters lesions are significantly correlated with lower legs ($r_s$=0.60, P=0.001), body ($r_s$=0.78, P<0.001), neck ($r_s$=0.84, P<0.001) and head lesions ($r_s$=0.67, P<0.001). Moreover, neck lesions are correlated with head ($r_s$=0.69, P<0.001) and body lesions ($r_s$=0.64, P<0.001).

Nasal discharges are positively correlated with ocular discharges ($r_s$=0.51, P=0.004). Additionally, 'BCS – very thin' is correlated with nasal ($r_s$=0.40, P=0.031) and ocular discharges ($r_s$=0.42, P= 0.025). Neck abscesses are moderately correlated with head ($r_s$=0.59, P=0.001) and body abscesses ($r_s$=0.46, P=0.013), similar to hindquarters and head abscesses ($r_s$=0.58, P= 0.001). Moreover, body abscesses are correlated with very thin animals ($r_s$=0.55, P=0.002).

Regarding farm size categories, Spearman's rank correlation analysis revealed that in small size farms, nasal discharges are very highly correlated with ocular discharges ($r_s$=0.91, P<0.001). In addition, in these farms there is a high correlation between very thin animals and nasal ($r_s$=0.70, P<0.024) and ocular discharges ($r_s$=0.82, P=0.004). ‘Hindquarters LAS’ is highly correlated with and ‘Body LAS’ ($r_s$=0.80, P=0.006) and ‘Lower legs LAS’ ($r_s$=0.72, P=0.018), similar to ‘Lower legs LAS’ and ‘Neck LAS’ ($r_s$=0.72, P=0.018). Neck abscesses are highly correlated with head abscesses ($r_s$=0.71, P=0.022) and ‘Body abscesses’ is highly correlated with ‘BCS – very thin’ ($r_s$=0.89, P=0.001). Similarly, there is a high correlation between ‘Improper disbudding’ and ‘Head LAS’ ($r_s$=0.81, P=0.008), and ‘Hindquarters LAS’ ($r_s$=0.78, P=0.014) as well.

In medium farms, ‘Claw overgrowth’ is highly correlated with ‘Knee LAS - score 2’ ($r_s$=0.83, P=0.006), and neck abscesses with udder abscesses ($r_s$=0.90, P=0.001), as well. There is a high correlation between head, neck ($r_s$=0.85, P=0.002), body ($r_s$=0.93, P<0.001), and hindquarters lesions ($r_s$=0.68, P=0.045). Body lesions are very highly correlated with neck lesions ($r_s$=0.95, P<0.001), and highly correlated with hindquarters lesions ($r_s$=0.83, P=0.003). Neck lesions are very highly correlated with hindquarters lesions ($r_s$=0.92, P<0.001).

In large farms, a high correlation was identified between the number of animals queuing at feeding rack and queuing at drinking ($r_s$=0.78, P=0.008). Equally, there is a positive correlation between dirty hindquarters and dirty lower legs ($r_s$=0.69, P=0.029). ‘Knee LAS - score 2’ is correlated with ‘Cleanliness – hindquarters’ ($r_s$=0.67, P=0.034) and with ‘BCS - very thin’ ($r_s$=0.69, P=0.028). Body lesions are significantly correlated with neck ($r_s$=0.64, P=0.047), lower legs ($r_s$=0.69, P=0.028) and hindquarters lesions ($r_s$=0.81, P=0.005). Moreover, hindquarters lesions are highly correlated with neck lesions ($r_s$=0.78, P=0.008).

### 2.2.5. Reliability studies

### 2.2.5.1. Inter-observer reliability (IOR)

#### a) Categorical data

According to Fleiss et al. (2003), the highest level of agreement for κ (“excellent”) was obtained for ‘Cleanliness- hindquarters’, ‘Cleanliness - lower legs’, ‘Udder asymmetry’, ‘Claw overgrowth’, ‘Body abscesses’, ‘Udder abscesses’, ‘Ocular discharge’ and ‘Diarrhoea’. Following Landis and Koch (1977), the highest level of agreement for $\kappa_w$ (“substantial”) was achieved for ‘BCS’ and ‘Knee LAS’. All the other assessed indicators had a level of agreement of “fair to good” (0.41–0.75) for κ (Fleiss et al., 2003), except for ‘Hindquarters abscesses’ and ‘Vulvar discharge’. Table 6 shows the results of inter-observer reliability of the indicators recorded on individual assessments.

| Animal-based indicator | | Agreement, tolerance 0 | Agreement, tolerance 1 | κ | $\kappa_w$ | $\kappa_w$ confidence boundaries |
|---|---|---|---|---|---|---|
| BCS | | 93.6 | 100 | 0.77** | 0.79*** | 0.70-0.88 |
| Cleanliness | Hindquarters | 93.6 | - | 0.79** | - | - |
| | Lower legs | 93.1 | - | 0.80** | - | - |
| | Udder | 97.8 | - | 0.59* | - | - |
| Lesions and swellings | Hindquarters | 91.4 | - | 0.67* | - | - |
| | Lower legs | 91.7 | - | 0.45* | - | - |
| | Knee | 96.1 | 100 | 0.78** | 0.79*** | 0.68-0.90 |
| | Head | 86.7 | - | 0.63* | - | - |
| | Body | 83.9 | - | 0.52* | - | - |
| | Neck | 85 | - | 0.55* | - | - |
| Udder asymmetry | | 99.4 | - | 0.95** | - | - |
| Claw overgrowth | | 95.6 | - | 0.91** | - | - |
| Abscesses | Hindquarters | 99.4 | - | 0[1] | - | - |
| | Body | 98.9 | - | 0.84** | - | - |
| | Udder | 99.7 | - | 0.93** | - | - |
| | Neck | 98.1 | - | 0.66* | - | - |
| | Head | 95 | - | 0.50* | - | - |
| Discharge | Nasal | 98.9 | - | 0.50* | - | - |
| | Ocular | 99.4 | - | 0.89** | - | - |
| | Vulvar | 99.7 | - | 0[2] | - | - |
| Diarrhoea | | 99.2 | - | 0.93** | - | - |

N=360, *Fair to good, **Excellent, ***Substantial
[1] Only 2 cases, and observers disagreed in both
[2] Only 1 case, and observers disagreed

Table 6 - Agreement and reliability evaluation (N=360, *Fair to good, **Excellent, ***Substantial) of the categorical data obtained from the assessments performed simultaneously by two assessors in 10 Portuguese farms, while testing the AWIN prototype for dairy goats.

### b) Continuous data

As depicted from Table 7, the ICC for inter-observer reliability for 'Improper disbudding', 'Queuing at feeding', 'Queuing at drinking', 'Hair coat condition' varied from 0.84 to 0.99 (95% CI: 0.48–1) indicating substantial agreement between observers. All the other indicators could not be computed.

| Animal-based indicator | | Absolute agreement[3] | CI | Consistency[4] | CI |
|---|---|---|---|---|---|
| Improper disbudding | | 0.99<br>P = 1.62e-10 | 0.98 < ICC < 1 | 0.99<br>P=4.67e-10 | 0.97<ICC<1 |
| Queuing | At feeding | 0.89<br>P = 7.76e-05 | 0.62 < ICC < 0.97 | 0.90<br>P=9.38e-5 | 0.64<ICC<0.97 |
| | At drinking | 0.99<br>P = 1.91e-09 | 0.96 < ICC < 0.1 | 0.99<br>P=4.99e-09 | 0.96<ICC<1 |
| Hair coat condition | | 0.85<br>P = 0.000457 | 0.51 < ICC < 0.96 | 0.84<br>P=0.00064 | . 0.48<ICC<0.96 |
| Oblivious | | - | - | - | - |
| Avoidance distance | Contact | - | - | - | - |
| | Acceptance | - | - | - | - |
| Kneeling | Feeding rack | - | - | - | - |
| | Pen | - | - | - | - |
| Severe lameness | | - | - | - | - |

Table 7 – Inter-observer reliability evaluation for continuous data obtained from the assessments performed simultaneously by two assessors in 10 Portuguese farms, while applying the AWIN prototype for dairy goats.

### c) QBA Analysis

The first principal component of the PCA from the QBA presented a moderate Pearson's correlation between the two observers (r=0.67; P=0.036). Concerning PC2, there was no significant correlation between observers (r=0.42; P=0.222). When considering the descriptors, the number of high Spearman's rank correlations between observers' scorings was low, being only verified for aggressive ($r_s$=0.79; P=0.006) and alert ($r_s$=0.81; P=0.005). Moderate Spearman's rank correlations were found between the scoring of agitated ($r_s$=0.68; P=0.031), fearful ($r_s$=0.67; P=0.035), frustrated ($r_s$=0.60; P=0.007) and lively ($r_s$=0.66; P=0.038). The correlations between observers' ratings regarding the remaining descriptors were not significant. Additional information on the correlations between QBA descriptors is presented in Table 20 (Annex 1).

[3] 'Absolute agreement' takes into account if good inter observer reliability is characterized by scores that are similar in absolute value.
[4] 'Consistency' considers if good inter observer reliability comes from scores that are similar in rank order.

### 2.2.5.2. Consistency over time of animal-based indicators

An average of 3.7 ± 1 months (SD≈30 days) passed between the two visits and no significant alterations, in management and housing conditions, were implemented during this period. The number of adult dairy goats on each farm ranged from 46 to 2000 animals, with a mean (±SD) of 512 (±613) adult dairy goats. Group assessment was accomplished in 1116 animals, and individual observations were made in 494 adult dairy goats. The mean number of animals in the evaluated pen was 153 ± 95 animals, and the animal sample varied from 32 to 61 adult dairy goats. In the course of these days, temperatures ranged from 15ºC to 26ºC and relative humidity from 42% to 86%.

**a) Variation in prevalences between visits**

According to the Wilcoxon signed rank test performed, only 'Head LAS' presented prevalence results significantly different between the two visits (P=0.037). The remaining animal-based indicators did not yield results significantly different between the two assessments. 'BCS - very fat', 'Diarrhoea', 'Cleanliness - hindquarters', 'Knee LAS - score 2', 'Head LAS' , 'Ocular discharge' and 'Claw overgrowth' showed a change in prevalences above 5% (%Δ>5%). Animal based-indicators as 'Improper disbudding', 'Oblivious', 'BCS – very thin', 'Hindquarters abscesses', 'Neck abscesses', 'Vulvar discharge', 'Severe lameness' and kneeling, both in pen and at feeding, presented a prevalence variation below 1% (%Δ<1%).'Panting score 2', 'Shivering score 2' prevalences demonstrated no variation between visits (%Δ=0), since there were no cases recorded. In Table 8, the variation in prevalence (and mean prevalence) of each indicator between two visits is shown. The recorded prevalences of the welfare indicators assessed during the two visits are given in Table 21 (Annex 1).

| Animal-based indicator | | Visit 1 | | | | | | Visit 2 | | | | | | Change (Visit 2 – Visit 1) | | Wilcoxon signed rank test P value* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Mean | Median | SD | Min | Max | N | Mean | Median | SD | Min | Max | Δ | %Δ | |
| **Pen-level Observations** | | | | | | | | | | | | | | | | |
| **Improper disbudding** | | 772 | 20 | 10.5 | 27.7 | 0.02 | 83.8 | 1102 | 19.1 | 13.8 | 19.2 | 4.2 | 62.5 | -0.9 | 0.7 | NS |
| **Queuing** | At feeding | 1131 | 23.2 | 17.3 | 18.8 | 5.8 | 61.4 | 1529 | 32.3 | 27.5 | 29.5 | 0 | 90 | 9.1 | 3.2 | NS |
| | At drinking | | 1.9 | 0 | 3.8 | 0 | 11.8 | | 3.6 | 0.7 | 6.1 | 0 | 16.1 | 1.7 | 1.2 | NS |
| **Hair coat condition** | | 1082 | 18.2 | 11.9 | 16.7 | 5 | 52 | 1483 | 10.1 | 8.3 | 5.6 | 3.3 | 20.4 | -8.1 | -2.5 | NS |
| **Oblivious** | | 1131 | 0.6 | 0 | 1.1 | 0 | 3.4 | 1529 | 0.1 | 0 | 0.2 | 0 | 0.5 | -0.5 | -0.4 | NS |
| **Panting** | Score 1 | 1131 | 1.1 | 0 | 3.2 | 0 | 10. 2 | 1529 | 3.1 | 0 | 5.9 | 0 | 18.3 | 2 | 2.3 | NS |
| | Score 2 | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NS |
| **Shivering** | Score 1 | 1131 | 9.2 | 0 | 29 | 0 | 91.8 | 1529 | 0 | 0 | 0 | 0 | 0 | -9.2 | -4.0 | NS |
| | Score 2 | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NS |
| **Latency to first contact (s)** | | 1032 | 52.4 | 27 | 88.3 | 0 | 266 | 1411 | 30.5 | 11.5 | 43.8 | 1 | 130 | -21.9 | - | NS |
| **Avoidance distance** | Contact | 1032 | 0.9 | 0.3 | 1.2 | 0 | 3.2 | 1411 | 1.4 | 0.6 | 1.9 | 0 | 5.7 | 0.5 | 1 | NS |
| | Accepta nce | | 1.8 | 1.4 | 1.9 | 0 | 4.8 | | 2.3 | 1.8 | 2.7 | 0 | 8.2 | 0.5 | 1 | NS |
| **Severe lameness** | | 1131 | 1.5 | 1 | 1.7 | 0 | 4.8 | 1529 | 0.9 | 0.7 | 1 | 0 | 3.3 | -0.6 | -0.1 | NS |
| **Kneeling** | Feeding | 1131 | 0 | 0 | 0 | 0 | 0 | 1529 | 1.2 | 0 | 3.7 | 0 | 11.7 | 1.2 | 0.9 | NS |
| | Pen | | 0.7 | 0 | 1.3 | 0 | 3.6 | | 0.3 | 0 | 0.8 | 0 | 2.5 | -0.4 | -0.5 | NS |

Table 8 - Differences in the prevalences of the animal-based indicators included in the AWIN prototype for dairy goats recorded during the two visits to Portuguese farms.

| Individual Assessment | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BCS** | Very thin | | 5.5 | 4.7 | 3.5 | 0 | 10.5 | | 5.6 | 4.5 | 5.9 | 0 | 18 | 0.1 | 0.7 | NS |
| | Very Fat | | 18.2 | 13.2 | 16.7 | 0 | 45.5 | | 30.8 | 35.2 | 15.4 | 1.6 | 48.1 | 12.6 | 9.8 | NS |
| **Diarrhoea** | | | 6.8 | 0 | 18.1 | 0 | 57.9 | | 0 | 0 | 0 | 0 | 0 | -6.8 | -6.4 | NS |
| **Udder Asymmetry** | | | 5 | 4.1 | 4.6 | 0 | 13.3 | | 7.3 | 6.8 | 4.5 | 0 | 15.3 | 2.3 | 2.5 | NS |
| **Cleanliness** | HQ | | 31.8 | 21.7 | 31.4 | 0 | 75 | | 13.9 | 5.9 | 16.2 | 0 | 39.7 | -17.9 | -8.4 | NS |
| | LL | | 19.2 | 6.3 | 25.9 | 0 | 64.6 | | 23.9 | 18.3 | 25.6 | 0 | 80.1 | 4.7 | -1.4 | NS |
| | Udder | | 1.6 | 0 | 2.2 | 0 | 6.3 | | 0.2 | 0 | 0.6 | 0 | 1.8 | -1.4 | -1.3 | NS |
| **Abscesses** | HQ | | 0.6 | 0 | 1.3 | 0 | 3.8 | | 0.8 | 0 | 1.4 | 0 | 3.6 | 0.2 | 0.1 | NS |
| | Body | | 10.9 | 3.6 | 14.3 | 0 | 43.8 | | 16.5 | 4.5 | 21.8 | 0 | 56.3 | 5.6 | 4.5 | NS |
| | Udder | 404 | 3 | 2.2 | 4 | 0 | 12.8 | 494 | 0.9 | 0 | 1.5 | 0 | 3.8 | -2.1 | -2.0 | NS |
| | Neck | | 3.5 | 3.8 | 4.1 | 0 | 12.5 | | 4.7 | 2.8 | 5.8 | 0 | 18.2 | 1.2 | 0.7 | NS |
| | Head | | 7.0 | 6.3 | 6.1 | 0 | 18.4 | | 3.8 | 3.3 | 3.6 | 0 | 10.3 | 3.2 | -3.8 | NS |
| **LAS** | HQ | | 2.7 | 0 | 4 | 0 | 10.5 | | 6.5 | 1.8 | 10.6 | 0 | 27.9 | 3.8 | 4.4 | NS |
| | LL | | 6.7 | 0 | 15.8 | 0 | 47.4 | | 8.9 | 1 | 14.2 | 0 | 41 | 2.2 | 4.0 | NS |
| | Knee (score 2) | | 10.7 | 3.4 | 15.6 | 0 | 46.7 | | 21.8 | 12.8 | 28.7 | 0 | 92.3 | 11.1 | 9.9 | NS |
| | Body | | 8.5 | 5.8 | 9.4 | 0 | 28.1 | | 9.9 | 2 | 16.7 | 0 | 41.9 | 1.4 | 1.9 | NS |
| | Neck | | 8.7 | 2.6 | 16 | 0 | 50 | | 4.8 | 3 | 5.7 | 0 | 16.3 | -3.9 | -3.3 | NS |
| | Head | | 7 | 6.3 | 6.1 | 0 | 18.4 | | 12 | 6.9 | 19.9 | 0 | 67.4 | 5 | -16.5 | 0.037 |

Table 8 - Differences in the prevalences of the animal-based indicators included in the AWIN prototype for dairy goats recorded during the two visits to Portuguese farms (continuation).

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Discharge** | Ocular | 404 | 7.8 | 6.3 | 8.6 | 0 | 27.3 | 494 | 2.8 | 0 | 6 | 0 | 19 | -5 | -7.0 | NS |
| | Nasal | | 3 | 3.3 | 3.4 | 0 | 9.4 | | 3.8 | 1.7 | 5.1 | 0 | 15.3 | 0.8 | 1.3 | NS |
| | Vulvar | | 0.3 | 0 | 0.8 | 0 | 2.6 | | 0 | 0 | 0 | 0 | 0 | 0.3 | -0.3 | NS |
| **Claw Overgrowth** | | | 43.6 | 45.8 | 28.3 | 0 | 89.5 | | 30.4 | 27.8 | 29.3 | 0 | 67.4 | -13.2 | -13.4 | NS |

NS non–significant
*P<0.05

Table 8 - Differences in the prevalences of the animal-based indicators included in the AWIN prototype for dairy goats recorded during the two visits to Portuguese farms. (continuation)

**b) Indicators relationships between visits**

Spearman's rank correlations between the two assessments of the same farm varied substantially and ranked from -0.01 to 0.79. 'Body abscesses' and 'Severe lameness' were above the threshold of 0.70 ($r_s$= 0.72 to 0.79; P<0.05), presenting the highest correlation coefficients between the two visits (Table 9). Based on the Spearman's correlations the other indicators assessed were not repeatable from one visit to another.

| Animal-based indicator | $r_s$ | P (two-tailed) |
|---|---|---|
| **Individual Assessment** | | |
| **Abscesses** Body | 0.72 | 0.019 |
| **Severe lameness** | 0.79 | 0.007 |

Table 9 - Significant correlations ($r_s$) between animal-based indicators assessed during two consecutive visits to the same ten farms, following the AWIN prototype for dairy goats.

### 2.2.6. Analysis of the seasonal visits by farm category

The initial (winter) and final (summer) assessments, performed at the same ten farms, produced diverse prevalences of the 24 recorded indicators, with large farms demonstrating, in general, higher prevalences. An analysis of the most prevalent indicators, concerning farm categories, is presented. Detailed information related to prevalence values of these animal-based indicators is given in Table 22 to 31 (Annex 1).

**a) Pen-level observations**

*Queuing at feeding and drinking*

'Queuing at feeding' prevalence increased in small farms, from 26.6% (winter) to 34.8% (summer), and in medium size farms, from 8.4% to 17.4%. In large farms, this indicator kept an identical prevalence, around 40%.

'Queuing at drinking' presented similar prevalences across seasons in the three farm categories. The overall prevalence varied from 3.2% to 4.4% through seasons, reaching a maximum of 16.1% in medium farms, in summer.

*Poor hair coat condition*

'Poor hair coat condition' prevalence showed a higher variation in small farms between seasons, decreasing from 30.3% to 9.1%. In medium and large farms prevalence values were similar in both seasons (about 7% and 13%, respectively).

*Improper disbudding*

In medium size farms, this indicator's prevalence decreased from 44.3% to 31.6% across seasons. In small and large farms, prevalence values were identical across seasons, varying from 4% to 1.7% in small farms and 13.8% to 14.1%, in large farms.

**b) Individual observations**

*Body condition*

The overall prevalence of overweight animals increased from 20% to 29.8%, and was mostly observed in small and large size farms, decreasing from 21.2% to 11.9% in medium farms. Although 'BCS- very thin' prevalence values did not show wide variation over seasons, in winter the highest prevalence of very thin animals was found in small farms (8.3%), and in summer in medium farms (10.1%).

*Cleanliness*

The prevalence values of hindquarters and lower legs cleanliness decreased in small and large size farms, increasing in medium farms, over seasons. In small farms, 'Cleanliness hindquarters' oscillated from 26% to 2.6%, and 'Cleanliness – lower legs' from 33.3% to 1.7%. In medium farms, 'Cleanliness - hindquarters' varied from 15.9% to 25.6%, and 'Cleanliness – lower legs' from 4.5% to 30.4%. In large farms, Cleanliness - hindquarters' varied from 51.7 % to 38.3% and 'Cleanliness – lower legs' from 41.5% to 36.4%, through seasons.

*Lesions and swellings*

The prevalence of lesions and swellings varied substantially in both seasons, and in the three farm categories. In small farms, the prevalences of lesions decreased in general, except for 'Knee LAS- score 2', which values increased (from 15.6% to 25.6%), and for 'Hindquarters LAS', 'Lower legs LAS' and 'Neck LAS' that were found at same prevalence. In medium farms, the prevalences of hindquarters, lower legs, knee (score 1) and body lesions increased, while head lesions decreased from 24.2% to 9.5%, knee (score 2) and neck lesions kept similar prevalence values (both around 5%). In large farms, the prevalence of lower legs, knee (score 1), neck and head lesions decreased, whilst hindquarters, body and knee lesions (score 2) increased. The highest overall prevalence was found for 'Knee LAS – score 1' in both seasons, reaching a maximum of 98.3% in medium farms in summer.

*Udder asymmetry*

'Udder asymmetry' prevalence had similar values among farm categories and through seasons. The overall prevalence varied from 5% to 7.5% across seasons, reaching a maximum of 15.3% in large farms, in summer.

*Claw overgrowth*

The prevalence of overgrown claws decreased through seasons in the three farm categories considered, showing a higher variation in small size farms, where it varied from 11.5% to zero.

*Abscesses*

The overall prevalence of abscesses in the different regions assessed had similar values in both seasons. The highest variation occurred in 'Body abscesses' in small farms, which prevalence value varied from 20.8%, in winter, to 32.5%, in summer. In medium farms, 'Body abscesses' oscillated from 8.3%, in winter, to 13.7%, in summer. In large intensive farms, 'Head abscesses' prevalence oscillated from 10.2% in winter, to 4.3% in summer.

*Nasal and ocular discharge*

'Ocular discharge' overall prevalence decreased from winter to summer (8.2% to 1.2%), reaching zero prevalence in small and medium farms. 'Nasal discharge' presented identical prevalences in small and medium farms in both seasons, reaching a maximum of 15.3% in summer in large farms.

*Severe lameness*

Although 'Severe lameness' was not one of the most prevalent indicators assessed in both seasons, considering it is one of the major welfare problem the results obtained among farms are represented in Figure 24. The overall prevalence of 'Severe Lameness' presented identical values in winter and summer, 1.2% and 1.1%, respectively (Table 32). There were no recorded cases of severe lameness in small farms, with this indicator prevalence reaching a maximum of 4.8% in large farms in winter.

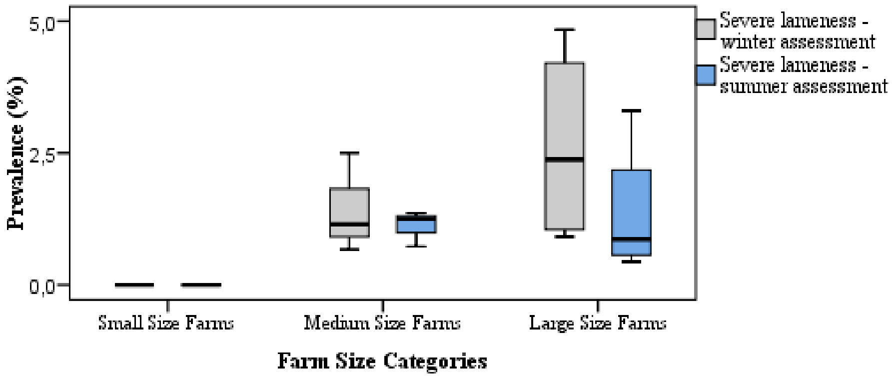


Figure 24 – 'Severe lameness' prevalence results obtained from the consecutive visits to 10 Portuguese intensive dairy goat farms, organized by farm size categories.

### 2.2.7. On-farm feasibility assessment: duration of the prototype

The mean time necessary to execute pen-level observations, assess queuing at feeding and drinking place, QBA, human-animal relationship and to perform clinical scoring varied from farm to farm, and was approximately 87 ± 33 min (1h 27 min ± 33 min), ranging from 43 min to 154 min. Table 10 summarizes the average time to accomplish the principal stages of the welfare assessment, recorded in 26 farms.

| Stages of the prototype | | N | Mean | Std. Deviation | Median | Min | Max |
|---|---|---|---|---|---|---|---|
| **Pen Observation (outside pen)*** | | 23 | **4** | 3.3 | 2 | 0 | 11 |
| **Pen Observation (inside pen)** | | 23 | **1.7** | 1 | 1 | 0 | 4 |
| **Queuing**** | At feeding | 24 | 15 | 0 | 15 | 15 | 15 |
| | At drinking | | 15 | 0 | 15 | 15 | 15 |
| | **Overall** | | **18.8** | 7.6 | 15 | 15 | 43 |
| **QBA** | | 26 | 11.1 | 2.9 | 10 | 10 | 20 |
| **HAR***** | Latency | 21 | 2.4 | 2.3 | 0.8 | 0.1 | 5 |
| | AD test | | 4.1 | 2.6 | 3.3 | 1 | 12.5 |
| | **Overall** | | **6.5** | 3.3 | 6.0 | 1.2 | 16.9 |
| **Clinical scoring**** | | 24 | **58.5** | 45.1 | 40 | 10 | 168 |
| **Overall** | | 18 | **87** | 33 | 75.7 | 43 | 153.9 |

*Pen observation (outside pen) time could only be recorded in 23 farms
**Queuing (at feeding rack and at drinking) and clinical scoring times could only be recorded in 24 farms
*** HAR tests time could only be recorded in 21 farms

Table 10 - Average time in minutes to perform the different stages of the AWIN on-farm welfare assessment prototype for dairy goats.

In small size farms, the time necessary to accomplish the different stages of the prototype was: 71 ± 29 min (1h 11 min ± 29 min; Table 33). Regarding medium farms, the required time to accomplish the stages of the prototype was 88 ± 25 min (1h 28 min ± 25 min; Table 34). In large size farms, the needed time to perform the stages of the prototype was 117 ± 36 min (1h 57 min ± 36 min; Table 35).

### 2.2.8. Resource and management based indicators

Certain resource and management-based indicators demonstrated variations between farm categories (Table 36).

The number of pens varied from one to 10 in the 30 visited farms, with an average number of pens with lactating goats of 1.4 pens in small size farms, 2.6 pens in medium size farms and 5.3 pens in large size farms.

The number of animals in pen ranged from 20 to 300 adult dairy goats in the three farm categories considered. The average number of animals in pen was 49, 123 and 133 animals, in small, medium and large farms, respectively.

Mean area per goat oscillated between 1.9 $m^2$ in large farms, to 3.0 $m^2$, in small size farms. Feed trough length was, in average, 17.4 m in small farms, 24.9 m in medium farms and 24.5 m in large farms. Alternatively (or sometimes additionally) to a long feed trough along the pen, some small farms had smaller feeders located inside the pens.

The number of drinkers per pen ranged from 0 to 10 drinkers through the three farm categories; the average water trough length had a maximum value of 2.2 m, in large farms, and a minimum of 1.2 m, in medium and small farms. Thus, the average number of goats per drinker and per water trough length varied from 16.3 to 28.8 animals and from 15.5 to 32.8 animals in the three farm categories, correspondingly.

The breed of each individual goat was not recorded, but the target population comprised different breeds, as demonstrated in Table 11. In small farms, breeds such as Serrana ecotype Transmontano, Alpine, Malagaña, Murciano-Granadina, Saanen, Florida, Charnequeira and crossbreds (Saanen with Alpine) could be found. Murciano-Granadina and Saanen were the most common breeds in medium size farms, with average numbers of 299 and 197 animals, respectively. In large farms, this was the case for Murciano-Granadina, Saanen and Alpine, with average numbers of 625, 696 and 678 animals, correspondingly.

| Breed | Farm Size Category<br>Small | | | Medium | | | Large | | |
|---|---|---|---|---|---|---|---|---|---|
| | Farms | Mean | Range | Farms | Mean | Range | Farms | Mean | Range |
| **Saanen** | 2 | 35 | 10-60 | 6 | 197 | 37-300 | 4 | 696.3 | 550-965 |
| **Alpine** | 3 | 53.3 | 10-100 | 1 | 35 | - | 2 | 677.5 | 55-1300 |
| **Crossbreds** | 3 | 36 | 18-60 | 1 | 22 | - | 1 | 330 | - |
| **Murciano-Granadino** | 3 | 36.7 | 3-85 | 6 | 298.7 | 35-440 | 6 | 625 | 500-700 |
| **Serrana ecotype Transmontano** | 3 | 83.3 | 80-90 | - | - | - | - | - | - |
| **Malagaña** | 1 | 50 | - | 1 | 60 | - | - | - | - |
| **Florida** | 2 | 19 | 3-35 | - | - | - | - | - | - |
| **Serpentina** | - | - | - | - | - | - | 1 | 15 | - |
| **Charnequeira** | 1 | 3 | - | - | - | - | - | - | - |

Table 11 – Breeds present on the 30 Portuguese assessed farms, where the AWIN prototype for dairy goats was tested.

Subsequently, an overview of the results obtained in the course of the on-farm testing of the prototype is presented (Table 12).

| Animal-based indicator | | N | Farms (%) | <5% | >5% | Δ% among farm categories[#] | κ | $κ_w$ | ICC | $r_s$** | Δ<5%*** | Stage | Range | Δt per farm category S | Δt per farm category M | Δt per farm category L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Prevalence (%) | | | | Reliability: IOR* | | | Reliability: COT | | Feasibility | | | | |
| **Pen-level observations** | | | | | | | | | | | | | | | | |
| **Improper disbudding**[+] | | 1778 | 19 (63.3) | | ✓ | 9.1-23.3 | | | S | | ✓ | **Pen Obs. (outside pen)°** | **0-11** | 0-10 | 2-9 | 2-11 |
| **Hair coat condition** | | 2715 | 30 (100) | | ✓ | 17.6-20.3 | | | S | | ✓ | | | | | |
| **Panting (score > 0)** | Score 1 | | 5 (16.7) | ✓ | | 0.1-14 | | | | | ✓ | | | | | |
| | Score 2 | | 0 (0) | ✓ | | 0 | | | | | ✓[1] | | | | | |
| **Shivering (score >0)** | Score 1 | | 4 (13.3) | ✓ | | 2.1-10.6 | | | | | ✓ | | | | | |
| | Score 2 | | 0 (0) | ✓ | | 0 | | | | | ✓[1] | | | | | |
| **Oblivious** | | | 8 (26.7) | ✓ | | 0.4-0.5 | | | - | | ✓ | | | | | |
| **Kneeling at feeding** | | | 0 (0) | ✓ | | 0 | | | - | | ✓ | | | | | |
| **Queuing at feeding** | | | 25 (83.3) | | ✓ | 12.9-39.7 | | | S | | ✓ | **Queuing°** | **15-43** | 15-22 | 15-43 | 15-24 |
| **Queuing at drinking** | | | 11 (36.7) | | ✓ | 0-6 | | | S | | ✓ | | | | | |
| **QBA** | | | | | | | | | | | | **QBA** | **10-20** | 10-20 | 10-20 | 10-16 |

Table 12 – Overview of the results obtained during the on-farm testing of the AWIN prototype in Portuguese dairy goat farms.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Avoidance distance** | Contact | 2715 | 12 (40) | ✓ | | 0.4-1.9 | | | - | | ✓ | **HAR°** | **1 -13** | 1-7 | 3-11 | 4-13 |
| | Acceptance | | 15 (50) | ✓ | | 0.6-2.1 | | | - | | ✓ | | | | | |
| **Latency to first contact (s)** | | | | | | | | | | | | | | | | |
| **Severe lameness** | | | 66.7 (20) | ✓ | | 0.6-2.7 | | | - | A | ✓ | **Pen Obs. (inside pen)** | **0-4** | 0-1 | 1-3 | 1-4 |
| **Kneeling in pen** | | | 8 (26.7) | ✓ | | 0.2-0.6 | | | - | | ✓ | | | | | |
| **Individual Assessment** | | | | | | | | | | | | | | | | |
| **BCS** | Very thin | 1172 | 17 (56.7) | | ✓ | 4.7-5.3 | E | S | | | ✓ | **Clinical Scoring°** | **10-168** | 10-75 | 30-168 | 40-168 |
| | Very Fat | | 25 (83.3) | | ✓ | 13.1-24.6 | | | | | | | | | | |
| **Cleanliness** | Hindquarters | | 21 (70) | | ✓ | 9.3-30 | E | | | | | | | | | |
| | Lower legs | | 19 (63.3) | | ✓ | 16.4-20.2 | E | | | | ✓ | | | | | |
| | Udder | | 9 (30) | ✓ | | 1.6-4.3 | F | | | | ✓ | | | | | |
| **Lesions and swellings** | Hindquarters | | 18 (60) | | ✓ | 5.3-13.5 | F | | | | ✓ | | | | | |
| | Lower legs | | 15 (50) | | ✓ | 0.7-16.3 | F | | | | ✓ | | | | | |
| | Knee – Score 1 | | 30 (100) | | ✓ | 78.5-93.8 | E | S | | | ✓ | | | | | |
| | Knee – Score 2 | | 17 (56.7) | | ✓ | 4.9-16.5 | | | | | | | | | | |

Table 12 – Overview of the results obtained during the on-farm testing of the AWIN prototype in Portuguese dairy goat farms. (continuation)

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lesions and swellings | Head | 2715 | 29 (96.7) | | ✓ | 11.1-30.7 | F | | | Clinical Scoring° | **10-168** | 10-75 | 30-168 | 40-168 |
| | Body | | 24 (80) | | ✓ | 13.8-17.4 | F | | ✓ | | | | | |
| | Neck | | 22 (73.3) | | ✓ | 4.6-18.7 | F | | ✓ | | | | | |
| Udder asymmetry | | | 24 (80) | | ✓ | 4.9-6.5 | E | | ✓ | | | | | |
| Claw overgrowth | | | 27 (90.0) | | ✓ | 11.8-41.9 | E | | | | | | | |
| Abscesses | Body | | 19 (63.3) | | ✓ | 7.5-14.1 | E | A | ✓ | | | | | |
| | Head | | 22 (73.3) | | ✓ | 3.4-12.8 | F | | ✓ | | | | | |
| | Neck | | 20 (66.7) | ✓ | | 2.5-6.3 | F | | ✓ | | | | | |
| | Udder | | 14 (46.7) | ✓ | | 1.3-2.9 | E | | ✓ | | | | | |
| | Hindquarters | | 9 (30) | ✓ | | 0.2-0.9 | 0 | | ✓ | | | | | |
| Discharge | Nasal | | 15 (50) | | ✓ | 2.3-8.1 | F | | ✓ | | | | | |
| | Ocular | | 16 (53.3) | | ✓ | 1.3-8.5 | E | | | | | | | |
| | Vulvar | | 3 (10) | ✓ | | 0.2-0.5 | 0 | | ✓ | | | | | |
| Diarrhoea | | | 6 (20) | ✓ | | 0.3-5.9 | E | | | | | | | |

+Improper disbudding could only be recorded in 23 farms

°Pen observation (outside pen) time could only be recorded in 23 farms; Queuing (at feeding rack and at drinking) and clinical scoring times could only be recorded in 24 farms; HAR tests time could only be recorded in 21 farms

#Prevalence variation per farm category: pen level observations - small farms, N= 473; medium farms, N = 1122; large farms, N=1120; individual assessment - small farms, N= 473; medium farms, N = 1122; large farms, N=1120

*IOR: F- Fair to good, E – Excellent, for κ based on Fleiss et al. (2003) and S – Substantial for $\kappa_w$ based on Landis and Koch (1977), and for ICC based on Shrout (1998); N = 360

**$r_s$ :(A)- Acceptable based on Martin and Bateson (2007); $r_s$= 0.67 to 0.79, P<0.05

***Δ<5%: Winter - pen level observations N= 1131, except 'Improper disbudding' N = 772, 'Hair coat condition' N=1082 and 'Avoidance distance test' N= 1032; individual assessment N= 404. Summer - pen level observations N= 1529, except 'Improper disbudding' N = 1102, 'Hair coat condition' N = 1483 and 'Avoidance distance test' N = 1411; individual assessment N= 494

[1]%Δ= 0, no cases recorded

Table 12 – Overview of the results obtained during the on-farm testing of the AWIN prototype in Portuguese dairy goat farms. (continuation)

## 2.3. Discussion

To the author's knowledge, there are no official on-farm protocols available for assessing the welfare of dairy goats (*Capra hircus*), and only the studies of Anzuino et al. (2010) and Muri et al. (2013) have published empirical data from overall welfare evaluation of this species. One of the main goals of the AWIN project was to produce a practical welfare assessment protocol centred on animal-based indicators that would give an accurate idea of the welfare status at farm level. However, as stated by EFSA (2012) and Blokhuis et al. (2013), animal-based indicators involve complex and time consuming observations of individual animals. Furthermore, indicators assessed on animals tend to be subjected to individual interpretation and their reliability may represent one of the main limitations in a welfare assessment tool, as mentioned in the studies of De Rosa et al. (2009) and Lensink et al. (2003). Welfare assessments performed directly at the production site, involve some particularities that have to be taken into account, as high stocking density, limited time for appraisal, husbandry constraints and farmers acceptability. There are some particularities of goats that render the welfare evaluation of this species challenging. As referred by Kilgour and Dalton (1984) and Houpt (2005; reviewed by Miranda-de la Lama & Mattiello, 2010) goats are gregarious, with a pronounced herd instinct, very curious and highly reactive, often exhibiting exploratory behaviour, which compromises the individual assessment. Several results based on the work developed by AWIN project in Portugal were compiled and presented in this study, being now subject to a joint reflection. An overall analysis will help to connect the different parts of this investigation, leading to a more fruitful and comprehensive discussion of the established aims.

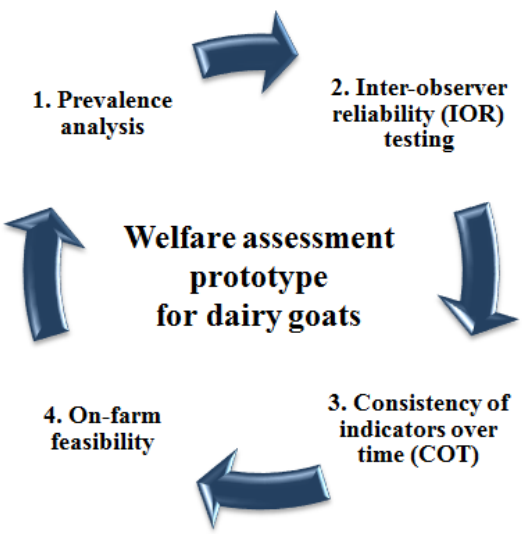



Figure 25 - Overview of the different areas of research addressed in this study, in order to produce a final welfare assessment protocol for intensively kept dairy goats.

As depicted in Figure 25, to test the overall feasibility of the welfare assessment prototype a prevalence analysis of the animal-based indicators was conducted. After this, the reliability was tested between observers and across consecutive visits. Finally, bearing in mind all the previous steps, the different indicators' assessment were considered for their feasibility at the farm.

**1. Prevalence analysis**

**a) Farm recruitment and sampling**

Following Dohoo, Martin and Stryhn (2003), the choice of a sample size implicates both statistical and non-statistical considerations. As examining a large number of farms is time consuming and costly, assessing a small sample is what is frequently available to researchers. However, a very important item that must be addressed when taking a sample is that the study population should be representative of the target population. From a total of 269 Portuguese intensive dairy goat farms, following the AWIN guidelines, a pre-determined sample of 30 sampling units (dairy goat farms) were drawn from the population, taking into account different size categories. The objective of the project was to test the prototype and not to compare farms or draw conclusions about the results. If that was the case, a more elaborate sampling method should have been carried out (Annex 4). As shown in Table 37 (Annex 4), almost all large farms ([>500]) were visited due to the small number in the country. However, for the other categories the sampling number is only a small proportion of the existing ones.

According to INE (2015), goat livestock, in 2013, was higher in Alentejo, Beira Interior, Ribatejo e Oeste and Trás-os-Montes[5]. Of the 30 visited farms, nine were located in the Centre region, nine in Lisboa e Vale do Tejo, seven in Alentejo and five in the North region, yielding a fairly accurate image of the country's reality.

**b) Final outcomes**

Some of the assessed indicators have similar prevalence values when compared with those included in the studies of Anzuino et al. (2010) and Muri et al. (2013), reflecting similar areas of concern. Therefore, throughout this discussion some comparisons are established, highlighting common problems in different countries. For instance, Anzuino et al. (2010) identified an overall prevalence of 3% of severely lame goats (score > 2), with a varying prevalence across farms. In Norwegian dairy goats, Muri et al. (2013) reported a 1.7% prevalence of goats with "any lameness". In the present study severe lameness was recorded at a prevalence of 1.8% which is very similar.

Similar to the Portuguese visited dairy goat farms, in British and Norwegian dairy goat farms claw overgrowth is a major problem, with severe claw overgrowth reaching a prevalence of

[5] Information regarding goat livestock distribution in Portugal is presented in Table 13 (Annex 1).

80% and 66%, respectively. Additionally, in both these studies nearly all animals had knee calluses, which is in agreement with the prevalence of knee lesions (score 1) found in this study.

**c) Relationships between animal-based welfare indicators**

When associating pen-level observations with individual assessments (sub-section 2.2.4.), dirty lower legs were correlated with dirty hindquarters, which can be explained by how often bedding is changed or added on each farm. Additionally, dirty lower legs were also correlated with knee lesions (score 2), that might result from prolonged kneeling behaviour exhibited by lame goats. Claw overgrowth was significantly correlated with severe lameness and knee lesions (score 2). Similarly, Anzuino et al. (2010) found that the prevalences of severely overgrown claws and lameness were correlated, suggesting that it probably revealed a general overall poor management on some farms. Claw overgrowth, penetrating injuries, trauma, inflammation of anatomical structures and presence of infectious agents that affect the limb joints are predisposing factors for lameness in goats, as referred by several authors (e.g. Mohamed, Badau & Kene 1996; Bokko & Chaudhari, 2001; Christodoulopoulos, 2009; Smith & Sherman, 2009), which confirms the relationships obtained between these indicators. Nevertheless, knee lesions can also be related with the quality of environment, as the type of floor and bedding quality. The relationship found between knee lesions (score 2) and very thin animals might be explained by the animals being unable to move properly to reach the feeding trough or feeding space in time in a very competitive environment, and thus having access to lower quality feed or even none. However, knee lesions might also involve pain, or be related to an infectious disease (arthritis) and so, being associated with weight loss.

A correlation between neck, head and body abscesses was also identified. The presence of external abscesses is most often associated with caseous lymphadenitis (CLA), defined by Smith and Sherman (2009) as a chronic contagious disease involving the enlargement and abscessation of one or more peripheral lymph nodes, supporting the correlations found.

Nasal discharges were correlated with ocular discharges, which, considering the period of assessment (winter), might point to the presence of a common infectious challenge or even a sub-optimal environment (i.e., low temperatures, which may have lead farmers to reduce ventilation rates). The presence of respiratory tract diseases might also explain the correlations found between very thin animals and the presence of ocular and nasal discharge, as goats clinically affected by a respiratory tract disease usually present weight loss, as stated by Smith and Sherman (2009).

**d) The overall results from Portuguese dairy goats farms by farm categories**

Further analysis of these findings revealed that the main welfare issues identified were associated to the farm size[6]. The prevalence of the animal-based indicators varied across farm size categories showing, in general, highest values in large farms. For example, claw overgrowth was a main issue, as previously mentioned, being detected in large and medium farms at a higher prevalence (38-42%) compared with smaller ones. Accordingly to Smith and Sherman (2009), this is probably due to a lack of claw wear when animals are housed on straw bedding and to less access to outdoor grazing[7]. In fact, in nine of the ten small farms visited, goats had access to outdoor grazing during nine months of the year, while in large and medium farms this was only verified in one and two farms, respectively, and for a period of time below two months. It is important to differentiate access to outdoor grazing from access to an exterior pen, which was largely verified in medium and large farms, since going to an exterior pen does not imply claw wear. Claw trimming was performed at different times before the visits, however, increased herd size is often associated with a reduction in the human:animal ratio and to less time to observe individual animals, as Stafford and Gregory (2008) mentioned in their research. In the present study this was also verified, with small farms presenting a higher human:animal ratio than large farms (Table 36). These results might indicate that, whatever management strategies are employed, claw health in large and medium farms is generally poorly controlled.

Wemelsfelder et al. (2009) defined QBA as a method that depends on the ability of observers to integrate observed details of behaviour, posture, and context into descriptions of an animal's style of behaving, or body language, using descriptors. These descriptors offer information that is directly relevant to animal welfare and that can be useful in addition to results obtained from quantitative indicators. A PCA analysis revealed two dimensions of goat behaviour labelled positive/negative mood (PC1) and high/low arousal (PC2). Agitated, alert, aggressive, fearful, as well as lively, sociable, content and relaxed showed the highest loadings on the first axis. The second one was defined by descriptors as curious, frustrated and irritated, coupled with descriptors as alert, suffering and bored, presenting a more ambiguous interpretation. The homogenous overall distribution of farms throughout the two axes might be supported by housing and management having a real effect on the animal's on-going behaviour, and these farms were only selected regarding their herd size, with all the animals being bred under an intensive production system. However, there is a small farm presenting a higher score in the first axis, meaning that these animals were assessed as more

[6] All data concerning the variation of indicators prevalence among farm categories can be found in sub-section 2.2.3.
[7] Table 36 (Annex 1) provides relevant information related to resource and management based indicators.

agitated, alert, aggressive and fearful. Regarding the second axis, one medium farm and one large farm showed higher scores, with the animals of these two farms being recorded as more curious, frustrated and irritated. These particular cases might be due to a low intensity of training on QBA, since it was only performed on a single class session. According to Napolitano, De Rosa, Serrapica and Braghieri (2015), in training, observers should discuss the meaning of each descriptor in order to standardize their evaluations, or for instance watch video clips representing the standard for each descriptor with the objective of memorizing the descriptors and their definitions through the observation of the clips. Moreover, to perform QBA it is important to know the full behavioural repertoire of the species under assessment, where the lack of capacity to recognize the behaviour of the species can lead to ambiguous results.

Considering area per goat (Table 36), it was observed that goats among all farm categories were housed at recommended stocking densities (1.5 $m^2$ per animal) according to several guidelines (Toussaint, 1997; Sevi, Casamassima, Pulina & Pazzona, 2009).

With regard to the relationship between indicators (sub-section 2.2.4.), it was found that in small farms there was a high correlation between nasal and ocular discharges. As mentioned, this welfare assessment was performed in winter, when the animals are more susceptible to respiratory tract diseases. Additionally, most of the small farms (80%) were situated in the North and Centre region of the country, where the animals have to deal with rough weather, which might substantiate these results.

Improper disbudding was correlated to head and hindquarters lesions. Partial horns (scurs) observed were likely to result from incorrect disbudding of goat kids. In small farms, it was verified that kids were disbudded later than in medium and large farms, with an average of 30 days (Table 36). Bowen (2014) recommends disbudding at five to seven days of age, to maximally prevent horn growth or the development of abnormal regrowth. However, as Anzuino et al. (2010) mention, the welfare implications of scurs have not been fully studied but it is probably a problem in adult animals, as they might injure other animals in the group or get caught in fences.

In medium farms, claw overgrowth was highly correlated with knee lesions (score 2), that might be explained by an association of factors, as high stocking densities linked to a low frequency of claw trimming and less time to detect conditions in individual animals. This overgrowth is a predisposing factor for lameness, and considering that lame goats spend more time kneeling, placing pressure on the knees, this can partly account for the formation of knee lesions, as already discussed by Anzuino et al. (2010). The results of a study performed by Ajuda, Vieira and Stilwell (2014) showed not only that claw overgrowth causes deep inflammation that can be detected by thermography, but also that claw trimming reduces the severity of this condition. Furthermore, high stocking densities might also explain the correlation found between head abscesses, the most frequent location of CLA, and the number of animals in pen, since poorly kept facilities and the natural curiosity of the animals may enhance contamination by *Corynebacterium pseudotuberculosis,* as referred by Smith and Sherman (2009).

Regarding large farms, the number of animals queuing at feeding and at drinking was highly correlated. High stocking densities, namely the high number of goats per feed space (Table 36), associated to the synchronous behaviour of these animals might explain this finding. It is important to mention that these correlations, found particularly in medium farms when expected to occur also in large ones, might be due to sampling issues or farm routines, as the number of animals per pen in both farm categories was similar.

Regarding behavioural observations to assess human-animal interactions, the latency period of the first goat that entered in contact with the assessor was shorter in large farms than in small ones, which can be explained by the breed differences among holding categories (sub-section 2.2.8.). Breeds as Saanen and Murciano-Granadina are reported to be docile and easy to handle, well-suited to intensive systems in several studies (e.g. Sinn & Rudenberg, 2008; Martínez et al., 2010; Escareño et al., 2012), being very common in large size farms in Portugal. In small farms, as mentioned before, goats had more access to outdoor grazing enhancing the expression of natural, foraging and exploratory behaviours and mostly having close contact with the stockperson, reacting negatively to the presence of other humans, which is in accordance with the work of Dwyer (2009) and Ekesbo (2011).

These particular findings echo different realities. The visits to different farm size categories provided the possibility to test if the prototype was flexible to be applied under different circumstances, but at the same time standardized enough to allow comparisons among holding. The protocol is intended to be functional in all farms irrespective of their herd sizes, resources or management procedures, although these different scenarios have to be taken into consideration. Thus, the development of a dairy goat welfare assessment protocol, combining different indicators, should aim to deliver sufficiently robust evaluations to provide a reliable overall picture of the animals' welfare state, regardless the farm size, preventing some welfare problems from remaining undetected. As far as the author is aware, there is no previous research in general overviews of farmed dairy goat welfare regarding comparison between farm sizes, with this study introducing a new line of research in goat's welfare assessment.

### e) Constraints and perspectives

Conducting this analysis to gain insight into the main areas of concern in Portuguese dairy goat farms was important for the process of developing a welfare assessment protocol. To be included in a suitable protocol, it was considered that welfare indicators had to fulfil the requirements of validity, repeatability and feasibility. To simplify matters, throughout the whole discussion the terms 'suggested protocol', 'final protocol' or 'protocol' will be used with the same meaning.

Due to the zero prevalence of panting and shivering score 2, it was decided to assess thermal stress in two levels: 'Score 0 - Absence of heat/cold stress' and 'Score 1 - Presence of heat/cold stress'. In the final protocol, the assessment of panting and shivering score 1 includes the cases scored in the prototype as panting and shivering score 2[8] (panting or shivering, respectively). Equally, AD test was eliminated, due to feasibility reasons. Not only another test to evaluate HAR (latency to first contact test) was already included in the protocol, but also the application of the AD test depended on the breed in assessment. For instance, breeds as Saanen and Murciano-Granadina accepted to be contacted or gently stroke more often, sometimes even complicating the assessment by grouping around the assessor, while breeds as Serrana ecotype Transmontano showed strong avoidance behaviour, making it difficult to carry out the test in a standardized way. Muri et al. (2013) also referred similar limitations regarding the AD test. Furthermore, factors as reproductive season and stocking density also may affect the assessment of this indicator, with the assessor not being able to perform the test due to lack of space inside the pen, or with the goats approaching the assessor more often when in oestrus.

Regarding 'Kneeling in the pen', its low prevalence, and the correlation found by Anzuino et al. (2010) between severe lameness and kneeling, supporting the findings of Mazurek et al. (2007), contributed to its exclusion from the protocol. It is acknowledged that the welfare implications of kneeling in the pen, severe lameness and claw overgrowth might overlap, and that animals affected by severe health conditions that would result in kneeling behaviour in pen, should in any case be detected by the assessment of severe lameness and overgrown claws.

In the evaluation of lesions and swellings, the correlations identified between hindquarters lesions, and lesions in all the other areas assessed, might suggest that these integument alterations result from inappropriate housing conditions. For instance, if goats scramble when resources are limited, or collide with rough edges. The correlations found between neck, body and head lesions also support this idea. Moreover, in medium farms, the correlations

[8] Thermal stress description is presented in sub-section 2.1.6.

identified between the number of animals in pen and body and head lesions might also contribute to the results obtained. Accordingly to Canali and Keeling (2009), some resource and management-based indicators can be considered as risk factors for lesions, so these should be analysed so that conclusions can be reached. It should also be mentioned that these lesions might result from a bad HAR, for example if the animals are driven or approached aggressively, they are more likely to panic and hit the facilities, causing an increased risk of injuries to both animals and stockpeople, as highlighted by Muri et al. (2013). Thus, the HAR can also provide helpful information in this regard. However, the on farm testing of this indicator resulted in some difficulties, related to breed and moulting, leading to an overall 'fair to good' level of inter-observer reliability for κ (Fleiss et al., 2003), which contributed to the elimination of lesions and swellings from the final protocol.

Since knee lesions (score 1) were a very common finding among farms, reaching a prevalence around 94% in small farms, which is in agreement with the British and Norwegian studies (Anzuino et al., 2010; Muri et al., 2013), the author decided to eliminate 'Knee LAS – Score 1' from the protocol as well, as they are likely to be normal under all husbandry practices, thus improving the protocol's reliability.

Cleanliness assessment was also challenging in some breeds. Together with the fact that its recording depended mostly on how often bedding was replaced or added, the author decided to exclude this indicator as it is from the protocol and propose another approach. In fact, besides being difficult to assess in some breeds, it was shown that in medium farms, in summer, the hindquarters and lower legs cleanliness reached higher prevalences when comparing to winter season, contrary to what was expected[9]. Therefore, the author proposes for this indicator to be visually assessed inside the pen, by recording the quality and quantity of bedding, as a resource-based indicator, since alternative promising animal-based indicators to assess 'Comfort around resting' criterion are not available. As a matter of fact, according to Battini et al. (2014) other potentially promising indicators to assess this criterion, such as resting, average distance (between lying animals) or nosing on/exploring another goat, are too time consuming and cannot be considered feasible for an on-farm protocol. Similarly, udder asymmetry assessment was removed from the protocol, due to its lack of validity as a welfare indicator, following the results presented by Ajuda and Stilwell (2014).

[9] An analysis of the seasonal visits by farm category can be found in sub-section 2.2.6.

## 2. Inter-observer reliability (IOR) testing

### a) Agreement and reliability evaluation

Reliability was assessed by examining test agreement between two different observers, following a similar approach found in other studies in the development of welfare indicators, as Mullan et al. (2011) and Phythian et al. (2013b). Since there are no global scientific criteria establishing the limits for an 'acceptable' agreement between observers, general guidelines were assumed[10] for κ (Fleiss et al., 2003), $\kappa_w$ (Landis & Koch, 1977), ICC (Shrout, 1998) and Pearson's (r) and Spearman's rank correlations ($r_s$) for QBA (Martin & Bateson, 2007). The interpretation of κ values must take into consideration the prevalence of the assessed indicator in the study population, since a population with few affected animals will deliver artificially low values of reliability, as mentioned by Hoehler (2000). Even though it may be considered a small number of observers when comparing to other studies (e.g. Mullan et al., 2011), the observer pool (n=2) was determined due to feasibility constraints.

Considering categorical data, for κ the level of agreement of indicators as 'Cleanliness - hindquarters', 'Cleanliness - lower legs', 'Udder asymmetry', 'Claw overgrowth', 'Body abscesses' ,'Udder abscesses', 'Ocular discharge' and ' Diarrhoea' was 'excellent'. For $\kappa_w$, the level of agreement of 'BCS', 'Knee LAS' was 'substantial'. These results show that the assessment of these indicators was highly reliable pointing to a very understandable and useable welfare assessment system. In addition, following Hewetson et al. (2006) and Meagher (2009) considerations, these high levels of inter-observer reliability offer further proof on the validity of these indicators. The levels of inter-observer reliability for all the other categorical data considered were interpreted as 'fair to good' agreement for κ, with the exception of 'Hindquarters abscesses' and 'Vulvar discharge', which were observed at a very low prevalence (below 1%), being in agreement with the findings of Anzuino et al. (2010). Regarding continuous data, ICC for 'Improper disbudding', 'Queuing at feeding' and 'Queuing at drinking' indicated substantial agreement between observers, meaning that the assessment was reproducible and repeatable between observers. Indicators as kneeling (both at feeding and pen), oblivious and avoidance distance could not be computed due to a low number of recorded cases, however, comparing the observers' assessments they were close to a complete match.

With regard to QBA's inter-observer reliability analysis, the low level of observer agreement on the second dimension (PC2) is most likely a result of the lack of intensive training, as mentioned above. Phythian, Michalopoulou, Duncan and Wemelsfelder (2013a) state that a key aspect on QBA training is the observers' concentration on the overall expressive qualities

[10] Sub-section 2.1.8. provides information on the thresholds assumed for the interpretation of the IOR results.

of behaviour, rather than on specific behavioural aspects or clinical signs they are familiar with, and it is possible that this may account to some extent for the non-significant correlations found between the two assessors on PC2, as descriptors as bored, content, curious, irritated, relaxed or sociable are more difficult to interpret. For the item designated as suffering, and considering the veterinary background of the observers, these results might be related to the low prevalence of goats presenting signs of suffering, which artificially deliver low values of reliability. For this descriptor, from five cases recorded the assessors disagreed in four.

The ‘fair to good’ level of agreement obtained for κ in indicators as lesions and swellings, neck and head abscesses, and nasal discharges, might also be explained by the training intensity, which probably led to a bad interpretation of the indicators assessment and delimitation of the different body areas considered. In order to keep the integrity of the assessment, and achieve a high level of agreement between observers, reducing inter-observer variation, in-depth training is recommended by several authors (Kristensen et al., 2006; Rushen et al., 2011; EFSA, 2012). Accordingly, these results highlight the need for proper training to overcome the levels of agreement reached and obtain an objective and consistent welfare assessment outcome. Gibbons et al. (2012) mention that good training is particularly important when the evaluation involves multiple observers who may have very different levels of experience working with animals. This training can consist on classroom presentations and exercises, providing instructions on data collection procedures, and then practical-field exercises on farm. Several strategies to familiarize observers with the assessment system can be included, as using photographs, video clips and regularly assessing the observers until they develop a uniform scoring. Furthermore, refining definitions or data recording design can also improve the reliability, as suggested by Knierim and Winkler (2009).

**b) Constraints and perspectives**

As stated before, due to the low prevalences of ‘Vulvar discharge’ and ‘Hindquarters abscesses’, and due to the fact that vulvar discharges are highly influenced by the reproductive season, it was decided to drop these indicators from the final protocol.

**3. Consistency of indicators over time (COT)**

**a) Variation in prevalences between visits**

As discussed by Meagher (2009) and Temple et al. (2013) test–retest reliability refers to the chance that the same data will be achieved if the test is repeated. Following Plesch, Broerkens, Laister, Winckler and Knierim (2010), a particular case of test-retest reliability is the consistency of assessments over time (COT) at farm level, meaning that results must be representative of the longer-term farm situation and not too sensitive to changes, either in the farming conditions, the weather or in the internal states of the animals, as long as the circumstances have not altered significantly. Capdeville and Vessier (2001) and Winckler et al. (2003) state that high levels of consistency are essential for welfare assessment systems at farm-level, guaranteeing objectivity for the farmer and credibility of the system, and according to Knierim and Winckler (2009) and Sørensen (2010), contributing to the reduction of recording costs, due to less farm visits being necessary since having indicators that do not change significantly over a long period of time, considering that farm conditions continue constant, will not require regularly repeated visits to obtain reliable estimates. Kirchner et al. (2013) state that the decision to reassess these farms would be based on the desired interval for the detection of actual changes in animal welfare.

Analysing the variation in the indicators prevalence between the two visits[11], the Wilcoxon signed rank test performed showed that only head lesions presented results significantly different (P=0.037). An overall consistency of results was apparent, with common findings such as improper disbudding or presence of external abscesses remaining common and those conditions that occurred less often, such as oblivious animals, remaining at low levels of prevalence. Some indicators' prevalence showed a change above 5% between the two visits (%Δ>5%) namely, 'BCS - very fat', 'Diarrhoea', 'Cleanliness - hindquarters', 'Knee LAS - score 2', 'Head lesions' , 'Ocular discharge' and 'Claw overgrowth'.

Blokhuis et al. (2013) express that the variability of an indicator across assessments may be caused by variation due to real differences on its prevalence, because of changes in management or feeding or due to an outbreak of a specific disease, as in the case of 'BCS - very fat' and 'Diarrhoea'. In the first assessment three farms presented an outbreak of diarrhoea, leading to these results. Furthermore, the effects of improvements in welfare state (and in the indicator in question), which might have resulted from the adoption of management or disease control measures by the farmer, as a consequence of the information received during the feedback of the findings from the first assessment, might also contribute to a low reliability of an indicator over time. In this study, this might have contributed for the 'Claw overgrowth' results, for instance. However, it is worth mentioning that the zero prevalence of overgrown claws found in small farms during the second assessment, might be related to a higher access to outdoor grazing during this season. Similarly, Temple et al. (2013) point that the variability of an indicator can be explained by seasonal effects and methodological constrictions (e.g. intra-observer effect or sampling strategy). For example, goats' dirtiness is influenced by seasonal effects and cleanliness routines, depending on how

[11] Table 21 (Annex 1) presents the differences in the prevalences of the animal-based indicators recorded during the two visits.

often bedding is replaced or added (Table 36). Seasonal effects can also explain the variability found in 'Ocular discharge'. Another possibility for low COT is the lack of intra-observer reliability. During the second sequence of visits an observer showed some difficulty in scoring knee lesions, therefore reinforcing the need for training. In a study performed by Gibbons et al. (2012) a five-day break in a training programme, to train observers to score injuries on dairy cows, resulted in decreased agreement for all injury scores, improving again in the next day after practice. This highlights the importance of continual practice and refresher course, mainly during the sensitive learning stage.

**b) Indicators relationships between visits**

Based on Spearman's rank correlations ($r_s$) only the presence of body abscesses and severely lame animals were repeatable from one visit to another. With concern to these indicators, attention should be given to farms that present these problems persistently. Similarly, the fact that body abscesses, due to CLA, and severe lameness are typically chronic conditions that persevere can support their consistency. Naturally, the interpretation of these findings should be made carefully, as each visit would have been subject to variations as in climate, season or pen in assessment. Even though this analysis of data from only two visits is unlikely to represent accurately the real variability over a longer period of time, it can still provide a little guidance on the variability of animal-based indicators.

As mentioned, in the present study Spearman's rank correlation ($r_s$) was applied to evaluate consistency of results over time. However, since this measure of reliability strongly depends on the variance of the recorded population of farms (Costa-Santos et al., 2011), the combined use of measures of reliability and agreement (as limits of agreement; Bland & Altman, 1986) is advised by several authors (de Vet et al.,2006; Dohoo et al.,2009). Nevertheless, since only two significant correlations between the two visits were found (at the 0.05-level), and were considered unambiguous, biologically conclusive and of importance in terms of animal welfare, the author decided to present the results. In addition, this study is a first approach to the subject and an in-depth analysis would have gone beyond the scope of this work. Furthermore, given the short period between the two visits (around four months) it was not obvious whether any significant changes seen between visits reflected natural variability in the herd performance, or the impacts of management changes. The continuous use of these animal-based indicators will deliver more data, providing increasing information on which indicators offer reliable results with long term collection, and which ones are most likely to drive change.

### c) Constraints and perspectives

Up to this point, and following Table 12 (section 2.2.), due to their low prevalence the assessment of panting and shivering (score 2), kneeling in pen, hindquarters abscesses and vulvar discharge was excluded from the final protocol. For feasibility reasons the avoidance distance test, cleanliness and lesions and swellings assessment were also removed from the protocol, with the lower levels of inter-observer reliability found for 'LAS' contributing as well for this decision. Although showing a high prevalence in the studied farms, udder asymmetry and knee lesions (score 1) were also drop from the final protocol, due to lack of validity.

## 4. On-farm feasibility

### a) On-farm feasibility assessment

Following Blokhuis et al. (2013), the implementation of any welfare assessment system arises in an environment that is intensely determined by economic, political, technological and socio-cultural factors that can all interrelate with each other. In this perspective, the feasibility of the prototype was examined in the first sequence of visits to the farms, to determine if the welfare assessment system was concise, effective and practical to perform in commercial situations. When testing the feasibility of the prototype, the focus was set on time requirements to complete each stage, peculiarities concerning characteristics of goats (e.g. natural curiosity), of the indicators (e.g. assessment procedure), and potential variability of on-farm conditions regarding farm size categories (e.g. number of pens). The mean time necessary to perform each stage of the prototype was recorded in 26 farms. In four farms assessors could not record the time needed to assess each animal-based indicator due to several reasons, for instance, due to the goats' natural curiosity leading the animals to reach the recording devices and making it impossible to record the time.

The number of animals per pen, the stocking densities and the animals' behaviour affected the time required to collect each welfare indicator[12]. 'Queuing' and 'Clinical scoring' were the most time consuming, with the average time to accomplish the different stages of the prototype increasing with the number of animals that had to be sampled. Although the mean time required to perform the protocol among farm categories did not differ considerably, in a future project one or more pens will be assessed, according to the number of pens in the farm, and therefore more animals, which will lead to different results. The length of the prototype also depended on the assessors' training and on the farm conditions, i.e., if goats could be restrained at their home pen, or animals could be inspected in the milking parlour.

[12] In sub-section 2.2.7. is presented the time necessary to accomplish the different stages of the prototype.

Occasionally, the exploratory behaviour of goats complicated the observations making the evaluation challenging. In some farms the collection of particular indicators, namely 'Improper disbudding' (seven farms) could not be carried out, due to high stocking densities associated with the disposition of the feeding places. However, regarding 'Improper disbudding', since it is the only promising indicator for the assessment of the 'Absence of pain induced by management procedures' criterion, presenting a substantial agreement between observers, it was decided to keep it in the protocol. Otherwise, the application of the prototype was easily performed.

The fact that some assessments (queuing) had to be carried out at certain times (after feeding distribution) reduced flexibility and limited the number of farms that could be evaluated in a given day. The time spent in resource and management-based indicators assessment questionnaire was highly influenced by the farmer's level of interest, being the only part that required the farmers input. However, as it provided important insight to the farm routines and risk factors underlying the indicators' prevalence, it was kept in the protocol.

**Stakeholders' concerns**

In the early stages of the AWIN project, when the project was being presented to stakeholders, several farmers expressed concern that applying the protocol would take too long and thus be too costly. This concern was made particularly clear at the goat stakeholders meeting organized in Portugal (Campo Branco; November 22, 2013), within one of the most important conferences of small ruminants in the country, where the indicators to be integrated were presented. Also concerns related to the potential bad image of certain breeds and production systems, as an outcome of a welfare assessment tool, were raised.

As one of the main goals of AWIN was the development of protocols with a high level of acceptability from the stakeholders, the opinion of farmers and technicians was also taken into consideration. The indicators chosen received generally positive responses and most farmers were surprised that it required little participation on their part in the data collection, thus not compromising their time. Furthermore, farmers showed a high level of interest in the animal-based indicators, especially in the behavioural ones and some comments and suggestions were made. It has to be mentioned that most farmers showed interest in the AWIN project and were keen on being informed about the results of the study afterwards, especially in knowing where their farms were situated in terms of animal welfare in relation to others.

**b) Lessons learned during prototype testing: proposal for a final protocol**

The choice of which indicators are to be finally incorporated in on-farm welfare assessment protocols rely on its validity, reliability and feasibility, as well as the objective of the assessment, the skills of the assessor, the conditions under which it is to be performed, the time available to collect data and financial constraints, as was underlined by Main (2009) and EFSA (2012). Regarding these considerations, it may not be necessary/feasible to perform a full assessment on each farm visit. As Blokhuis et al. (2013) refer, in order to encourage acceptance it is essential to adjust the workload and time requested, whilst guaranteeing that the holistic nature of the assessment remains, providing an overall and reliable view of animal welfare, which reinforces the earlier perception of the AWIN project towards a two-step approach.

**The two-step approach**

The flow of data collection of the first and second level welfare assessment suggested protocols, starts at the time of feed distribution (main meal) from outside the pen, and continues inside the pen, as performed while testing the prototype. After applying the protocol, the farm manager is asked to answer the resource and management-based indicators assessment questionnaire.

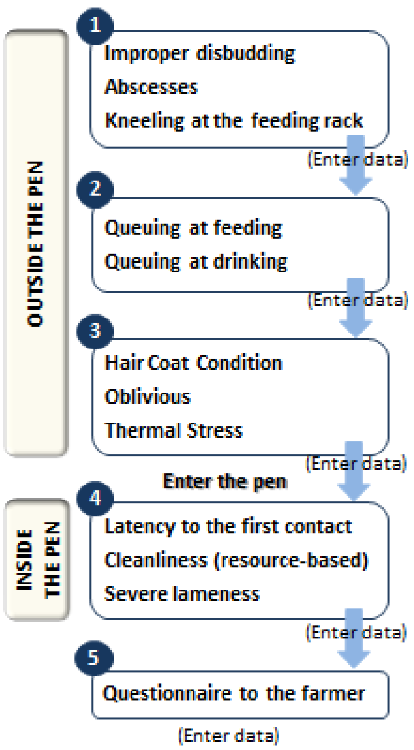



Figure 26 - Flow of data collection of first level welfare assessment suggested protocol for intensively kept dairy goats.

As depicted from Figure 26, in a first step, only pen-level observations are performed, using relevant animal-based indicators that presented an overall high level of agreement between observers, and a resource-based indicator. The welfare assessment starts with the recording of indicators as ‘Improper disbudding’, ‘Abscesses’ (visual assessment of head, neck and shoulders of each goat from the feeding barrier), ‘Kneeling at the feeding rack’, ‘Queuing at feeding’, ‘Queuing at drinking’, ‘Hair coat condition’, ‘Oblivious’, ‘Panting Score’ and

'Shivering Score'. Afterward, the assessor enters the pen performs 'Latency to first contact' test, and assesses 'Cleanliness (resource based)' and 'Severe lameness'. This first level welfare assessment protocol only requires the presence of a single assessor and observations are carried out only in one pen with adult dairy goats, even if more pens with animals are present. To minimise the risk of "false negative", the pen considered at higher risk is selected, regarding aspects as density, feeding space/animal ratio, drinking place/animal ratio and presence of both horned and hornless goats. Those farms that did not reach an acceptable outcome in all indicators, or that did not comply with the current legislation are then subject to a second-level protocol (Figure 27).

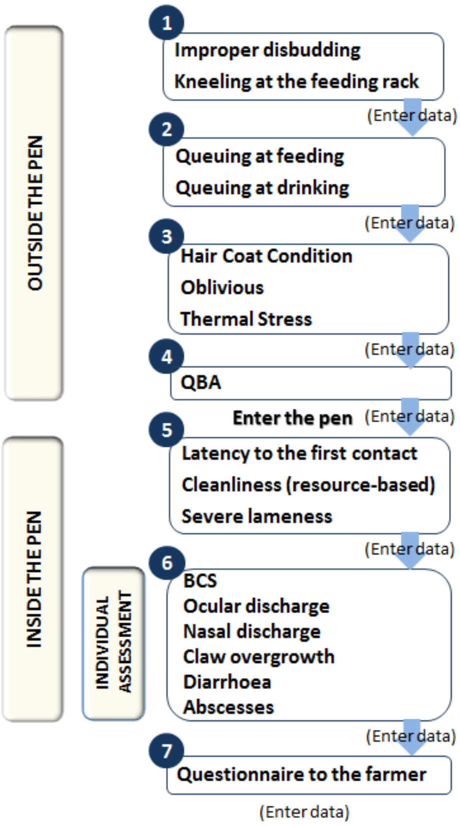



Figure 27 - Flow of data collection of second level welfare assessment suggested protocol for intensively kept dairy goats.

In this second-level protocol, two assessors have to be present and a more comprehensive protocol is applied, involving an individual close examination of animals (individual assessment), the assessment of more indicators, and of those that needed more training (e.g. 'Nasal discharge', 'Head abscesses'[13], 'Neck abscesses' and 'QBA'). The flow of data collection is similar to the flow adopted during prototype testing. The number of pens (excluding infirmary, culling or maternity pens), and animals to be examined will depend on the size of the farm: in those with two to seven pens the assessment will be performed in two,

[13] The assessment of 'Abscesses' in a second-level protocol is performed by detailed individual observations of the animals, as mentioned in sub-section 2.1.6.

in farms with eight to 10 pens the assessment of three will be needed, and in the ones with more than 11 pens, at least 25% of these will have to be assessed. Furthermore, for feasibility reasons, the location for individual evaluation may vary, the goats can be restrained in their home pen, or the animals can be assessed in the milking parlour (either during milking or not), depending on farm conditions and routines, which allows a more flexible assessment.

Although the ‘Queuing’ stage involves an observation period of at least 15 minutes, from the research performed in this study it was apparent that this indicator cannot be replaced by resource-based indicators. In large farms queuing animals at the feeding place presented a higher prevalence (around 40%), however when comparing the number of feed spaces, the feed trough length, and the number of animals per pen, between medium and large farms they had similar values. This finding emphasizes the need for an animal-based indicator to assess the ‘Absence of prolonged hunger’ criterion. Additionally, it is important to stress that some indicators, such as BCS, hair coat condition and queuing (both at feeding and drinking) can be used to assess different criteria and provide information about several welfare aspects, being beneficial in order to save time.

Despite of existing few validated animal-based indicators for the assessment of dairy goats’ welfare comparatively to other species, for most of the 12 criteria considered, following the Welfare Quality® approach, it was possible to identify one or several animal-based indicators that meet the required conditions to obtain an operational on-farm welfare assessment system that in the end, can serve multiple practical purposes, such as a research, legislative, certification and as advisory tool.

## CHAPTER III – Concluding remarks and further research perspectives

This study describes the results of the application of the on-farm welfare assessment prototype for dairy goats, developed by the AWIN project, where the reliability and feasibility of the animal-based indicators involved were tested. The application of the prototype, in 30 Portuguese farms, contributed to an increased awareness of the main welfare issues affecting intensively kept dairy goats.

The main areas of concern were claw overgrowth, queuing at feeding, overweight animals, poor hair coat condition and improper disbudding. Some of the assessed indicators presented similar prevalences to those included in previous studies (e.g. Anzuino et al., 2010; Muri et al., 2013), underlining common problems in different countries. The relationships between animal-based welfare indicators revealed the importance of the living conditions and management aspects, and how these characteristics reflect on the animals' welfare, for instance by means of their cleanliness, lesions or claw overgrowth. Furthermore, our results show that the identified welfare issues are related to farm sizes, with larger farms heading higher concerns.

The animal-based indicators under study presented moderate to high levels of agreement between observers, with the exception of QBA, emphasizing the need for in-depth training when considering this indicator. Some items due to their very low prevalence showed artificially low values of reliability (e.g. 'Vulvar discharge', 'Hindquarters abscesses'), or could not be computed (e.g. 'Oblivious', 'Severe lameness'), which points toward the need for further studies on their on-farm prevalence and reliability. In addition, to standardize the implementation of the assessment protocol, studies on intra-observer reliability of the considered indicators are also required, as the consistency of items over time is essential in welfare assessment protocols that need to be applied at any time.

The analysis of the variation in the indicators prevalence between seasonal visits revealed an overall consistency of results, with common findings remaining stable (e.g. improper disbudding) and those conditions that happened less frequently, continuing at low levels of prevalence (e.g. kneeling, both in pen and at feeding). Some of the indicators prevalence presented a change above 5% over this period of time (e.g. 'BCS - very fat', 'Claw overgrowth'), but only the prevalence of 'Head lesions' showed results significantly different between visits. From further analysis, based on Spearman's rank correlations, only the presence of body abscesses and of severely lame animals were repeatable from one visit to another.

Several factors, as stocking densities, the animals' behaviour, farm routines and the assessor's skills, affected the time required to assess each welfare indicator. From all the stages of the prototype ‘Queuing’ and ‘Clinical scoring’ were the most time consuming. As expected, the mean time necessary to apply the prototype was longer in large farms (about two hours), although it did not differ considerably among farm categories, since only a single pen was assessed in all the farm categories under study. In spite of the assessment of animal-based indicators being more complex and time consuming, from this research it was fairly straightforward that for most criteria these indicators cannot be replaced by resource-based indicators. These were only considered when there were no potentially feasible animal-based indicators available (e.g. ‘Cleanliness’).

An overall analysis of all the parts of this research project led to the suggestion of a final protocol, which the author believes has the potential to work not only as a legislative and regulatory basis but also as a certification tool, following a two-step approach, due to the amount of time and effort required for a comprehensive welfare assessment on farm.

The prototype was a preliminary step in the direction of creating a standardized scheme for on-farm welfare assessment of dairy goats in Europe, comprising animal-based indicators. Even though the AWIN project developed an official on-farm welfare assessment protocol for intensively kept dairy goats, it did not cover all the questions and every aspect of this particular species, actually raising new questions, discussion points and demands. Considering the increasing societal request for farm animal welfare, the next step would be to develop and implement specific protocols for other categories, as kids and bucks, different purposes, as meat production, and production systems, such as extensive systems. The outcomes could therefore be compared to assess if different levels of animal welfare are related to different categories and levels of intensification, allowing the identification of priority categories for legislation, for instance. It would also facilitate the analysis of associations between welfare problems and pinpoint their associated risk factors. As at present there is scarce information on the welfare of goats and more is needed in order to fill the gaps. The continuous application of the protocols through time will deliver more data on this subject, providing increased information. In fact, at the moment a mobile application software (app) is being developed for on-farm data collection, which will allow assessing welfare indicators in dairy goats in a simple and standardized manner. A fundamental issue is how this information will be managed, therefore a promising area of research is the development and management of a central database that will allow storing all data collected, allowing for the extension of knowledge. Subsequently, these data could be used to continuously update stakeholders, help farmers acknowledge their progress and compare their position with European averages, for instance, introducing the concept of benchmarking in this field. Like in all welfare assessment schemes, these protocols will also need regular updating and adjustment in view of new scientific evidence, societal developments, and practical experiences obtained during their application, to accurately reflect the state of the animal rather than just the nature and quality of its living conditions.

Further research on the identification of valid, reliable and significant ‘iceberg’ indicators, i.e., key indicators that are expected to reveal major problems on a farm is also required.

Another area of research would be to study potential problematic aspects at a particular farm, identifying risk factors and help determining the frequency of visits necessary to obtain a consistent welfare assessment. For instance, if a high prevalence of overweight animals is identified as a major problem on a specific farm, the assessment of the resource-based indicators (e.g. quality and number of feed delivery per day, feed trough length, number of feed spaces per pen) will be needed to identify the source of the problem and the best strategy to solve it. This approach would lead to more efficient assessments, focused primarily or even exclusively on major welfare risks. Furthermore, it would also allow tailoring possible advice or recommendations to the farm-specific circumstances, which can encourage cost-benefit studies of welfare improvement strategies.

An additional path of research would be the development of methodologies to reduce the workload and time involved on animal welfare assessment at farm level. For instance, by the development and validation of new and more practical animal-based welfare indicators, in order to simplify the protocols or produce shorter but efficient tools to implement the system.

## References


Abdi, H. & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459. doi:10.1002/wics.101

Ajuda, I. & Stilwell, G. (2014). Using farm routine procedures to reduce pain in dairy goats. [abstract] [electronic version]. In A.J. Zanella (Ed.), *Proceedings of Animal Welfare Indicators 3rd Annual Conference, Prague, 13-15th May 2014*, p. 12. Retrieved Jan. 22, 2015 from http://www.animal-welfare-indicators.net/site/index.php/awin-events

Ajuda, I., Vieira, A. & Stilwell, G. (2014, June). *Are there differences in dairy goats' claws' temperature, before and after trimming?* Paper presented at the IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lisbon, Portugal.

Alamer, M. (2006). Physiological responses of Saudi Arabia indigenous goats to water deprivation. *Small Ruminant Research*, 63(1-2), 100–109. doi:10.1016/j.smallrumres.2005.02.004

Alamer, M. (2009). Effect of water restriction on lactation performance of Aardi goats under heat stress conditions. *Small Ruminant Research*, 84(1-3), 76–81. doi:10.1016/j.smallrumres.2009.06.009

Alawa, J., Ngele, M., & Ogwu, D. (2000). Chronic caprine mastitis in Nigerian goat breeds: microbiological flora and histopathological findings. *Small Ruminant Research*, 35(3), 203–207. doi:10.1016/s0921-4488(99)00099-1

Albenzio, M., Taibi, L., Caroprese, M., De Rosa, G., Muscio, A. & Sevi, A. (2003). Immune response, udder health and productive traits of machine milked and suckling ewes. *Small Ruminant Research*, 48(3), 189–200. doi:10.1016/s0921-4488(03)00005-1

Altinçekiç, S.O. & Koyuncu, M. (2012). Small ruminant health and welfare. *Krmiva*, 54(6), 205-216. Retrieved from hrcak.srce.hr/file/152602

Amon, T., Amon, B., Ofner, E. & Boxberger, J. (2001). Precision of assessment of animal welfare by the 'TGI 35 L' Austrian Needs Index. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 51(30), 114-117. doi: 10.1080/090647001316923180

Andersen, I. L. & Bøe, K. E. (2007). Resting pattern and social interactions in goats—The impact of size and organisation of lying space. *Applied Animal Behaviour Science*, 108(1-2), 89–103. doi:10.1016/j.applanim.2006.10.015

Andersen, I.L., Roussel, S., Ropstad, E., Braastad, B.O., Steinheim, G., Janczak, A.M., Jørgensen, G.H.M. & Bøe, K.E. (2008). Social instability increases aggression in groups of dairy goats, but with minor consequences for the goats' growth, kid production and development. *Applied Animal Behaviour Science*, 114(1-2), 132–148. doi:10.1016/j.applanim.2008.01.007

Andreasen, S.N. & Forkman B. (2012). The welfare of dairy cows is improved in relation to cleanliness and integument alterations on the hocks and lameness when sand is used as stall surface. *Journal of Dairy Science*, 95(9):4961-4967. doi: 10.3168/jds.2011-5169

Anil, S., Anil, L. & Deen, J. (2009). Effect of lameness in pigs in terms of “Five Freedoms”. *Journal of Applied Animal Welfare Science*, 12(2), 144–145. doi:10.1080/10888700902720094

Animal Welfare Indicators (n.d.). *Animal Welfare Indicators: Overall organization*. Retrieved Jan. 14, 2015 from http://www.animal-welfare-indicators.net/site/index.php/project-overview

Animal Welfare Indicators (2014a). *AWIN newsletter – March 2014*. Retrieved Jan. 20, 2015 from http://www.animal-welfare-indicators.net/site/index.php/mar-2014

Animal Welfare Indicators (2014b). *AWIN welfare assessment for goats: AWIN training guide*. Unpublished Manuscript.

Anzuino, K, Bell, N.J., Bazeley, K.J. & Nicol, C.J. (2010). Assessment of welfare on 24 commercial UK dairy goat farms based on direct observations. *Veterinary Record*, 167(20), 774–780. doi:10.1136/vr.c5892

Appleby, M.C. (2003). The European Union ban on conventional cages for laying hens: History and prospects. *Journal of Applied Animal Welfare Science*, 6(2), 103–121. doi:10.1207/s15327604jaws0602_03

Aziz, M.A. (2010). Present status of the world goat populations and their productivity. *Lohmann Information* 45(2), p. 42. Retrieved from http://lohmann-information.com/content/l_i_45_artikel17.pdf

Barbosa, M. (1993). Goat's milk research in Portugal. *Le Lait*, 73 (5-6), 425-429. doi: http://dx.doi.org/10.1051/lait:19935-640

Barroso, F. G., Alados, C. L. & Boza, J. (2000). Social hierarchy in the domestic goat: effect on food habits and production. *Applied Animal Behaviour Science*, 69(1), 35–53. doi:10.1016/s0168-1591(00)00113-1

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(2-3), 179–192. doi:10.1016/s0301-6226(99)00067-6

Bartussek, H. (2001). An historical account of the development of the Animal Needs Index ANI-35L as part of the attempt to promote and regulate farm animal welfare in Austria: An example of the interaction between animal welfare science and society. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 51(30), 34–41. doi:10.1080/090647001316923036

Battini, M., Vieira, A., Barbieri, S., Ajuda, I., Stilwell, G. & Mattiello, S. (2014). Invited review: Animal-based indicators for on-farm welfare assessment for dairy goats. *Journal of Dairy Science*, 97(11), 6625–6648. doi:10.3168/jds.2013-7493

Battini, M., Peric, T., Ajuda, I., Vieira, A., Grosso, L., Barbieri, S. Stilwell, G., Prandi, A., Comin, A., Tubaro, F. & Mattiello, S. (2015). Hair coat condition: A valid and reliable indicator for on-farm welfare assessment in adult dairy goats. *Small Ruminant Research*, 123(2-3), 197–203. doi:10.1016/j.smallrumres.2014.12.009

Baumont, R., Prache, S., Meuret, M. & Morand-Fehr, P. (2000). How forage characteristics influence behaviour and intake in small ruminants: a review. *Livestock Production Science*, 64(1), 15–28. doi:10.1016/s0301-6226(00)00172-x

Bell, D.E, Raiffa, H & Tversky, A. (1988). *Decision making: descriptive, normative and prescriptive interactions*. Cambridge, UK: Cambridge University Press.

Bennett, R.M. and Appleby, M. (2010) Animal welfare policy in the European Union. In: Oskam, A., Meester, G. and Silvis, H. (eds.) *EU policy for agriculture, food and rural areas*. Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 243-251. ISBN 987-90-8686-118-7

Berg, W., Jolly, A., Rambeloarivony, H., Andrianome, V. & Rasamimanana, H. (2009). A scoring system for coat and tail condition in ringtailed lemurs, Lemur catta. *American Journal of Primatology*, 71(3), 183–190. doi:10.1002/ajp.20652

Bergonier, D., Berthelot, X. & Poumarat, F. (1997). Contagious agalactia of small ruminants: Current knowledge concerning epidemiology, diagnosis and control. *Rev. sci. tech. Off. int. Epiz.*,;16(3), 848–873.

Bland, J. M., & Altman, D.G. (1996). Statistics Notes: Measurement error and correlation coefficients. *BMJ*, 313(7048), 41–42. doi:10.1136/bmj.313.7048.41

Blandford, D. & Fulponi, L. (1999). Emerging public concerns in agriculture: domestic policies and international trade commitments. *Eur Rev Agric Econ*, 26(3), 409–424. doi:10.1093/erae/26.3.409

Blokhuis, H.J., Jones, R.B., Geers, R., Miele, M. & Veissier, I. (2003). Measuring and monitoring animal welfare: Transparency in the food product quality chain. *Animal Welfare*, 12(4), 445 – 455. Retrieved from http://www.researchgate.net/publication/40793867_Measuring_and_monitoring_animal_welfare_Transparency_in_the_food_product_quality_chain

Blokhuis, H.J., Veissier, I., Miele, M. & Jones, B. (2010). The Welfare Quality® project and beyond: Safeguarding farm animal well-being. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 60(3), 129–140. doi:10.1080/09064702.2010.523480

Blokhuis, H., Miele, M., Veissier, I., & Jones, B. (Eds). (2013). *Improving farm animal welfare. Science and society working together: The Welfare Quality approach.* Wageningen, The Netherlands: Wageningen Academic Publisher. ISBN: 978-90-8686-216-0

Bøe, K.E., Berg, S., & Andersen, I.L. (2006). Resting behaviour and displacements in ewes—effects of reduced lying space and pen shape. *Applied Animal Behaviour Science*, 98(3-4), 249–259. doi:10.1016/j.applanim.2005.10.001

Bøe, K.E. & Ehrlenbruch, R. (2013). Thermoregulatory behavior of dairy goats at low temperatures and the use of outdoor yards. *Can. J. Anim. Sci*.; 93(1), 35–41. doi:10.4141/cjas2012-028

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(3), 375–397. doi:10.1016/j.physbeh.2007.02.003

Bokko, B.P. & Chaudhari, S.U.R. (2001). Prevalence of lameness in sheep in the north east region of Nigeria. *International Journal of Agriculture & Biology*, 3(4), 519–521. Retrieved from http://www.fspublishers.org/published_papers/8834_..pdf

Botreau, R., Veissier, I, Butterworth, A., Bracke, M.B.M & Keeling, L.J. (2007a). Definition of criteria for overall assessment of animal welfare. *Animal Welfare*, 16, 225-228.

Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M. B. M., Capdeville, J. & Veissier, I. (2007b). Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods. *Animal*, 1(08), 1179-1187. doi:10.1017/s1751731107000535

Botreau, R., Bracke, M. B. M., Perny, P., Butterworth, A., Capdeville, J., Van Reenen, C. G., & Veissier, I. (2007c). Aggregation of measures to produce an overall assessment of animal welfare. Part 2: analysis of constraints. *Animal*, 1(8), 1188-1197. doi:10.1017/s1751731107000547

Botreau, R., Capdeville, J., Perny, P. & Veissier, I. (2008). Multicriteria evaluation of animal welfare at farm level: An application of mcda methodologies. *Foundation of Computing and Decision Sciences*, 33, 1–18.

Botreau, R., Veissier, I & Perny, P. (2009). Overall assessment of animal welfare: Strategy adopted in Welfare Quality®. *Animal Welfare*, 18, 363-370. *Foundation of Computing and Decision Sciences*, 33, 287-316.

Boyazoglu, J. & Morand-Fehr, P. (2001). Mediterranean dairy sheep and goat products and their quality. *Small Ruminant Research*, 40(1), 1–11. doi:10.1016/s0921-4488(00)00203-0

Boyazoglu, J., Hatziminaoglou, I. & Morand-Fehr, P. (2005). The role of the goat in society: Past, present and perspectives for the future. *Small Ruminant Research*, 60(1-2), 13–23. doi:10.1016/j.smallrumres.2005.06.003

Bracke, M.B.M. (Ed.). (2009). *Animal welfare in a global perspective: A survey of foreign agricultural services and case studies on poultry, aquaculture and wildlife*. Wageningen, The Netherlands: Wageningen Academic Publishers

Bracke, M. B. M., Spruijt, B. M. & Metz, J. H. M. (1999). Overall animal welfare reviewed: Part 3: Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science*, 47, 307–322.

Bracke, M.B.M., Spruijt, B.M., Metz, J.H.M. & Schouten, W.G.P. (2002). Decision support system for overall welfare assessment in pregnant sows a: Model structure and weighting procedure. *Journal of Animal Science*, 80, 1819-1834. doi:/2002.8071819x

Brambell Committee. (1965). *Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems. Command report 2836*. London, The United Kingdom: Her Majesty's Stationery Office.

Brenner, H. & Kliebsch, U. (1996). Dependence of weighted kappa coefficients on the number of categories. *Epidemiology*, 7(2), 199–202. doi:10.1097/00001648-199603000-00016

Broom, D. M. (1986). Indicators of poor welfare. *British Veterinary Journal*, 142(6), 524–526. doi:10.1016/0007-1935(86)90109-0

Broom, D. M. (1991). Animal welfare: Concepts and measurement. *Journal of Animal Science*, 69, 4167-4175.

Broom, D.M. (2006). Behaviour and welfare in relation to pathology. *Applied Animal Behaviour Science*, 97(1), 73–83. doi:10.1016/j.applanim.2005.11.019

Broom, D.M. (2007). Welfare in relation to feelings, stress and health. *Revista electrónica de Veterinária*, 8(12B), 1695-7504.

Broom, D.M. & Johnson, K.G. (1993). *Stress and animal welfare*. London: Chapman & Hall. ISBN: 0-412-39580-0

Broom, D. M. & Corke, M. J. (2002). Effects of Disease on Farm Animal Welfare. *Acta Vet. Brno*, 71(1), 133–136. doi:10.2754/avb200271010133

Brozos, C., Mavrogianni, V.S. & Fthenakis, G.C. (2011). Treatment and control of peri-parturient metabolic diseases: pregnancy toxemia, hypocalcemia, hypomagnesemia. *Veterinary Clinics of North America: Food Animal Practice*, 27(1), 105–113. doi:10.1016/j.cvfa.2010.10.004

Bruno-de-Sousa, C., Martinez, A.M., Ginja, C., Santos-Silva, F., Carolino, M.I., Delgado, J.V. & Gama, L.T. (2011). Genetic diversity and population structure in Portuguese goat breeds. *Livestock Science*, 135(2-3), 131–139. doi:10.1016/j.livsci.2010.06.159

Caldeira, R.M., Belo, A.T., Santos, C.C., Vazques, M.I & Portugal, A.V. (2007). The effect of long-term feed restriction and over-nutrition on body condition score, blood metabolites and hormonal profiles in ewes. *Small Ruminant Research*, 68(3), 242–255. doi:10.1016/j.smallrumres.2005.08.026

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. doi:10.1037/h0046016

Canali, E. & Keeling, L. (2009). Welfare Quality® project: from scientific research to on farm assessment of animal welfare. *Italian Journal of Animal Science*, 8(2), 900-903.

Canali, E. & Barbieri, S. (2014). Welfare Assessment Protocols - WP1 Highlights, Talks. [abstract] [electronic version]. In A.J. Zanella (Ed.), *Proceedings of Animal Welfare Indicators 3rd Annual Conference, Prague, 13-15th May 2014*, pp. 17-18. Retrieved Jan. 22, 2015 from http://www.animal-welfare-indicators.net/site/index.php/awin-events

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*, 51(30), 62–68. doi:10.1080/090647001316923081

Carenzi, C. & Verga, M. (2009). Animal welfare: review of the scientific concept and definition. *Italian Journal of Animal Science*, 8(1), 21-30.

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*, 8(1s), 343. doi:10.4081/ijas.2009.s1.343

Castel, J.M., Ruiz, F.A., Mena, Y. & Sánchez-Rodríguez, M. (2010). Present situation and future perspectives for goat production systems in Spain. *Small Ruminant Research*, 89(2-3), 207–210. doi:10.1016/j.smallrumres.2009.12.045

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. doi:10.1037/h0026256

Colditz, I.G., Ferguson, D.M., Collins, T., Matthews, L. & Hemsworth, P.H. (2014). A prototype tool to enable farmers to measure and improve the welfare performance of the farm animal enterprise: The unified field index. *Animals*, 4(3), 446–462. doi:10.3390/ani4030446

Coleman, J.S.M. (2010). Principal Components Analysis. In N.J. Salkind. (Ed.) *Encyclopedia of research design*. London: SAGE Publications, Inc.

Condon, D. & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52-64. doi:10.1016/j.intell.2014.01.004

Contreras, A., Sierra, D., Sánchez, A., Corrales, J.C., Marco, J.C., Paape, M.J. & Gonzalo, C. (2007). Mastitis in small ruminants. *Small Ruminant Research*, 68(1-2), 145–153. doi:10.1016/j.smallrumres.2006.09.011

Costa, E.D., Murray, L., Dai, F., Canali, E., & Minero, M. (2014). Equine on-farm welfare assessment: A review of animal-based indicators. *Animal Welfare*, 23(3), 323–341. doi:10.7120/09627286.23.3.323

Costa-Santos, C., Bernardes, J., Ayres-de-Campos, D., Costa, A. & Costa, C. (2011). The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. *Journal of Clinical Epidemiology*, 64(3), 264–269. doi:10.1016/j.jclinepi.2009.11.010

Christodoulopoulos, G. (2009). Foot lameness in dairy goats. *Research in Veterinary Science*, 86(2), 281–284. doi:10.1016/j.rvsc.2008.07.013

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. doi:10.1037/h0040957

Dalla Villa, P., Matthews, L.R.., Alessandrini, B., Messori S. &. Migliorati G. (2014). Drivers for animal welfare policies in Europe. *Rev. sci. tech. Off. int. Epiz*., 33 (1), 39-46.

Dalmau, A., Geverink, N.A., Van Nuffel, A., van Steenbergen, L., Van Reenen, K., Hautekiet, V, Vermeulen, K., Velarde, A. & Tuyttens, F.A.M. (2010). Repeatability of lameness, fear and slipping scores to assess animal welfare upon arrival in pig slaughterhouses. *Animal*, 4(5), 804–809. doi:10.1017/S1751731110000066

Dar, A. A., Jadhav, R.K., Dimri, U., Khan, A.A., Khan, H.M. & Sharma, M.C. (2014). Effects of physiological status and seasonal variation on plasma mineral profile of sheep in Kashmir valley. *Scientific Research and Essays*, 9(4), 69-76. doi: 10.5897/SRE2013.5694

Darcan, N., Cedden, F. & Cankaya, S. (2009). Spraying effects on some physiological and behavioural traits of goats in a subtropical climate. *Italian Journal of Animal Science*, 7(1), 77. doi:10.4081/ijas.2008.77

Dawkins, M.S. (2000). Animal minds and animal emotions. *American Zoologist*, 40(6), 883–888. doi: 10.1093/icb/40.6.883

Dawkins, M.S. (2004). Using behaviour to assess animal welfare. *Animal Welfare*, 13(1), 3-7(5).

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(3), 193–209. doi:10.1016/j.applanim.2005.05.006

De Rosa, G., Grasso, F., Pacelli, C., Napolitano, F. & Winckler, C. (2009). The welfare of dairy buffalo. *Italian Journal of Animal Science*, 8(1), 103. doi:10.4081/ijas.2009.s1.103

De Vet, H.C., Terwee, C.B., Knol, D.L., Bouter, L.M. (2006) When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59(10), 1033–1039. doi:10.1016/j.jclinepi.2005.10.015

De Vries, M., Bokkers, E.A.M.,van Schaik, G., Botreau, R., Engel, B., Dijkstra, T. & de Boer, I.J.M.. (2013). Evaluating results of the Welfare Quality multi-criteria evaluation model for classification of dairy cattle welfare at the herd level. *Journal of. Dairy Science*, 96(10), 6264–6273. doi:10.3168/jds.2012-6129

Decreto-Lei n. º 64/2000, de 22 de Abril. (2000). *Diário Da República - I Série-A, n.º 95- 22 de Abril de 2000*. Ministério da Agricultura, do Desenvolvimento Rural e das Pescas. Lisboa.

Decreto-Lei n.º 113/2013, de 7 de Agosto. (2013). *Diário Da República - I Série-A, n.º 151- 7 de Agosto de 2013*. Ministério da Agricultura, do Desenvolvimento Rural e das Pescas. Lisboa.

Devendra, C. (2013). Investments on Pro-poor Development Projects on Goats: Ensuring Success for Improved Livelihoods. *Asian-Australasian Journal of Animal Sciences*, 26(1), 1–18. doi:10.5713/ajas.2013.r.01

Directiva 98/58/CE do Conselho, de 20 de Julho de 1998. *Jornal Oficial das Comunidades Europeias nº221/23*. O Conselho da União Europeia. Bruxelas

Dohoo, I.R., Martin, S.W. & Stryhn, H. (2003) *Veterinary Epidemiologic Research*. Charlottetown, Canada: Atlantic Veterinary College.

Dubeuf, J.P., Morand-Fehr, P & Rubino, R. (2004). Situation, changes and future of goat industry around the world. *Small Ruminant Research*, 51(2), 165–173. doi:10.1016/j.smallrumres.2003.08.007

Dubeuf, J.P. (2010). Characteristics and diversity of the dairy goat production systems and industry around the world. *Tecnol. & Ciên. Agropec.*, 4(4), 25-31. Retrieved from http://emepa.org.br/revista/volumes/tca_v4_n4_dez/tca03_characteristic.pdf

Duncan, I.J.H. (2005). Science-based assessment of animal welfare: Farm animals. *Rev. sci. tech. Off. int. Epiz.*, 24 (2), 483-492.

Duncan, I.J.H. (2006). The changing concept of animal sentience. *Applied Animal Behaviour Science*, 100, 11–19. doi:10.1016/j.applanim.2006.04.011

Duncan, I.J. & Petherick, J. C. (1991). The implications of cognitive processes for animal welfare. *Journal of Animal Science*, 69, 5017-5022. Retrieved from http://jas.fass.org/content/69/12/5017

Dwyer, C. (2009). The behaviour of sheep and goats. In P. Jensen (Ed.), *The ethology of domestic animals: An introductory text, modular texts*. (2nd ed.). (pp.161-176). Wallingford, Oxfordshire, The United Kingdom: CABI Publishing. ISBN: 0851996027

Ekesbo, I. (Ed.). (2011). *Farm animal behaviour: Characteristics for assessment of health and welfare*. Oxfordshire, The United Kingdom: CABI Publishing. ISBN: 10: 1845937708

Escareño, L., Salinas-Gonzalez, H., Wurzinger, M.,Iñiguez, L., Sölkner, J. & Meza-Herrera, C. (2012). Dairy goat production systems. *Tropical Animal Health and Production*, 45(1), 17-34. doi: http://dx.doi.org/10.1007/s11250-014-0555-z.

Ehrlenbruch, R., Pollen, T. Andersen, I.L. & Bøe, K.E. (2010). Competition for water at feeding time: The effect of increasing number of individuals per water dispenser. *Applied Animal Behaviour Science*, 126(3-4), 105–108. doi:10.1016/j.applanim.2010.06.004

Ellegaard, L., Cunningham, A., Edwards, S., Grand, N., Nevalainen, T., Prescott, M. & Schuurman, T. (2010). Welfare of the minipig with special reference to use in regulatory toxicology studies. *Journal of Pharmacological and Toxicological Methods*, 62(3), 167–183. doi:10.1016/j.vascn.2010.05.006

European Commission (2012). *Commission staff working paper impact assessment – Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the European Union strategy for the protection and welfare of animals.* Brussels: EC. Retrieved from http://ec.europa.eu/food/animal/welfare/actionplan/docs/impact_assesment_19012012_en.pdf

European Food Safety Authority (2012). Scientific Opinion on the use of animal-based measures to assess welfare of dairy cows. *The EFSA Journal*, 10(1), 2554. Retrieved from http://www.efsa.europa.eu/it/efsajournal/doc/2554.pdf

European Food Safety Authority (2012). Statement on the use of animal-based measures to assess the welfare of animals. *The EFSA Journal*, 10(6), 2767. Retrieved from http://www.efsa.europa.eu/en/search/doc/2767.pdf

Eze, C.A. (2002). Technical note: Lameness and reproductive performance in small ruminants in Nsukka Area of the Enugu State, Nigeria. *Small Ruminant Research*, 44(3), 263–267. doi:10.1016/s0921-4488(02)00030-5

Food and Agriculture Organization of the United Nations. (2015). *Dairy production and products: Small ruminants*. Retrieved Jan. 5, 2015, from http://www.fao.org/agriculture/dairy-gateway/milk-production/dairy-animals/small-ruminants/en/#.VQDOBOHHLQM

Food and Agriculture Organization of the United Nations Statistics Division. (2015). *FAOSTAT*. Retrieved Jan. 5, 2015, from http://faostat3.fao.org/browse/Q/QA/E

Farm Animal Welfare Council (2009). *Farm Animal Welfare Council: Five freedoms*. Retrieved Jan. 22, 2015, from http://webarchive.nationalarchives.gov.uk/20121007104210/http:/www.fawc.org.uk/freedoms.htm

Fioni, L. (2014). *Validazione di indicatori di comfort termico per la valutazione del benessere negli allevamenti caprini da latte*. BSc Thesis. Milan, Italy: University of Milan.

Fleiss, J.L., Levin B., Paik, M.C. (2003). *Statistical methods for rates and proportions*. (3rd ed.) Hoboken: John Wiley & Sons.

Forkman, B. (1996). The social facilitation of drinking: what is facilitated, and who is affected? *Ethology*, 102(2), 252–258. doi:10.1111/j.1439-0310.1996.tb01122.x

Fraser, D. (1995). Science, values and animal welfare: Exploring the 'inextricable connection'. *Animal Welfare*, 4(2), 103-117.

Fraser, D. (2003). Emerging animal welfare standards and their implications for animal hygiene. In *Proceedings of XIth ISAH Congress in Animal Hygiene, 23-27 February 2003, Mexico City*, pp. 61-66.

Fraser, D. (2006). Animal welfare assurance programs in food production: A framework for assessing the options. *Animal Welfare*, 15, 93-104.

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. Retrieved from www.innocua.net/web/download-1287/fraser1997.pdf

Gaias, G. (2012). *Body condition score and body composition of Sarda dairy ewes*. Tesi di Dottorato in Scienze dei Sistemi Agrari e Forestali e delle Produzioni Alimentari Indirizz Scienze e Tecnologie Zootecniche. Sassari, Italy: Università Degli Studi di Sassari. Retrieved from http://eprints.uniss.it/8252/1/Gaias_G_Body_condition_score_body.pdf

Gamer, M., Lemon, J., Fellows, I. & Singh, P. (2012). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84.

Gaughan, J. B. (2003). Using panting scores to assess heat load in cattle. Interactions between climate and animal production. *EAAP Technical Series*, 7, 99–107. Rome, Italy: EAAP.

Gibbons, J., Vasseur, E., Rushen, J. & de Passillé, A. (2012). A training programme to ensure high repeatability of injury scoring of dairy cows. *Animal Welfare,* 21(3), 379-388. doi: 10.7120/09627286.21.3.379

Görgülü, M., Boğa, M., Şahin, A., Serbester, U., Kutlu, H.R., Şahinler, S. (2008) Diet selection and eating behaviour of lactating goats subjected to time restricted feeding in choice and single feeding system. *Small Ruminant Research*, 78(1-3), 41–47. doi:10.1016/j.smallrumres.2008.04.004

Grandin, T. (Ed.) *Improving animal welfare: A practical approach*. Wallingford, Oxfordshire: CABI. doi:10.1079/9781845935412.0032

Haenlein, G. F. (2004). Goat milk in human nutrition. *Small Ruminant Research*, 51(2), 155–163. doi:10.1016/j.smallrumres.2003.08.010

Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1): 23–34.

Harwood, D. (2006). *Goat health and welfare. A veterinary guide*. Wiltshire, The United Kingdom: Crowood Press Ltd:

Hemsworth, P. (2003). Human–animal interactions in livestock production. *Applied Animal Behaviour Science*, 81(3), 185–198. doi:10.1016/s0168-1591(02)00280-0

Hemsworth, P.H., Coleman, G.J., Barnett, J.L. & Borg, S. (2000). Relationships between human-animal interactions and productivity of commercial dairy cows. *Journal of Animal Science*; 78, 2821–2831.

Hervieu, J. & Morand-Fehr, P. (1999). Comment noter l'état corporel des chèvres. *Réussir Chèvre*. 26–32.

Hess, L., Bolos, A., Hoffmann, R. & Surry, Y. (2014). Is animal welfare better on small farms? Evidence from veterinary inspections on Swedish farms. In *Proceedings of EAAE 2014 Congress 'Agri-Food and Rural Innovations for Healthier Societies,* *Ljubljana, Slovenia, August 26-29*. Retrieved Feb. 8, 2015 from http://ageconsearch.umn.edu/handle/182781

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(25), 852–858. doi:10.1136/vr.158.25.852

Hill, N.P., Murphy, P.E., Nelson, A.J., Mouttotou, N., Green, L.E., & Morgan, K.L. (1997). Lameness and foot lesions in adult British dairy goats. *Veterinary Record*, 141(16), 412–416. doi:10.1136/vr.141.16.412

Hoehler, F. K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology*, 53(5), 499–503. doi:10.1016/s0895-4356(99)00174-2

Holmes, B.W. & Moore, Y.F. (1981). Metabolisable energy required by feral goats for maintenance and the effects of cold climatic conditions on their heat production. In *Proceedings of the New Zealand Society of Animal Production*, 41, pp. 163-166

Hörning, B. (2001). The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 51(30), 42–47. doi:10.1080/090647001316923045

Hristov, S., Maksimovic, N., Stankovic, B., Zujovic, M., Pantelic, V., Stanisic, N. & Zlatanovic, Z. (2012). The most significant stressors in intensive sheep production. *Biotechnology in Animal Husbandry*, 28(4), 649–658. doi:10.2298/bah1204649h

Hughes, J. (2001). A system for assessing cow cleanliness. *In Practice*, 23(9), 517–524. doi: 10.1136/inpract.23.9.517

Ingenbleek, P.T.M., Immink. V.M., Spoolder, H.A.M., Bokma, M.H. & Keeling, L.J. (2012). EU animal welfare policy: Developing a comprehensive policy framework. *Food Policy*, 37(6), 690–699. doi:10.1016/j.foodpol.2012.07.001

Institut National de la Recherche Agronomique (1988). *Alimentation des bovins, ovins, caprins*. Paris: INRA.

Instituto Nacional de Estatística (2011). *Recenseamento Agrícola 2009: Análise dos principais resultados*. Lisboa: INE.

Instituto Nacional de Estatística (2015). *Portal Nacional de Estatística: Base de dados*. Retrieved Jan. 5, 2015, from http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0000963&contexto=bd&selTab=tab2

Jackson, J.E. (1991). *A user's guide to principal components*. New York: Wiley.

Jackson, K.M.A. & Hackett, D. (2007). A note: The effects of human handling on heart girth, behaviour and milk quality in dairy goats. *Applied Animal Behaviour Science*, 108(3-4), 332–336. doi:10.1016/j.applanim.2007.01.011

Jago, J., Krohn, C., & Matthews, L. (1999). The influence of feeding and handling on the development of the human–animal interactions in young cattle. *Applied Animal Behaviour Science*, 62(2-3), 137–151. doi:10.1016/s0168-1591(98)00219-6

Johnsen, P. F., Johannesson, T., & Sandøe, P. (2001). Assessment of farm animal welfare at herd level: many goals, many methods. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 51(30), 26-33. doi: http://dx.doi.org/10.1080/090647001316923027

Jones, A.C. & Gosling, S.D. (2005). Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. *Applied Animal Behaviour Science*, 95(1-2), 1–53. doi:10.1016/j.applanim.2005.04.008

Jørgensen, G.H.M., Andersen, I.L. & Bøe, K.E. (2007). Feed intake and social interactions in dairy goats - The effects of feeding space and type of roughage. *Applied Animal Behaviour Science*, 107(3-4):239-251. doi: 10.1016/j.applanim.2006.10.007

KilBride, A.L., Gillman, C.E. & Green, L.E. (2009a). A cross-sectional study of the prevalence of lameness in finishing pigs, gilts and pregnant sows and associations with limb lesions and floor types on commercial farms in England. *BMC Vet Res,* 5(1), 30. doi:10.1186/1746-6148-5-30

KilBride, A.L., Gillman, C.E., Ossent, P. & Green, L.E. (2009b). Impact of flooring on the health and welfare of pigs. *In Practice*, 31(8), 390–395. doi:10.1136/inpract.31.8.390

Kirchner, M.K. (2012). *Implementation of an animal-based welfare assessment system in beef bull farms*. Doctoral Thesis. Vienna: Division of Livestock Science, Department for Sustainable Agricultural Systems, University of Natural Resources and Life Sciences.

Kirchner, M.K., Schulze Westerath, H., Knierim, U., Tessitore, E., Cozzi, G. & Winckler, C. (2013). On-farm animal welfare assessment in beef bulls: consistency over time of single measures and aggregated Welfare Quality® scores. *Animal*, 8(03), 461–469. doi:10.1017/s1751731113002267

Klaas, I.C., Enevoldsen, C., Vaarst, M. & Houe, H. (2004). Systematic clinical examinations for identification of latent udder health types in Danish dairy herds. *Journal of Dairy Science,* 87(5), 1217–1228. doi:10.3168/jds.s0022-0302(04)73272-5

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.

Koyuncu, M. & Altınçekiç, S.O. (2013). Importance of body condition score in dairy goats. *Macedonian Journal of Animal Science*, 3(2), 167–173.

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(9), 3721–3728. doi:10.3168/jds.s0022-0302(06)72413-4

Landis, J.R. & Koch, G.G: (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33(1), 159-174. doi:10.2307/2529310

Laporte-Broux, B., Duvaux-Ponter, C., Roussel, S., Promp, J., Chavatte-Palmer, P. & Ponter, A.A. (2011). Restricted feeding of goats during the last third of gestation modifies both metabolic parameters and behavior. *Livestock Science*, 138(1-3), 74–88. doi:10.1016/j.livsci.2010.12.008

Lassen, J., Sandøe, P. & Forkman, B. (2006). Happy pigs are dirty!: conflicting perspectives on animal welfare. *Livestock Science*, 103(3), 221–230. doi: 10.1016/j.livsci.2006.05.008

Le Jaouen, J.C. & Toussaint, G. (1993). Le lait de chèvre en Europe. *Le Lait*, 73(5-6), 407–415. doi:10.1051/lait:19935-638

Le Neindre, P. (2009). Some perspectives of the assessment of farm animal welfare in the European Union. *Agrociencia*, 8(3), 28-34.

Leitner, G., Silanikove, N. & Merin, U. (2008). Estimate of milk and curd yield loss of sheep and goats with intrammamary infection and its relation to somatic cell count. *Small Ruminant Research*, 74(1-3), 221–225. doi:10.1016/j.smallrumres.2007.02.009

Lengarite, M.I., Mbugua, P.N., Gachuiri, C.K. & Kabuage, L.W. (2012). Mineral status of sheep and goats grazing in the arid rangelands of northern Kenya. *Pakistan Journal of Nutrition*, 11(4), 383-390. doi:10.3923/pjn.2012.383.390

Lensink, J, Boissy, A. & Veissier, I. (2000). The relationship between farmers' attitude and behaviour towards calves, and productivity of veal units. *Annales de Zootechnie*, 49(4), 313-327. doi:10.1051/animres:2000122

Lensink, B.J., van Reenen, C.G., Engel, B., Rodenburg, T. Veissier, I. (2003). Repeatability and reliability of an approach test to determine calves' responsiveness to humans: "a brief report". *Applied Animal Behaviour Science*, 83(4), 325–330. doi:10.1016/s0168-1591(03)00138-2

Lima, M.S., Pascoal, R.A., & Stilwell, G.T. (2012a). Glycaemia as a sign of the viability of the foetuses in the last days of gestation in dairy goats with pregnancy toxaemia. *Irish Veterinary Journal*, 65(1), 1. doi:10.1186/2046-0481-65-1

Lima, M.S., Pascoal, R.A., Stilwell, G.T., Hjerpe, C.A. (2012b) Clinical findings, blood chemistry values, and epidemiologic data from dairy goats with pregnancy toxemia. *Bovine Practice*, 46, 102–110,

Lu, C.D. (1989). Effects of heat stress on goat production. *Small Ruminant Research*, 2(2), 151–162. doi:10.1016/0921-4488(89)90040-0

Lundmark, F., Berg, C., Schmid, O., Behdadi, D. & Röcklinsberg, H. (2014). Intentions and values in animal welfare legislation and standards. *J Agric Environ Ethics*, 27(6), 991–1017. doi:10.1007/s10806-014-9512-0

Maciel, C.T. & Bock, B. (2013). Modern politics in animal welfare: The changing character of governance of animal welfare and the role of private standards. *Int. Jrnl. of Soc. of Agr. & Food*, 20(2), 219–235. Retrieved from http://edepot.wur.nl/266159

Main, D.C.J. (2009). Application of welfare assessment to commercial livestock production. *Journal of Applied Animal Welfare Science*, 12(2), 97-104. doi: 10.1080/10888700902719658

Main, D.C.J., Webster, A.J.F. & Green, L.E. (2001). Animal Welfare Assessment in Farm Assurance Schemes. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 51(030), 108–113. doi:10.1080/090647001316923171

Main, D.C.J., Whay, H.R., Green, L.E., & Webster, A.J.F. (2003). Effect of the RSPCA Freedom Food scheme on the welfare of dairy cattle. *Veterinary Record*, 153(8), 227–231. doi:10.1136/vr.153.8.227

Main, D. C. J., Mullan, S., Atkinson, C., Cooper, M., Wrathall, J. H. M. & Blokhuis, H. J. (2014). Best practice framework for animal welfare certification schemes. *Trends in Food Science & Technology*, 37(2), 127–136. doi:10.1016/j.tifs.2014.03.009

Martin, P. & Bateson, P. (2007). *Measuring behaviour: An introductory guide*. Cambridge; New York: Cambridge University Press.

Martínez, A.M., Vega-Pla, J.L., León, J.M., Camacho, M.E., Delgado, J.V. & Ribeiro, M.N. (2010). Is the Murciano-Granadina a single goat breed? A molecular genetics approach. *Arq. Bras. Med. Vet. Zootec.*, 62(5), 1191–1198. doi:10.1590/s0102-09352010000500023

Mason, G. & Mendl, M. (1993). Why is there no simple way of measuring animal welfare? *Animal Welfare*, 2, 301-319.

Mattiello, S., Battini, M., Andreoli, E., Minero, M., Barbieri, S. & Canali, E. (2010). Avoidance distance test in goats: A comparison with its application in cows. *Small Ruminant Research*, 91(2-3), 215–218. doi:10.1016/j.smallrumres.2010.03.002

Mattiello, S., Battini, M., Barbieri, S., Vieira, A., Stilwell, G., Canali, E. & Zanella, A. (2015). Un protocollo per la valutazione del benessere nella capra: I risultati del progetto AWIN. *Large Animal Review*, 1(4), 90 – 93. Retrieved from http://www.sipaoc.it/documenti/2014/XXI%20Congresso%20SIPAOC%202014%20-%20Atti%20congressuali.pdf

Mazurek, M., Marie, M. & Desor, D. (2007). Potential animal-centred indicators of dairy goat welfare. *Animal Welfare*, 16, 161–164

Mazurek, M., McGee, M., Crowe, M.A., Prendiville, D.J., Boivin, X., & Earley, B. (2011). Consistency and stability of behavioural fear responses of heifers to different fear-eliciting situations involving humans. *Applied Animal Behaviour Science*; 131(1-2), 21–28. doi:10.1016/j.applanim.2011.01.004

McGregor, B. A. (2002). *Extent and source of short and cotted mohair*: *A report for the Rural Industries Research and Development Corporation*. Barton, Australia: RIRDC. Retrieved from https://rirdc.infoservices.com.au/downloads/02-108

McGregor, B.A. & Butler, K.L. (2008). Relationship of body condition score, live weight, stocking rate and grazing system to the mortality of Angora goats from hypothermia and their use in the assessment of welfare risks. *Australian Veterinary Journal*, 86(1-2), 12-17. doi:10.1111/j.1751-0813.2007.00249.x

Mellor, D.J. & Webster, J.R. (2014). Development of animal welfare understanding drives change in animal welfare standards. *Rev. sci. tech. Off. int. Epiz.*, 33(1), 121-130.

Merck Veterinary Manual (2014). *Perinatal management of goats: Health-management interaction: goats: Merck Veterinary Manual*. Retrieved Feb. 8, 2015 from http://www.merckmanuals.com/vet/management_and_nutrition/health-management_interaction_goats/perinatal_management_of_goats.html

Merck Veterinary Manual (2014). *Nutritional requirements of goats: Nutrition: Goats: Merck Veterinary Manual*. Retrieved Feb. 12, 2015 from http://www.merckmanuals.com/vet/management_and_nutrition/nutrition_goats/nutritional_requirements_of_goats.html?qt=poor%20hair%20goats&alt=sh

Miele, M., Veissier, I., Evans, A. & Botreau, R. (2011). Animal welfare: Establishing a dialogue between science and society. *Animal Welfare*, 20, 103-117.

Millman, S.T., Duncan, I.J.H., Stauffacher, M. & Stookey, J.M. (2004). The impact of applied ethologists and the international society for applied ethology in improving animal welfare. *Applied Animal Behaviour Science*, 86 (3/4), 299-311. doi: http://dx.doi.org/10.1016/j.applanim.2004.02.008

Minero, M., Tosi, M. A., Canali, E. & Wemelsfelder, F. (2009). Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science*, 116, 74–81. doi: http://dx.doi.org/10.1016/j.applanim.2008.07.001

Miranda-de la Lama, G.C. & Mattiello, S. (2010). The importance of social behaviour for goat welfare in livestock farming. *Small Ruminant Research*, 90(1-3), 1–10. doi:10.1016/j.smallrumres.2010.01.006

Meagher, R.K. (2009). Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science,* 119(1-2), 1–14. doi:10.1016/j.applanim.2009.02.026

Mellor, D.J. & Webster, J.R. (2014). Development of animal welfare understanding drives change in minimum welfare standards. *Rev sci tech Off int Epiz.* 33(1), 121-130. Retrieved from in http://www.oie.int/doc/ged/D13659.PDF

Miranda-de la Lama, G.C. & Mattiello, S. (2010). The importance of social behaviour for goat welfare in livestock farming. *Small Ruminant Research*, 90(1-3), 1–10. doi:10.1016/j.smallrumres.2010.01.006

Mohammed, A., Badau, U.A. & Kene, R.O.C. (1996). Lameness in sheep and goats in relation to hoof condition in sahel zone of Nigeria. *Bull. Anim. Hlth. Prod. Afr.*, 44: 97–100.

Morand-Fehr, P. (2005). Recent developments in goat nutrition and application: A review. *Small Ruminant Research*, 60(1-2), 25–43. doi:10.1016/j.smallrumres.2005.06.004

Morand-Fehr, P., Boutonnet, J.P., Devendra, C., Dubeuf, J.P., Haenlein, G.F.W., Holst, P., Mowlemg, L. & Capote, J. (2004). Strategy for goat farming in the 21st century. *Small Ruminant Research*, 51(2), 175–183. doi:10.1016/j.smallrumres.2003.08.013

Mullan, S., Edwards, S.A., Butterworth, A, Whay, H.R., & Main, D.C.J. (2011). Inter-observer reliability testing of pig welfare outcome measures proposed for inclusion within farm assurance schemes. *The Veterinary Journal*, 190(2), 100–109. doi:10.1016/j.tvjl.2011.01.012

Muri, K., Stubsjøen, S.M. & Valle, P.S. (2013). Development and testing of an on-farm welfare assessment protocol for dairy goats. *Animal Welfare*, 22(3), 385-400. doi: 10.7120/09627286.22.3.385

Murray, C.M., Eberly, L.E., & Pusey, A.E. (2006). Foraging strategies as a function of season and rank among wild female chimpanzees (Pan troglodytes). *Behavioral Ecology*, 17(6), 1020–1028. doi:10.1093/beheco/arl042

Napolitano, F., De Rosa, G., Girolami, A., Scavone, M. & Braghieri, A. (2011). Avoidance distance in sheep: Test–retest reliability and relationship with stockmen attitude. *Small Ruminant Research*, 99(2-3), 81–86. doi:10.1016/j.smallrumres.2011.03.044

Napolitano, F., De Rosa, G., Serrapica, M. & Braghieri, A. (2015). A continuous recording approach to qualitative behaviour assessment in dairy buffaloes (Bubalus bubalis). *Applied Animal Behaviour Science*. doi:10.1016/j.applanim.2015.01.017

Nonga, H.E. Makungu, M., Bittegeko, S.B.P. &. Mpanduji, D.G (2009). Occurrences and management of lameness in goats: A case study of Magadu farm, Morogoro, Tanzania. *Small Ruminant Research*, 82(2-3), 149–151. doi:10.1016/j.smallrumres.2009.02.001

O'Callaghan, K A., Cripps, P.J., Downham, D.Y. & Murray, R.D. (2003). Subjective and objective assessment of pain and discomfort due to lameness in dairy cattle. *Animal Welfare*, 12, 605-610.

Olofsson, J. (1999). Competition for total mixed diets fed for ad libitum intake using one or four cows per feeding station. *Journal of Dairy Science*, 82(1), 69–79. doi:10.3168/jds.s0022-0302(99)75210-0

Paterna, A., Contreras, A., Gómez-Martín, A., Amores, J., Tatay-Dualde, J., Prats-van der Ham, M. & De la Fe, C. (2014). The diagnosis of mastitis and contagious agalactia in dairy goats. *Small Ruminant Research*, 121(1), 36–41. doi:10.1016/j.smallrumres.2013.12.002

Phythian, C.J., Cripps, P.J., Michalopoulou, E., Jones, P.H., Grove-White, D., Clarkson, M.J., Winter, A.C., Stubbings, L.A & Duncan, J.S. (2012). Reliability of indicators of sheep welfare assessed by a group observation method. *The Veterinary Journal*, 193(1), 257–263. doi:10.1016/j.tvjl.2011.12.006

Phythian, C.J.,Michalopoulou, E., Duncan J. & Wemelsfelder, F. (2013a). Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science*, 144, 73– 79. doi: http://dx.doi.org/10.1016/j.applanim.2012.11.011

Phythian, C.J., Toft, N., Cripps, P.J, Michalopoulou, E., Winter, A.C., Jones, P.H., Grove-White, D. & Duncan, J.S. (2013b). Inter-observer agreement, diagnostic sensitivity and specificity of animal-based indicators of young lamb welfare. *Animal*, 7(7), 1182–1190. doi:10.1017/s1751731113000487

Pirisi, A., Lauret, A. & Dubeuf, J.P. (2007). Basic and incentive payments for goat and sheep milk in relation to quality. *Small Ruminant Research*, 68(1-2), 167–178. doi:10.1016/j.smallrumres.2006.09.009

Plesch, G., Broerkens, N., Laister, S., Winckler, C. & Knierim, U. (2010). Reliability and feasibility of selected measures concerning resting behaviour for the on-farm welfare assessment in dairy cows. *Applied Animal Behaviour Science*, 126(1-2), 19–26. doi:10.1016/j.applanim.2010.05.003

Plummer, P.J. & Schleining, J.A. (2013). Assessment and management of pain in small ruminants and camelids. *Veterinary Clinics of North America: Food Animal Practice*, 29(1), 185–208. doi:10.1016/j.cvfa.2012.11.004

Popescu, A. (2013). Study regarding the trends in the world and European goat milk production. *Lucrări Ştiinţifice-Seria Zootehnie*, 59, 127-132.

Reneau, J.K., Seykora, A.J., Heins, B.J., Endres, M.I., Farnsworth, R.J. & Bey, R.F. (2005). Association between hygiene scores and somatic cell scores in dairy cattle. *Journal of the American Veterinary Medical Association*, 227(8), 1297–1301. doi:10.2460/javma.2005.227.1297

Rook, A.J. & Penning, P.D. (1991). Synchronisation of eating, ruminating and idling activity by grazing sheep. *Applied Animal Behaviour Science*, 32(2-3), 157–166. doi:10.1016/s0168-1591(05)80039-5

Rossi R. & Scharrer E. (1992). Circadian patterns of drinking and eating in pygmy goats. *Physiology and Behaviour*; 51(5), 895-897. doi:10.1016/0031-9384(92)90067-c

Rousing, T., Bonde, M. & Sørensen, J.T. (2001). Aggregating welfare indicators into an operational welfare assessment system: A bottom-up approach. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 51(30), 53–57. doi:10.1080/090647001300004790

Rousing, T. & Waiblinger, S. (2004). Evaluation of on-farm methods for testing the human–animal relationship in dairy herds with cubicle loose housing systems—test–retest and inter-observer reliability and consistency to familiarity of test person. *Applied Animal Behaviour Science*, 85(3-4), 215–231. doi:10.1016/j.applanim.2003.09.014

Rushen, J. (2003). Changing concepts of farm animal welfare: bridging the gap between applied and basic research. *Applied Animal Behaviour Science*, 81(3), 199–214. doi:10.1016/s0168-1591(02)00281-2

Rushen, J., Butterworth, A. & Swanson, J. C. (2011). Animal behavior and well-being symposium: Farm animal welfare assurance: Science and application. *Journal of Animal Science*, 89, 1219-1228. doi: 10.2527/jas.2010-3589

Russel, A.J.F., Doney, J.M. & Gunn, R.G. (1969). Subjective assessment of body fat in live sheep. *The Journal of Agricultural Science,* 72(03), 451. doi:10.1017/s0021859600024874

Rutherford, K.M.D., Donald, R.D., Lawrence, A.B. & Wemelsfelder, F. (2012). Qualitative Behavioural Assessment of emotionality in pigs. *Applied Animal Behaviour Science*, 139(3-4), 218–224. doi:10.1016/j.applanim.2012.04.004

Sahlu, T. & Goetsch, A.L. (2005). A foresight on goat research. *Small Ruminant Research,* 60(1-2), 7–12. doi:10.1016/j.smallrumres.2005.06.002

Sandøe, P., Munksgaard, L., Bådsgård, N.P. & Jensen, K.H. (1997). How to manage the management factor – assessing animal welfare at the farm level. In J.T. Sørensen (Ed.), *Proceedings of the fourth International Symposium on the Livestock Farming Systems: Livestock Farming Systems - more than food production, Foulum, Denmark, 22-23 August 1996*, EAAP Publication, 89, pp. 221-230. Wageningen, The Netherlands: Wageningen Pers.

Santucci, P., Branca, A., Napoleone, M., Bouche R., Aumont, G., Poisot F. & Alexandre, G. (1991). Body condition scoring of goats in extensive conditions. In Morand-Fehr (Ed.), *Goat Nutrition*. (pp.240–255). Wageningen, The Netherlands: Wageningen Academic Publishers.

Sarkar, M., Rahman, S., Sarker, B., Anisuzzaman, A., Begum, N., & Mondal, M. (2010). Epidemiology and pathology of ectoparasitic infestations in black Bengal goats in Gaibandha and Mymensingh districts of Bangladesh. *Bangladesh Journal of Veterinary Medicine*, 8(1): 41 – 50. doi:10.3329/bjvm.v8i1.7718

Schreiner, D.A. & Ruegg, P.L. (2003). Relationship between udder and leg hygiene scores and subclinical mastitis. *Journal of Dairy Science*, 86(11), 3460–3465. doi:10.3168/jds.s0022-0302(03)73950-2

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 Science,* 51(30), 5–10. doi:10.1080/090647001316922983

Scott, E.M., Fitzpatrick, J.L., Nolan, A.M., Reid, J. & Wiseman, M L. (2003). Evaluation of welfare state based on interpretation of multiple indices. *Animal Welfare,* 12(4) 467-468.

Scott, K., Chennells, D.J., Armstrong, D., Taylor, L., Gill, B.P. & Edwards, S.A. (2007). The welfare of finishing pigs under different housing and feeding systems: Liquid versus fry feeding in fully-slatted and straw-based housing. *Animal Welfare*, 16, 53–62.

Scott, I. & Mazhindu, D. (2014). *Statistics for health care professionals: An introduction*. ($2^{nd}$ ed.). London: SAGE Publications.

Sejian, V. & Srivastava, R.S. (2010). Pineal–adrenal–immune system relationship under thermal stress: effect on physiological, endocrine, and non-specific immune response in goats. *Journal of Physiology and Biochemistry*, 66(4), 339-349. doi:10.1007/s13105-010-0040-8

Sevi, A., Annicchiarico, G., Albenzio, M., Taibi, L., Muscio, A. & Dell'Aquila, S. (2001). Effects of solar radiation and feeding time on behavior, immune response and production of lactating ewes under high ambient temperature. *Journal of Dairy Science*, 84(3), 629–640. doi:10.3168/jds.s0022-0302(01)74518-3

Sevi, A., Casamassima, D., Pulina, G. & Pazzona, A. (2009). Factors of welfare reduction in dairy sheep and goats. *Italian Journal of Animal Science*, 8(1s), 81. doi:10.4081/ijas.2009.s1.81

Short, R.E., Grings, E.E., MacNeill, M.D., Heitschimidt, R.K. & Haferkamp, M.R. (1996). Effects of supplement, and sire breed of calf during fall grazing period on cow and calf performance. *Journal of Animal Science*, 74, 1701-1710.

Shinde, A.K., Verma, D.L. & Singh, N.P. (2004). Social dominance-subordinate relationships on a flock of Marwari goats. *The Indian Journal of Animal Sciences*; 74: 216–219.

Shrout, P.E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, 7(3), 301–317. doi:10.1191/096228098672090967

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

Silanikove, N. (1992). Effects of water scarcity and hot environment on appetite and digestion in ruminants: A review. *Livestock Science*, 30(3), 175-193. doi: 10.1016/S0301-6226(06)80009-6

Silanikove, N. (2000). Effects of heat stress on the welfare of extensively managed domestic ruminants. *Livestock Production Science*, 67 (1-2), 1–18. doi:10.1016/s0301-6226(00)00162-7

Silva, J. Santos e; Fitas da Cruz, Vasco; Barbosa, José Carlos (2007). A strategy to prevent damages in Portuguese small ruminants due to hot climate conditions. In *Proceedings of the 6th International Seminar FAO-CIHEAM Network on Sheep and Goats*. *Ponte de Lima*, Portugal: DRAPN / FAO-CIHEAM. Retrieved from http://hdl.handle.net/10198/3252

Sim, J. & Wright, C.C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257-268. Retrieved from http://ptjournal.apta.org/content/85/3/257.full

Sinn, R. & Rudenberg, P. (2008). *Raising goats for milk and meat*. Little Rock, Arkansas, The United States of America: Heifer International.

Smith, B.P. & Magdesian, K.G. (Eds.) (2008). Alterations in alimentary and hepatic function, In B.P. Smith. *Large animal internal medicine*. (4th edition). (pp.96-116) St. Louis, Missouri: Mosby Elsevier

Smith, M.C. & Sherman, D.M. (2009). *Goat Medicine.* (2nd ed.). Ames, Iowa, The United States of America: Wiley-Blackwell. doi:10.1002/9780813818825

Smulders, D., Verbeke, G. & Geers, R. (2006). Validation of a behavioral observation tool to assess pig welfare. *Physiology & Behavior*, 89(3), 438–447. doi:10.1016/j.physbeh.2006.07.002

Sørensen, D.B. (2010). Never wrestle with a pig ... *Laboratory Animals*, 44(2), 159–161. doi:10.1258/la.2010.009100

Sørensen, J.T, Sandøe, P. & Halberg, N. (2001). Animal welfare as one among several values to be considered at farm level: The idea of an ethical account for livestock farming. *Acta Agriculturae Scandinavica, Section A — Animal Science,* 51(30), 11–16. doi:10.1080/090647001316922992

Sørensen, J.T., Rousing, T., Møller, 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., Hörning, B., Waiblinger, S. & Wemelsfelder, F. (2003). Integrating parameters to assess on-farm welfare. *Animal Welfare*, 12, 529-534.

Stafford, K. & Gregory, N. (2008). Implications of intensification of pastoral animal production on animal welfare. *New Zealand Veterinary Journal*, 56(6), 274–280. doi:10.1080/00480169.2008.36847

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

Sundrum, A., Andersson, R. & Postler, G. (1994). *Tiergerechtheitsindex – 200, 1994: ein leitfaden zur beurteilung von haltungssystemen. [Animal needs index 200/1994 – A manual for the assessment of housing systems]*. Bonn: Germany: Köllen Druck und Verlag GmbH.

Temple, D. (2012). *Animal welfare assessment on intensive and extensive pig farms*. Doctoral Thesis. Barcelona: Departament de Ciència Animal i del Aliments, Facultat de Veterinària de Barcelona, Universitat Autònoma de Barcelona -.

Temple, D., Manteca, X., Dalmau, A., & Velarde, A. (2013). Assessment of test–retest reliability of animal-based measures on growing pig farms. *Livestock Science*, 151(1), 35–45. doi:10.1016/j.livsci.2012.10.012

Tiberio, M.L. & Diniz, F. (2014). Sheep and goat production in Portugal: A dynamic view. *Modern Economy*, 05(06), 703–722. doi:10.4236/me.2014.56066

Toussaint, G. (1997). The housing of milk goats. *Livestock Production Science*; 49(2), 151–164. doi:10.1016/s0301-6226(97)00011-0

Tregenza, T. (1994). Common misconceptions in applying the ideal free distribution. *Animal Behaviour*, 47(2), 485–487. doi:10.1006/anbe.1994.1068

Van Horne, P.L.M., & Achterbosch, T.J. (2008). Animal welfare in poultry production systems: Impact of EU standards on world trade. *World's Poultry Science Journal*, 64(01). doi:10.1017/s0043933907001705

Vapnek, J. & Chapman, M. (2010). *Legislative and regulatory options for animal welfare.* Rome: Food and Agriculture Organization of the United Nations. Retrieved from http://www.fao.org/docrep/013/i1907e/i1907e00.pdf

Veissier, I., Beaumont, C, Lévy, F. (2007). Les recherches sur le bien-être animal: buts, méthodologie et finalité. *INRA Prod. Anim.*, 20 (1), 3-10.

Veissier, I, Butterworth, A., Bock, B. & Roe, E. (2008). European approaches to ensure good animal welfare. *Applied Animal Behaviour Science*, 113(4), 279–297. doi: http://dx.doi.org/10.1016/j.applanim.2008.01.008

Veissier, I., Aubert, A. & Boissy, A. (2012). Animal welfare: A result of animal background and perception of its environment. *Animal Frontiers*, 2(3), 7–15. doi:10.2527/af.2012-0043

Veit, H.P., Mccarthy, F., Friedericks, J., Cashin, M., Angert, R. (1993). A survey of goat, and cattle diseases in the Artibonite Valley, Haiti, West Indies. *Rev. Elev. Med. Vet. Pay*, 46, 27–38.

Viñuela-Fernández, I., Jones, E., Welsh, E.M. &. Fleetwood-Walker, S.M. (2007). Pain mechanisms and their implication for the management of pain in farm and companion animals. *The Veterinary Journal*, 174(2), 227–239. doi:10.1016/j.tvjl.2007.02.002

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*, 30, 73–77. doi: http://dx.doi.org/10.1080/090647001316923108

Waiblinger, S., Menke, C. & Coleman, G. (2002). The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Applied Animal Behaviour Science*, 79(3), 195–219. doi:10.1016/s0168-1591(02)00155-7

Waiblinger, S., Menke, C., & Fölsch, D. (2003). Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. *Applied Animal Behaviour Science*, 84(1), 23–39. doi:10.1016/s0168-1591(03)00148-5

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K., Jones, R.B. (2006). Assessing the human–animal relationship in farmed species: A critical review. *Applied Animal Behaviour Science*, 101(3-4), 185–242. doi:10.1016/j.applanim.2006.02.001

Weary, D.M., Niel, L., Flower, F.C., & Fraser, D. (2006). Identifying and preventing pain in animals. *Applied Animal Behaviour Science*, 100(1-2), 64–76. doi:10.1016/j.applanim.2006.04.013

Weary, D.M., Huzzey, J.M., & von Keyserlingk, M.A.G. (2008). Board-invited review: Using behavior to predict and identify ill health in animals. *Journal of Animal Science*, 87(2), 770–777. doi:10.2527/jas.2008-1297

Webster, A.J.F. (2001). Farm animal welfare: the five freedoms and the free market. *The Veterinary Journal*, 161, 229–237. doi: http://dx.doi.org/10.1053/tvjl.2000.0563

Webster, J. (2005). The assessment and implementation of animal welfare: Theory into practice. *Rev sci tech Off int Epiz.*, 24(2), 723-734.

Webster, A.J.F. (2009). The virtuous bicycle: A delivery vehicle for improved farm animal welfare. *Animal Welfare*, 18, 141–147.

Webster, A.J.F., Main, D.C.J., Whay, H.R., 2004. Welfare assessment: indices from clinical observation. *Animal Welfare*, 13, 93–98.

Weir, J.P. (2005). Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *The Journal of Strength and Conditioning Research*, 19(1), 231. doi:10.1519/15184.1

Welfare Quality®. (2009). *Welfare Quality® assessment protocol for cattle*. Lelystad, Netherlands: Welfare Quality® Consortium.

Wemelsfelder, F. (1997). The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science*, 53(1-2), 75–88. doi:10.1016/s0168-1591(96)01152-5

Wemelsfelder, F. (2007). How animals communicate quality of life: the qualitative assessment of behavior. *Animal Welfare*, 16(S), 25-31.

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*, 51(30), 21–25. doi:10.1080/090647001300004763

Wemelsfelder, F., Millard, F., de Rosa, G. & Napolitano, F. (2009). Qualitative behaviour assessment. In B. Forkman, & L. Keeling (Eds.) *Welfare Quality® Report No. 11 — Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*, pp.215–224. Cardiff, The United Kingdom: Cardiff University.

Wemelsfelder, F. & Mullan, S. (2014). Applying ethological and health indicators to practical animal welfare assessment. *Rev. sci. tech. Off. int. Epiz.*, 33 (1), 111-120.

Whay, H.R. (2007). The journey to animal welfare improvement. *Animal Welfare*, 16(2), 117-122.

Whay, H.R., Waterman, A.E. & Webster, A.J.F. (1997). Associations between locomotion, claw lesions and nociceptive threshold in dairy heifers during the peri-partum period. *The Veterinary ]ourna1*, 154(2), 155–161. doi:10.1016/s1090-0233(97)80053-6

Whay, H.R., Main, D.C.J., Green, L.E. & Webster, A.J.F. (2003a). Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records. *Veterinary Record*, 153(7), 197–202. doi:10.1136/vr.153.7.197

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(4), 611-617.

Winckler, C.; Capdeville, J.; Gebresenbet, G.; Hörning, B.; Roiha, U.; Tosi, M. & Waiblinger, S. (2003). Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. *Animal Welfare,* 12(4), 619-624.

Winckler, C. (2006). On-farm welfare assessment in cattle: From basic concepts to feasible assessment systems. In *Proc. XXIV World Buiatrics Congress, Nice, France,* pp.493–500. Wilmington, The United States of America: Gloyd Group. Retrieved from http://www.ivis.org/proceedings/wbc/wbc2006/Winckler.pdf?LA=1

Winter, A.C. (2008). Lameness in sheep. *Small Ruminant Research*, 76(1-2), 149–153. doi:10.1016/j.smallrumres.2007.12.008

Winter, A.C. (2011). Treatment and control of hoof disorders in sheep and goats. *Veterinary Clinics of North America: Food Animal Practice*, 27(1), 187–192. doi:10.1016/j.cvfa.2010.10.018

Yangilar, F. (2013). As a potentially functional food: Goats' milk and products. *Journal of Food and Nutrition Research*, 1(4), 68-81. doi: 10.12691/jfnr-1-4-6

Yeates, J.W. & Main, D.C.J. (2008). Assessment of positive welfare: A review. *The Veterinary Journal*, 175(3), 293–300. doi:10.1016/j.tvjl.2007.05.009

Zupan, M., Bojkovski, D., Štuhec, I., Kompan, D. (2010). Foraging behaviour of sheep at pasture with different types of vegetation in a paddock. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 96(2), 103–109.

## Annex 1 – Tables

Table 13 - Goat livestock ($x10^3$) by Geographic localization (Agrarian area) and Category (goat livestock); data reference period: 2013 (INE, 2015).

| Geographic localization (Agrarian area) | Goat livestock | | | | |
|---|---|---|---|---|---|
| | Total ($x10^3$) | Goats and kids which have been mated ($x10^3$) | Goats ($x10^3$) | Kids which have been mated ($x10^3$) | Other goats ($x10^3$) |
| **Portugal** | **398** | **334** | **313** | **21** | **64** |
| **Entre Douro e Minho** | 49 | 42 | 40 | 2 | 7 |
| **Trás-os-Montes** | 51 | 45 | 43 | 2 | 6 |
| **Beira Litoral** | 48 | 41 | 40 | 2 | 7 |
| **Beira Interior** | 68 | 59 | 57 | 2 | 9 |
| **Ribatejo e Oeste** | 51 | 41 | 37 | 4 | 10 |
| **Alentejo** | 103 | 83 | 78 | 5 | 20 |
| **Algarve** | 15 | 12 | 11 | 1 | 3 |
| **Açores** | 7 | 6 | 5 | 1 | 1 |
| **Madeira** | 5 | 5 | 3 | 1 | 1 |

| Animal-based indicator | | Side | | N | Goats (%) | Farms (%) |
|---|---|---|---|---|---|---|
| **Individual Assessment** | | | | | | |
| **Cleanliness** | Hindquarters | Left | | 1172 | 229 (19.5) | 21 (70) |
| **Cleanliness** | Hindquarters | Right | | 1172 | 227 (19.4) | 20 (66.7) |
| **Cleanliness** | Lower legs | Left | | 1172 | 210 (17.9) | 21 (70) |
| **Cleanliness** | Lower legs | Right | | 1172 | 207 (17.7) | 20 (66.7) |
| **Abscesses** | Hindquarters | Left | | 1172 | 7 (0.6) | 5 (16.7) |
| **Abscesses** | Hindquarters | Right | | 1172 | 8 (0.7) | 7 (23.3) |
| **Abscesses** | Body | Left | | 1172 | 126 (10.8) | 17 (56.7) |
| **Abscesses** | Body | Right | | 1172 | 134 (11.4) | 19 (63.3) |
| **Abscesses** | Neck | Left | | 1172 | 48 (4.1) | 16 (53.3) |
| **Abscesses** | Neck | Right | | 1172 | 58 (4.9) | 15 (50) |
| **Claw overgrowth** | | Left | | 1172 | 402 (34.3) | 25 (83.3) |
| **Claw overgrowth** | | Right | | 1172 | 410 (35) | 27 (90) |
| **Lesions and swellings** | Hindquarters | Left | | 1172 | 124 (10.6) | 16 (53.3) |
| **Lesions and swellings** | Hindquarters | Right | | 1172 | 123 (10.5) | 17 (56.7) |
| **Lesions and swellings** | Lower legs | Left | | 1172 | 112 (9.6) | 13 (43.3) |
| **Lesions and swellings** | Lower legs | Right | | 1172 | 118 (10.1) | 14 (46.7) |
| **Lesions and swellings** | Body | Left | | 1172 | 183 (15.6) | 22 (73.3) |
| **Lesions and swellings** | Body | Right | | 1172 | 169 (14.4) | 24 (80) |
| **Lesions and swellings** | Neck | Left | | 1172 | 169 (14.4) | 20 (66.7) |
| **Lesions and swellings** | Neck | Right | | 1172 | 161 (13.7) | 20 (66.7) |
| **Lesions and swellings** | Knee (score >0) | Left | Score 1 | 1172 | 968 (82.6) | 30 (100) |
| **Lesions and swellings** | Knee (score >0) | Left | Score 2 | 1172 | 135 (11.5) | 16 (53.3) |
| **Lesions and swellings** | Knee (score >0) | Right | Score 1 | 1172 | 958 (81.7) | 30 (100) |
| **Lesions and swellings** | Knee (score >0) | Right | Score 2 | 1172 | 139 (11.9) | 18 (60) |

Table 14 – Prevalence results of cleanliness, abscesses, claw overgrowth and presence of lesions and swellings, scored on the right and left side of the animals, from the 30 intensive Portuguese farms where the AWIN prototype was tested.

Table 15 – Overall prevalences of the most prevalent indicators included in the AWIN prototype for dairy goats: results of the 30 intensive Portuguese farms.

<table>
<tr><th colspan="3" rowspan="2">Animal-based indicator</th><th rowspan="2">N</th><th rowspan="2">Goats (%)</th><th rowspan="2">Farms (%)</th><th colspan="3">Variation in sample prevalence (% of goats) across the farms</th></tr>
<tr><th>Median</th><th>IQR</th><th>Max</th></tr>
<tr><td colspan="9">Pen-level observations</td></tr>
<tr><td colspan="3">Improper disbudding*</td><td>1778</td><td>256 (14.4)</td><td>19 (63.3)</td><td>8</td><td>2.1 –16.9</td><td>83.8</td></tr>
<tr><td rowspan="2">Queuing</td><td colspan="2">At feeding</td><td rowspan="2">2715</td><td>721 (26.6)</td><td>25 (83.3)</td><td>16.6</td><td>6.8-39.1</td><td>75</td></tr>
<tr><td colspan="2">At drinking</td><td>130 (4.8)</td><td>11 (36.7)</td><td>0</td><td>0-5.4</td><td>28.1</td></tr>
<tr><td colspan="3">Hair coat condition</td><td>2715</td><td>508 (18.7)</td><td>30 (100)</td><td>16.5</td><td>9.9-32.2</td><td>70</td></tr>
<tr><td colspan="9">Individual assessment</td></tr>
<tr><td rowspan="2">BCS</td><td colspan="2">Very thin</td><td rowspan="2">1172</td><td>60 (5.1)</td><td>17 (56.7)</td><td>3.5</td><td>0-7.1</td><td>26.9</td></tr>
<tr><td colspan="2">Very Fat</td><td>220 (18.8)</td><td>25 (83.3)</td><td>15.5</td><td>7.5-28.4</td><td>53.3</td></tr>
<tr><td rowspan="2">Cleanliness</td><td colspan="2">Hindquarters</td><td rowspan="2">1172</td><td>229 (19.5)</td><td>21 (70)</td><td>5.5</td><td>0-31.5</td><td>75</td></tr>
<tr><td colspan="2">Lower legs</td><td>210 (17.9)</td><td>19 (63.3)</td><td>3.1</td><td>0-12.6</td><td>64.6</td></tr>
<tr><td rowspan="7">Lesions and swellings</td><td colspan="2">Hindquarters</td><td rowspan="7">1172</td><td>124 (10.6)</td><td>18 (60)</td><td>2.6</td><td>0-22.3</td><td>48</td></tr>
<tr><td colspan="2">Lower legs</td><td>112 (9.6)</td><td>15 (50)</td><td>0</td><td>0-16.3</td><td>47.4</td></tr>
<tr><td rowspan="2">Knee</td><td>Score 1</td><td>968 (82.6)</td><td>30 (100)</td><td>93.7</td><td>73.5-98.1</td><td>100</td></tr>
<tr><td>Score 2</td><td>135 (11.5)</td><td>17 (56.7)</td><td>2.4</td><td>0-13.1</td><td>57.8</td></tr>
<tr><td colspan="2">Head</td><td>299 (25.5)</td><td>29 (96.7)</td><td>13.3</td><td>6.5-46</td><td>83.3</td></tr>
<tr><td colspan="2">Body</td><td>183 (15.6)</td><td>24 (80)</td><td>9</td><td>0-26.8</td><td>50</td></tr>
<tr><td colspan="2">Neck</td><td>169 (14.4)</td><td>22 (73.3)</td><td>8.5</td><td>0-19.7</td><td>66.7</td></tr>
<tr><td colspan="3">Udder asymmetry</td><td>1172</td><td>68 (5.8)</td><td>24 (80)</td><td>5.1</td><td>2.7-8.3</td><td>20</td></tr>
<tr><td colspan="3">Claw overgrowth</td><td>1172</td><td>402 (34.3)</td><td>27 (90.0)</td><td>27.5</td><td>4.6-53.4</td><td>89.5</td></tr>
<tr><td rowspan="2">Abscesses</td><td colspan="2">Head</td><td rowspan="2">1172</td><td>91 (7.8)</td><td>22 (73.3)</td><td>4.3</td><td>0-10.2</td><td>37</td></tr>
<tr><td colspan="2">Body</td><td>126 (10.8)</td><td>19 (63.3)</td><td>2.8</td><td>0-14.2</td><td>50</td></tr>
<tr><td rowspan="2">Discharge</td><td colspan="2">Nasal</td><td rowspan="2">1172</td><td>62 (5.3)</td><td>15 (50)</td><td>1.1</td><td>0-5.8</td><td>47.9</td></tr>
<tr><td colspan="2">Ocular</td><td>67 (5.7)</td><td>16 (53.3)</td><td>2.3</td><td>0-6.3</td><td>27.3</td></tr>
</table>

*Improper disbudding could be recorded on only 23 farms (1778 goats)

Table 16 – Overall prevalences of the less prevalent indicators of AWIN prototype for dairy goats, tested in 30 intensive Portuguese farms.

| Animal-based indicator | | N | Goats (%) | Farms (%) | Variation in sample prevalence (% of goats) across the farms | | |
|---|---|---|---|---|---|---|---|
| | | | | | Median | IQR | Max |
| **Pen-level observations** | | | | | | | |
| **Oblivious** | | 2715 | 13 (0.5) | 8 (26.7) | 0 | 0-0.5 | 4.3 |
| **Panting (score > 0)** | Score 1 | 2715 | 69 (2.5) | 5 (16,7) | 0 | 0 | 100 |
| | Score 2 | | 0 (0) | 0 (0) | 0 | 0 | 0 |
| **Shivering (score >0)** | Score 1 | 2715 | 74 (2.7) | 4 (13.3) | 0 | 0 | 91.8 |
| | Score 2 | | 0 (0) | 0 (0) | | 0 | |
| **Avoidance distance** | Contact | 2715 | 27 (0.1) | 12 (40) | 0 | 0-1.4 | 3.9 |
| | Acceptance | | 36 (1.3) | 15 (50) | 0.4 | 0-2.3 | 4.8 |
| **Severe lameness** | | 2715 | 48 (1.8) | 66.7 (20) | 1.4 | 0-2.6 | 16 |
| **Kneeling** | Feeding rack | 2715 | 0 (0) | 0 (0) | 0 | 0 | 2.9 |
| | Pen | | 13 (0.5) | 8 (26.7) | 0 | 0-0.4 | 3.6 |
| **Individual assessment** | | | | | | | |
| **Cleanliness** | Udder | 1172 | 37 (3.2) | 9 (30) | 0 | 0-2.8 | 23.7 |
| **Abscesses** | Neck | 1172 | 48 (4.1) | 20(66.7) | 2.6 | 0-5.4 | 24.1 |
| | Udder | | 29 (2.5) | 14(46.7) | 0 | 0-3 | 12.8 |
| | Hindquarters | | 7 (0.6) | 9 (30) | 0 | 0 | 5.7 |
| **Discharge** | Vulvar | 1172 | 3 (0.3) | 3 (10) | 0 | 0 | 2.6 |
| **Diarrhoea** | | 1172 | 31 (2.6) | 6 (20) | 0 | 0 | 57.9 |

Table 17 – Latency to first contact test's results from the 30 Portuguese intensive dairy goat farms assessed during AWIN prototype testing, organized by farm category.

| Animal-based indicator | N | Mean | Std. Deviation | Median | Minimum | Maximum | Percentiles | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 25 | 50 | 75 |
| **Latency to first contact (s)** | 2715 | 125.9 | 129.2 | 49.5 | 0 | 300 | 17.5 | 49.5 | 300 |
| **Small Sized Farms** | | | | | | | | | |
| **Latency to first contact (s)** | 473 | 139.7 | 139.2 | 61.5 | 10 | 300 | 19.8 | 61.5 | 300 |
| **Medium Sized Farms** | | | | | | | | | |
| **Latency to first contact (s)** | 1122 | 156.9 | 131.4 | 122.5 | 7 | 300 | 38.5 | 122.5 | 300 |
| **Large Sized Farms** | | | | | | | | | |
| **Latency to first contact (s)** | 1120 | 81.1 | 117 | 21 | 0 | 300 | 5.3 | 21 | 188.8 |

| Animal-based indicator | | | N | Goats (%) | Farm Size Category: Small | | | Medium | | | Large | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| **Pen-level observations** | | | | | | | | | | | | | |
| **Improper disbudding*** | | | 1778 | 256 (14.4) | 453 | 6 (60) | 41 (9.1) | 557 | 6 (60) | 130 (23.3) | 768 | 7 (70) | 85 (11.1) |
| **Queuing** | At feeding | | 2715 | 721 (26.6) | 473 | 8 (80) | 131 (27.7) | 1122 | 8 (80) | 145 (12.9) | 1120 | 9 (90) | 445 (39.7) |
| | At drinking | | | 130 (4.8) | | 0 (0) | 0 (0) | | 7 (70) | 67 (6.0) | | 4 (40) | 63 (5.6) |
| **Poor hair coat condition** | | | 2715 | 508 (18.7) | 473 | 10 (100) | 84 (17.8) | 1122 | 10 (100) | 197 (17.6) | 1120 | 10 (100) | 227 (20.3) |
| **Individual Assessment** | | | | | | | | | | | | | |
| **BCS** | Very thin | | 1172 | 60 (5.1) | 305 | 5 (50) | 16 (5.3) | 407 | 7 (70) | 19 (4.7) | 460 | 5 (50) | 25 (5.4) |
| | Very Fat | | | 220 (18.8) | | 6 (60) | 40 (13.1) | | 9 (90) | 67 (16.5) | | 10 (100) | 113 (24.6) |
| **Cleanliness** | Hindquarters | | 1172 | 229 (19.5) | 305 | 8 (80) | 53 (17.4) | 407 | 5 (50) | 38 (9.3) | 460 | 8 (80) | 138 (30.0) |
| | Lower legs | | | 210 (17.9) | | 6 (60) | 50 (16.4) | | 7 (70) | 67 (16.5) | | 8 (80) | 93 (20.2) |
| **Lesions and swellings** | Hindquarters | | 1172 | 124 (10.6) | 305 | 4 (40) | 16 (5.2) | 407 | 7 (70) | 46 (11.3) | 460 | 7 (70) | 62 (13.5) |
| | Lower legs | | | 112 (9.6) | | 2 (20) | 2 (0.7) | | 6 (60) | 35 (8.6) | | 7 (70) | 75 (16.3) |
| | Knee | Score 1 | | 968 (82.6) | | 10 (100) | 286 (93.8) | | 10 (100) | 321 (78.9) | | 10 (100) | 361 (78.5) |
| | | Score 2 | | 135 (11.5) | | 4 (40) | 15 (4.9) | | 6 (60) | 44 (10.8) | | 7 (70) | 76 (16.5) |

Table 18 – Most prevalent indicators of the 30 Portuguese intensive dairy goat farms visited for prototype testing: each indicator's prevalence is organized according to farm category.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Lesions and swellings** | Head | 1172 | 299 (25.5) | 305 | 9 (90) | 34 (11.1) | 407 | 10 (100) | 125 (30.7) | 460 | 10 (100) | 140 (30.4) |
| | Body | | 183 (15.6) | | 7 (70) | 42 (13.8) | | 7 (70) | 71 (17.4) | | 10 (100) | 70 (15.2) |
| | Neck | | 169 (14.4) | | 5 (50) | 14 (4.6) | | 8 (80) | 76 (18.7) | | 9 (90) | 79 (17.2) |
| **Udder asymmetry** | | 1172 | 68 (5.8) | 305 | 7 (70) | 15 (4.9) | 407 | 9 (90) | 23 (5.7) | 460 | 8 (80) | 30 (6.5) |
| **Claw overgrowth** | | 1172 | 402 (34.3) | 305 | 7 (70) | 36 (11.8) | 407 | 10 (100) | 155 (38.1) | 460 | 10 (100) | 211 (41.9) |
| **Abscesses** | Head | 1172 | 91 (7.8) | 305 | 5 (50) | 18 (5.9) | 407 | 7 (70) | 14 (3.4) | 460 | 10 (100) | 59 (12.8) |
| | Body | | 126 (10.8) | | 5 (50) | 23 (7.5) | | 7 (70) | 38 (9.3) | | 8 (80) | 65 (14.1) |
| **Discharge** | Nasal | 1172 | 62 (5.3) | 305 | 4 (40) | 7 (2.3) | 407 | 5 (50) | 33 (8.1) | 460 | 6 (60) | 22 (4.8) |
| | Ocular | | 67 (5.7) | | 3 (30) | 4 (1.3) | | 6 (60) | 24 (5.9) | | 7 (70) | 39 (8.5) |

[*]Improper disbudding could be recorded on only 23 farms (1778 goats)

Table 18 – Most prevalent indicators of the 30 Portuguese intensive dairy goat farms visited for prototype testing: each indicator's prevalence is organized according to farm category. (continuation)

| Animal-based indicator | | N | Goats (%) | Farm Size Category: Small N | Small Farms (%) | Small Goats (%) | Medium N | Medium Farms (%) | Medium Goats (%) | Large N | Large Farms (%) | Large Goats (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Pen-level observations** | | | | | | | | | | | | |
| **Oblivious** | | 2715 | 13 (0.5) | 473 | 1 (10) | 2 (0.4) | 1122 | 4 (40) | 6 (0.5) | 1120 | 3 (30) | 5 (0.4) |
| **Panting (score > 0)** | Score 1 | 2715 | 69 (2.5) | 473 | 3 (30) | 66 (14.0) | 1122 | 1 (10) | 1 (0.1) | 1120 | 1 (10) | 2 (0.2) |
| | Score 2 | | 0 (0) | | 0 (0) | 0 (0) | | 0 (0) | 0 (0) | | 0 (0) | 0 (0) |
| **Shivering (score > 0)** | Score 1 | 2715 | 74 (2.7) | 473 | 2 (20) | 50 (10.6) | 1122 | 2 (20) | 24 (2.1) | 1120 | 0 (0) | 0 (0) |
| | Score 2 | | 0 (0) | | 0 (0) | 0 (0) | | 0 (0) | 0 (0) | | 0 (0) | 0 (0) |
| **Avoidance distance** | Contact | 2715 | 27 (0.1) | 473 | 2 (20) | 2 (0.4) | 1122 | 3 (30) | 4 (0.4) | 1120 | 7(70) | 21 (1.9) |
| | Acceptance | | 36 (1.3) | | 4 (40) | 5 (1.0) | | 3 (30) | 7 (0.6) | | 8 (80) | 24 (2.1) |
| **Severe lameness** | | 2715 | 48 (1.8) | 473 | 3 (30) | 3 (0.6) | 1122 | 8 (80) | 15 (1.3) | 1120 | 9 (90) | 30 (2.7) |
| **Kneeling** | Feeding rack | 2715 | 0 (0) | 473 | 0 (0) | 0 (0) | 1122 | 0 (0) | 0 (0) | 1120 | 0 (0) | 0 (0) |
| | Pen | | 13 (0.5) | | 1 (10) | 1 (0.2) | | 5 (50) | 5 (0.5) | | 2 (20) | 7 (0.6) |
| **Individual Assessment** | | | | | | | | | | | | |
| **Cleanliness** | Udder | 1172 | 37 (3.2) | 305 | 3 (30) | 5 (1.6) | 407 | 3 (30) | 12 (2.9) | 460 | 3 (30) | 20 (4.3) |

Table 19 - Less prevalent indicators of the 30 Portuguese intensive dairy goat farms visited: each indicator's prevalence is organized regarding the three farm categories considered.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Abscesses** | Udder | 1172 | 29 (2.5) | 305 | 3 (30) | 4 (1.3) | 407 | 5 (50) | 12 (2.9) | 460 | 6 (60) | 13 (2.8) |
| | Neck | | 48 (4.1) | | 5 (50) | 9 (3.0) | | 5 (50) | 10 (2.5) | | 10 (100) | 29 (6.3) |
| | Hindquarters | | 7 (0.6) | | 1 (10) | 2 (0.7) | | 4 (40) | 1 (0.2) | | 4 (40) | 4 (0.9) |
| **Discharge** | Vulvar | 1172 | 3 (0.3) | 305 | 0 (0) | 0 (0) | 407 | 2 (20) | 2 (0.5) | 460 | 1 (10) | 1 (0.2) |
| **Diarrhoea** | | 1172 | 31 (2.6) | 305 | 1 (10) | 1 (0.3) | 407 | 3 (30) | 24 (5.9) | 460 | 2 (20) | 6 (1.3) |

Table 19 – Less prevalent indicators of the 30 Portuguese intensive dairy goat farms visited: each indicator's prevalence is organized regarding the three farm categories considered. (continuation)

Table 20 – QBA's inter-observer reliability analysis: data obtained from the assessments performed simultaneously by two assessors in 10 Portuguese farms, while applying the AWIN prototype for dairy goats (significant correlations between observers at the 0.05-level are presented in bold and gray).

| Dimension/descriptor | r | $r_s$ | P |
|---|---|---|---|
| PC1 | **0.67** | | 0.036 |
| PC2 | 0.42 | | 0.222 |
| Aggressive | | **0.79** | 0.006 |
| Agitated | | **0.68** | 0.031 |
| Alert | | **0.81** | 0.005 |
| Bored | | 0.20 | 0.58 |
| Content | | 0.26 | 0.48 |
| Curious | | -0.115 | 0.75 |
| Fearful | | **0.67** | 0.035 |
| Frustrated | | **0.60** | 0.007 |
| Irritated | | 0.24 | 0.50 |
| Lively | | **0.66** | 0.038 |
| Relaxed | | 0.32 | 0.36 |
| Sociable | | -0.26 | 0.47 |
| Suffering | | 0.09 | 0.80 |

| Animal-based indicator | | Visit 1 | | Visit 2 | | Change (Visit 2 - Visit 1) |
|---|---|---|---|---|---|---|
| | | N | Goats (%) | N | Goats (%) | %Δ |
| **Pen-level observations** | | | | | | |
| **Improper disbudding** | | 772 | 19.2 | 1102 | 19.9 | 0.7 |
| **Queuing** | At feeding | 1131 | 25.1 | 1529 | 28.3 | 3.2 |
| | At drinking | | 3.2 | | 4.4 | 1.2 |
| **Hair coat condition** | | 1082 | 12.6 | 1483 | 10.1 | -2.5 |
| **Oblivious** | | 1131 | 0.5 | 1529 | 0.1 | -0.4 |
| **Panting** | Score 1 | | 0.6 | | 2.9 | 2.3 |
| | Score 2 | | 0 | | 0 | 0 |
| **Shivering** | Score 1 | | 4.0 | | 0 | -4.0 |
| | Score 2 | | 0 | | 0 | 0 |
| **Avoidance distance** | Contact | 1032 | 1.4 | 1411 | 2.4 | 1.0 |
| | Acceptance | | 2.1 | | 3.1 | 1.0 |

Table 21 – Prevalence of the animal-based indicators, included in the AWIN prototype for dairy goats, assessed during the two visits to 10 Portuguese farms.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Severe lameness** | | 1131 | 1.2 | 1529 | 1.1 | -0.1 |
| **Kneeling** | Feeding rack. | | 0 | | 0.9 | 0.9 |
| | Pen | | 0.8 | | 0.3 | -0.5 |
| **Individual Assessment** | | | | | | |
| **BCS** | Very thin | 404 | 5.2 | 494 | 5.9 | 0.7 |
| | Very fat | | 20 | | 29.8 | 9.8 |
| **Diarrhoea** | | | 6.4 | | 0 | -6.4 |
| **Udder asymmetry** | | | 5.0 | | 7.5 | 2.5 |
| **Cleanliness** | HQ | | 33.9 | | 25.5 | -8.4 |
| | LL | | 27.5 | | 26.1 | -1.4 |
| | Udder | | 1.5 | | 0.2 | -1.3 |

Table 21– Prevalence of the animal-based indicators, included in the AWIN prototype for dairy goats, assessed during the two visits to 10 Portuguese farms (continuation).

| | | | | | | |
|---|---|---|---|---|---|---|
| **Abscesses** | HQ | 404 | 0.7 | 494 | 0.8 | 0.1 |
| | Body | | 9.9 | | 14.4 | 4.5 |
| | Udder | | 3.0 | | 1.0 | -2.0 |
| | Neck | | 4.0 | | 4.7 | 0.7 |
| | Head | | 7.4 | | 3.6 | -3.8 |
| **LAS** | HQ | | 2.5 | | 6.9 | 4.4 |
| | LL | | 5.9 | | 9.9 | 4.0 |
| | Knee (score 2) | | 10.6 | | 20.5 | 9.9 |
| | Body | | 8.4 | | 10.3 | 1.9 |
| | Neck | | 8.2 | | 4.9 | -3.3 |
| | Head | | 28.0 | | 11.5 | -16.5 |
| **Discharge** | Ocular | | 8.2 | | 1.2 | -7.0 |
| | Nasal | | 2.7 | | 4.1 | 1.3 |
| | Vulvar | | 0.3 | | 0 | -0.3 |
| **Claw overgrowth** | | | 45.8 | | 32.4 | -13.4 |

Table 21 – Prevalence of the animal-based indicators, included in the AWIN prototype for dairy goats, assessed during the two visits to 10 Portuguese farms (continuation).

Table 22 - Prevalence of 'Queuing at feeding' and 'Queuing at drinking' across seasons, assessed during two consecutive visits to 10 Portuguese farms, while testing the AWIN prototype for dairy goats.

| | Animal-based indicator | | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Small | | | Medium | | | Large | | |
| | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | Queuing | At feeding | 1131 | 284 (25.1) | 158 | 3 (100) | 42 (26.6) | 467 | 3 (100) | 39 (8.4) | 506 | 4 (100) | 203 (40.1) |
| | | At drinking | | 36 (3.2) | 158 | 0 (0) | 0 (0) | 467 | 7 (70) | 7 (1.5) | 506 | 4 (40) | 29 (5.7) |
| *Summer* | Queuing | At feeding | 1529 | 433 (28.3) | 178 | 3 (100) | 62 (34.8) | 665 | 3 (100) | 116 (17.4) | 686 | 4 (100) | 255 (37.2) |
| | | At drinking | | 67 (4.4) | 178 | 0 (0) | 0 (0) | 665 | 2 (50) | 6 (0.9) | 686 | 3 (100) | 61 (8.9) |

Table 23 - Prevalence of 'Hair coat condition' in winter and summer, recorded during two consecutive visits to 10 Portuguese farms where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Small | | | Medium | | | Large | | |
| | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | **Hair coat condition** | 1082 | 136 (12.6) | 109 | 2(66.7) | 33 (30.3) | 467 | 3 (100) | 35 (7.5) | 506 | 4 (100) | 68 (13.4) |
| *Summer* | **Hair coat condition** | 1483 | 150 (10.1) | 132 | 2 (66,7) | 12 (9.1) | 665 | 4 (100) | 48 (7.2) | 686 | 3 (100) | 90 (13.1) |

Table 24 - Prevalence results of 'Improper disbudding' obtained from two visits (in winter and summer) to 10 Portuguese farms, where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Small | | | Medium | | | Large | | |
| | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| ***Winter*** | **Improper disbudding** | 772 | 148 (19.2) | 99 | 2 (66.7) | 4 (4) | 167 | 2 (66.7) | 74 (44.3) | 506 | 4 (100) | 70(13.8) |
| ***Summer*** | **Improper disbudding** | 1102 | 220 (19.9) | 118 | 2 (66.7) | 8 (1.7) | 418 | 2 (66.7) | 132 (31.6) | 566 | 4 (100) | 80 (14.1) |

Table 25 - Prevalence results of ‘BCS’ obtained from two visits (in winter and summer) to 10 Portuguese farms, where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | | N | Goats (%) | Farm Size Category: Small | | | Medium | | | Large | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | **BCS** | Very thin | 404 | 21 (5.2) | 96 | 3 (100) | 8 (8.3) | 132 | 3 (100) | 8 (6.1) | 176 | 3 (75) | 5 (2.8) |
| | | Very Fat | | 81 (20) | | 2 (66.7) | 13 (13.5) | | 2(66.7) | 28 (21.2) | | 4(100) | 40 (22.7) |
| *Summer* | **BCS** | Very thin | 494 | 29 (5.9) | 117 | 2 (66.7) | 5 (4.3) | 168 | 2 (66.7) | 17 (10.1) | 209 | 3 (75) | 7 (3.4) |
| | | Very Fat | | 147 (29.8) | | 3 (100) | 48 (41.0) | | 3 (100) | 20 (11.9) | | 4 (100) | 79 (37.8) |

Table 26 - Prevalence of 'Cleanliness' in winter and summer, assessed during two consecutive visits to 10 Portuguese farms where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | | N | Goats (%) | Farm Size Category: Small N | Small Farms (%) | Small Goats (%) | Medium N | Medium Farms (%) | Medium Goats (%) | Large N | Large Farms (%) | Large Goats (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Winter* | **Cleanliness** | Hindquarters | 404 | 137 (33.9) | 96 | 2 (66.7) | 25 (26) | 132 | 2(66.7 ) | 21 (15.9) | 176 | 3 (75) | 91 (51.7) |
| | | Lower legs | | 111 (27.5) | | 2 (66.7) | 32 (33.3) | | 2 (66.7) | 6 (4.5) | | 3 (75) | 73 (41.5) |
| *Summer* | **Cleanliness** | Hindquarters | 494 | 126 (25.5) | 117 | 1 (33.3) | 3 (2.6) | 168 | 3 (100) | 43 (25.6) | 209 | 4 (100) | 80 (38.3) |
| | | Lower legs | | 129 (26.1) | | 1 (33.3) | 2 (1.7) | | 3 (100) | 51 (30.4) | | 4 (100) | 76 (36.4) |

| | Animal-based indicator | | | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Small | | | Medium | | | Large | | |
| | | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | Lesions and swellings | Hindquarters | | 404 | 10 (2.5) | 96 | 4 (40) | 2 (2.1) | 132 | 7 (70) | 1 (0.8) | 176 | 7 (70) | 7 (4) |
| | | Lower legs | | | 24 (5.9) | | 2 (20) | 0 (0) | | 6 (60) | 0 (0) | | 7 (70) | 24 (13.6) |
| | | Knee | Score 1 | | 335 (82.9) | | 3 (100) | 80 (83.3) | | 3 (100) | 116 (87.9) | | 4 (100) | 139 (79) |
| | | | Score 2 | | 43 (10.6) | | 2 (66.7) | 15 (15.6) | | 3 (100) | 5 (3.8) | | 3 (75) | 23 (13.1) |
| | | Head | | | 113 (28) | | 3 (100) | 13 (13.5) | | 3 (100) | 32 (24.2) | | 4 (100) | 68 (38.6) |
| | | Body | | | 34 (8.4) | | 2(66.7) | 10 (10.4) | | 2 (66.7) | 13 (9.8) | | 3(75) | 11 (6.3) |
| | | Neck | | | 33 (8.2) | | 0 (0) | 0(0) | | 3 (100) | 7 (5.3) | | 3 (75) | 26 (14.8) |

Table 27 - Prevalence of 'LAS' in the regions considered in AWIN prototype for dairy goats: results from two visits, in winter and summer, to 10 Portuguese farms.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Hindquarters | | | 34 (6.9) | | 0 (0) | 0 (0) | | 2 (66.7) | 18 (10.7) | | 4 (100) | 16 (7.7) |
| | | Lower legs | | | 49 (9.9) | | 0 (0) | 0 (0) | | 3 (100) | 37 (22) | | 2 (50) | 12 (5.7) |
| | | Knee | Score 1 | | 367 (74.3) | | 3 (100) | 67 (57.3) | | 3 (100) | 158 (94) | | 4 (100) | 142 (67.9) |
| ***Summer*** | **Lesions and swellings** | | Score 2 | 494 | 101 (20.5) | 117 | 3 (100) | 30 (25.6) | 168 | 2 (66.7) | 9 (5.4) | 209 | 4 (100) | 62 (29.7) |
| | | Head | | | 57 (11.5) | | 2 (66.7) | 3 (2.6) | | 3 (100) | 16 (9.5) | | 3 (75) | 38 (18.2) |
| | | Body | | | 51 (10.3) | | 1 (33.3) | 2 (1.7) | | 3 (100) | 27 (16.1) | | 3 (75) | 22 (10.5) |
| | | Neck | | | 24 (4.9) | | 1 (33.3) | 2 (1.7) | | 2 (66.7) | 8 (4.8) | | 3 (75) | 14 (6.7) |

Table 27 - Prevalence of 'LAS' in the regions considered in AWIN prototype for dairy goats: results from two visits, in winter and summer, to 10 Portuguese farms (continuation).

Table 28 - Prevalence of ‘Udder asymmetry’ in winter and summer: assessed during two consecutive visits to 10 Portuguese farms where AWIN prototype for dairy goats was applied.

| | Animal-based indicator | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Small | | | Medium | | | Large | | |
| | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | **Udder asymmetry** | 404 | 20 (5) | 96 | 3 (100) | 6 (6.3) | 132 | 2 (66.7) | 5 (3.8) | 176 | 3 (75) | 9 (5.1) |
| *Summer* | **Udder asymmetry** | 494 | 37 (7.5) | 117 | 2 (66.7) | 8 (6.8) | 168 | 3 (100) | 10 (6) | 209 | 4 (100) | 19 (9.1) |

Table 29 - Prevalence of ‘Claw overgrowth’ in winter and summer, assessed during two consecutive visits to 10 Portuguese farms where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Small | | | Medium | | | Large | | |
| | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | **Claw overgrowth** | 404 | 185 (45.8) | 96 | 2 (66.7) | 11 (11.5) | 132 | 3(100) | 65 (49.2) | 176 | 4 (100) | 109 (61.9) |
| *Summer* | **Claw overgrowth** | 494 | 160 (32.4) | 117 | 0 (0) | 0 (0) | 168 | 3 (100) | 62 (36.9) | 209 | 4 (100) | 98 (46.9) |

| | Animal-based indicator | | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Small | | | Medium | | | Large | | |
| | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| ***Winter*** | **Abscesses** | Head | 404 | 30 (7.4) | 96 | 2 (66.7) | 5 (5.2) | 132 | 2 (66.7) | 7 (5.3) | 176 | 4 (100) | 18 (10.2) |
| | | Neck | | 16 (4) | | 1 (33.3) | 4 (4.2) | | 1 (33.3) | 4 (3) | | 4(100) | 8 (4.5) |
| | | Body | | 40 (9.9) | | 3 (100) | 20 (20.8) | | 2 (66.7) | 11 (8.3) | | 3 (75) | 9 (5.1) |
| | | Hindquarters | | 3 (0.7) | | 0 (0) | 0 (0) | | 1 (33.3) | 1 (0.8) | | 1 (25) | 2 (1.1) |
| | | Udder | | 12 (3) | | 1 (33.3) | 2 (2.1) | | 2 (66.7) | 6 (4.5) | | 3 (75) | 4 (2.3) |

Table 30 - Prevalence of 'Abscesses' in the regions considered in AWIN prototype for dairy goats: results from two visits, in winter and summer, to 10 Portuguese farms.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Summer* | **Abscesses** | Head | 494 | 18 (3.6) | 117 | 2 (66.7) | 5 (4.3) | 168 | 2 (66.7) | 4 (2.4) | 209 | 3 (75) | 9 (4.3) |
| | | Neck | | 23 (4.7) | | 2 (66.7) | 6 (5.1) | | 2 (66.7) | 3 (1.8) | | 3 (75) | 14 (6.7) |
| | | Body | | 71 (14.4) | | 3 (100) | 38 (32.5) | | 1 (33.3) | 23 (13.7) | | 2 (50) | 10 (4.8) |
| | | Hindquarters | | 4 (0.8) | | 1 (33.3) | 1 (0.9) | | 1 (33.3) | 1 (0.6) | | 1 (25) | 2 (1) |
| | | Udder | | 5 (1) | | 0 (0) | 0 (0) | | 1 (33.3) | 1 (0.6) | | 2 (50) | 4 (1.9) |

Table 30 - Prevalence of ‘Abscesses’ in the regions considered in AWIN prototype for dairy goats: results from two visits, in winter and summer, to 10 Portuguese farms (continuation).

Table 31 - Prevalence results of 'Discharge' from two consecutive visits to 10 Portuguese farms, where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Small | | | Medium | | | Large | | |
| | | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| ***Winter*** | **Discharge** | Nasal | 404 | 11 (2.7) | 96 | 2 (66.7) | 4 (4.2) | 132 | 1 (33.3) | 3 (2.3) | 176 | 2 (50) | 4 (2.3) |
| | | Ocular | | 33 (8.2) | | 2 (66.7) | 3 (3.1) | | 2 (66.7) | 16 (12.1) | | 3 (75) | 14 (8) |
| ***Summer*** | **Discharge** | Nasal | 494 | 20 (4.1) | 117 | 1 (33.3) | 2 (1.7) | 168 | 1 (33.3) | 2 (1.2) | 209 | 3 (75) | 16 (7.7) |
| | | Ocular | | 6 (1.2) | | 0 (0) | 0 (0) | | 0 (0) | 0 (0) | | 3 (75) | 6 (2.9) |

Table 32 - Prevalence results of ‘Severe lameness’ from two consecutive visits to 10 Portuguese farms, where AWIN prototype for dairy goats was tested.

| | Animal-based indicator | N | Goats (%) | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Small | | | Medium | | | Large | | |
| | | | | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) | N | Farms (%) | Goats (%) |
| *Winter* | **Severe lameness** | 1131 | 14 (1.2) | 158 | 0 (0) | 0 (0) | 467 | 2 (66.7) | 3 (0.6) | 506 | 4 (100) | 11 (2.1) |
| *Summer* | **Severe lameness** | 1529 | 16 (1.1) | 178 | 0 (0) | 0 (0) | 665 | 3 (100) | 8 (1,2) | 686 | 3 (100) | 8 (1.2) |

Table 33 - Average time in minutes to perform the different stages of the AWIN on-farm welfare assessment prototype for dairy goats in small farms.

| Small size farms | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Stages of the prototype** | | **N** | **Mean** | **Std. Deviation** | **Median** | **Min** | **Max** |
| **Pen Observation (outside pen)*** | | 9 | **3** | 3.5 | 2 | 0 | 10 |
| **Pen Observation (inside pen)** | | 10 | **9** | 0.32 | 1 | 0 | 1 |
| **Queuing**** | At feeding | 8 | 15 | 0 | 15 | 15 | 15 |
| | At drinking | | 15 | 0 | 15 | 15 | 15 |
| | Overall | | **17.1** | 2.7 | 16 | 15 | 22 |
| **QBA** | | 10 | 11 | 3.2 | 10 | 10 | 20 |
| **HAR***** | Latency | 10 | 2.3 | 2.3 | 1 | 0.2 | 1 |
| | AD test | | 2.7 | 1.4 | 2.5 | 1 | 5 |
| | Overall | | **5.1** | 1.9 | 5.3 | 1.2 | 7 |
| **Clinical scoring**** | | 10 | **32.4** | 29 | 20.1 | 10 | 75 |
| **Overall** | | 8 | **71.4** | 29 | 61.2 | 43 | 131 |

*Pen observation (outside pen) time could only be recorded in 9 farms
**Queuing (at feeding rack and at drinking) and clinical scoring times could only be recorded in 8 farms
*** HAR tests time could only be recorded in 10 farms

Table 34 - Average time in minutes to perform the different stages of the AWIN on-farm welfare assessment prototype for dairy goats in medium farms.

| Medium size farms | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Stages of the prototype** | | **N** | **Mean** | **Std. Deviation** | **Median** | **Min** | **Max** |
| **Pen Observation (outside pen)*** | | 7 | **4.7** | 3.5 | 2 | 2 | 9 |
| **Pen Observation (inside pen)** | | 9 | **1.8** | 0.7 | 2 | 1 | 3 |
| **Queuing**** | At feeding | 9 | 15 | 0 | 15 | 15 | 15 |
| | At drinking | | 15 | 0 | 15 | 15 | 15 |
| | Overall | | **21.6** | 11.7 | 15 | 15 | 43 |
| **QBA** | | 9 | 11.2 | 3.3 | 10 | 10 | 20 |
| **HAR***** | Latency | 7 | 2.4 | 2.4 | 0.8 | 0.1 | 5 |
| | AD test | | 4.3 | 1.8 | 4 | 2 | 7 |
| | Overall | | **6.7** | 2.7 | 7.4 | 2.8 | 11 |
| **Clinical scoring**** | | 8 | **59.8** | 46.1 | 40 | 30 | 168 |
| **Overall** | | 6 | **87.9** | 24.9 | 76.7 | 60.8 | 120.1 |

*Pen observation (outside pen) time could only be recorded in 7 farms
**Queuing (at feeding rack and at drinking) and clinical scoring times could only be recorded in 9 farms
*** HAR tests time could only be recorded in 7 farms

Table 35 - Average time in minutes to perform the different stages of the AWIN on-farm welfare assessment prototype for dairy goats in large farms.

| Large size farms | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Stages of the prototype** | | **N** | **Mean** | **Std. Deviation** | **Median** | **Min** | **Max** |
| **Pen Observation (outside pen)*** | | 7 | **4.7** | 3.8 | 2 | 2 | 11 |
| **Pen Observation (inside pen)** | | 7 | **2.9** | 1.1 | 3 | 1 | 4 |
| **Queuing**** | At feeding | 7 | 15 | 0 | 15 | 15 | 15 |
| | At drinking | | 15 | 0 | 15 | 15 | 15 |
| | Overall | | **17** | 3.6 | 15 | 15 | 24 |
| **QBA** | | 7 | **11.1** | 2.3 | 10 | 10 | 16 |
| **HAR***** | Latency | 4 | 2.5 | 2.5 | 2.4 | 0.32 | 5 |
| | AD test | | 7.1 | 3.8 | 6 | 4 | 12.5 |
| | Overall | | **9.7** | 5.1 | 8.2 | 5.3 | 16.9 |
| **Clinical scoring**** | | 6 | **100.3** | 46.6 | 99.5 | 40 | 168 |
| **Overall** | | 4 | **117.2** | 36 | 119.7 | 75.3 | 153.9 |

*Pen observation (outside pen) time could only be recorded in 7 farms
**Queuing (at feeding rack and at drinking) and clinical scoring times could only be recorded in 7 farms
*** HAR tests time could only be recorded in 4 farms

| Indicators | | Farm Size Category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Small | | | Medium | | | Large | | |
| | | Farms | Mean | Range | Farms | Mean | Range | Farms | Mean | Range |
| **Number of pens** | | 10 | 1.2 | 1-3 | 9** | 3.1 | 1-9 | 9** | 5.5 | 1-10 |
| **Number of pens with lactating goats** | | 10 | 1.4 | 1-3 | 10 | 2.6 | 1-5 | 10 | 5.3 | 2-9 |
| **Number of animals in pen** | | 10 | 48.8 | 20-70 | 10 | 123.2 | 35-300 | 10 | 132.7 | 25-225 |
| **Area per goat ($m^2$)** | | 10 | 3 | 1.08-7.8 | 10 | 3.5 | 0.2-18.8 | 10 | 1.9 | 0.9-2.8 |
| **Number of feed spaces per pen** | | 9* | 5.7 | 0-24 | 10 | 1.9 | 1-20 | 10 | 1.8 | 1-34 |
| **Goats per feed space** | | 9* | 8.3 | 1.5-59 | 10 | 31.4 | 0.4-130 | 10 | 49.5 | 6.6-280 |
| **Feed trough length (m)º** | | 9* | 17.4 | 0-48 | 10 | 24.9 | 5-80 | 10 | 24.5 | 3.0-68 |
| **Goats per feed trough length** | | 9* | 1.7 | 0.8-7.4 | 10 | 11.7 | 1.5-26.7 | 10 | 2.8 | 1.2-50 |
| **Number of drinkers per pen** | | 9* | 2 | 0-10 | 10 | 1.6 | 1-4 | 10 | 2.2 | 1-7 |
| **Goats per drinker** | | 9* | 16.3 | 7-59 | 10 | 28.8 | 35-200 | 10 | 24.9 | 8-112.5 |
| **Water trough length** | | 9* | 1.2 | 0-5 | 10 | 1.2 | 0-2 | 10 | 2.2 | 0.2-8 |
| **Goats per water trough length** | | 9* | 15.5 | 14-59 | 10 | 32.8 | 27.7-178 | 10 | 30.2 | 12.5-157.1 |
| **Access to outdoor grazing*** ** | Months per year | 9 | 9.2 | 0-12 | 2 | 1.8 | 0-12 | 1 | 1.2 | 0-12 |
| **Disbudding** **** | Age (days) | 4 | 30 | 0-180 | 6 | 19.8 | 0-90 | 7 | 21.8 | 0-120 |
| **Bedding (days)** | Added | 10 | 5.1 | 1-30 | 9** | 4.9 | 1-15 | 9** | 4.9 | 1-10 |
| | Replaced | 9* | 86.7 | 15-210 | 10 | 103.6 | 1-180 | 9** | 89.8 | 8-365 |
| **Human:animal ratio** | | 10 | 0.027 | 0.01-0.08 | 10 | 0.006 | 0-0.01 | 10 | 0.004 | 0-0.01 |

*These indicators could only be recorded in nine small farms
** These indicators could only be recorded in nine medium and large farms
***Access to outdoor grazing was only verified in nine, two and one small, medium and large farms, correspondingly
****Disbudding was only performed in four, six and seven small, medium and large farms, respectively
ºMeters per vertically separated feed space. In pens with horizontal rail feed bunk: m along the rail

Table 36 - Resource and management-based indicators of the 30 Portuguese farms where the AWIN prototype for dairy goats was tested.

## Annex 2 – Figures

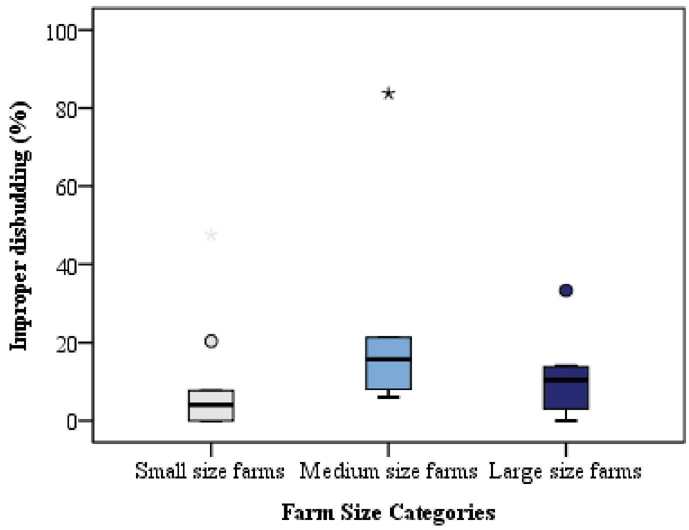



Figure 28 - Variation in 'Improper disbudding' prevalence in the 30 dairy farms visited, by farm categories. Points (o) and asterisks (*; extreme values) represent outliers.

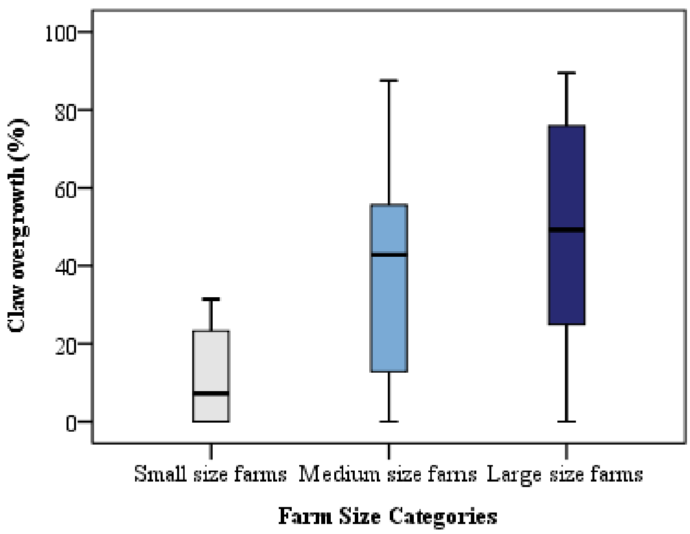



Figure 29 - Claw overgrowth' prevalences by farm categories from the 30 dairy farms visited.

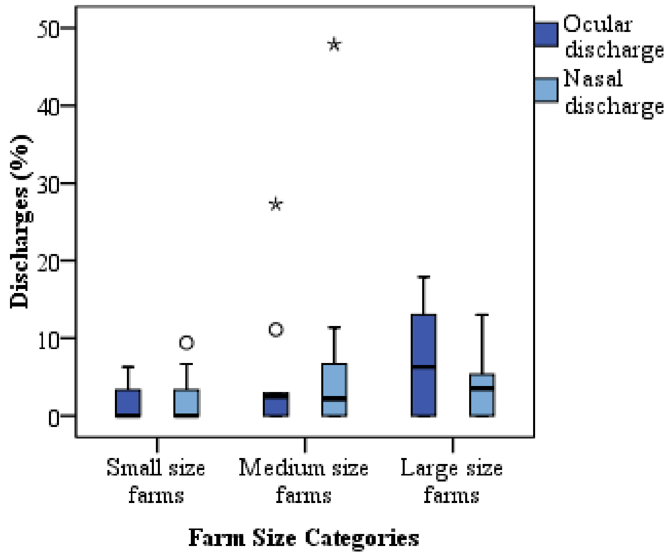


Figure 30 - Variation in ocular and nasal discharges prevalences in the 30 dairy farms visited, by farm categories. Points (o) and asterisks (*; extreme values) represent outliers.

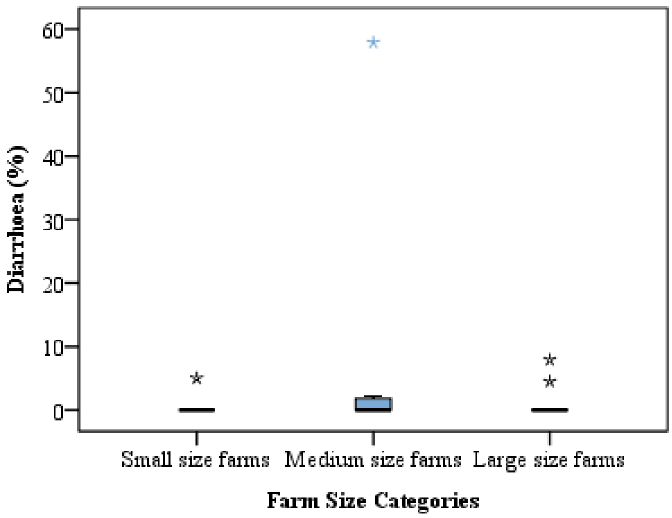


Figure 31 - ‘Diarrhoea’ prevalences by farm categories from the 30 dairy farms visited. Asterisks (*; extreme values) represent outliers.

## Annex 3 – Field data sheets

Figure 32 - Pen-level observations' recording sheet of the AWIN prototype for dairy goats: improper disbudding (stage 1), queuing at feeding and drinking (stage 2) and hair coat condition, oblivious, thermal stress and kneeling at feeding rack assessment (stage 3).

Date: ___/___/___ Observer: ______________

Farm: __________________

Number of pens: ___ ID of pen in assessment: ___ Number of animals in the pen: ___

**ALWAYS COLLECT TIME NEEDED FOR EACH INDICATOR!!!**

| | | |
|---|---|---|
| IMPROPER DISBUDDING | Number of BAD disbudded animals | |

| QUEUING AT FEEDING RACK | |
|---|---|
| *Scan* | *Number of queuing animals* |
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |

| QUEUING AT WATER PLACE | |
|---|---|
| *Scan* | *Number of queuing animals* |
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |

| | | | |
|---|---|---|---|
| HAIR COAT CONDITION | Number of animals with POOR hair coat condition | | |
| KNEELING | Number of animals in kneeling position | *Feeding rack* | |
| | | *Pen* | |
| OBLIVION | Number of oblivious animals | | |
| ABNORMAL LYING POSTURE | Number of abnormal lying posture animals | | |

| THERMAL STRESS | | | | |
|---|---|---|---|---|
| *Panting score* | Number of animals | | *Shivering score* | Number of animals |
| Score 0 | | | Score 0 | |
| Score 1 | | | Score 1 | |
| Score 2 | | | Score 2 | |
| | | | | |

**PANTING SCORE**

**Score 0** – Normal respiration: the mouth is closed, the flank moves regularly (slightly visible). The legs are frequently held near the body.

**Score 1** – Elevated respiration: from slightly to moderate panting with mouth closed. May be present small amount of drool or saliva. The posture is functional to heat loss, e.g. the neck is frequently extended, the legs may be held far from the body.

**Score 2** – Panting: from heavy to severe open-mouthed panting. The mouth is open accompanied by protruding tongue and excessive salivation. The neck is frequently extended, the legs may be held far from the body.

**SHIVERING SCORE**

**Score 0** – Hair coat is flat on the back, no signs of cold stress are visible.

**Score 1** – The hair is bristling on the back. The goat has a thick coat.

**Score 2** – The goat is shivering and may take a posture with arched back and head down.

Figure 33 - Pen-level observations' recording sheet of the AWIN prototype for dairy goats: Qualitative Behaviour Assessment (QBA; stage 4).

***Farm:*** ___________

**AGGRESSIVE**

**AGITATED**

**ALERT**

**BORED**

**CONTENT**

**CURIOUS**

**FEARFUL**

**FRUSTRATED**

**IRRITATED**

**LIVELY**

**RELAXED**

**SOCIABLE**

**SUFFERING**

Figure 34 - Pen-level observations' recording sheet of the AWIN prototype for dairy goats: group assessment on human-goat relationship (stage 5).

| | |
|---|---|
| **LATENCY TO FIRST CONTACT (seconds)** | |

| **AVOIDANCE DISTANCE TEST** | | |
|---|---|---|
| *Goat (TAG)* | *Contact (Y/N)* | *Acceptance (Y/N)* |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

## Clinical Scoring sheet

Date: __/__/__ Observer:

Farm:

Number of pens: ___ ID of pen in assessment: ___ Number of animals in the pen: ___

Following the sampling strategy in attachment, always assess the same animals for the indicators present in this chart.

| | | Rear | | Udder | | | Hind Q | | | | | | Body | | | | Knee | | Lower legs | | | | | | Neck | | | | Head | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Goat ID | BCS (-1/0/1) | Diarrhoea (0/1) | Vulvar discharge (0/1) | Asymmetry (0/1) | Cleanli-ness (0/1) | Abscesses (0/1) | Cleanli-ness (0/1) | | Lesions and swellings (0/1) | | Abscesses (0/1) | | Lesions and swellings (0/1) | | Abscesses (0/1) | | Lesions and swellings (0/1/2) | | Cleanl-iness (0/1) | | Lesions and swellings (0/1) | | Claw over-growth (0/1) | | Lesions and swellings (0/1) | | Abscesses (0/1) | | Lesions and swellings (0/1) | Absces-ses (0/1) | Ocular discharge (0/1) | Nasal discharge (0/1) |
| | | | | | | | L | R | L | R | L | R | L | R | L | R | L | R | L | R | L | R | L | R | L | R | L | R | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

**ID:** Identification; **BCS:** -1-Very Thin, 0-Regular, 1-Very fat; **Disease:** 0-Absence, 1-Presence; **Cleanliness:** 0- No dirt, wet hair, or minor splashing, 1- Separate or continuous plaques of dirt; **Lesions:** 0- Absence of lesions/swellings (skin damage with/without hair loss), 1- Presence of lesions/swellings (skin damage with/without hair loss); **Udder Asymmetry:** 0-No asymmetry, 1-Asymmetry; **Knee Lesions:** 0 - No lesions, hair loss or skin thickening, 1 - Skin damage with/without hair loss and reddened skin, but no enlargement of any joint, 2 - Skin damage with hair loss, and enlargement of at least one joint, showing a thick callus; **Claw overgrowth:** 0- without overgrown claws, 1- with overgrown claws; **Abscesses:** 0- Absence, 1-Presence

Figure 35 – Individual assessment's recording sheet of the AWIN prototype for dairy goats: clinical scoring (stage 6).

**Clinical Scoring sheet**

Date: __/__/__ Observer:

Farm:

Number of pens: ___ ID of pen in assessment: ___ Number of animals in the pen: ___

1. Record the number of severely lame animals /Score 1) in the pen: _____
2. Record the number of kneeling animals: at the feeding rack: _____ and at the pen:_____

**ID:** Identification; **BCS:** -1-Very Thin, 0-Regular, 1-Very fat; **Disease:** 0-Absence, 1-Presence; **Cleanliness:** 0- No dirt, wet hair, or minor splashing, 1- Separate or continuous plaques of dirt; **Lesions:** 0-Absence of lesions/swellings (skin damage with/without hair loss), 1- Presence of lesions/swellings (skin damage with/without hair loss); **Udder Asymmetry:** 0-No asymmetry, 1-Asymmetry; **Knee Lesions:** 0 - No lesions, hair loss or skin thickening, 1 - Skin damage with/without hair loss and reddened skin, but no enlargement of any joint, 2 - Skin damage with hair loss, and enlargement of at least one joint, showing a thick callus; **Claw overgrowth:** 0- without overgrown claws, 1- with overgrown claws; **Abscesses:** 0- Absence, 1-Presence

Figure 35 – Individual assessment's recording sheet of the AWIN prototype for dairy goats: clinical scoring (stage 6). (continuation)

## Questionnaire: On-farm collection of welfare indicators

**Assessment of management and resource based indicators for lactating dairy goats**

**Date:** ____/____/____

### I. Identification of the farm

1. Farm identification: ______________________________
2. Geographic location: ______________________________
3. Weather conditions at the farm on the day of the assessment: 3.1 Temperature: __________ 3.2 Relative humidity: __________

______________________________

### II. Identification of the farm manager

1. Age: ______________________________
2. Gender: ☐Female ☐Male
3. Role on the farm: ______________________________
4. Education level: ☐Primary ☐Secondary ☐Higher
5. Participation to specific courses on animal welfare (Y/N): ______________________________
6. Years running the operation: ______________________________

### III. Resource-based check-list

| | Question | Answer |
|---|---|---|
| 1. | Activity of the farm manager:<br>1. Part-time<br>2. Full-time | |
| 2. | Workers on the farm:<br>1. Number of permanent workers:<br>2. Number of seasonal workers: | |
| 3. | Milkers:<br>1. Number: | |
| 4. | Breeds present at the farm:<br>1. Saanen (n. of adult dairy goats):<br>2. Alpine (n. of adult dairy goats):<br>3. Crossbreds (n. of adult dairy goats):<br>4. Other breeds (specify:______________________________) | |

Figure 36 – Resource and management-based indicators assessment questionnaire made to the farmers at stage 7 and 8 of the AWIN prototype.

5. Number of pens with lactating dairy goats: __
Please specify for each pen:

| | N. of animals | Length | Width | Number of feeding spaces | Feed trough length | Number of functioning water troughs | Water trough n° and length | Presence of horned and dehorned goats (Y/N) |
|---|---|---|---|---|---|---|---|---|
| Pen 1 | | | | | | | | |
| Pen 2 | | | | | | | | |
| Pen 3 | | | | | | | | |
| Pen 4 | | | | | | | | |
| Pen 5 | | | | | | | | |
| Pen 6 | | | | | | | | |
| Pen 7 | | | | | | | | |
| Pen 8 | | | | | | | | |
| Pen 9 | | | | | | | | |
| Pen 10 | | | | | | | | |
| Pen 11 | | | | | | | | |

5.1 Flooring material:
- ☐ concrete
- ☐ soil
- ☐ wood bars
- ☐ rubber bars
- ☐ other, please specify: ______________________________________________

5.2 Bedding material:
- ☐ straw
- ☐ wood shavings
- ☐ no bedding
- ☐ other, please specify: ______________________________________________

6. Presence of outdoor grazing or exterior pen (Y/N):
If YES:
1. How many days/year?
2. How many hours/day?

7. Type of milking parlour:
1. ☐ Manual ☐ Automatic
2. N° milking points: _________
3. Is the milking machine checked regularly (at least once/year) (Y/N):

8. Number of replacement goats per year?

9. With respect to the previous year, the number of lactating goats in the farm is:
- ☐ increasing
- ☐ more or less the same
- ☐ decreasing

Figure 36 – Resource and management-based indicators assessment questionnaire made to the farmers at stage 7 and 8 of the AWIN prototype. (continuation)

## IV. Management-based questionnaire

| | Question | Answer |
|---|---|---|
| 1. | Pen grouping strategy:<br>☐ No specific strategy<br>☐ Production level<br>☐ Days in milking<br>☐ Goats' age<br>☐ Other, please specify: | |
| 2. | Is regrouping done (Y/N): If Yes, how often: | |
| 3. | Bedding:<br>3.1How often is new bedding replaced:<br>3.2How often is new bedding added: | |
| 4. | Feed:<br>4.1 Number of feed delivery/day:<br>4.2 Roughage/concentrate ratio:<br>4.3 Roughage always available (Y/N):<br>4.4 Concentrate always available (Y/N): | |
| 5. | Water:<br>5.1 Water origin:<br>☐ Public services<br>☐ Well, stream<br>☐ Other, please specify:<br>5.2 Routine for water analysis:<br>☐ No routine, done whenever necessary<br>☐ Every 3 months<br>☐ Every 6 months<br>☐ Every year<br>☐ Other, please specify: | |
| 6. | Frequency of claw trimming:<br>No routine, done whenever necessary<br>☐ Every 3 months<br>☐ Every 6 months<br>☐ Every year<br>☐ Other, please specify:<br>6.1 Date of last claw trimming of goats in the selected pen: | |
| 7. | Age at first kidding: | |
| 8. | Are all goats subjected to a dry period? (Y/N)<br>8.1 If yes, for how long are the goats dry?<br>8.2 If no, which percentage of goats are not dried? | |
| 9. | Disbudding/dehorning routine:<br>9.1 Is disbudding done routinely? (Y/N)<br>9.2 Pain management in disbudding (Y/N):<br>9.3 Age of the animals at disbudding:<br>9.4 Is dehorning done routinely? (Y/N)<br>9.5 Pain management in dehorning (Y/N): | |
| 10. | Infirmary - are sick animals isolated? (Y/N): | |
| 11. | How important do you think it is to gently touch the goats? (1 – not important to 5 – very important) | |
| 12. | How important do you think it is to talk to the goats during milking? (1 – not important to 5 – very important)? | |
| 13. | Do you use a stick when you enter the pen? (Y/N)<br>13.1 If Y, why?<br>☐ moving the animals<br>☐ beating the animals<br>☐ safety tool<br>☐ other (________________________________________) | |

Figure 36 – Resource and management-based indicators assessment questionnaire made to the farmers at stage 7 and 8 of the AWIN prototype. (continuation)

| | |
|---|---|
| 14. | When you enter the pen, what do most goats do?<br>☐ they approach you<br>☐ they stay still<br>☐ they go away |
| 15. | How much do you think pain matters to goats?<br>(1-very little to 5-very much) |

## V. Data-driven indicators (overall farm)

| | Question | Answer |
|---|---|---|
| 1. | Average age within milking goats: | |
| 2. | Mean parity: | |
| 3. | Mean days in milking: | |
| 4 | Mean SCC | |
| 5. | Culling strategy:<br>☐ age of the animals<br>☐ diseases or injuries<br>☐ low milk production<br>☐ other (________________________________) | |
| 6. | Do you usually sell goats to other dairy farms? (Y/N): | |
| 7. | Herd average annual milk yield (litres) in 2013 | |
| 8. | Annual mean milk content in 2013:<br>8.1. fat<br>8.2 protein | |
| 9. | Mortality in 2013:<br>9.1 Adults<br>9.2 Newborns | |
| 10. | History of sanitary slaughters in the last 3 years? (Y/N) | |

Figure 36 – Resource and management-based indicators assessment questionnaire made to the farmers at stage 7 and 8 of the AWIN prototype. (continuation)

## Annex 4 - Sample size determination

From a total of 3058 Portuguese dairy goat farms, 269 farms are under intensive production system (DGAV, personal communication). It is recommendable that the number of sampling units to test (dairy goat farms), for each farm size category generated in the present study ([50-99], [100-499] and [>500]), be drawn assuming for instance, an absolute error of 5% and calculating the minimum sample size for a 95 % confidence level. Using a software for sample size estimation (e.g. Win Episcope 2.0) and assuming a 50% expected ratio (as before data collection reference prevalences were not available), the minimum number of sampling units to be tested is shown in Table 37:

| Categories | [50-99] | [100-499] | [>500] | Total |
|---|---|---|---|---|
| **Population** | 92 | 161 | 16 | 269 |
| **Sample** | 75 | 114 | 16 | 205 |
| **Pre-determined sample** | 10 | 10 | 10 | 30 |

Table 37 - Minimal number of sampling units to be tested by farm size category in contrast with the pre-determined sample.