Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:49
Faculty of Veterinary Medicine and Animal Science

Can more animal-based parameters improve the official animal welfare control for dairy cows?

Birgitta Staaf Larsson

Can more animal-based parameters improve the official animal welfare control for dairy cows?

Birgitta Staaf Larsson
Faculty of Veterinary Medicine and Animal Science
Department of Animal Biosciences
Uppsala

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Uppsala 2025

Acta Universitatis Agriculturae Sueciae
2025:49

Cover: The author together with a cow.
Photo: Sara Larsson

ISSN 1652-6880
ISBN (print version) 978-91-8046-484-0
ISBN (electronic version) 978-91-8046-534-2
https://doi.org/10.54612/a.48bfuob511

Swedish University of Agricultural Sciences, Department of Animal Biosciences, Uppsala, Sweden

Print: SLU Grafisk service, Uppsala 2025

Can more animal-based parameters improve the official animal welfare control for dairy cows?

Abstract

The Swedish official animal welfare control (OC) mainly uses resource-based, and a few animal-based measures, that are neither quantified nor qualified. In Paper I, OC was compared with two other protocols using mainly animal-based measures: Ask the Cow (AC) and Welfare Quality (WQ). The data were collected from 41 farms (mean herd size 65 cows, range 12-268). Totally, in all herds, the OC had only 55 remarks out of 1,763 possible. If farms were ranked (worst to best) within OC, AC and WQ, respectively, OC showed no or a moderate correlation ($r = 0.40$, $P = 0.009$) to observations in WQ and AC, respectively. For example, dirtiness was reported only on two farms in OC while dirtiness was reported at all herds in WQ. In AC, on average, 18% of the animals were reported dirty. Paper II aimed to study to what extent body condition and skin soiling in cattle are rated different depending on the profession, education and experience. Swedish animal professionals and undergraduate students in animal science ($n = 569$) rated 6 to 40 photos concerning animal body condition, and skin soiling. Students rated the degree of soiling more seriously and recommended stricter corrective actions. In Paper III, thermal comfort was measured by using two handheld skin temperature sensors. Skin temperatures in 21 dairy cows of two breeds were recorded monthly for one year and compared to indoor and rectal temperatures. Skin temperatures did not correlate well with rectal temperature but correlated strongly with indoor temperature and could be a quick monitoring tool. In paper IV, assessment of hydration status was measured by variations in milk osmolality within and between days, within individual cows, between two breeds and over season and lactation stages as well as the correlation to ambient temperature and relative humidity. There were significant variations in milk osmolality, and if osmolality could be regularly monitored, it could be a tool to detect dehydration. The answer of the thesis aim is yes, the official animal welfare control can be improved by using more and quantified animal-based parameters.

Keywords: animal-based measures, animal welfare assessment, animal welfare control, body condition, cleanliness, dairy cattle, dehydration, milk osmolality, resource-based measures, skin soiling, thermal comfort

Kan användandet av fler djurbaserade parametrar förbättra djurskyddskontrollen?

Sammanfattning

Den svenska officiella djurskyddskontrollen (OK) använder huvudsakligen resursbaserade mått och bara ett fåtal djurbaserade, som varken mäts kvantitativt eller kvalitativt. I artikel I jämfördes OK med två andra protokoll, som huvudsakligen använder djurbaserade mått: Fråga kon (FK) och Welfare Quality (WQ). Data samlades in från 41 gårdar (snittstorlek 65 kor, intervall 12–268). Totalt, i alla besättningar, hade OK endast 55 anmärkningar av 1763 möjliga. Vid rankning av gårdarna inom OK, FK och WQ, visade OK ingen eller måttlig korrelation (r = 0,40, P = 0,009) med observationer i WQ respektive FK. Smutsighet rapporterades till exempel endast på två gårdar i OK, medan smutsighet observerades i alla besättningar i WQ. I FK rapporterades i genomsnitt 18 % av djuren som smutsiga. I artikel II var syftet att studera i vilken utsträckning bedömning av hull och smutsighet hos nötkreatur varierar beroende på yrke, utbildning och yrkeserfarenhet. Svenska experter inom djurområdet samt studenter inom husdjursvetenskap (n = 569) bedömde från 6 till 40 foton med avseende på hull och smutsighet. Studenterna bedömde graden av nedsmutsning mer allvarligt och rekommenderade striktare åtgärder. I artikel III mättes termisk komfort med hjälp av två hudtemperaturmätare. Hudtemperaturen hos 21 mjölkkor av två raser registrerades varje månad under ett år och jämfördes med inomhus- och rektaltemperaturer. Hudtemperaturerna korrelerade inte väl med rektaltemperaturen men korrelerade starkt med inomhustemperaturen och skulle kunna användas som snabb övervakning av kors temperaturförhållanden. I artikel IV beskrevs variationen i mjölkosmolalitet (som markör för vätskestatus) inom enskilda kor, inom och mellan dagar, över säsong och laktation, mellan två raser samt korrelationen med omgivningstemperatur och relativ fuktighet. Det fanns signifikanta variationer i mjölkens osmolalitet. Om osmolaliteten kunde övervakas regelbundet skulle det kunna vara ett verktyg för att upptäcka uttorkning. Svaret på frågan i avhandlingens titel är ja, den officiella djurskyddskontrollen kan förbättras genom att använda fler och mer kvantifierade djurbaserade parametrar.

Nyckelord: djurbaserade parametrar, djurvälfärdsbedömning, djurskyddskontroll, hull, mjölkkor, nedsmutsning, mjölkosmolalitet, renlighet, resursbaserade parametrar, termisk komfort, uttorkning

Preface

I am an animal lover and always have been. My first close experience with cows was at our family´s summer cottage, where I collected the family´s daily milk at a nearby dairy farm. Then I joined the ´Swedish Blue Star´ courses learning more about dairy cattle. When I was 20, I met my husband, and we decided to move to the countryside. For 25 years we lived on a small farm. During the years, we had hens, cats, rabbits, guinea pigs, a Giant Schnauzer and periodically horses. Just beside our house we had grazing dairy cows. I worked for five years as an animal welfare inspector and worked for a couple of years on animal transport issues at the Swedish Board of Agriculture, Sweden. In 2002 I got a post as a university lecturer at the Swedish University of Agricultural Sciences (SLU) in Skara, Sweden, working with continuing education in animal protection and welfare as well as I was one of the designers of the Bachelor Program in Ethology and Animal Welfare. In the spring of 2012, I held a position as an animal welfare inspector at the County Administrative Board of Västra Götaland, Sweden, and later the same year I got a post at the Swedish Centre for Animal Welfare at SLU in Uppsala, Sweden. My urge was to develop my research skills in animal welfare and the Swedish Association for the Protection of Animals made it possible by awarding me research funding for PhD studies, for which I am very grateful, thank you!

Dedication

To Henrik, my beloved husband for 35 years, who always believes in me and encourages me to follow my dreams. I love you!

Contents

List of publications 15

Abbreviations 17

Abbreviations related to legislation 19

1. Introduction 21
   - 1.1 The Concept of Animal Welfare 23
     - 1.1.1 A brief history 23
     - 1.1.2 The European Commission 23
     - 1.1.3 The Swedish Animal Welfare Act 24
     - 1.1.4 The Swedish official control of animal welfare 24
   - 1.2 Improved animal welfare measurements 25
     - 1.2.1 European interest in measuring animal welfare 26
     - 1.2.2 Welfare Quality 26
     - 1.2.3 Ask the Cow 27
     - 1.2.4 Other farm assurance schemes 28
     - 1.2.5 Importance of measuring body condition 29
     - 1.2.6 Importance of measuring cleanliness 29
     - 1.2.7 Importance of thermal comfort 30
     - 1.2.8 Importance of adequate water intake 31

2. Aims of the thesis 35

3. Material and methods 37
   - 3.1 Paper I - Comparison of assessment protocols 37
     - 3.1.1 Data processing, ranking and statistical analyses 37
   - 3.2 Paper II - Visual assessment of body condition and skin soiling 38
     - 3.2.1 Statistical analyses 38
   - 3.3 Paper III - Methods for assessing skin temperature 39
     - 3.3.1 Statistical analyses 40
   - 3.4 Paper IV - Milk osmolality to assess hydration status 40
     - 3.4.1 Statistical analysis 41

4. Results 43
   - 4.1 Paper I 43

4.1.1 Ranking of the herds within the various systems 43
4.1.2 Assessment of body condition 45
4.1.3 Assessment of cleanliness 46
4.1.4 Assessment of skin lesions.. 46
4.1.5 The WQ assessment regarding the four principles 46
4.2 Paper II 46
4.2.1 Results.. 46
4.3 Paper III 47
4.3.1 Skin temperature measurements 47
4.4 Paper IV 48
4.4.1 Variation in milk osmolality 48

5. Discussion 49
5.1 The Swedish official control and comparison with other animal welfare assessment schemes 51
5.1.1 Freedom from thirst 54
5.1.2 Body condition scoring 55
5.1.3 Cleanliness 56
5.1.4 Sickness and lesions 57
5.1.5 Principle Component Analyses....................................... 57
5.1.6 Lying down vs standing up 57
5.1.7 Natural behaviour 58
5.1.8 Animal welfare implications 59
5.1.9 Skin temperature vs rectal temperature 60
5.1.10 Skin temperature vs indoor temperature 60
5.1.11 Usefulness in the official control 61
5.1.12 Osmolality as an indicator of the hydration status 62
5.2 Animal Welfare assessment improvements................................ 63
5.2.1 Incentives for improved animal welfare 63

6. Main conclusions 65

7. Future research 67

References.. 69

Popular science summary 79

Populärvetenskaplig sammanfattning 83

Acknowledgements 87

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Staaf Larsson, B., Jansson, A., Holmberg, M., Winblad von Walter, L., Stéen, M., & Dahlborn, K. (2024). A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds. Acta Agriculturae Scandinavica, Section A — Animal Science, 73:3-4, 153-164. https://doi.org/10.1080/09064702.2024.2317708

II. Staaf Larsson, B., Petersson, E., Stéen, M. & Hultgren, J. (2020). Visual assessment of body condition and skin soiling in cattle by professionals and undergraduate students using photo slides. Acta Agriculturae Scandinavica, Section A — Animal Science, 70:1, 31-40. https://doi.org/10.1080/09064702.2020.1849380

III. Staaf Larsson, B., Dahlborn, K., & Jansson, A. (2024). Methods for assessing skin temperature in two breeds of dairy cows and their correlation to indoor and rectal temperature. Acta Agriculturae Scandinavica, Section A — Animal Science, 74(1), 43–50. https://doi.org/10.1080/09064702.2024.2435339

IV. Staaf Larsson, B., Jansson, A., Dahlborn, K. Milk osmolality as a practical tool in official control to assess hydration status in dairy cows – A pilot study (submitted)

The contribution of Birgitta Staaf Larsson to the papers included in this thesis was as follows:

I. Drafted the manuscript with regular input from co-authors and support from supervisors. Corresponded with the journal.

II. Planned the project and collected the data. Drafted the manuscript with regular input from co-authors and support from supervisors. Corresponded with the journal.

III. Actively participated in the study planning. Collected the data and did some of the statistical analyses. Finalised the manuscript with regular input from co-authors and supervisors. Corresponded with the journal.

IV. Actively participated in the study planning. Collected the data and made the analyses. Performed the statistical analysis with help from the supervisor and a statistician. Drafted the manuscript and finalised it with regular input from co-authors and supervisors. Corresponded with the journal.

Abbreviations

AC	Ask the Cow
ADT	Avoidance distance test
AWA	Animal Welfare Act
AWO	Animal Welfare Ordinance
BCS	Body condition score
BTSF	Better Training for Safer Food
CAB	County Administrative Board
ECM	Energy corrected milk
ECTS	European Credit Transfer System
EFSA	European Food Safety Authority
EU	European Union
EURCAWs	European Centres for Animal Welfare
IRT	Infrared thermometer
MTT	Medical thermistor thermometer
NMR	National Milk Recording
OC	Official control
PCA	Principal component analysis
QBA	Qualitative Behaviour Assessment
RH	Relative humidity
SD	Standard deviation
SDG	Sustainable Development Goals
SH	Swedish Holstein
SLU	Swedish University of Agricultural Sciences
SRB	Swedish Red Breed
THI	Temperature-humidity index

WOAH	World Organization for Animal Health
WQ	Welfare Quality

Abbreviations related to legislation

Directive 98/58/EC	Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. Official Journal of the European Union L 221, 8.8.1998
Directive 2010/63/EU	Directive 2010/63/EU of the European Parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes (EUT L 276, 20.10.2010, p. 33, Celex 32010L0063)
Regulation (EU) 2017/625	Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products. (EUT L 95, 7.4.2017, p. 1, Celex 32017R0625)
SFS 2018:1192	Djurskyddslagen SFS 2018:1192 [The Swedish Animal Welfare Act]
SFS 2019:66	Djurskyddsförordningen SFS 2019:66 [The Swedish Animal Welfare Ordinance]
SJVFS 2022:13 (L 44)	Statens jordbruksverks föreskrifter (SJVFS 2022:13) om offentlig kontroll på djurskyddsområdet [The Swedish Board of Agriculture´s regulations (SJVFS 2022:13) on official control in the area of animal welfare, Case No L 44]
SJVFS 2019:18 (L 104)	Statens jordbruksverks föreskrifter och allmänna råd (SJVFS 2019:18) om nötkreaturhållning inom lantbruket m.m., senast omtryckt genom 2024:7, saknr L104 [The Swedish Board of Agriculture´s regulations and general advice (SJVFS 2019:18) on cattle husbandry etc., last reprinted by 2024:7, Case No L104]

1. Introduction

Animal welfare in Sweden is regulated through national and European Union (EU) legislation (Swedish Board of Agriculture, 2025). Animal welfare inspectors carry out the official animal welfare control, focusing mainly on preventive and potential risks in animal housing and management (Blokhuis *et al*., 2010; Broom, 2017). Inspection protocols can include three categories of assessments: resource-, management- and animal-based, although the latter currently to a limited degree (Figure 1). Resource-based measures are based on observations of the animal´s environment and resources, such as space allowance and air quality, while management-based measures include caretaking strategies and animal handling (Keeling, 2009). Animal-based measures are to study the animal´s response or a situation that influences the animal.

Figure 1. Swedish and European protocols for animal welfare inspections include three categories of assessments: resource-, management- and animal-based parameters, although the latter is currently limited in scope. Examples of registrations and measurements are shown in the figure.

Blokhuis *et al*. (2010) state that using resource-based measures requires extensive detailed legislation and does not stimulate farmers to make innovations. However, they also say that evaluating animal welfare only with animal-based welfare parameters cannot fully replace the mostly resource-based parameters used in animal welfare control, not to mislead farmers into investing in equipment or buildings that are not appropriate. Krueger *et al*. (2020) concluded in a study comparing three assessment programs, Welfare Quality (WQ), FARM (the U.S. National Dairy Farmers Assuring Responsible Management Program) and the Code (the New Zealand Code of Welfare: Dairy Cattle), that it is important to consider both the animals´

environment, the management and direct animal measurements. Lundmark *et al*. (2016) compared the Swedish legislation with three private standards (the Swedish milk industry control program Arlagården, www.arla.com, Seal of Quality http://sigill.se and KRAV organic standard www.krav.se/en/) focusing on dairy cows in Sweden. Like the Swedish legislation, all three private standards used more resource-based and management-based requirements than animal-based ones. However, the organic standard (KRAV) measures other things with the goal of achieving higher welfare than the official legislation.

The animal´s health status and well-being are important, as well as the animal´s feelings (affective state) and the ability to express natural behaviours (Fraser *et al.,* 1997). According to Blokhuis *et al.* (2003), animal production has become increasingly industrialised, with quantity often taking priority over quality. At the same time, consumers demand food produced from healthy animals with good welfare kept in an appropriate environment (Blokhuis *et al.,* 2003; European Commission, 2016). Buller *et al.* (2018) raised the importance of including animal welfare in international governance associated with agricultural sustainability. Ventura *et al*. (2021) pointed out that best practices for animal welfare are inevitably to be implemented to achieve sustainable agriculture. Keeling *et al*. (2022) investigated the importance of animal welfare in achieving the United Nations Sustainable Development Goals (SDG). They concluded that increased animal welfare may also positively impact the SDGs.

To check animal-based measures is to study the animal´s response or a situation that influences the animal. Animal-based measurement can be performed directly on the animal and/or indirectly and includes animal records (Berthe *et al*., 2012; Welfare Quality, 2009). The measurements may, for example, intend to (i) assess the degree of impaired functioning associated with injury, disease, and malnutrition; (ii) provide information on animals´ needs and affective states such as hunger, pain and fear; or (iii) assess the physiological, behavioural and immunological changes or effects that animals show in response to various challenges (Berthe *et al*., 2012).

Apart from official control, animal-based measures may also facilitate benchmarking (Goossens *et al*., 2008), thus contributing to a gradual improvement in animal welfare.

1.1 The Concept of Animal Welfare

1.1.1 A brief history

In 1964, Ruth Harrison wrote the book ´Animal Machines´, which explains how traditional farming became factory farming. The book was an eye opener for the public and an intensive debate about animal production followed. The Technical Committee of Enquiry into the Welfare of Livestock Kept under Intensive Conditions (Brambell, 1965) stated that domestic animals can suffer and experience feelings, and animals should be protected as far as possible from conditions that may cause suffering. Therefore, the Farm Animal Welfare Council (FAWC, 1993) developed the Five Freedoms, which are internationally accepted animal welfare standards. Fraser *et al*. (1997) raised the importance of considering an animal´s quality of life, including a natural life, being protected from fear, pain, and other negative states and being in good order. Mellor and Reid (1994) developed the Five Freedoms to include the mental state, and Boissy *et al*. (2007) introduced the concept of positive welfare, further elaborated by Keeling *et al*. (2021) and Rault *et al*. (2025) not only to include positive emotions in animals, but also a good life and happiness.

The World Organization for Animal Health (WOAH) has traditionally been focused on animal health. However, animal welfare has gained increased attention. The WOAH code states, ´Animal welfare means the physical and mental state of an animal in relation to the conditions in which it lives and dies´ (Article 7.1.1. WOAH Terrestrial Animal Health Code). The article further explains ´An animal experiences good welfare if the animal is healthy, comfortable, well nourished, safe, is not suffering from unpleasant states such as pain, fear and distress, and is able to express behaviours that are important for its physical and mental state.´

1.1.2 The European Commission

In the Treaty on European Union, Maastricht Treaty (European Union, 1992), a Declaration on the protection of animals was included, stating that ´when drafting and implementing Community legislation on the common agricultural policy, transport, the internal market and research, to pay full regard to the welfare requirements of animals.´ In the Lisbon Treaty (European Commission, 2007), amended from the Treaty of Amsterdam (European Commission, 1997), animals were considered sentient beings.

The Council Directive concerning the protection of animals bred or kept for farming purposes, implemented in 1998, states that Member States shall make provisions to ensure that the owners or keepers take all reasonable steps to ensure the welfare of their animals.

In the communication on the EU Strategy for the Protection and Welfare of Animals 2012-2015, the Commission addressed the importance of compliance with the legislation (European Commission, 2012). The European Commission (2012) favours increased use of output-based measures, i.e., animal-based, reflecting animal welfare per se. In the EU Farm to Fork strategy, animal welfare is important and is also what the citizens want (European Commission, 2020). The European Commission (2022) inquired about specific indicators to monitor whether the animal welfare conditions are improving, remaining stable or worsening in the EU Member States, which are currently lacking. The Commission states that when authorities use indicators, their controls provide a better assessment of the real state of the welfare of animals.

1.1.3 The Swedish Animal Welfare Act

The first Swedish Animal Welfare Act (AWA) goes back to 1944 (AWA 1944:219) and was replaced in 1988 by AWA 1988:534. The Swedish Government appointed an investigation team in 2009 to develop a basis for a new AWA. The Government was keen to explore a system of animal-based welfare parameters as an important complement to include in the legislation, and that animal welfare parameters could be used as a valuable tool for the farmer to monitor the animals him-/herself. Sweden adopted the new AWA (2018:1192), which entered into force on 1 April 2019. This Act and the decrees issued by the law consist of minimum requirements and shall protect animals against unnecessary suffering and disease. The law sets out conditions necessary to ensure that animal welfare problems do not arise.

1.1.4 The Swedish official control of animal welfare

The responsibility for ensuring that regulations on animal welfare from the Parliament and the Government are implemented lies with the 21 County Administrative Boards (CAB) in Sweden (AWA 2018:1192). The CAB veterinarians are responsible for the veterinary public health, animal health personnel, animal welfare and health control, and of the appropriate training for the animal welfare inspectors, who carry out the official animal welfare control in each county. The animal welfare inspectors´ task is to check that Swedish and EU animal welfare legislation is followed. Animal welfare inspectors usually have a graduate degree in agronomy, biology, animal science or in environmental and health science. Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products[1] also highlights the need for compliance verification. The controls are often carried out as planned farm visits and occasionally due to complaints from e.g. veterinarians or the public. Animal facilities including e.g. space allowance, lighting, noise and air quality, are checked during an inspection.

The animal welfare inspectors use a species-specific checklist (Supplementary Material S1 in Staaf Larsson *et al*., 2024; Swedish Board of Agriculture, 2024a) and, as a help, written guidance (Swedish Board of Agriculture, 2024b) developed by the Swedish Board of Agriculture. The checklists are primarily based on resource-based legislation but also include some animal-based measures (e.g., cleanliness and body condition). Animal welfare inspectors indicate ´Yes´, ´No´, ´Not checked´ and ´Not applicable´ for the various checkpoints (Swedish Board of Agriculture, 2024a). The inspection takes approximately 1-2 hours on the farm. After the control, a report is delivered to the farmer concerning the outcome of the control of the animals and the farmers´ compliance with the legislation. If the inspection shows that there are not many or serious deficiencies, the control report sometimes only contains some advice to follow. If the farmer has committed serious shortcomings in how he/she cares for the animals, or if the requirements in the legislation have not been met, an injunction may be issued with or without a fine (Swedish Board of Agriculture, 2022).

1.2 Improved animal welfare measurements

The validity, reliability and feasibility are important when choosing animal welfare parameters. In addition, accuracy and precision are furthermore important for measuring progression in the long term and for comparisons

[1]EUT L 95, 7.4.2017, s. 1, Celex 32017R0625.

between farms. The validity of animal-based parameters is superior to resource-based and management-based ones, as animal-based parameters measure the direct outcome on the animals (Knierim & Winckler, 2009). However, reliability can be somewhat problematic when assessing animal-based parameters, as an agreement between and within observers can be challenging (Knierim & Winckler, 2009) when measuring subjective methods, e.g., parameters where behaviours are evaluated. The parameters become more objective when using animal-based measurements such as temperature or osmolality. Feasibility must also be considered, as the time consumption for an assessment can be a limiting factor.

1.2.1 European interest in measuring animal welfare

To increase knowledge about animal welfare, the EU Commission funded the Welfare Quality (WQ) project, which ran from 2004-2009. The Commission adopted a Community Action Plan on the Protection and Welfare of Animals with initiatives and measures to improve the protection and welfare of animals for 2006-2010. In the European Parliament resolution on evaluation and assessment of the Animal Welfare Action Plan 2006-2010, the European Parliament called for a review of animal welfare standards in relation to consumer and producer interests (European Parliament, 2015). The Parliament requested the assessment and further development of the WQ assessment protocol concerning its practical application and the ability to improve the harmonisation of animal welfare in the EU.

1.2.2 Welfare Quality

The Welfare Quality (WQ) protocol was developed in 2009 by the Welfare Quality Network (Welfare Quality, 2009). Welfare Quality is a system that mainly uses animal-based measures, and it is based on four principles (good feeding, good housing, good health and appropriate behaviour). These principles, divided into 12 criteria, assesses various parameters (Blokhuis, 2008). Each principle is rated from 0-100%, with results of 0-20% being categorised as not classified (defined in Welfare Quality (2009) as animal welfare is low and considered unacceptable), 20-60% meaning that results are acceptable, 60-80% is better than average and > 80% is excellent. The results of the four combined principles are calculated and weighed due to their importance, in assigning the herd to a welfare category (Welfare Quality, 2009). For a more detailed description of the calculations, see

Welfare Quality Protocol (2009). According to the WQ assessment protocol for dairy cattle, in herds with less than 30 cows, all cows are assessed, and with an increasing number of animals, up to 300 cows, a maximum of 73 animals are assessed. A WQ assessment starts with an ´avoidance distance test´ (ADT) assessing the human-animal interaction. For more information, see Welfare Quality (2009). In addition, the assessor studies aggressive interactions (head butting, displacement and chasing), lying behaviour, body condition score (BCS) and lesions. The assessor also performs a Qualitative Behaviour Assessment (QBA), which aims to reflect how animals behave and interact with each other. See Paper 1 or Welfare Quality (2009) for a detailed description. The number of water cups or troughs is noted, as well as the water flow and cleanliness of the water sources. The WQ protocol includes a ´management questionnaire´ that provides details such as number of cows, heifers and dry cows, breed, grazing period, outdoor access and information on disbudding/dehorning and tail docking. Additionally, information is collected on the number of cows suffering from dystocia, downer cows and the mortality rate.

A three-scaled BCS is used in WQ: normal, very lean and very fat (Welfare Quality, 2009). For cleanliness, only one side of the cow´s body is checked, followed by the whole hindquarters from behind. See Paper 1 or Welfare Quality (2009) for a detailed description. Welfare Quality assesses lying time, the time it takes for the cow to lie down from standing until she bends over and lowers her knee, her hindquarters rest on the ground and her front legs are extended. At least six cows are checked for lying time (Welfare Quality, 2009). Welfare Quality assesses only adult dairy cows, and the assessment duration varies between four to eight hours (Blokhuis *et al.*, 2010), depending on the herd size. The Welfare Quality protocol is mainly used for research purposes but has also been used in, for example, Better Training for Safer Food courses, funded by the Commission, which assess compliance with EU legislation.

1.2.3 Ask the Cow

In Sweden, 74% of the dairy farms are affiliated with Växa´s milk recording scheme (Växa, 2025). In addition to the milk recording scheme, Växa Sweden also offers an animal welfare assessment. The Växa protocol ´Ask the Cow´ mainly contains animal-based parameters. A random sample of animals (35 cows, 35 young stock and 35 calves) is systematically observed and assessed. For cows, the information is recorded at the group level, i.e., how many cows are lying down or standing up in the stall and if there is competition for feed and access to water. Most cows should lie in the stall and ruminate for good animal welfare. At the individual cow level, rising behaviour, BCS, cleanliness, hoof appearance, lesions and lameness are recorded. For young animals and calves, BCS, cleanliness, lesions and health are recorded (Supplementary Material S2 in Staaf Larsson *et al*., 2024).

The result is compared with the data of other herds and presented graphically with an image of two daisy flowers, one for the cows and one for the young animals and calves. The results are shown as whole petals if the results are good, but improvements are needed if the flowers have cut petals. Sandgren *et al*. (2009) used nine animal-based measurements: cleanliness and BCS of cows, young animals and calves, in combination with lameness, injuries/inflammations, and rising behaviour. The protocol has been regularly used in farm assessments by Växa and was also used in a study by Jeppsson *et al.* (2024). After the assessment, the AC assessor provides the farmer with an action plan with suggestions for improving animal welfare, followed by a second visit. Ask the Cow mainly uses animal-based measures, and a farm visit takes approximately 3-4 hours.

1.2.4 Other farm assurance schemes

Animal welfare assessment may have several purposes, such as food labelling, classification, certification, and control of legislation compliance or for advisory purposes. Many certification schemes evaluate animal welfare (e.g., Assured Dairy Farms, Defra, Freedom Foods, RSPCA and Soil Association). A certification scheme would include the aim of welfare improvement in addition to demonstrating compliance with the relevant standard, and following the principles plan, do, check, and improve (Main *et al*., 2014). The standards should at least include the relevant legislation and codes of practice. Main *et al*. (2014) enlighten that a scheme should stimulate interest among the farmers and propose proactive discussions about the commercial value with the on-farm welfare assessment. Programs that can provide tangible improvements in welfare, should have a coordinated approach to resource and outcome assessment (Main *et al*., 2014). Most welfare science has focused on the welfare harms associated with not providing specific resources rather than the welfare benefit of providing a resource (Main *et al*., 2014).

In Austria and Germany, the Animal Needs Index, ANI, was developed, as a pragmatic system based on a consensus of people from different perspectives to make it more accepted (Bartussek, 1999). Inspections are likely to have a positive effect on animal welfare (Lomellini-Dereclenne *et al*., 2017), and increased knowledge of the legal requirements among farmers will stimulate improvements (Anneberg *et al*., 2013; Lomellini-Dereclenne *et al*., 2017). In France, inspectors use 32 points to verify compliance with animal welfare legislation (Lomellini-Dereclenne *et al*., 2017). The French study found that eight of these points significantly impact the animal welfare level on farms, with the most significant risks being when farm records do not comply with legal requirements, lack of basic care, including inadequate feeding, and lack of appropriate farmer skills. Lomellini-Dereclenne *et al*. (2017) concluded, after reviewing 11,346 reports from 9,327 farms that farms did not keep records, that they were approximately four times more likely to be seriously non-compliant with the regulations. The researchers suggested improvements to refine the checklist, harmonise interpretations, increase farmers´ knowledge of legal requirements, and develop improvement plans (Lomellini-Dereclenne *et al*., 2017).

1.2.5 Importance of measuring body condition

Body condition scoring is an important animal welfare measure, and farms with poor animal welfare often have lean animals. Complaints about the BCS and skin soiling are common animal welfare issues in cattle husbandry (Keeling, 2009). Extreme thinness, as well as obesity, increase the risk of diseases like milk fever, retained placenta, endometritis, ketosis, abomasal displacement and dystocia in cows (Gillund *et al*., 2001; Roche & Berry, 2006). Thinness has been associated with low milk production (Roche *et al*., 2007), low conception rate (Pryce *et al*., 2001) and an increased risk for sole ulcer and white line disease (Green *et al*., 2014). Green *et al*. (2014) showed that cows with a BCS < 2.5 (on a scale from 1 = thin to 5 = fat) are more likely to become lame. Over-conditioning may also cause reduced milk yield (Gillund *et al*., 2001).

1.2.6 Importance of measuring cleanliness

There are several reasons for assessing skin soiling in farm animals. The most prevalent recorded non-compliance at official Swedish animal welfare controls is soiled animals (Keeling, 2009; Lundmark Hedman *et al*., 2021).

Improved cattle cleanliness has many benefits, such as strengthened food safety, reduced mastitis incidence (Hughes, 2001), and improved animal comfort. In addition, the fact that intact hides at slaughter increase profit (Nafstad, 1999) is an effective driver towards enhanced animal welfare. Etching of the skin, infections and irritation can be complications of chronic faecal soiling (Nafstad, 1999). Soiled cows are more likely to get mastitis (Breen *et al*., 2009). Removing dry lumps of manure from the skin is painful, faecal soiling causes skin lesions (Hauge *et al*., 2012), and damaged hides can be an economic setback for the farmer after slaughter (Nafstad, 1999). Soiled animals can also threaten food safety (Hughes, 2001). Several risk factors for soiling are related to building design, management and stockmanship (Radeski *et al*., 2015). Hughes (2001) introduced a cleanliness scoring system, evaluating four separate skin areas on the cow to indicate the cause of soiling.

The consistency of the manure strongly influences the risk of soiling, i.e. high-yielding cows have a muddier manure than dry cows and, therefore, become soiled more easily. Ward *et al*. (2002) concluded that since cows in early lactation produce large amounts of loose faeces it is difficult to maintain a satisfactorily clean surface.

1.2.7 Importance of thermal comfort

The Swedish Board of Agriculture´s regulations and general advice (SJVFS 2019:18) on cattle husbandry etc. state that ´In barns, animals should have a climate adapted to the type of animal and how it is kept (thermal comfort)´. However, there is no advice on assessing the animals´ thermal comfort or body temperature (Swedish Board of Agriculture, 2024b). In the checklist, the statement ´The barn has a climate and air quality that is adapted to the species and the form of animal husbandry´ is to be answered with ´Yes´, ´No´, ´Not checked´ or ´Not applicable´ (Swedish Board of Agriculture, 2024a).

The thermoneutral zone is the environmental temperature range where animals do not need to perform active strategies to maintain normal body temperature and is defined by lower (the animal shivers) and upper (the animal sweats and/or pants) critical ambient temperatures (Sjaastad *et al*., 2016). Within the thermoneutral zone, the animal regulates body temperature by shifting blood flow to/from the skin, which causes alterations in skin temperature. Changes in skin surface temperature accordingly reflect changes in skin blood flow in response to alterations in environmental temperature (Scoley *et al*., 2019). According to Armstrong (1994), also in countries with moderate climates, dairy cows are exposed to periods of heat stress, leading to reduced feed intake, increased water intake, changed metabolic rate and maintenance requirements, increased evaporated water loss, increased respiration rate, changed blood hormone concentration, and increased body temperature.

Skin temperature has the potential to be an indicator of whether an animal is at the borders of its thermoneutral zone, i.e. near the initiation of active thermoregulation like shivering or panting/sweating. Several options are available for measuring skin temperature on farm animals (Nogami *et al*., 2014; Scoley *et al*., 2019; Furukawa *et al*., 2024). Measuring rectal temperature is the gold standard for assessment of body temperature and fever in animals (Sun *et al*., 2021). Both farmers and animal welfare inspectors require quick, reliable and cost-effective methods. Measuring rectal temperature is time-consuming and requires physical contact with the animal. It is, accordingly, not feasible for farmers or animal welfare inspectors to perform at the herd level. Therefore, there is a need for simpler and less invasive methods for the accurate determination of body temperature in dairy cows under farm conditions.

1.2.8 Importance of adequate water intake

Freedom from thirst is commonly included in welfare assessments, and the first statement in the Five Freedoms is ´Freedom from hunger and thirst´ (FAWC, 1993). Mellor and Reid (1994) later developed the ´Five Domains´ including the first Domain, ´Thirst/hunger/malnutrition´, evaluated in a five-grade scale from Grade O ´Water/fluid is available in quantities which satisfy thirst´ to Grade X ´Water/fluid or food restrictions or excesses where the predicted endpoint is death´.

According to the Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes, animals shall be sufficiently fed and watered. The animal welfare inspector controls that feeding and water systems are designed, dimensioned, and positioned so that the system allows the cow to have a peaceful and natural intake of food and water. The inspectors also control that the requirements for access to water and water quality are met. Currently, there are no animal-based measures to assess the hydration status of dairy cows. No animal-based parameter is used in any welfare assessment protocol to ensure that animals receive sufficient water. If the water supply is assessed at all in welfare protocols, checking is done using resource-based parameters, such as the number of water bowls and sometimes their cleanliness and water flow capacity. Checking can indicate water availability at the herd level, but there is no information on whether the herd or individual cows are suffering from thirst.

The osmolality of body fluids, such as blood or milk, indicates whether an animal is suffering from insufficient water intake, leading to dehydration or too much water in the body, called hyper hydration. The animals´ thirst is stimulated via osmoreceptors in the brain that sense when body fluids´ osmolality/osmolarity (= solute concentration) increases, which is associated with dehydration (Sjaastad *et al*., 2016). Water deprivation and the intake of hypertonic fluids increase the osmolality of the extracellular body fluids. High osmolality also stimulates osmoreceptors in the brain (third ventricle) to release antidiuretic hormone (ADH/vasopressin), which, through an increase in aquaporins, reabsorbs water in the kidneys and concentrates the urine.

Dairy cows consume up to approximately 100 litres of water daily to maintain milk production (Murphy, 1992; Jensen & Vestergaard, 2021). In lactating animals, there is continuous water loss from the body through urine, faeces, milk and evaporation from the skin and respiratory tract. Water sources include drinking water, feed and metabolic water. Important factors that influence water consumption in dairy cows are diet composition, dry matter intake, milk production, ambient temperature and humidity (Murphy, 1992), and thus the temperature-humidity index (THI) (McDonald *et al*., 2020). At conditions above the thermoneutral zone, animals lose water through active evaporation (sweating and panting) to regulate body temperature, which leads to even greater fluid losses. The need for drinking water in dairy goats increases by approximately 100% when the ambient temperature increases from approximately 20°C to >30°C (Olsson & Dahlborn, 1989). Murphy *et al*. (1983) measured water intake in dairy cows compared to the lowest and highest daily temperatures. They found an increase in water intake of 1.2 kg per degree (°C) at the lowest daily temperature. With climate change, water availability is becoming an increasing concern and access to water may become an increasing problem.

Other factors affecting drinking behaviour are eating patterns, whether water is offered in a trough or water bowls, flow rates in water bowls, water temperature and, when water bowls are shared, animal dominance (Murphy, 1992). In a study on Holstein cows provided water ad lib from two water bins (of 40 L) per 20 cows, the cows visited the drinker 19 times/day on average and spent 81 seconds per visit at the drinker (McDonald *et al*., 2020). They showed that the behaviour around the drinker was affected by an increase in THI, i.e., the higher the THI, the more competitive behaviour ($R^2 = 0.75$, $P = 0.0004$). Their study indicates that the watering and housing system is important for the drinking behaviour of dairy cows and that cows may drink both more or less than they need in such housing systems. Cows may drink less often from troughs than from bowls and can drink at a rate of 15 kg/min (Castle & Thomas, 1975), and the water intake is affected by the size and height of the water trough (Pinheiro Machado Filho *et al*., 2004).

Milk osmolality follows the osmolality in blood plasma (Linzell & Peaker, 1971; Olsson & Dahlborn, 1989), and milk osmolality reflects the individual animals´ level of dehydration (Shipe, 1959). The degree of dehydration in individual cows can thus be assessed, i.e., the degree to which a cow ingests enough water for their milk and other bodily functions and to avoid thirst. Since blood sampling is too invasive to use in official control, milk sampling is a better option and could also be used to control the water supply regularly. Milk osmolality is routinely measured at the farm level, at the dairy when milk is delivered, by determining the freezing point of the milk (Büttel *et al*., 2008). In Sweden, milk osmolality is measured at herd level by taking tank samples to check, using freezing point depression, that the milk is not diluted with water. In Sweden, 74% of dairy farms are affiliated with the Swedish monitoring program ´Cow control´ (Växa, 2025), including monthly milk sample collections. Analyses of milk osmolality can be used to see if the water supply is sufficient in dairy cows (Shipe, 1959). The usual method for measuring osmolality is with the freezing point methodology, which states that the higher the content of solutes, the lower the freezing point. The freezing point increases with higher water content in the milk (Bjerg *et al*., 2005), while at the same time, the osmolality values decrease.

2. Aims of the thesis

This thesis aimed to study whether including more animal-based parameters will improve the official animal welfare control for dairy cows. More specific aims were to compare the outcome of different animal welfare protocols in the same herds and the effect of assessors´ background, as well as to investigate the potential of a few new animal-based objective assessments not currently included in any protocol.

The specific aims and hypothesis of the thesis were:

- Paper I – The aim was to compare how 41 dairy farms were ranked using the three different animal welfare protocols with the starting hypothesis that the best and worst farms in OC would also be identified as the best and worst in AC and WQ, respectively.
- Paper II – The aim was to investigate how different categories of professionals and students assess cleanliness and the BCS. A second aim was to estimate the interrater agreement for these assessments. We hypothesised that previous animal related education and professional experience with animal welfare result in a more uniform and consistent assessment, compared to no such education and less experience, and that long professional animal welfare practice improves interrater agreement.
- Paper III – The aim was to evaluate the feasibility of using two handheld skin temperature sensors during field conditions and the correlation of the values obtained with indoor temperature and the gold standard of body temperature, i.e., rectal temperature. The hypothesis was that these measurements have potential as future tools for farmers and inspectors in official animal welfare controls and that there might be breed differences.
- Paper IV – The aim was to describe variations in milk osmolality within and between days, within individual cows, and over season and lactation stages as well as the correlation to ambient temperature and relative humidity. In addition, the collection procedure and individual cows´ deviation from the tank milk osmolality were evaluated. The hypothesis was that milk osmolality has potential as a future tool for farmers and inspectors in official animal welfare controls.

3. Material and methods

In this section, a summary of methods and experimental procedures is given. Detailed descriptions can be found in the corresponding sections within each paper (I-IV). For correlations and comparisons, P-values < 0.05 were considered significant. Results are presented as mean ± SD if not stated otherwise.

3.1 Paper I - Comparison of assessment protocols

In this study, 41 dairy farms were ranked using three different protocols: 1) the Swedish official animal welfare control (OC; Swedish Board of Agriculture), 2) Ask the Cow (AC; Växa Sverige, the Swedish Dairy Association), and 3) Welfare Quality (WQ; Welfare Quality Network) protocols for dairy cows. Twenty-four persons performed the animal welfare assessments altogether, 10 animal welfare inspectors from CAB carried out the OC (each 1-10 controls) as ordinary state control, six AC-trained persons from the Swedish Dairy Association conducted the AC assessments (each 2-11 assessments), and eight WQ-trained animal welfare inspectors from CAB the WQ assessments (each 5-11 assessments). In each herd, all three types of assessments (OC, AC and WQ) were performed simultaneously on the same day. The collected data comprised 44 parameters from the OC, 51 from AC, and 59 from WQ.

3.1.1 Data processing, ranking and statistical analyses

The AC ranking was based on the result of 12 animal-based parameters (proportion of lean, fat, dirty, severely dirty and lame cows, proportion of cows with long hooves, asymmetric hooves, lesions, severe lesions, rising problems, and proportion of cows lying outside the cubicles and standing in the cubicles). In WQ, the overall ranking was based on the sum of scores for the four welfare principles: good feeding, good housing, good health and appropriate behaviour. In OC, the number of negative remarks was the determining factor, i.e., fewer complaints resulting in a lower (better) ranking.

The proportions of lean and fat cows were compared using a t-test. The proportions of dirty cows and lesions were not analysed statistically since the registration methods differed. Correlation analysis was performed between remarks on OC and rankings in WQ and AC protocols and between remarks/rankings and NMR data (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA) using the PROC CORR function and regression equations were created in Excel (Microsoft Excel, Microsoft Corp., Redmond, Washington, USA). Rankings of tie-stalls and loose-housing systems according to the different protocols were compared using a Wilcoxon two-sample test (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA).

Principal component analysis (PCA) was performed in SIMCA 17 (Sartorius Stedim Data Analytics AB, Umeå, Sweden) as a multivariate complement to the statistical analyses.

3.2 Paper II - Visual assessment of body condition and skin soiling

The aim was to study differences in the assessment of BCS and cleanliness in cattle depending on the assessors´ profession, education and professional experience. During 15 sessions, 569 individuals evaluated BCS and cleanliness and recommended actions on photos of dirty cattle. The assessors were animal welfare inspectors, veterinarians/advisers, animal welfare researchers, students or were categorised as other professions. They were also asked about their amount of animal-related education and whether their practical experience in the profession was more or less than three years.

3.2.1 Statistical analyses

Analyses were made in Stata IC, version 15 (StataCorp, College Station, Texas, USA). To facilitate comparisons between slides, all recorded scores were transformed to standardised scores on a 3-level ordinal scale -1–0–1, calculated as the deviation of each value from the overall sample mode for the slide, as −1 for original scores below the mode, 0 for original scores equal to the mode and 1 for original scores above the mode. The standardised scores were then modelled statistically using generalised ordinal logistic regression. The participants were arranged in clusters identified by profession categories. The ratings were expected to be independent between profession clusters but not necessarily within them, which affected the standard errors of coefficients. For each profession category and assessment domain, the agreement between participants was estimated by observed percent agreement (joint probability of agreement), a generalised kappa statistic adapted to ordinal data, multiple observers and incomplete designs (Abraira & Pérez de Vargas, 1999), and Gwet´s Agreement $Coefficient_1$, treating slides as subjects and using the original scores. Associated 95% confidence intervals were calculated for kappa and Gwet´s Agreement $Coefficient_1$.

Figure 2. One of the photos for evaluating BCS.

3.3 Paper III - Methods for assessing skin temperature

In Paper III, 21 cows from the breeds SH and SRB at Lövsta Research Centre, SLU, Uppsala, Sweden, were included. The feasibility of two handheld skin temperature sensors was evaluated, and the correlation of the values obtained with indoor temperature and the gold standard of body temperature, i.e., rectal temperature. Potential differences in rectal and skin temperature between the breeds were also studied. Temperature measurements were conducted for four days in February and then monthly during the remaining study.

Skin measurements were made at three places: 10 cm below the hip, at the vulva and the neck, using an infrared thermometer (IRT) and a medical thermistor thermometer (MTT, conduction) at the hip and the neck.

Figure 3. Medical thermistor thermometer, infrared thermometer and rectal thermometer that were used in Paper III. Photo: Birgitta Staaf Larsson

3.3.1 Statistical analyses

The data were analysed with a MIXED model (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA) with temperature as the dependent variable (y), date and breed as fixed factors and individual as random factors. Correlations were analysed using Pearson correlation analysis (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA).

3.4 Paper IV - Milk osmolality to assess hydration status

In Paper IV, milk samples from 21 cows of two breeds (SH and SRB) in SLU´s research barn were collected (n = 941), and osmolality was analysed. Milk samples were collected monthly during one lactation, and daily fluctuations in osmolality were studied for four days at the beginning of lactation. The cows were housed in a loose-housing system and milked twice daily in an automatic milking rotary, developed by DeLaval (Tumba, Sweden).

Milk samples were taken by hand milking from the 20 dairy cows every two hours for four days in February from 05:00 h in the morning to 17:00 h in the evening; the first and the last samples were collected during regular milking in the automatic milking rotary. After that, monthly samples were collected at the time of regular milking for one year. The time when the samples were collected was registered for each cow. Bulk samples from the milk tank were also analysed to compare osmolality at the herd level.

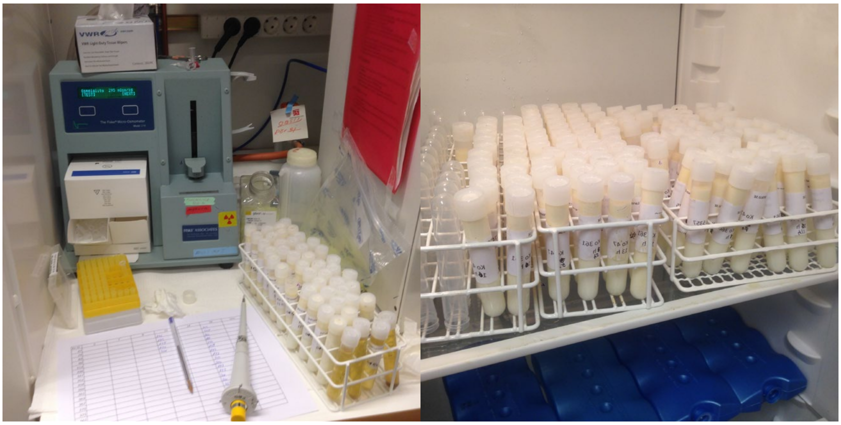

Figure 4. Milk osmolality was analysed using a Fiske® 210 Micro-Sample Osmometer (Fiske® Associates, Norwood, Massachusetts, USA). Photo: Birgitta Staaf Larsson

3.4.1 Statistical analysis

The data on variations during the day and year were analysed using the MIXED procedure in SAS. Milk osmolality was the dependent variable, and breed was included as a fixed effect. Repeated measures within cows were specified by the mean time of the sample for the daily fluctuations, respectively days (12 different dates) for the monthly variations and cow ID as the subject. A spatial power structure was used for the model´s error term, allowing the observed time points to be at different distances. Values were compared pairwise using Tukey´s adjustment for multiple comparisons. The assumptions of the model were checked by visual inspection of residual plots. Paired t-tests were performed to compare the individual monthly osmolality values with the osmolality of the tank milk. Correlations were then analysed using Pearson correlation analysis (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA). P values <0.05 were considered significant. Descriptive results were presented as individual values, means ± SD and range. The scatter plots and regression equations were made in Excel (Microsoft Excel, Microsoft Corp., Redmond, Washington, USA).

4. Results

In this chapter, the most important results are presented. More details can be found in each paper.

4.1 Paper I

4.1.1 Ranking of the herds within the various systems

The OC was carried out based on, in total 1,763 inspection points in the checklist. Of these, the 41 herds received a total of only 55 remarks. The number of remarks varied from zero in 18 herds to five remarks in four, with a mean of 1 ± 2. The most common remark (n = 11) was the lack of a backup system to guarantee sufficient air regeneration in a mechanical ventilated barn and/or lack of alarm equipment in case of a ventilation breakdown. The second most common remark (in eight barns) was that the regulation on calves being kept in single compartments was not met. These complaints could either be that the calves´ stall did not comply with the legislation or that calves were kept in a single stall after eight weeks of age, which is not permitted by the legislation. In AC, there were 624 remarks out of 1,876 possible (33%). In WQ, there were 680 remarks out of 1,083 (63%) possible, counting the measurements considering good – not good, e.g., clean or not clean. In general, tie-stalls received a poor ranking compared to loose-housing in the WQ system.

The correlation coefficient (r) in rankings between OC and AC was 0.40 (P = 0.009), but no correlation was found between OC and WQ, nor AC compared to WQ (Figure 5).

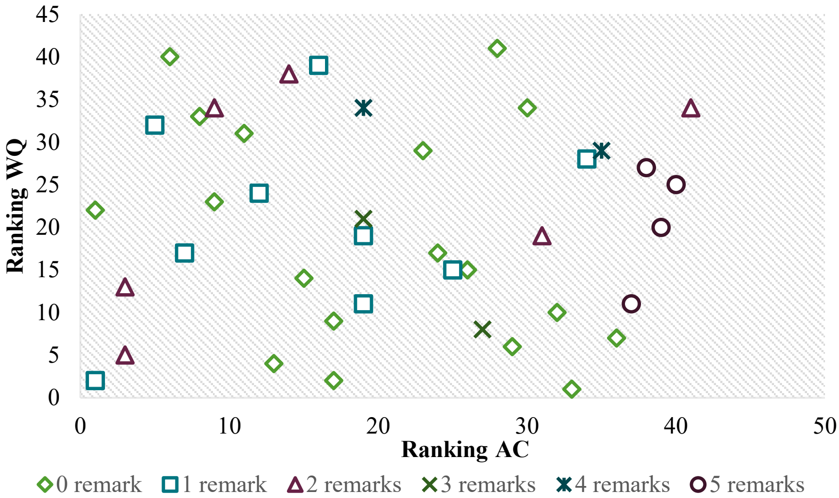

Figure 5. Ranking in Welfare Quality (WQ) compared to the rank in Ask the Cow (AC). Each farm is displayed as a dot, with different symbols according to the number of remarks in the official control.

There were no significant correlations between animal welfare ranking in any assessment systems and the number of cows, kg ECM, calf mortality, mastitis or percentage of heifers with more than 70 days between calving and first insemination. However, the percentage of cows that failed to become pregnant within 120 days from calving was correlated (r = 0.36, P = 0.02) with the outcome of the OC checklist.

A PCA plot of all parameters from the three protocols and the number of cows per herd from the NMR revealed parameters of importance for the individual loadings and overall outcome (Figure 6). Days on pasture strongly influenced the results of the PCA analysis, with ´no lame cows´ as the closest parameter. The 11 terms for positive state in QBA in WQ (active, relaxed, calm, content, friendly, playful, positively occupied, lively, inquisitive, sociable and happy) clustered in the lower right quadrant of the plot, together with ´touch´ from the avoidance distance test and number of cows per herd (Figure 6). The four terms for the negative state in the QBA in WQ (bored, uneasy, indifferent and frustrated) clustered in the upper right quadrant, together with ´dirtiness´ (dirty udder, upper leg, lower leg), ´the assessor can approach to within 50 cm but cannot touch the animal´ and the AC parameters ´asymmetrical hooves´, ´dirty cows´ and ´lesions´. ´Hairless patches´ and ´no lesions´ were close to each other in the lower part of the upper right quadrant (Figure 6). The WQ parameters ´severely lame´ and ´moderately lame´, ´swellings´ and ´number of water bowls´ were close to the centre of the PCA plot (Figure 6).

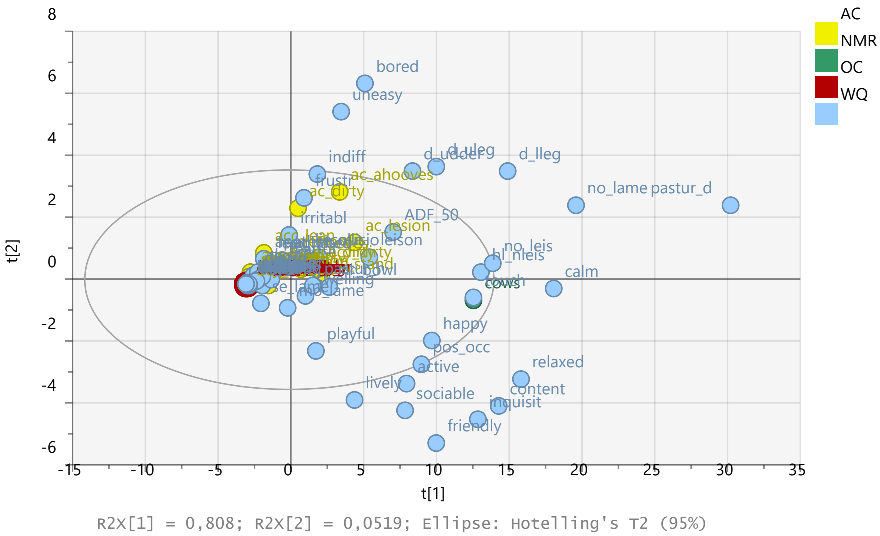

Figure 6. Principal component analysis score plot of parameters assessed in the official oontrol (OC, red dots), Ask the Cow (AC, yellow dots) and Welfare Quality (WQ, blue dots) protocols, and number of cows per herd (green dots) according to the National Milk Recording (NMR) database. The ellipse shows the 95% confidence interval. Parameters loading close to the origin were less important, while parameters loading further away contributed more to the model. Autofit generated three components [R2X(cum) = 0.889, Q2X(cum) = 0.842], of which the score plot illustrates the first two [R2X(cum) = 0.860, Q2X(cum) = 0.827]. pos_occ = positively occupied, inquisit = inquisitive, indiff = indifferent, frustr = frustrated, irritabl = irritable, hl_nleis = hairless patches, no_leis = no lesions, no_lame = no lame cows, pastur_d = number of days at pasture, d_lleg = dirty lower leg, d_uleg = dirty upper leg, d_udder = dirty udder, ADF_50 = the assessor can approach closer than 50 cm but cannot touch the animal, ac_dirty = dirty cows, ac_ahooves = asymmetric hooves, ac_lesion = lesions, se_lame = severe lame, mo_lame = moderate lame, nr_bowl = number of water bowls.

4.1.2 Assessment of body condition

All the herds had animals with acceptable BCS, according to the OC. The total proportion of lean cows in AC was 3 ± 4% and in WQ 7 ± 10%. The amounts of fat cows were 6 ± 8% for AC and 4 ± 5% for WQ assessments, respectively.

4.1.3 Assessment of cleanliness

Two herds had remarks on the cleanliness of the cows in the OC. The AC assessment of the cleanliness showed that, on average, 18 ± 25% of the animals were dirty, and 31 out of 41 herds had dirty cows. Almost all the cows (97%) were dirty in two herds, and seven farms had severely dirty cows (mean value 1 ± 4% of the animals on all farms). According to the AC assessment, ten herds (24%) had no dirty animals at all.

In the WQ assessments, all herds had cows with dirty legs (mean 68 ± 26% of animals) and flanks (mean 50 ± 26% of animals), and on average, 43 ± 26% of cows also had dirty udders. None of the cows had dirty udders in one herd and were generally clean on both legs and flanks.

4.1.4 Assessment of skin lesions

The outcome of the assessment of skin lesions was similar in AC and WQ in the poorest-ranked herds but in less agreement for the best-ranked herds.

4.1.5 The WQ assessment regarding the four principles

The results of the four principles were calculated and weighed due to their importance (Welfare Quality, 2009) in allocating the herd to an animal welfare category. The herds were categorised as acceptable (11/41) or better than average (27/41). The three remaining herds did not receive a final categorisation due to missing values. The overall main complaints focused on feeding. Referred herds were mainly those within tie-stalls where cows only had access to one water source, or herds with lean animals. Regarding animal behaviour, most comments concerned herds with tethered animals without access to daily exercise and too short grazing periods.

4.2 Paper II

4.2.1 Results

The more extensive animal-related education, the less likely the animal welfare inspectors were to give higher BCS. There were no significant differences in skin soiling scores between education levels or professional experience. On the other hand, a significant effect ($P \leq 0.05$) of profession was shown. Students had a higher overall probability of rating the degree of soiling higher than animal welfare inspectors and veterinarians/advisers, regardless of professional experience. Students recommended stricter control actions than welfare inspectors and veterinarians/advisers did, and veterinarians/advisers recommended less strict control actions than animal welfare inspectors.

Agreement between participants within profession categories was fair to moderate, with kappa values between 0.29 and 0.37, and Gwet´s Agreement Coefficient$_1$ between 0.31 and 0.48. Participants with at least 3 years of professional experience had slightly better agreement for BCS than less experienced assessors, but the difference was not conclusive for soiling and control actions. The percent agreement between Swedish animal welfare inspectors was 0.56, 0.58 and 0.50 for BCS, skin soiling and recommended action respectively, and for veterinarians the agreement was 0.57, 0.58 and 0.53 respectively.

4.3 Paper III

4.3.1 Skin temperature measurements

It took about 2 seconds per cow to measure with the IR device (IRT) and more than one minute per cow with the conduction technique (MTT). Skin temperatures did not correlate well with rectal temperature (IRThip: $r = 0.15$, $P = 0.03$; IRTvulva: $r = 0.17$, $P = 0.02$; IRTneck: $r = 0.15$, $P = 0.04$; MTThip: $r = 0.15$, $P = 0.04$; MTTneck: $r = 0.18$, $P = 0.02$) but correlated strongly ($P<0.0001$) with indoor temperature (IRThip: $r = 0.7$, IRTvulva: $r = 0.6$, IRTneck: $r = 0.7$, MTThip: $r = 0.6$, MTTneck: $r = 0.5$). There was a significant effect of the month on rectal temperature ($P = 0.02$), with the lowest mean values (38.3 ± 0.1°C) recorded in October and the highest (38.8 ± 0.1°C) in August. There was also a significant effect of month on skin temperature ($P < 0.0001$). There was no effect of breed on rectal temperature and generally no effect of breed on skin temperature, except from temperature recorded at the hip, which was higher for cows of the SH breed than for SRB cows when measured with IRT (29.7 ± 0.2 vs. 29.1 ± 0.2°C, respectively, $P = 0.02$), and lower when measured with MTT (34.7 ± 0.2 vs. 35.1 ± 0.1°C, respectively, $P = 0.02$). Neck temperature recorded by IR best reflected indoor temperature ($r = 0.73$, $P < 0.0001$), with no difference between breeds.

4.4 Paper IV

4.4.1 Variation in milk osmolality

The mean milk osmolality during the four days in February was 299 ± 5 mOsm/kg. Individual osmolality ranged from 279 to 317 mOsm/kg, and there were significant variations in milk osmolality between cows (the mean of all collected samples was 297-304 mOsm/kg, 2%). The osmolality varied significantly within days (14-26 mOsm/kg, 5%) and between months (10-38 mOsm/kg, 7%) for each cow. Swedish Red Breed cows had higher (1%, P=0.02) osmolality (301±1 mOsm/kg) than Swedish Holstein cows (299±1 mOsm/kg). The differences observed are of the magnitude 1-9%, with variations between breeds being the smallest and between months/lactation stages and individuals the largest.

5. Discussion

This thesis aimed to study if additional and more detailed animal-based measures can improve the official Swedish animal welfare control by including detailed descriptions of the level of, e.g., dirty, lean or fat animals. In addition, new measurements on thermal comfort and the hydration level in individual dairy cows were evaluated. Paper I, comparing three protocols, showed that much more information can be obtained if more detailed information on cleanliness and BCS is registered and by adding registrations of lesions and wounds as well as to evaluate lameness. The advantages of these measurements are that the level of lean or fat animals, cleanliness, lesions, wounds and lameness on different farms will be known. The information can also be shown at the national level and if there is a progression over time at the farm and national levels. Today, it is only registered if there are lean, fat or dirty animals, not how many nor to what extent. However, you always must consider the disadvantages, such as the more measures to be taken, the more time is needed to conduct the animal welfare control. Nevertheless, the advantage is that a lot more information is received.

Our hypothesis in Paper I was that the outcome, in terms of best and worst farms, would be similar by all three protocols OC, AC and WQ. However, there were substantial differences in the outcomes, even though there was some correlation between OC and AC. Since skin lesions, wounds and lameness are not measured in OC, the suggestion is to include registrations of those animal-based measurements in the OC checklist to improve the official control. In addition, registering the BCS, the level of dirtiness and the number of animals affected would further improve the OC protocol. However, the outcome of the animal-based assessments depends on the methods used and how well the assessors are calibrated. It is also important to use animal-based measures that are evaluated to be feasible and reliable.

In Paper II, it turned out that individuals with similar backgrounds agreed quite well in their assessments, but individuals with different professions and experiences were less harmonised. In addition, students scored cattle dirtier than other professions. It can be discussed whether this is due to increased interest in animal welfare among younger individuals, a lack of familiarity with animal husbandry, and thus less tolerance to malpractices, or whether it is due to other reasons. To illuminate this, solid education and calibration courses are needed. Today, the continuous education is mainly done by the CABs, although I believe that SLU would be the best to give such education. As explained in Lundmark Hedman *et al*. (2025), continuous education is also given at the European level, funded by the EU Commission, as Better Training for Safer Food (BTSF) courses, as well as that the European Centres for Animal Welfare (EURCAWs) provides guidance on how to conduct the inspections. The established four EURCAWs are responsible for pigs, poultry and other small farm animals, ruminants & equines and aquatic animals, respectively. I have been involved, both in BTSF courses (regarding pigs) and in the EURCAW-*Ruminants & Equines* (Brunet *et al*. 2024). Having a common understanding and calibration is important to make a uniform assessment throughout the EU. Participants´ knowledge has improved through these initiatives and investments, and the consistency within and between the EU member states will hopefully improve.

Further research is needed to develop best practices for training inspectors in animal welfare assessment, e.g., determination of a certain evaluated point where the fat content is easy to measure, to be checked in combination with ultrasonic measurement, as measured in Schröder and Staufenbiel (2006). Another option is to use digital camera systems for continuous and automated monitoring of BCS, which was found to be a more accurate assessment than manual BCS observations (Bell *et al*., 2018)

In the third study (Paper III), the skin temperature measured with the IR sensor correlated to the ambient conditions. It could thus be a valuable tool for measuring thermal stress in individuals. The IR sensor was quick to record and had limited interaction with the cow and could, therefore, be interesting for future studies investigating animals outside the thermoneutral zone.

Study IV concluded that variations in milk osmolality between individuals, days and breeds exceed what can be interpreted as thirst (+1-2%). Therefore, knowledge about the individual level is needed for milk osmolality to be of practical use. Sampling time may be a limitation during welfare inspections. However, if osmolality could be regularly monitored through monthly samples collected by the Cow Control or, even better, by the milking equipment, this would benefit the animals´ welfare and the work of the inspector and the farmer, but more research is needed.

5.1 The Swedish official control and comparison with other animal welfare assessment schemes

Paper I showed that the ranking of animal welfare at the herd level in our study differed substantially between the systems. This was not unexpected since different parameters are used in the various systems. However, our study hypothesised that the same farms would be ranked best and worst, respectively, irrespective of protocol. Nevertheless, there was an apparent mismatch in the rankings, although two of the protocols showed a moderate correlation. The mismatch was partly due to differences in which parameters were used, but also the quality of the assessments, and likely also their accuracy. A standard animal welfare protocol needs to be developed to assess welfare and guide farmers. Sandgren *et al*. (2009) found in their study that two cow fertility measures (calving difficulties and the proportion of heifers with more than 70 days between calving and first insemination) and calf mortality in the NMR corresponded with 77% of the herds′ animal welfare rating. Allendorf and Wettemann (2015) showed in a study on 115 dairy farms in western Germany that cow losses, replacement rate and calving interval corresponded negatively with dairy productivity. In our study, we found a positive correlation between the number of remarks in OC and the percentage of cows with 120 days between calving and pregnancy, i.e., the more remarks, the more cows with longer intervals from calving to pregnancy. If the percentage of cows with 120 days between calving and pregnancy is shown to be a valid indicator for compromised animal welfare, this could be used as a risk-based indicator for more frequent animal welfare control, but more research is needed.

Even though on some occasions when AC and WQ almost had the same assessment criteria, the outcome of animal welfare at herd level could differ. The reason for this is likely that the assessment of animals will differ in a herd since the selected animals will probably not be the same. The differences in the outcome may partly be due to the dissimilar systems, with 12-35 assessed cows in AC and 12-73 in WQ, respectively, and different assessors evaluated the cows differently. A smaller percentage of the herd will be evaluated in cases of larger herd sizes. Therefore, the risk is that the evaluation will be less representative. On the other hand, it is too time-consuming to evaluate a whole herd when there are too many animals. Accurately measuring all animals when there are many can take more than a day, which is likely not an efficient use of the CABs’ resources.

Nevertheless, to solve this problem, one option could be to focus on the animals in the worst condition, although the herd´s average will not be evaluated.

When assessing the cleanliness or skin lesions and swellings, the entire body is checked in AC, but only one side of the body is assessed in WQ. In the OC system, only 55 remarks were made even though AC and WQ, assessing the same herds, showed high numbers of lean cows, cows with skin lesions and dirty cows. Skin lesions, wounds and lameness are not registered in the OC checklist, where the only question is whether sick or injured cattle are given the necessary care and a separate compartment if needed. The results show few remarks concerning BCS and cleanliness when using OC, while AC and WQ performed at the same time identify numerous animals with remarks.

There was a weak to moderate positive correlation ($r = 0.40$, $P = 0.009$) between the OC checklist and AC outcome, and the herds with the most deficiencies generally coincide with those with the herds with the poorest animal welfare ranking in AC. According to the PCA analysis, behaviours, only measured in WQ, were the most important parameters to identify the animal welfare outcome. The animal-based measures used in programs like AC and WQ provide considerably more information about the animals´ well-being and welfare than the mainly resource-based measurements recorded today in the OC.

Except to some extent in AC, animal behaviour is only checked in WQ, and these measurements were the most important in identifying the outcome, according to the PCA analysis. The Swedish AWA (2018:1192), chapter 2, article 2, states that animals shall be kept and cared for in a suitable environment for animals and in such a way that their well-being is promoted, that they can perform behaviours they are highly motivated for, and behaviour disorders should be prevented. However, no behavioural assessment is in the OC checklist. To include the whole WQ QBA assessment may be too time consuming, but a few of those measures could be of value, e.g., evaluating if the cows are calm and relaxed. More research is needed to study which parameters are the most appropriate for checking natural behaviour. Another important issue for natural behaviour is the ability to move freely. Since 2010, building new barns for tied animals has been prohibited, but old barns are still allowed. In 2024, 36% of the herds, including 17% of the dairy cows in Sweden were held in tied-up systems (Växa, 2025).

All the OC observations were listed and compared with the sum of all observations by the OC in dairy herds in Sweden in 2011 (pers. com. A. Barchiesi, Swedish Board of Agriculture). Most remarks were associated with resource-based measures, and the most common was the lack of a backup system to guarantee sufficient air regeneration in a mechanically ventilated barn and/or the lack of alarm equipment in case of a ventilation breakdown in the calf barn. This resourced-based measurement was the most common complaint with 29% of all remarks in Sweden in 2011 and was also the most prominent OC remark (27%) in this study.

The outcome of animal-based assessments can depend highly on the observer, which animal variables that are included, and the methods used to assess the variables. In my opinion, behaviour assessment, for example, may easily be disturbed and influenced by the overall feeling of the farm acquired from the first impression. It is a challenge to conduct some of the measurements unbiased. All variables included in assessment protocols should be relevant and easy to perform to reduce observer influence or to use behaviours that easily could be answered ´Yes´ or ´No´, e.g., if allogrooming is observed. Allogrooming is perceived as a natural social behaviour (de Freslon *et al*., 2020). Using brushes can be another option for measuring (de Vries *et al*., 2007; Keeling *et al*., 2021). In Danish legislation, brushes for dairy cows are compulsory (Staaf Larsson, 2023). Paper II shows that veterinarians and animal welfare inspectors, who often visit farms, consider animals cleaner than, e.g., students. However, the proposed animal-based measures measuring the hydration status and thermal comfort, evaluated in this thesis, are more reliable and not dependent on the assessor since they are objectively measured. Luckily, in our study in Paper I, the inter-observer agreement test in WQ before the study showed a mean percentage agreement at 88 ± 17%.

The levels of BCS, cleanliness and skin lesions assessed in both AC and WQ were considered unacceptable on some farms, whereas in the OC, most of these animal welfare condition parameters were not even noted with a remark. The OC only assesses if sick cattle or cattle with skin lesions are given necessary care, if needed, in a separate stall. In the OC system, animal welfare assessment is done partly (to a low extent) with animal-based measures (e.g., evaluating cleanliness and BCS) without quantifying and without detailed documentation in the assessment protocol about the number of animals and levels of body condition. However, the control should cover every single animal. The result showed numerous remarks about the animals´ cleanliness according to WQ and AC, but only a few were condemned in the OC checklist. Therefore, animal-based measures ought to get more attention in the OC. The OC neglects much information about individual animal welfare by focusing on resource-based measures. In a report by Directorate-General (DG) SANTE from an audit carried out in Sweden in May 2019, the European Commission (2019) concluded that the OC results, reported in terms of compliance/non-compliance, are missing valuable information about the severity, extent, duration and recurrence of non-compliance. The European Commission (2022) investigated the use of animal welfare indicators in the EU Member States. One mission was to find suitable indicators to monitor whether the animal welfare conditions are improving, remaining stable, or worsening in the different Member States. The European Commission concluded that when animal welfare indicators are used, the control provides a better assessment of the actual welfare of animals, and most of the EU MS considered that indicators would be animal-based (European Commission, 2022).

The WQ system only assesses dairy cows (adult animals) compared to AC and the OC, which also assess calves and young stock. Corresponding assessments are not included in the WQ. Both the OC and AC systems check young stock and calves, which is important from an animal welfare point of view and because the legislation covers all animals.

All assessed herds were categorised as acceptable or better than average. As in the present study, farmers who voluntary participate will presumably not be the worst farms. Therefore, we can presume that the animal welfare level in these visited farms is, on average, better than the mean for all Swedish farms. However, all three systems studied could identify herds at high risk for poor animal welfare, but the herds were ranked differently within the OC, AC and WQ assessment systems.

5.1.1 Freedom from thirst

All three (OC, AC and WQ) assessment systems studied used a resource-based measure to assess the water supply. As explained earlier, in the OC, the demands on water supply and water quality are checked. In the other two systems (AC, WQ), the number of water cups or water troughs and the hygiene of the water source are recorded. Additionally, WQ measures the water flow. The WQ system gives tie-stalls a low rank for the principles of good nutrition because cows do not have access to more than one water source. Two cows share one water cup in herds with tied cows, and subsequently the WQ gives low parameter values. Otten *et al*. (2020) concluded that insufficient water supply was the main area of concern in their study. Heath *et al*. (2014a/b) found that the one resource-based parameter (absence from prolonged thirst) measured in WQ correctly classified 88% of the farms, compared to a whole WQ assessment, when studying 92 dairy farms in England and Wales. It is debatable whether the water parameter impacts the results too much, as this might be confounded by the animals being tied up. However, our study shows that the resource-based water parameter was not the best measure to distinguish the level of animal welfare. In the study by Heath *et al*. (2014a/b), no animal-based measure was used to check the actual hydration status of the animals.

In the Swedish checklist used in the official control, there are two questions regarding water, i.e., 1) if the requirements for access to water and the quality of the water are met and 2) if ´feeding and water systems are designed, dimensioned and placed so that they allow a calm and natural intake of feed and water´; both shall be answered by ´Yes´, ´No´ or ´Not controlled´. If the water supply is assessed in welfare protocols, this is done using resource-based parameters, e.g., the number of water bowls and, occasionally, their cleanliness and water flow capacity. The parameters might indicate water accessibility at the herd level, but there is no information regarding whether the herd or individual cows suffer from thirst.

5.1.2 Body condition scoring

Within the OC system, the inspector checks if the BCS is acceptable. Animal welfare inspectors are trained to use the 5-grade scale BCS, from very lean (1) to very fat (5), according to Wildman *et al*. (1982). However, there is just a general question about the BCS of all animals in the OC protocol. The WQ score the body condition on a 3-grade scale; normal, very lean and very fat (Welfare Quality, 2009). Compared to other scales (Wildman *et al*., 1982; Gillund *et al*., 1999), scoring from lean to fat, WQ scale the condition, 0 as normal, 1 as very lean and 2 as very fat. The WQ BCS scale confuses participants in WQ training courses compared to other BCS schemes (Wildman *et al*., 1982; Gillund *et al*., 1999). The other schemes estimate increases with rising BCS. In addition, if a mean value of the BCS will be calculated, the WQ system is not appropriate. However, aiming for a low value as possible (as in the WQ system), this approach is relevant.

In a study by Radeski *et al*. (2015) on 11 Macedonian dairy farms, the mean percentage of very lean cows was over 40%, in contrast with a study by Tremetsberger *et al*. (2019), where very lean cows seldom (3%) occurred. Our results were like the latter, with 3% lean cows registered in the AC system and 7% in WQ. According to Swedish statistics, the deficiencies in BCS have increased over the past eight years from approximately 8% of the farms to about 15% (Swedish Board of Agriculture, 2025). However, neither the number of affected cattle nor the extent of the deficiencies are recorded.

5.1.3 Cleanliness

The animal welfare inspectors at the OC are trained to use a 4-grade scale: 1) clean animal, 2) moderately dirty animal, 3) heavily dirty animal, and 4) very heavily dirty animal. However, as for BCS, there is just a general question about the cleanliness of all animals in the OC protocol. Despite a considerably large proportion of animals considered dirty in this study, by both AC and WQ, OC noted only two herds with dirty animals. The three assessment systems use different measurements, and the animal-based measurements are controlled in different ways, as well as the management-based and resource-based. Hauge *et al*. (2012) concluded that housing, feeding and management are essential for keeping animals clean. Preventive work has failed if dirty or soiled animals (Lundmark *et al*., 2016). Tremetsberger *et al*. (2015) found that udder health and cleanliness improved significantly when improving husbandry practices, such as improved udder cleaning routines.

In the yearly report from the Swedish Board of Agriculture (2025) about the outcome of animal welfare control, deficiencies in cleanliness increased from 24% during the years 2018-2023 to 32% last year. The reasons explained are increased prices and lack of good quality straw material. The deficiencies were mainly reported on small farms, especially those with tied animals. However, the statistics do not show how many animals were dirty or to what extent (Swedish Board of Agriculture, 2025).

5.1.4 Sickness and lesions

In the OC checklist, the only question concerning sick animals or animals with lesions is if the animals are given necessary care and have access to a separate compartment. The checklist lacks any items about sick animals or animals with lesions, as well as questions about how many animals are sick or injured. It is particularly pointed out in the AWA (2018:1192), chapter 2, article 1, that animals shall be treated well and protected from unnecessary suffering and disease. The OC checklist does not thoroughly meet legislators´ intentions. This notion is further strengthened by Webster (2012), who believes that the individual animal must be assessed to guarantee the individual animal's welfare.

5.1.5 Principle Component Analyses

The SIMCA analysis was performed to visualise which parameters within each system had the highest impact on the explained variability. The PCA plot of all parameters from the three protocols and the number of cows per herd from the NMR revealed parameters of importance for the individual loadings and overall outcome. Days on pasture strongly influenced the outcome of the PCA analysis, with ´no lame cows´ as the closest parameter. Previous studies have also found access to pasture to be associated with reduced incidence of lameness (Haskell *et al*., 2006; Chapinal *et al*., 2013; de Vries *et al*., 2015). The analysis showed that the behaviour parameters in WQ highly influenced the outcome in contrast to OC and AC, which do not measure behaviour. The 11 terms for positive state in QBA in WQ clustered in the lower right quadrant of the plot, together with ´touch´ from the avoidance distance test and number of cows per herd, which all seem to reflect contented and calm animals. On the other hand, the four terms for the negative state in the QBA in WQ clustered in the upper right quadrant, together with different WQ parameters on dirtiness, ´the assessor can approach to within 50 cm but cannot touch the animal´ and the AC parameters ´asymmetrical hooves´, ´dirty cows´ and ´lesions´. All these are negative parameters of animal welfare, clearly separated from the positive ones (Figure 6).

5.1.6 Lying down vs standing up

Observing how the cow stands up or lies down indicates how fit she is and if the cubicles are appropriately adapted to the cows. Ask the Cow, and WQ have chosen different approaches to assess how easily the cows change their positions between standing and lying. While AC measures the cows´ rising behaviour, WQ measures the cows´ lying down. The two ways of quantifying the behaviours do not make the two systems exactly comparable, but if the cow has trouble lying down, she likely has problems getting up as well.

Krohn and Munksgaard (1993) studied the time it takes for a cow to lie down. In contrast to our assessment of the lying down time, they measured the entire time it took to lie down from the initial examination. The actual movement took 10 seconds, and the time from the initial examination of the place to lie down until lying took 89 ± 6 (first lactation) to 127 ± 10 (third lactation) seconds until lying. Their study found that the duration of the lying down pattern was significantly longer in tie-stalls, and interruptions of the lying down movement occurred more frequently in tie-stalls than in the loose-housing systems. In our study, the mean time for a cow to lie down varied between 4.6 and 9.8 (mean 6.4) seconds in each herd. During a herd visit, it is easier to get a cow that is lying down to stand up, as controlled in the AC system, than to wait for a cow to voluntarily lie down, as assessed in WQ. The AC assessors force the cow to stand up with a gentle clap on the hindquarters. In WQ, the lying down time for only six cows per herd is measured, and the question is whether such few animals represent a large farm. However, Brörkens *et al*. (2009) concluded that even though only six cows per herd were assessed, a representative value for the farm was achieved. On the other hand, rising behaviour needed to be assessed in about 18 to 19 cows per herd to give a reliable result. In the study, the researchers waited for the cows to lie down and rise voluntarily, which was a tricky and time-consuming assessment method. Checking the cows´ rising behaviour can be a valuable tool for assessing both whether the animals have health problems, but especially if the pen design is appropriate. In my opinion, checking the laying down behaviour is too time-consuming, as the cows must lie down voluntarily.

5.1.7 Natural behaviour

The AWA (2018:1192) states that animals shall be kept and managed in a satisfactory environment that promotes their welfare and allows them to perform such natural behaviours for which they are highly motivated. According to Algers (1990), the behaviour should result in functional feedback, i.e., reduce the animal´s motivation to perform the abovementioned behaviour. No tools, measurements or quality evaluation make it possible for the OC today to assess animal´s natural needs and behaviours. The animals´ behaviours are limited, by law, to those that relate to the welfare requirements of motion, rest and comfort, enrichment, foraging and socialising.

Within the WQ, the parameters that influence feeding, health, housing, and animal behaviour are assessed, the latter of which is unique to WQ compared to other assessment systems. One challenge is that the behaviour tests are time-consuming (de Vries *et al*., 2013). All assessors in our study responded that WQ is too time-consuming to implement thoroughly in the OC.

Evaluating the behaviour assessments and getting consensus among assessors may also be challenging. Compared to measuring resource-based parameters, such as the lengths of the stalls or the air quality, measuring behaviour demands experienced assessors with appropriate knowledge about the behaviour in question. Knierim and Winckler (2009) discuss the low repeatability challenge when assessing farm animal welfare. The reliability of the behavioural measures can thus be questioned. It is a challenge to conduct the measurements unbiased. Nevertheless, using behaviours that easily could be answered ´Yes´ or ´No´ will decrease this obstacle. More research on this topic is needed to test the actual reliability of the behaviour assessment.

5.1.8 Animal welfare implications

The Swedish AWA states that animals shall be kept and cared for in a good environment and in such a way that their well-being is promoted and that they should be able to perform behaviours they are highly motivated for (natural behaviours), and abnormal behaviours should be prevented. Altogether, these are all important factors for the animals´ well-being. Our results show that none of the assessment protocols detect all needed to evaluate animal welfare according to the demands stated in the legislation. To correctly assess animal welfare, scientifically evaluated animal-based measures should be included and used complementary in the OC to identify suffering in individual animals as well as herd-level problems.

The EU Commission aims to harmonise, coordinate and implement equivalent animal welfare control in the EU. Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls[2] states that the controls may include specific animal welfare indicators based on measurable performance criteria, and the design of such indicators based on scientific and technical evidence to verify compliance with the requirements. The interest in animal-based welfare parameters in the EU Parliament and the Commission will probably mean that Sweden must improve the control system. In addition to appropriate legislation, controls must be carried out to check compliance and prevent animals from suffering.

5.1.9 Skin temperature vs rectal temperature

All skin temperature measurements showed a very weak correlation with rectal temperature, indicating that measurements at the positions evaluated (neck, hip, vulva) cannot be used to assess changes in actual body temperature (rectal temperature) at the individual level in healthy cows (no fever) at the ambient temperature range prevailing in the study period (12.4-23.7°C). This finding is not surprising since the mechanisms used by dairy cows to maintain constant body core temperature are both sensitive and fine-tuned, with thermoreceptors in skin and organs responding to temperature changes of less than 0.1°C and sending signals to the hypothalamus to adjust peripheral vasculature (Sjaastad *et al*., 2016). Blood is directed to/away from the core, and core temperature is maintained. The potential of the handheld temperature sensors to detect modified body temperature (e.g., elevated rectal temperatures in a heated environment outside the thermoneutral zone) remains to be determined since climate conditions were not extreme in the present study. In our study, there were few significant changes in rectal temperature over an indoor ambient temperature range of 12.7 to 23.7°C, and no cow was observed with fever, which limited our possibilities to link skin temperatures to fever and heat stress (defined as significantly elevated rectal temperatures).

5.1.10 Skin temperature vs indoor temperature

All skin temperature measurements showed a strong positive correlation with indoor temperature, reflecting the effects of radiation from the surroundings and the effects of vasodilation or vasoconstriction. Jansson *et al*. (2021) also

[2]EUT L 95, 7.4.2017, s. 1, Celex 32017R0625.

found that the surroundings greatly influenced the temperature measured on the animal. In our study, the IRT measurements at the neck showed the strongest correlation with indoor temperature. In May-September, when indoor temperature was within the range of 17.6-23.7°C, IRT values at the neck exceeded 31.5°C. During the rest of the year, when indoor temperatures were lower, the IRT values at the neck were always below 30.2°C. IRT at the neck gives a quick measurement with limited interaction with the cow and could, therefore, be investigated further as a tool for monitoring thermoregulation. IRT cameras can also be permanently installed for temperature monitoring, e.g., when approaching a water station (Schaefer *et al*., 2012) or a milking parlour.

There were variations in measured temperature values at all body sites. Skin temperature measurements using MTT showed a variation of 2.7°C, with the highest temperatures obtained during the warmest period of the year (June-September) and not at the start of peak lactation (March to May) when metabolic rate is expected to be highest. IRT measurements on humans and animals are known to be influenced by the surrounding environment (Houdas & Guieu, 1975; Jansson *et al*., 2021). Skin temperature measurements using IRT showed the highest overall variation, with the largest variation (5.4°C) in temperature measured at the hip during June-September.

5.1.11 Usefulness in the official control

Temperature measurements must be performed quickly and safely in official animal welfare controls, and in large herds. The MTT sensor took much longer to obtain each measurement (more than one minute) of the two handheld methods tested in the present study. It required physical contact with the cow and the operator to adjust the equipment, which was a challenge on some occasions. This technique is, therefore, not optimal as a routine tool for animal welfare control. The IRT device was quick (~2 seconds per measurement) and required no physical contact with the cow and, therefore, has greater potential in this context. This type of equipment is also reasonably cheap (~200 Euro, Google search on 15 Oct 2023) and can be used without great financial risk in dirty indoor conditions. In conclusion, the IRT sensor quickly recorded with limited interaction with the cow. It could be interesting for future studies investigating animals outside the thermoneutral zone and developing tools for use in the official control.

Technology may have substantial possibilities in the future. Already today, there are, e.g., bolus equipment measuring core temperature, drinking behaviour, rumination and activity (e.g., smaXtec, 2025).

5.1.12 Osmolality as an indicator of the hydration status

When planning this study, the purpose was to investigate how osmolality changes in relation to water intake. Due to technical problems, this was not done. Unfortunately, lacking this information has greatly influenced the results achieved and the possibility of making interpretations concerning individual fluid status. Therefore, we could only measure the osmolality variations during the day without knowing the time and amount of water intake. However, our results were interesting, as the variation within cows was much higher than expected, even though the cows had free access to water. We could also study how the osmolality changes in totally water-restricted cows, although such a study is much more invasive for the animals.

The differences in osmolality observed were 1-9%, with variations between breeds being the smallest and between months/lactation stages and individuals the largest. It is concluded that there are variations in milk osmolality between individuals, days and breeds that exceed what can be interpreted as thirst stimulating (+1-2%); therefore, knowledge about the individual level is needed for milk osmolality to be of practical use. Sampling time may be a limitation during welfare inspections. However, if milk osmolality analyses could be done in the milking machine, the milk osmolality could be monitored automatically. Milk osmolality could be of value for continuous monitoring of the individual cow and herd levels to detect irregularities. It may also be used as part of official control, especially when water deficiencies are suspected, but further studies are needed.

Our study shows how the osmolality changes during the day and during lactation. The variation in osmolality in milk from one cow during a day fluctuated in various amounts between cows; the standard error of the means within cows varied between 2.0 and 7.3 during the four days in February, probably due to the amount and time of water intake. There were significant differences between the monthly values. In November, the osmolality was higher than in the other months in our 20 cows and the tank milk. According to the Lövsta Research Centre manager, the barn had power failures affecting the water supply, which could have influenced water intake and, thereby, the osmolality levels. In contrast, in December, the osmolality levels were lower than usual.

If the osmolality from each cow was regularly analysed, e.g., when sampling the monthly milk samples in the Cow Control, the knowledge about the individual cow will increase and cows with high osmolality values will be identified. If there are any suspicions about abnormalities in the water supply, milk samples may also be taken at the animal welfare control. In June, with a tank osmolality at 290.4 mOsm/kg, the osmolality deviated the most in our studied cows, with a maximum difference of up to 19.6 mOsm/kg (+6.7%) in one individual.

5.2 Animal Welfare assessment improvements

5.2.1 Incentives for improved animal welfare

Webster (2012) investigated incentives and constraints for improved animal welfare by farmers and consumer demands for welfare-friendly products in the United Kingdom. The complaints Webster (2012) got from the farmers were that there was too much inspection, too little reward, and too few signs of improvement. Consumers´ behaviour when buying food of animal origin was based more on emotions than cognition (Webster, 2012). The WQ project objectives include consumer concerns and the market demand for good animal welfare and food quality. To make it possible to evaluate animal welfare-friendly food, the system must be reliable and trustable for the consumers.

Anneberg *et al*. (2012) investigated how Danish livestock farmers perceived animal welfare controls. The scientists investigated the farmers´ experience of animal welfare control related to the legislation and discussed animal welfare assessment with the farmers and the need for a third-party audit. The results showed that the farmers thought the animal welfare controls were necessary and inevitable to find the farmers not complying with the legislation and to protect the animals. The controls also improve the customers´ trust in the farmers´ production. However, all farmers felt disturbed by the controls and thought that the controls were, in general, unfair. Some farmers believed the control protected against ´home blindness´ and saw the controls as a possible way to improve animal welfare. They also believed the controls provide information about new legislation. Many studies (Anneberg *et al*., 2012; Blokhuis, 2008; Main *et al*., 2003) highlight the importance of farmers´ education in increasing awareness of animal welfare.

6. Main conclusions

The herds in Paper I were ranked differently in the animal welfare systems OC, AC and WQ, which can partially be explained by the different parameters used. In the OC, only 55 remarks were made in total, even though high numbers of lean cows, cows with lesions and dirty cows were reported in both AC and WQ. Herds with tie-stalls were ranked low in terms of animal welfare in WQ since the cows were prevented from moving freely and had access to just one water bowl. The animal-based measures used in programs like AC and WQ provide considerably more information about the animals´ well-being and welfare than the resource-based measurements recorded today in the OC. To improve the animal welfare assessment, scientifically evaluated animal-based measures should be included and used complementary in the OC to identify suffering in individual animals and herd-level problems.

Individuals from different professional backgrounds, with different levels of experience, view and score skin soiling and body condition in cattle differently. In contrast, the scorings of assessors with similar backgrounds agree well. Undergraduate students in animal science rate cattle soiling stricter than professional assessors, and they seem prepared to take stricter corrective actions against animal soiling at an official control inspection.

The variation in milk osmolality between and within cows is considerable. The fact that each sample may have a considerable variation in repeated analyses has to be considered if a single sample of milk osmolality is to be used as an animal-based indicator of the hydration status.

Concluding the outcome of this thesis, the main advice is to include detailed animal-based measurements for body condition, cleanliness, lesions and swellings in the official control. Not only if there is non-compliance but also the number of animals affected and the severity of the situation.

The proposals of this thesis are to add in the OC:

- Include information on the BCS
- Include the severity and number of dirty animals
- Include the number of animals with lesions and the severity
- Include severity and the number of lame animals
- Include the number of sick animals
- Add information on the housing system

- Add information on whether animals are allogrooming or not
- Check the cows´ rising behaviour

In addition

- Measuring thermal comfort with IR sensors can be an option in the future
- Measuring milk osmolality by taking milk samples can be valuable in cases of suspected water shortages, and regular measurement of osmolality in the milking machine in the future would be of great value

7. Future research

Further research about inter-rater agreements and how to develop effective training programmes for animal welfare inspectors is needed. A common understanding of animal welfare requirements is important (Sørensen & Fraser, 2010; Anneberg *et al*., 2012). Berg and Lundmark Hedman (2020) highlighted the importance of training programs, guidelines, and checklists for the inspectors and the farmers to know and understand the requirements.

Milk osmolality could be of value for continuous monitoring of the individual cow and herd levels to detect irregularities. It may also be used as part of official control. If analyses could be done in the milking machine, the milk osmolality could be monitored automatically, especially when water deficiencies are suspected. However, further studies are needed.

More research is needed to study if the percentage of cows with 120 days between calving and pregnancy is a valid indicator for compromised animal welfare that can be used as a risk-based indicator.

References

Abraira, V. & Pérez de Vargas, A. (1999). Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs. *Qüestiió*. 23, 561-571.

Algers, B. (1990). Naturligt beteende – ett naturligt begrepp? *Svensk veterinärtidning.* Vol. 42 nr 12. 517–519

Allendorf, J.J. & Wettemann, P.J.C. (2015). Does animal welfare influence dairy farm efficiency? A two-stage approach. *Journal of Dairy Science.* 98, (11), 7730-40

Anneberg, I., Vaarst, M. & Sandøe, P. (2013). To inspect, to motivate - or to do both? A dilemma for on-farm inspection of animal welfare. *Animal Welfare.* 22, (2), 185-94.

Anneberg, I., Vaarst, M. & Sørensen, J.T. (2012). The experience of animal welfare inspections as perceived by Danish livestock farmers: A qualitative research approach. *Livestock Science*. 147(1–3), 49–58.

Armstrong, D.V. (1994). Heat stress interaction with shade and cooling. *Journal of Dairy Science.* 77, 2044-50.

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

Bell, M.J., Maak, M., Sorley, M. & Proud, R. (2018) Comparison of Methods for Monitoring the Body Condition of Dairy Cows. *Frontiers in Sustainable Food Systems*. 2:80. doi: 10.3389/fsufs.2018.00080

Berg, C. & Lundmark Hedman, F. (2020). Compliance with animal welfare regulations: Drivers and consequences. *CAB Reviews*. 2020, 15, 1-10. doi: 10/1079/PAVSNNR202015025, http://www.cabi.org/cabreviews

Berthe, F., Vannier, P., Have, P., Serratosa, J., Bastino. E., Broom, D.M., Hartung, J. & Sharp, J.M. (2012). Special issue: The role of EFSA in assessing and promoting animal health and welfare. *EFSA Journal.* 2012;10(10):s1002. [10 pp.]. doi:10.2903/j.efsa.2012.s1002 Available online: www.efsa.europa.eu/efsajournal

Bjerg, M., Rasmussen, M. D. & Nielsen, M.O. (2005). Changes in freezing point of blood and milk during dehydration and rehydration in lactating cows. *Journal of Dairy Sciences*. 88, (9), 3174-85.

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), 445-55.

Blokhuis, H.J. (2008) International cooperation in animal welfare: the Welfare Quality project. *Acta Veterinaria Scandinavica*. 50: S10

Blokhuis, H.J., Veissier, I., Miele, M. & Jones, B. (2010). The Welfare Quality project and beyond: Safeguarding farm animal well-being. *Acta Agricultural Scandinavica Section A-Animal Science*. 60, (3), 129-40.

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I, Langbein, J., Bakken, M., Veissier, I & Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*. 92, 375–397. doi: 10.1016/j.physbeh.2007.02.003

Brambell, F.W.R. (1965). Report of the Technical Committee of Enquiry into the Welfare of Livestock Kept under Intensive Conditions; HMSO: London, UK.

Breen, J. E., Green, M. J. & Bradley, A. J. (2009). Quarter and cow risk factors associated with the occurrence of clinical mastitis in dairy cows in the United Kingdom. *Journal of Dairy Science*. 92:2551–2561

Broom, D.M. (2017). Animal welfare in the European Union. Study for the PETI Committee. *European Parliament*. Brussels, Belgium. https://op.europa.eu/sv/publication-detail/-/publication/74df7b49-ffe7-11e6-8a35-01aa75ed71a1/language-en

Brunet, V., Blokhuis, H., Staaf Larsson, B., D'Albenzio, S., Winckler, C., Schenkenfelder, J., Hanlon, A., Sossidou, E., Tuominen, L. & Veissier I. (2024). Improving welfare by sharing knowledge: Functioning of an EU Reference Centre for Animal Welfare. Conference: 9th International Conference on the Welfare Assessment of Animals at Farm Level (WAFL). https://www.researchgate.net/publication/383671158_Improving_welfare_by_sharing_knowledge_Functioning_of_an_EU_Reference_Centre_for_Animal_Welfare#fullTextFileContent [accessed Jun 13 2025].

Brörkens, N, Plesch G, Laister S, Zucca D, Winckler C, Minero M & Knierim U. (2009). Reliability testing concerning behaviour around resting in cattle in dairy cows and beef bulls. In Forkman B & Keeling L (eds) Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves *Welfare Quality Reports*. No. 11 pp 7-24. Cardiff University

Buller, H., Blokhuis, H., Jensen, P. & Keeling, L. (2018). Towards Farm Animal Welfare and Sustainability. *Animals*. 8, (6)

Büttel, B., Fuchs, M. & Holz, B. (2008). Freezing point osmometry of milk to determine the additional water content - an issue in general quality control and German food regulation. *Chemistry Central Journal*. 2, 7.

Castle, M.E., & Thomas T.P. (1975). Water-intake of British Friesian cows on rations containing various forages. Anim. Prod. 20, (APR), 181-9.

de Freslon, I., Peralta, J.M., Strappini, A.C. & Monti, G. (2020) Understanding Allogrooming Through a Dynamic Social Network Approach: An Example in a Group of Dairy Cows. Front. Vet. Sci. 7:535. doi: 10.3389/fvets.2020.00535 de Vries, M., Engel, B., den Uijl, I., van Schaik, G., Dijkstra, T., de Boer, I.J.M. & Bokkers, E.A.M. (2013). Assessment time of the Welfare Quality protocol for dairy cattle. *Animal Welfare*. 22: 85-93

de Vries, M., Bokkers, E.A.M., van Reenen, C.G., Engel, B., van Schaik, G., Dijkstra, T. & de Boer, I.J.M. (2015). Housing and management factors associated with indicators of dairy cattle welfare, Preventive Veterinary Medicine Volume 118, Issue 1, Pages 80-92, ISSN 0167-5877, https://doi.org/10.1016/j.prevetmed.2014.11.016.

European Commission. (1997). Treaty of Amsterdam; Amending the Treaty on European Union and the Treaties establishing the European Communities and certain related acts (97/C 340/01) eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:11997D/TXT

European Commission. (2007). Treaty of Lisbon; Amending the Treaty on European Union and the Treaty establishing the European Community (2007/C 306/0) https://publications.europa.eu/resource/cellar/688a7a98-3110-4ffe-a6b3-8972d8445325.0007.01/DOC_19

European Commission. (2012). 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 2012-2015

European Commission. (2016). Attitudes of Europeans towards Animal Welfare Special Eurobarometer 442. Accessed 15 September 2020, available at: http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/71348.

European Commission. (2019). Final report of an audit carried out in Sweden from 13 May 2019 to 17 May 2019 in order to evaluate the use of quality controls and indicators for animal welfare Directorate-General for Health and Food Safety Health and food audits and analysis DG(SANTE) 2019-6593

European Commission. (2020) Farm to Fork strategy: For a fair, healthy and environmentally-friendly food system. In DG SANTE/Unit ´Food Information and Composition, Food Waste´; European Commission: Brussels, Belgium. [Google Scholar]

European Commission. (2022). DG Health and Food Safety Overview report (2021-7319) The use of indicators for animal welfare at farm level. Luxembourg: Publications Office of the European Union. ISBN 978-92-76-49326-6 doi:10.2875/85740

European Parliament. (2015). The European Parliament´s communication on the strategy European Parliament. B8-1281/2015, European Parliament resolution on a new animal welfare strategy for 2016-2020, (2015/2957[RSP]). Texts adopted - A new animal welfare strategy for 2016-2020 - Thursday, 26 November 2015

European Union. (1992). Treaty on European Union (Maastricht Treaty) Maastricht Treaty
FAWC. (1993) Report on Priorities for Animal Welfare Research and Development. Farm Animal Welfare Council.
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, (3), 187-205
Furukawa, E., Ozawa, T., Higaki, S., Suda, T., Sasaki, Y., Murayama, K., Noguchi, M. & Yoshioka, K. (2024). Changes in skin temperature and behaviors of preweaning Holstein calves in a hot environment monitored by a multimodal tail-attached device. Journal of Dairy Science, 5(4), 332–337. doi:10.3168/jdsc.2023-0515
Gillund, P., Reksen, O., Karlberg, K., Randby, Å.T., Engeland, I. & Lutnæs, B. (1999). Utprøving av en holdvurderingsmetode på NKF-kyr. *Norsk Veterinærtidskrift*. 111:623-632 (Evaluation of a method for body condition scoring of Norwegian Red Dairy Cows)
Green, L.E., Huxley, J.N., Banks, C. & Green, M.J. (2014). Temporal associations between low body condition, lameness and milk yield in a UK dairy herd. *Preventive Veterinary Medicine*. 113, (1), 63-71.
Goossens, X., Sobry, L., Ödberg, F., Tuyttens, F., Maes, D., De Smet, S., Nevens F., Opsomer, G., Lommelen, F. & Geers, R. (2008) A population-based on-farm evaluation protocol for comparing the welfare of pigs between farms. *Journal of Dairy Science*. 17: 35–41
Harrison, R. Animal Machines: The New Factory Farming Industry. (1964). 186 p. Vincent Stuart Publishers: London, UK
Haskell, M. J., Rennie, L. J., Bowell, V., Bell, M. J., Lawrence, B., (2006). Housing system, milk production, and zero-grazing effects on lameness and leg injury in dairy cows. *Journal of Dairy Science*. 89(11), 4259–4266. http://doi.org/10.3168/jds.S0022-0302(06)72472-9.
Hauge, S.J., Kielland, C., Ringdal, G., Skjerve, E. & Nafstad, O. (2012). Factors associated with cattle cleanliness on Norwegian dairy farms. *Journal of Dairy Science*. 95, (5), 2485-96.
Heath, C.A.E., Browne, W.J., Mullan, S. & Main, D.C.J. (2014a). Navigating the iceberg: reducing the number of parameters within the Welfare Quality assessment protocol for dairy cows. *Animal*. 8, (12), 1978-86
Heath, C.A.E., Lin, Y., Mullan, S., Browne, W.J. & Main, D.C.J. (2014b) Implementing Welfare Quality in UK assurance schemes: evaluating the challenges. *Animal Welfare*. 23, (1), 95-107
Houdas, Y. & Guieu, J. (1975). Environmental factors affecting skin temperatures. *Bibliotheca radiologica* 6, 157–165.
Hughes, J. (2001). A system for assessing cow cleanliness. In Practice, 23, 517-524.

Jansson, A., Lindgren, G., Velie, B.D. & Solé, M. (2021). An investigation into factors influencing basal eye temperature in the domestic horse (Equus caballus) when measured using infrared thermography in field conditions. *Physiology & Behavior* 228, Article 113218

Jensen, M.B. & Vestergaard, M. (2021). Invited review: Freedom from thirst-Do dairy cows and calves have sufficient access to drinking water? *Journal of Dairy Science*. 104, (11), 11368-85

Jeppsson, K.H., Magnusson, M., Nilsson, S.B., Ekman, L., Winblad von Walter, L., Jansson, L.E., Landin, H., Rosander, A. & Bergsten, C. (2024). Comparisons of recycled manure solids and wood shavings/sawdust as bedding material - implications for animal welfare, herd health, milk quality and bedding costs in Swedish dairy herds. *Journal of Dairy Science.* 107(8):5779–5793. https://doi.org/10.3168/jds.2023-24192

Keeling, L.J. (2009). An analysis of animal-based versus resource-based comments in official animal welfare inspection reports from organic and conventional farms in Sweden. *Animal Welfare*. 18, 391-397.

Keeling, L.J., Marier, E.A., Olmos Antillón, G., Blokhuis, H.J., Staaf Larsson, B. & Stuardo, L. (2022) A global study to identify a potential basis for policy options when integrating animal welfare into the UN Sustainable Development Goals. *Frontiers in Animal Science*. 3:974687. https://doi.org/10.3389/fanim.2022.974687

Keeling, L.J., Winckler, C., Hintze, S. & Forkman, B. (2021) Towards a Positive Welfare Protocol for Cattle: A Critical Review of Indicators and Suggestion of How We Might Proceed. *Frontiers in Animal Science*. 2:753080. doi: 10.3389/fanim.2021.753080

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(4), 451–458

Krohn, C.C. & Munksgaard, L. (1993). Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments II. Lying and lying-down behaviour. *Applied Animal Behaviour Science*. 37, 1–16

Krueger, A., Cruickshank, J., Trevisi, E. & Bionaz, M. (2020). Systems for evaluation of welfare on dairy farms. *Journal of Dairy Research*. 87(S1), 13–19. https://doi.org/10.1017/S0022029920000461

Linzell, J.L. & Peaker, M. (1971). Mechanisms of milk secretion. *Physiological Reviews*. 51, 564–597.

Lomellini-Dereclenne, A.C., Miele. M., Mounier, L. & Veissier, I. (2017). Implementation of the European legislation to protect farm animals: A case study on French inspections to find solutions to improve compliance. *Animal Welfare*. 26, 311–321

Lundmark, F., Röcklinsberg, H., Wahlberg, B. & Berg, C. (2016). Content and structure of Swedish animal welfare legislation and private standards for dairy cattle. *Acta Agricultural Scandinavica Section A-Animal Science*. 66, (1), 35-42

Lundmark Hedman, F., Andersson, M.; Kinch, V., Lindholm, A., Nordqvist, A. & Westin, R. (2021) Cattle Cleanliness from the View of Swedish Farmers and Official Animal Welfare Inspectors. *Animals*. 11, 945. https://doi.org/10.3390/ani11040945.

Lundmark Hedman, F., Ewerlöf, I.R., Frössling, J. & Berg, C. (2025) Official and private animal welfare inspectors' perception of their own on-site inspections. *Frontiers in Veterinary Science*. 12:1575471. doi: 10.3389/fvets.2025.1575471

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 Food Science Technology*. 37, (2), 127-36

Main DCJ, Whay HR, Green LE & Webster AJF. (2003). Effect of the RSPCA Freedom Food scheme on the welfare of dairy cattle. *Veterinary Record.* 153, 227-231

McDonald, P.V., von Keyserlingk, M.A.G. & Weary, D.M. (2020). Hot weather increases competition between dairy cows at the drinker. *Journal of Dairy Science*. 103, (4), 3447-58.

Mellor, D.J. & Reid, C.S.W. (1994) Concepts of animal well-being and predicting the impact of procedures on experimental animals. In: Baker R, Jenkin G and Mellor DJ (eds) Improving the Well-being of Animals in the Research Environment pp 3-18. Australian and New Zealand Council for the Care of Animals in Research and Teaching: Glen Osmond, SA, Australia

Murphy. (1992). Water metabolism of dairy cattle. *Journal of Dairy Science*. 75, (1), 326-33

Murphy, M.R., Davis, C.L. & McCoy, G.C. (1983). Factors affecting water-consumption by Holstein cows in early lactation. *Journal of Dairy Science*. 66, (1), 35-8.

Nafstad, O. (1999). Hide damages and quality defects in cattle leather. Skader og kvalitetsfeil på norske storfehuder. *Norsk Veterinærtidsskrift*. 111, (5), 311-319.

Nogami, H., Okada, H., Miyamoto, T., Maeda, R. & Itoh, T. (2014). Wearable wireless temperature sensor nodes appressed to base of a calf's tail. *Sensors and Materials*. 26(8), 539–545.

Olsson, K. & Dahlborn K. (1989). Fluid balance during heat stress in lactating goats. *Quarterly Journal of Experimental Physiology*. 74, 645-659.

Otten, N.D., Rousing, T., de Oliveira, V.H.S., Reiten, M., Michelsen, A.M., Hakansson, F., Lund, V.P., Houe, H., Denwood, M., Sorensen, J.T., Forkman, B. & Kirchner, M.K. (2020). A weak point analysis of welfare in Danish dairy herds using two different welfare assessment systems. *Animal Welfare*. 29, (2), 197-207.

Pinheiro Machado Filho, L. C., D. L. Teixeira, D. M. Weary, M. A. G. von Keyserlingk, & M. J. Hötzel. (2004). Designing better water troughs: Dairy cows prefer and drink more from larger troughs. *Applied Animal Behaviour Science*. 89:185–193. https://doi.org/10.1016/j.applanim.2004.07.002

Pryce, J.E., Coffey, M.P. & Simm, G. (2001). The relationship between body condition score and reproductive performance. *Journal of Dairy Science*. 84, (6), 1508-15.

Radeski, M., Janevski, A. & Ilieski, V. (2015). Screening of selected indicators of dairy cattle welfare in Macedonia. *Macedonian Veterinary Review*. 38 (1): 43-51.

Rault, J-L., Bateson, M., Boissy, A., Forkman, B., Grinde, B., Gygax, L., Lynning Harfeld, J., Hintze, S., Keeling, L.J., Kostal, L., Lawrence, A.B., Mendl, M.T., Miele, M., Newberry, R.C., Sandøe, P., Špinka, M., Taylor, A.H., Webb, L.E., Whalin, L. & Bak Jensen, M. (2025). A consensus on the definition of positive animal welfare. *Biology Letters.* 21: 20240382. https://doi.org/10.1098/rsbl.2024.0382

Roche, J.R. & Berry, D.P. (2006). Periparturient climatic, animal, and management factors influencing the incidence of milk fever in grazing systems. *Journal of Dairy Science*. 89, 2775-2783.

Roche, J.R., Lee, J.M., Macdonald, K.A. & Berry, D.P. (2007). Relationships among body condition score, body weight, and milk production variables in pasture-based dairy cows. *Journal of Dairy Science*. 90, (8), 3802-15.

Sandgren, C.H., Lindberg, A. & Keeling, L.J. (2009). Using a national database to identify herds with poor welfare. *Animal Welfare*. 18:523-532

Schröder, U. J. & Staufenbiel, R. (2006). Invited review: Methods to determine body fat reserves in the dairy cow with special regard to ultrasonographic measurement of backfat thickness. *Journal of Dairy Science*. 89, 1–14. doi:10.3168/jds.S0022-0302(06)72064-1

Scoley, G., Gordon, A. & Morrison, S.J. (2019). The effect of calf jacket usage on performance, behaviour and physiological responses of group-housed dairy calves. Animal, 13(12), 2876–2884. doi:10.1017/S1751731119001071

Schaefer, A. L., Cook, N. J., Bench, C., Chabot, J. B., Colyn, J., Liu, T., Okine, E. K., Stewart, M. & Webster, J. R. (2012). The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. Research in Veterinary Science, 93(2), 928–935. doi:10.1016/j.rvsc.2011.09.021

Shipe, W. F. 1959. The freezing point of milk - a review. *Journal of Dairy Science*. 42, (11), 1745-62.

Sjaastad, Ø.V., Hove, K. & Sand, O. (2016). Physiology of domestic animals. Scandinavian Veterinary Press, Oslo, Norway. Third edition: ISBN 978-82-91743-41-7 smaXtec. (2025) smaXtec: Early detection for dairy cows with bolus technology (Accessed 2025-04-15)

Sørensen, J.T. & Fraser, D. (2010). On-farm welfare assessment for regulatory purposes: issues and possible solutions. *Livestock Science*. 131, 1-7.

Staaf Larsson, B. (2023). Q2E - Good husbandry practice for cattle. EURCAW-*Ruminants & Equines*. Zenodo. https://doi.org/10.5281/zenodo.7788782

Staaf Larsson, B., Jansson, A., Holmberg, M., Winblad von Walter, L., Stéen, M., & Dahlborn, K. (2024). A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 73:3-4, 153-164. https://doi.org/10.1080/09064702.2024.2317708

Sun, D., Webb, L., van der Tol, P.P.J. & van Reenen, K. (2021). A systematic review of automatic health monitoring in calves: Glimpsing the future from current practice. *Frontiers in Veterinary Science*, 8, 761468. doi:10.3389/fvets.2021.761468

Swedish Board of Agriculture (2022). Vägledning för offentlig djurskyddskontroll [Guidance for official animal welfare control]. Jönköping: Swedish Board of Agriculture Vägledning om djurskyddskontroll

Swedish Board of Agriculture (2024a). Checklista - Nötkreatur [Checklist - Cattle]. Jönköping: Swedish Board of Agriculture Checklista-notkreatur-tgau.pdf

Swedish Board of Agriculture (2024b). Bilaga till checklista - Nötkreatur [Appendix to checklist - Cattle]. Jönköping: Swedish Board of Agriculture Bilaga till checklista - Nötkreatur

Swedish Board of Agriculture (2025). Nationell djurskyddsrapport 2024 - En redovisning av kontrollmyndigheternas arbete [National Animal Welfare Report 2024 - A report on the work of the control authorities] Jönköping: Swedish Board of Agriculture ovr737.pdf

Tremetsberger, L., Leeb, C. & Winckler, C. (2015). Animal health and welfare planning improves udder health and cleanliness but not leg health in Austrian dairy herds. *Journal of Dairy Science*. 98, (10), 6801-11

Tremetsberger, L., Winckler, C. & Kantelhardt, J. (2019). Animal health and welfare state and technical efficiency of dairy farms: possible synergies. *Animal Welfare*. 28, (3), 345-52

Webster, J. (2012). Critical control points in the delivery of improved animal welfare. *Animal Welfare*. 21(S1): 117-123

Welfare Quality®. (2009). Welfare Quality® assessment protocol for cattle. *Welfare Quality® Consortium*. Lelystad, Netherlands https://www.welfarequalitynetwork.net/media/1319/dairy-cattle-protocol.pdf

Ventura, G., Lorenzi, V., Mazza, F., Clemente, G.A., Iacomino, C., Bertocchi, L. & Fusi, F. (2021) Best Farming Practices for the Welfare of Dairy Cows,

Heifers and Calves. *Animals*. 11(9):2645. https://doi.org/10.3390/ani11092645
Wildman, E.E., Jones, G.M., Wagner, P.E., Boman, R.L., Troutt, H.F. & Lesch, T.N. (1982). A dairy cow body condition scoring system and its relationship to selected production characteristics. *Journal of Dairy Science*. 65, 495–501
Växa (2025) Växa Husdjursstatistik. Cattle statistics 2025 www.vxa.se/statistik
Ward, W.R., Hughes, J.W., Faull, W.B., Cripps, P.J., Sutherland, J.P. & Sutherst, J.E. (2002). Observational study of temperature, moisture, pH and bacteria in straw bedding, and faecal consistency, cleanliness and mastitis in cows in four dairy herds. *Veterinary Record* 151, 199-206

Popular science summary

Animal welfare control can be carried out either by checking what resources we offer the animals and/or by assessing the animals themselves, so-called animal-based parameters. Today, the animal welfare inspector assesses the herd during a farm visit according to the Swedish Board of Agriculture´s checklist. This mainly assesses resource- and management-based parameters (i.e. what spaces and resources the animals have and how the animals are cared for) and the animal-based parameters that are checked are neither quantitatively nor qualitatively assessed. The checklist contains the answer options ´Yes´, ´No´, ´Not checked´ and ´Not applicable´. The answers therefore do not show to what extent the criterion is met, i.e. for example regarding the question of clean animals, neither the degree of dirtiness nor the number of dirty animals. This thesis is about how animal welfare control can be improved by using more animal-based parameters and how the animal-based parameters that are currently in the checklist can provide more detailed information. The European Commission has also requested more and more detailed animal-based parameters to measure animal welfare, both to be able to assess animal welfare at farm level and what overall animal welfare a country has, and to be able to follow changes over time and whether there are any improvements. If good methods for assessing animal welfare are agreed upon, these could also be used by animal keepers themselves and in animal welfare labelling of animal-based foods.

In Paper I, results from two well-established protocols for assessing animal welfare in dairy cows, Welfare Quality and Växa´s ´Ask the Cow´, which mainly contain animal-based parameters, have been compared with the official control on 41 farms. Ask the Cow is used to show the animal keeper how their animal welfare is compared to other farms and what can be improved. Welfare Quality is mainly used in research to evaluate animal welfare on farms. This study was conducted to investigate whether protocols with more and more detailed animal-based parameters differ in assessment compared to the Swedish Board of Agriculture checklist, which contains mainly resource- and management-based parameters. Paper II investigates whether assessments of BCS and dirty animals differ between different professional groups, and whether experience and animal-related training affect assessments. Data were collected at 15 training sessions during the years 2009 to 2016. A total of 569 people; animal welfare inspectors, veterinarians/advisors, animal welfare scientists, animal keepers and students in animal science assessed 6 to 40 photos with regard to BCS, dirtiness and recommended actions regarding dirty animals.

The legislation requires that animals have thermal comfort and that animals receive sufficient water, but no proposal on how this should be checked in connection with animal welfare inspections at farm level. In Papers III and IV I have therefore studied potential methods for measuring body temperature in dairy cows and milk osmolality = the number of particles in the body fluids. High values indicate dehydration and low values indicate that the animal has just drunk large amounts of water. The method is currently used to show at tank milk level whether the milk has been diluted with water. Blood sampling is too invasive to be used in the official control; therefore, milk sampling is a better alternative that could be used as a control of water supply at the individual level. The data collection for Papers III and IV was done on 20 cows for four days at the beginning of lactation and then once a month for a year.

The results from Paper 1 show that of the 41 farms visited, the eight farms that received the most remarks in the official control were ranked in the range of 19–40 in Ask Cow and 5–37 in Welfare Quality, where 1 represents the best and 41 (Ask Cow) and 38 (Welfare Quality) the worst farm, respectively. The spread was therefore large. Studies have shown that much more information about animal welfare can be obtained if more detailed information about BCS, dirtiness and injuries is recorded. For example, in the official animal welfare control, dirtiness was reported on only two farms, while dirtiness was observed in all herds in Welfare Quality and according to Ask the Cow, 31 of the 41 farms had dirty cows. No lean or fat cows were noted on any of the 41 farms when assessed using the checklist, while there were 23 farms with lean and 26 farms with fat cows according to Ask the Cow, and 33 farms with lean and 23 farms with fat cows according to Welfare Quality

The study on different professional categories (Paper II) showed that the more extensive the animal science education, the less likely the animal welfare officers were to assess a high hull. The students assessed the cattle as dirtier than the animal welfare inspectors and veterinarians/advisors did. Students also recommended more stringent measures than the animal welfare inspectors and veterinarians/advisors did, and veterinarians/advisors recommended took less strict measures than animal welfare inspectors did.

The study on measuring thermal comfort on dairy cows (Paper III) showed that the rectal temperature was not correlated with the indoor temperature. That the rectal temperature was constant was expected because cows, which are even-tempered animals, regulate their temperature, i.e. strive to maintain their body temperature regardless of what happens to the outside temperature. However, all skin temperature values showed a strong correlation with indoor temperature. When the ambient temperature increases, blood flow to the skin increases to more easily release excess heat, and when the ambient temperature decreases, blood flow to the skin decreases to preserve heat inside the body. It may therefore be possible to measure thermal comfort of dairy cows using skin temperature sensors during official control. The fluid balance of cows can also be measured by analysing the osmolality in milk (Paper IV), although there were large individual daily variations, between cows and between months. There is also potential to measure both the body condition (which is already done on some farms), dirtiness and the cows´ body temperature continuously through technology that is rapidly developing. In the long term, it should also be possible to measure the milk osmolality in the milk robot, together with the measurements of milk temperature, cell count, colour and milk quantity that are already done today. This thesis has shown that with methods that are practically feasible during an animal welfare control, much more information about animal welfare can be obtained on farms regardless of the level of technology on the farm. The conclusion is that much more information can be obtained if assessments of, for example, body condition and dirtiness are made at the individual level. Two new methods have also been developed and tested to be able to assess the hydration status and thermal comfort respectively.

Populärvetenskaplig sammanfattning

Djurskyddskontroller kan utföras antingen genom att kontrollera vilka resurser vi erbjuder djuren och/eller genom att bedöma själva djuren, så kallade djurbaserade parametrar. Idag bedömer djurskyddshandläggaren besättningen vid ett gårdsbesök enligt Jordbruksverkets checklista. Denna bedömer främst resurs- och skötselbaserade parametrar (d.v.s. vilka utrymmen och resurser djuren har och hur djuren sköts) och de djurbaserade parametrar som kontrolleras är varken kvantitativt eller kvalitativt bedömda. I checklistan finns svarsalternativen ”Ja”, ”Nej”, ”Ej kontrollerat” och ”Ej aktuellt”. Svaren visar alltså inte i vilken utsträckning kriteriet är uppfyllt, d.v.s. till exempel vad gäller frågan om rena djur, varken grad av smutsighet eller antalet smutsiga djur. Denna avhandling handlar om hur djurskyddskontrollen kan förbättras genom att använda fler djurbaserade parametrar och hur de djurbaserade parametrar som finns i checklistan idag kan ge en mer detaljerad information. Även EU-kommissionen har efterfrågat fler och mer detaljerade djurbaserade parametrar för att mäta djurvälfärden, både för att kunna bedöma djurvälfärden på gårdsnivå och vilken övergripande djurvälfärd ett land har samt för att kunna se förändringar över tid och om det sker några förbättringar. Om man enas om bra metoder för att bedöma djurvälfärden skulle dessa även kunna användas av djurhållarna själva samt vid djurvälfärdsmärkning av animaliska livsmedel.

I artikel I har resultat från två väletablerade protokoll för bedömning av djurvälfärd hos mjölkkor, Welfare Quality och Växas ”Fråga Kon”, vilka huvudsakligen innehåller djurbaserade parametrar, jämförts med den offentliga kontrollen på 41 gårdar. Fråga Kon används för att visa djurhållaren hur djurvälfärden är på gården jämfört med andra gårdar och vad som kan förbättras. Welfare Quality används främst inom forskningen för att utvärdera gårdars djurvälfärd. Denna studie gjordes för att undersöka om protokoll med fler och mer detaljerade djurbaserade parametrar skiljer sig i bedömningen jämfört med Jordbruksverkets checklista, som innehåller mest resurs- och skötselbaserade parametrar. Artikel II undersöker om bedömningar av hull och smutsiga djur skiljer sig åt mellan olika yrkesgrupper, samt om erfarenhet och en djurrelaterad utbildning påverkar bedömningarna. Data samlades in vid 15 utbildningstillfällen under åren 2009 till 2016. Totalt 569 personer; djurskyddshandläggare, veterinärer/

rådgivare inom mjölknäringen, djurskyddsforskare, djurhållare samt studenter inom djurvetenskap bedömde 6 till 40 foton med avseende på hull, smutsighet och rekommenderade åtgärder med anledning av smutsiga djur.

I lagstiftningen finns det krav på att djuren ska ha termisk komfort och att de ska få tillräckligt med vatten, men inget förslag på hur detta ska kontrolleras i samband med djurskyddskontroll på gårdsnivå. I artikel III och IV har jag därför studerat potentiella metoder för att mäta kors kroppstemperatur och mjölkens osmolalitet = antalet partiklar i kroppsvätskor. Höga värden visar på uttorkning och låga värden att djuret precis druckit stora mängder vatten. Metoden används idag på tankmjölksnivå för att kunna visa om mjölken spätts ut med vatten. Blodprovstagning är för ingripande för att användas i den offentliga kontrollen, därför är provtagning av mjölk ett bättre alternativ som skulle kunna användas som kontroll av vattenförsörjningen på individnivå. Insamlingen av data till artikel III och IV gjordes på 20 kor under fyra dagar i början av laktationen samt därefter en gång per månad under ett års tid.

Resultaten från artikel I visar att av de 41 gårdar som besöktes, rankades de åtta gårdar som fick flest anmärkningar i den offentliga kontrollen i intervallet 19–40 i Fråga Kon och 5–37 i Welfare Quality, där 1 representerar den bästa och 41 (Fråga Kon) respektive 38 (Welfare Quality) den sämsta gården. Spridningen var alltså stor. Studierna har visat att mycket mer information om djurvälfärden går att få fram om man registrerar mer detaljerad information om hull, smutsighet och skador. I den offentliga djurskyddskontrollen rapporterades till exempel smuts bara på två gårdar medan smuts observerades vid alla besättningar i Welfare Quality och enligt Fråga Kon hade 31 av de 41 gårdarna smutsiga kor. Det noterades inga magra eller feta kor hos någon av de 41 gårdarna när de bedömdes med checklistan, medan det fanns 23 gårdar med magra och 26 gårdar med feta kor enligt Fråga Kon, respektive 33 gårdar med magra och 23 gårdar med feta kor enligt Welfare Quality.

Studien om olika yrkeskategorier (artikel II) visade att ju mer omfattande djurvetenskaplig utbildning, desto mindre benägna var djurskyddshandläggarna att bedöma ett högt hull. Studenterna bedömde att nötkreaturen var smutsigare än vad djurskyddshandläggarna och veterinärerna/rådgivarna gjorde. Studenter rekommenderade också strängare åtgärder än djurskyddshandläggare och veterinärer/rådgivare gjorde, och veterinärer/rådgivare rekommenderade mindre strikta åtgärder än vad djurskyddshandläggare gjorde.

Studien om att mäta kors termiska komfort (artikel III) visade att rektaltemperaturen inte var korrelerad med inomhustemperaturen. Att rektaltemperaturen var konstant var förväntat eftersom kor, som är jämnvarma djur, temperaturreglerar, d.v.s. eftersträvar att behålla sin kroppstemperatur oavsett vad som händer med yttertemperaturen. Alla hudtemperaturvärden visade däremot en stark korrelation med inomhustemperaturen. Vid ökad omgivningstemperatur ökar genomblödningen till huden för att lättare avge överskottsvärme och vid sänkt omgivningstemperatur minskar genomblödningen till huden för att bevara värmen inne i kroppen. Det kan därmed vara möjligt att mäta kors termiska komfort med hjälp av hudtemperatursensorer vid den offentliga kontrollen. Kors vätskestatus är också möjlig att mäta genom att analysera mjölkens osmolalitet (artikel IV), även om det var stora individuella dygnsvariationer, mellan kor och mellan månader. Det finns också potential att mäta både hull (vilket redan görs på en del gårdar), smutsighet och kornas kroppstemperatur kontinuerligt genom den teknik som snabbt utvecklas. Även osmolaliteten i mjölk bör på sikt kunna mätas i mjölkroboten, tillsammans med de mätningar av mjölktemperatur, celltal, färg och mjölkmängd som redan görs idag. Denna avhandling har visat att med metoder som är praktiskt genomförbara vid en djurskyddskontroll kan mycket mer information om djurvälfärden tas fram på gårdar oavsett nivå på gårdens teknik. Slutsatsen är att mycket mer information går att få fram om bedömningar av t.ex. hull och smutsighet görs på individnivå. Två nya metoder har också tagits fram och testats för att kunna bedöma kors vätskestatus respektive termisk komfort.

Acknowledgements

I am so grateful to the Swedish Association for the Protection of Animals for providing funding for my doctoral studies, and especially to the Chairman of the Association Lillemor Wodmar, who believed in me from the first time we met. Although the thesis has taken much longer than we initially hoped, a lot has been done during the years. I am also grateful to the Government through the Swedish Centre for Animal Welfare, SLU, for funding the first study.

Big thanks to my first main supervisor Associate Professor Margareta Stéen who encouraged me and believed in me from the very beginning, for all good discussions and for being a good friend. Big thanks to my current main supervisor Professor Anna Jansson for always being helpful with your broad knowledge, especially in physiology and statistics. Big thanks also to my co-supervisor Professor Kristina Dahlborn for helping me with data sharing, study planning and impressive knowledge, especially about milk production physiology. Thanks to my co-supervisor Associate Professor Jan Hultgren for your thoroughness and great knowledge, especially in statistics. I would also like to thank co-authors Louise Winblad von Walter and Mia Holmberg for their contributions to the first paper and to Erik Petersson, who contributed to Paper II.

In Paper I, I am grateful to the farmers who allowed us to carry out the project and a big thank you to the animal welfare inspectors and all other participants who made the study possible. Thanks to Christoph Winckler and Lukas Tremetsberger, University of Natural Resources and Life Sciences (BOKU) Austria, who performed the WQ training. Thanks also to Isabelle Veissier and Anne Lamadon French National Institute for Agriculture, Food, and Environment (INRAE) France, who conducted the WQ calculations. I am also grateful to Mary McAfee for English editing.

Thanks to all colleagues at the former Department of Anatomy, Physiology and Biochemistry for all the help, fruitful conversations and coffee breaks, and thanks to all the PhD students I have met during the years. Thanks to my colleagues at the Swedish Centre for Animal Welfare. Thanks also to Claudia von Brömssen, always helpful with the statistics and to all other staff who have helped me during the years at SLU, and thanks to all researchers in the VHC building for interesting and fruitful discussions. During my time as a PhD student, I had the possibility to participate as the

PhD student representative in the Faculty Board of the Faculty of Veterinary Medicine and Animal Science, in SLU Future Animals, Nature and Health, and in the VMF PhD Council, all of which were greatly appreciated.

Thanks to the staff at Lövsta for advice during the planning of my field study and a big thank you to Jessica Isaksson who helped me keep track when I collected all the samples at Lövsta. I am also grateful for the cooperation of the participants in the scoring sessions.

I would also like to thank my beloved children. Sara, thank you for your encouragement during our years we shared the apartment in Uppsala and for the legislative discussions. Sofia, thank you for your encouragement as a goal-oriented person in your career and for all the redeeming laughs. David, thank you very much for discussing statistics and for all the constructive conversations about the methodology. Thanks also to my children´s life partners Rickard, Viktor and Lucy and to my beloved grandchildren Valter, Cornelia, William and Clara, that were born during my PhD studies, for understanding that I did not have much time with you during parts of my thesis work.

Finally, thanks to my beloved husband Henrik. Several summer vacations have been fully or partially devoted for my thesis work, without you complaining a word. This thesis would not have been completed without your encouragement and patience. I love you!

I

ACTA AGRICULTURAE SCANDINAVICA, SECTION A — ANIMAL SCIENCE
2024, VOL. 73, NOS. 3–4, 153–164
https://doi.org/10.1080/09064702.2024.2317708

A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds

B. Staaf Larsson [a], A. Jansson [a], M. Holmberg [b], L. Winblad von Walter [c], M. Stéen [a] and K. Dahlborn [a]

[a]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden; [b]National Veterinary Institute, SVA, Uppsala, Sweden; [c]Växa Sverige, Stockholm, Sweden

ABSTRACT
This study compared the outcome of the Swedish Official Control (OC) with that of two other well-established protocols for assessing animal welfare in dairy cows, 'Ask the Cow' (AC) and Welfare Quality (WQ). Forty-one farms already scheduled for an official control were assessed by the three protocols on the same day. The hypothesis was that farms would be ranked similarly in terms of best and worst, irrespective of the protocol used. A second aim was to explore whether any of the animal-based quantified measures in AC and WQ could be candidates to improve OC. The eight farms with most remarks in OC (3–5) were ranked in the range 19–40 in AC and 5–37 in WQ. The only correlation observed ($r = 0.40$, $P = 0.009$) was between the rankings in AC and OC. Candidate measurements to improve OC taken from AC and WQ are quantifying individual body condition and cleanliness and recording skin lesions.

ARTICLE HISTORY
Received 11 October 2023
Accepted 7 February 2024

KEYWORDS
Animal-based measures; animal welfare assessment; dairy cattle; resource-based measure

Introduction

A high level of animal protection and good animal welfare is important from an animal perspective, but also because consumers demand high-quality food produced by healthy animals kept in a sound environment (Berthe et al., 2012; KilBride et al., 2012). However, there is neither an international protocol nor a gold standard for animal welfare assessment of dairy cows, although the Welfare Quality (WQ) protocol has been used in several scientific studies carried out in different countries (e.g. Radeski et al., 2015; Van Os et al., 2018; Gieseke et al., 2022; Barry et al., 2023). Concerning housing and management of dairy cows, there are yet no specific European Union (EU) regulations in place, but general rules on keeping farm animals are included in Council Directive 98/58/EC (European Commission, 1998). In Sweden, the Animal Welfare Act (2018:1192; Swedish Government, 2018) and Ordinance (2019:66; Swedish Government, 2019) are applicable to animals kept by humans, with specific paragraphs regarding dairy cows. In addition, specific regulations concerning cattle are in place (SJVFS 2019:18, Case No L 104, Swedish Board of Agriculture, 2019).

Assessments of animal welfare are generally based on management-, resource- or animal-related parameters. The first two, also called input-based parameters, refer to animal care and the environment influencing the animals. The outcome, i.e. how the animals are influenced by their environment, is assessed using animal-based parameters (Keeling, 2009; Radeski et al., 2015). Animal-based measures are commonly applied in animal welfare science (Keeling, 2009; Sandgren et al., 2009). In contrast, Swedish animal welfare legislation has a preventative focus (Lundmark et al., 2016), and therefore the official animal welfare control (OC) conducted by County Administration Boards (CABs) in Sweden primarily involves resource-based measures. This is in accordance with current EU legislation, which also relies on providing resources and management (Blokhuis et al., 2010).

In 2010, the Swedish Dairy Association launched an animal-based scoring system called 'Ask the Cow' (AC). This protocol is used as a benchmarking advisory tool and focuses on welfare of cows, young stock and calves. As mentioned, another protocol in use in Europe is the WQ system, which originated from an EU-funded research project running from 2004 to 2009 with the aim of increasing animal welfare in the food production chain. In this study, OC was compared with the outcome of WQ and AC protocols. To our knowledge, this is the first study to apply different protocols in practice to the same set of farms in Sweden. The aim was to compare

CONTACT B. Staaf Larsson Birgitta.Staaf.Larsson@slu.se
Supplemental data for this article can be accessed at https://doi.org/10.1080/09064702.2024.2317708

how 41 dairy farms were ranked using the three different animal welfare protocols with the starting hypothesis that the best and worse farms in OC also would be identified as the best and worst in AC and WQ, respectively. A second aim was to investigate whether the ranking obtained correlated with data on production, reproduction and mastitis on the farms, with the hypothesis that some correlations with production data would be observed, as earlier reported (Sandgren et al., 2009; Otten et al., 2020). Such information could be used to support implementation of good farm practices. If the AC and WQ protocols indicated poor welfare where OC did not, changes in the Swedish official control could be warranted.

Material and methods

Participating counties and farms

To make the work practically feasible, it was carried out in collaboration with the Swedish CABs, i.e. the regional authorities responsible for the OC (checking animal welfare at farms). At one of their meetings, we had the opportunity to present the project, after which four CABs representing most of Sweden geographically voluntarily agreed to participate in the project. Collaboration with CABs also provided the opportunity to train animal welfare inspectors so that they could implement the WQ protocol. The collaboration with the CABs made it possible to visit the farms and carry out all three assessments (OC, AC, WQ) at the same time on the same day. The assessments were performed by three assessors in parallel. The study was carried out in spring 2011, before the cows had access to summer pasture. Forty-one farms were recruited from the four CABs. Ten dairy farms in each of the four CAB areas that were already scheduled for an OC inspection were selected for inclusion, plus an additional farm in one county. Accordingly, since the farms already were planned to undergo an animal welfare control by the CABs, farmers did not participate on a voluntary basis primarily, which could induce a bias towards farmers with special interest in animal welfare, and potentially more homogenous farms from an animal welfare perspective. Nevertheless, the farmers agreed to be assessed by the WQ and AC assessment, and all farmers confirmed in writing their willingness to participate. All farms were part of the Swedish National Milk Recording (NMR) system, but had not been assessed previously by the AC or WQ protocol.

Three animal welfare protocols were compared: (1) OC: Swedish official animal welfare control (Swedish Board of Agriculture, Supplementary Material S1); (2) AC: Ask the Cow (Swedish Dairy Association, Supplementary Material S2); and (3) WQ: Welfare Quality (Welfare Quality Network, Supplementary Material S3). An overview of the parameters assessed in these three different animal welfare protocols is provided in Table 1. More information can be found below and in

Table 1. Comparison of the Official Control (OC), Ask the Cow (AC) and Welfare Quality (WC) protocols in terms of animal-based, management-based and resource-based measures included in assessments on farms with dairy cows.

Category	Parameter	OC[1]	AC	WQ[2]
Animal-based	Body condition	Yes	Yes*	Yes
	Cleanliness	Yes	Yes*	Yes
	Hooves	Yes	Yes	No
	Hairless patches, and lesions/swellings	No	Yes*	Yes
	Lameness	No	Yes	Yes
	Standing/lying in the stall	No	Yes	No
	Rising behaviour	No	Yes	No
	Lying behaviour	No	No	Yes
	Behaviour	No	To a low extent	Yes
	Vermin, parasites	No	Yes*	Yes
	Health	Partly	Yes*	Yes
	Competition at the feeding table	Yes	Yes	Yes
	Calves, young stock	Yes	Yes*	No
	Avoidance distance	No	No	Yes
Management-based	Competent personnel	Yes	No	No
	Daily observations	Yes	No	No
	Daily/yearly cleaning	Yes	No	No
Resource-based	Water access and hygiene	Yes	Yes	Yes
	Space requirements	Yes	No	No
	The stall interior and floor	Yes	Indirect	Indirect
	Air quality	Yes	No	No
	Day light	Yes	No	No
	Bedding material	Yes	No	No
	Noise	Yes	No	No

[1]All questions in the OC checklist are answered 'Yes', 'No', 'Not checked' or 'Not applicable', and relate to all animals on the farm.
[2]Assessment is done on 35 cows, and in some cases (*) on 35 young stock and 35 calves in addition.
[3]Number of animals is 30–73, depending on the size of the herd.

Supplementary Material S1–S3. Two of the farms were not included in the WQ assessment, since an avoidance distance test (see below) could not be carried out due to the stall interior (not enough space in front of the cows to approach them properly). Therefore, the WQ assessment was incomplete for those farms. Another farm was excluded from the WQ ranking due to missing data. Since it turned out that the maximum remarks achieved in OC were five, no detailed ranking was feasible, therefore the level of remarks were compared in OC instead of the rankings.

Training of assessors for the study

The OC inspectors and the AC assessors were experienced and worked regularly with the respective assessment program, so no additional training was provided in these cases. Both OC and AC assessors regularly perform calibration exercises. The WQ protocol had not been used in Sweden before this study, so training on the system was provided for experienced OC animal welfare inspectors from the four participating CABs. This comprised a three-day training course given at the animal research facility at the Swedish University of Agricultural Sciences (SLU), Uppsala, and included practical guidance on live animals, exercises involving assessments of animals on video clips and an inter-observer agreement test. The inter-observer agreement test included 28 WQ parameters, performed on live animals (20 cows), regarding body condition score (BCS), cleanliness, health issues, skin condition, lesions and locomotion (Supplementary Table S1). The two instructors, experienced scientists in animal welfare, decided the correct value for each cow and each parameter (silver standard). If the instructors did not agree, the measure was excluded from the calculations. The inter-observer agreement test on the WQ parameters was in agreement with the silver standard to 88 ± 17% (Range 17–100; Supplementary Table S1).

A total of 24 individuals performed the animal welfare assessments: 10 animal welfare inspectors from the participating CABs carried out OC on farms as routine official inspections, six AC-trained assessors from the Swedish Dairy Association conducted the AC assessments and eight WQ-trained animal welfare inspectors from the participating CABs conducted the WQ assessments (Table 2). Assessors were only responsible for one protocol each (OC, AC or WQ), and all three assessments of each herd were performed at the same time on the same day. Assessor to be used at the different farms was decided by the regional CAB and for AC by the Swedish Dairy Association.

Animal welfare assessment according to the protocols

Swedish official animal welfare control (OC)

During an OC inspection, animal facilities are checked in terms of e.g. space allowance, lighting, noise and air quality (Supplementary Material S1). In the OC checklists, each parameter corresponds to a statutory requirement, i.e. the animal welfare inspector ticks 'Yes' (in compliance with the legislation), 'No', 'Not checked' or 'Not applicable' for the different parameters (Swedish Board of Agriculture, 2022). Questions about water supply to animals in the OC checklist focus on whether the system is designed, dimensioned and positioned so that it allows calm and natural intake, and whether the requirements on access to water and water quality are met (see Supplementary Material S1). The assessment takes around one to two hours, depending on herd size and identified deficiencies.

Ask the Cow (AC)

In the AC protocol, assessments are made at both herd and individual level. At herd level, the Cow Comfort Index (CCI), determined as proportion of cows in cubicles/stalls that are lying down, is assessed, since lying is suggested to be important for good animal welfare (Jensen et al., 2005). Competition between animals for feed is rated as low, medium or high risk of competition behaviour. The number of water bowls or centimetres of water trough is documented, as is the hygiene status of water sources. In addition, the presence of abnormal behaviour, such as stereotypies, is recorded. At the individual cow level, rising behaviour is assessed, where the cow should be able to rise without difficulty with no hesitation longer than five seconds. Body condition score, cleanliness, hoof condition, skin lesions and lameness

Table 2. Assessors (A–Y) applying the official control (OC), Ask the Cow (AC) and the Welfare Quality (WQ) protocols on the 41 participating dairy farms (1–41). Number of farms assessed by each assessor is given in brackets.

Assessment protocol	Farm 1–10	Farm 11–20	Farm 21–31	Farm 32–41
OC	A(5), B(2), C(1), D(1), E(1)	F(10)	G(9), H(2)	I(7), J(3)
AC	K(10),	L(10)	M(11)	N(4), O(4), *P*(2)
WQ	Q(5), R(5)	S(10) + T(4)[1]	U(11) + V(11)[1]	X(10) + Y(2)[1]

[1]The assessments were made in collaboration.

are recorded (see Supplementary Material S2). A farm visit takes around three to four hours.

Welfare Quality (WQ)

The WQ protocol is based on four principles: good feeding, good housing, good health and appropriate behaviour. A WQ assessment starts with an 'avoidance distance test', assessing the human–animal interaction. The assessor approaches cows standing at the feeding table until signs of animal withdrawal emerge or until the assessor can touch the muzzle of the cow (Welfare Quality, 2009; Radeski et al., 2015). If the cow avoids the assessor, the observed distance between assessor and cow is estimated. The assessor also assesses social behaviour (taking into account only aggressive interactions such as head butt, displacement and chasing up), lying behaviour, body condition, cleanliness and lesions. In addition, a qualitative behaviour assessment of how the animals behave and interact with each other is performed by observing the entire herd for 20 min and assessing 20 different parameters on a min–max scale, to gain an overall view of the herd and the expressive quality of the activity at group level. Parameters used at herd level include observations on whether the animals are active, relaxed, fearful, agitated, calm, content, indifferent, frustrated, friendly, bored, playful, positively occupied, lively, inquisitive, irritable, calmless/uneasy, sociable, apathetic, happy or distressed (Welfare Quality, 2009). The number of water bowls or length of water trough is documented (should be more than one water point per cow, according to WQ) and water flow and cleanliness of water sources are noted. For a short version of the WQ assessment protocol, see Supplementary Material S3. The duration of the assessment varies between four and eight hours (Blokhuis et al., 2010), depending on the size of the herd. WQ only assesses dairy cows (adult animals) and contains no protocols for assessing calves and young stock (Brscic et al., 2019).

Herds, production data and health data

The number of cows per herd in the 41 herds ranged from 12 to 268. The largest herd had almost twice as many cows as the second largest ($n = 139$), and the median herd size was 55 cows (mean 65 cows). The average herd size in Sweden at the time was 66 cows (Swedish Board of Agriculture, 2023). Twenty-two of the participating herds had tie-stalls and 19 had a loose-housing system. Yearly milk production was 9554 ± 897 kg energy-corrected milk (ECM) per cow (range 7676–11,855 kg ECM/cow). Cow- and herd-level data were obtained from the NMR database, including information about milk production, reproduction, health and mastitis history for individual cows.

Data processing and statistical analyses

The AC ranking was calculated from the outcome of 12 parameters: proportion of lean, fat, dirty, severely dirty and lame cows, proportion of cows with long hooves, asymmetric hooves, lesions, severe lesions, rising problems, and proportion of cows lying outside the cubicles and standing in the cubicles, based on the 35 assessed cows. The overall ranking of the herds according to the WQ protocol was based on the sum of scores for the four welfare principles: good feeding, good housing, good health and appropriate behaviour. These calculations were made by staff at the French National Institute for Agriculture, Food, and Environment (INRAE, France), who originally took part in developing the WQ protocol for dairy cows. Based on the outcomes of the WQ and AC assessments, the herds were ranked 1–38 or 41 (where 1 was the best and 38/41 the worst). If two herds received the same value, this was considered a tied ranking and the next number in the series was excluded. The ranking of farms according to OC was based on the number of negative remarks, i.e. less complains resulting in a lower (better) ranking.

The proportions of lean and fat cows were analysed by a *t*-test. The proportions of dirty cows and lesions were not analysed statistically, since the methods for registration differed. Correlation analysis was performed between remarks on OC and rankings in WQ and AC protocols and between remarks/rankings and NMR data (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA) using the PROC CORR function and regression equations were created in Excel (Microsoft Excel, Microsoft Corp., Redmond, Washington, USA). Rankings of tie-stalls and loose-housing systems according to the different protocols were compared using a Wilcoxon two-sample test (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA). For correlations and comparisons, P-values <0.05 were considered significant. Results are presented as mean ± SD if not stated otherwise.

Results

Ranking of herds and correlations between protocols

The number of (negative) remarks per herd in OC ranged from 0 to 5 and the eight farms with most remarks (3–5 remarks) were ranked in the interval 19–40 in AC and 5–37 in WQ (Table 3). There were 18 farms with no remarks

Table 3. Number of negative remarks made about the 41 dairy farms in the Official Control (OC) and farm ranking according to the Ask the Cow (AC) and Welfare Quality (WQ) protocols. There are three missing values for WQ, two since avoidance distance not could be assessed on two farms and one due to missing values, so the total value could not be calculated.

Remarks in OC	Rank AC	Rank WQ
0	1	16
0	6	35
0	8	23
0	9	17
0	11	29
0	12	21
0	13	3
0	17	
0	17	6
0	19	14
0	23	18
0	24	20
0	26	12
0	28	36
0	30	33
0	32	8
0	33	1
0	36	
1	1	7
1	5	25
1	7	13
1	15	2
1	16	38
1	19	28
1	25	31
1	29	4
1	34	15
2	3	10
2	3	9
2	9	30
2	14	34
2	31	11
2	41	32
3	19	22
3	27	5
4	19	27
4	40	26
5	35	19
5	37	
5	38	24
5	39	37

in OC, and these were within the ranking range 1–36 in both AC and WQ (Table 3). There was a weak to moderate positive correlation ($r = 0.40$, $P = 0.009$) between remarks in OC and the AC ranking, but no correlation ($P > 0.05$) between OC remarks and the WQ ranking, or between the AC and WQ rankings (Figure 1).

The 41 herds as a whole received 55 remarks in OC out of 1763 possible (41 farms × 44 parameters, i.e. 3%), the most common ($n = 11$) being lack of a back-up system to ensure sufficient air regeneration in a mechanically ventilated animal house and/or lack of an alarm system to signal a ventilation breakdown. The second most common remark ($n = 8$) concerned calf housing, e.g. calves were kept in individual pens after eight weeks of age, which is not permitted under EU legislation (Council Directive 2008/119/EC of 18 December 2008, European Commission, 2008), and thus not under Swedish legislation. In AC, the 41 herds received 624 remarks out of 1876 possible (33%), with the most common remarks concerning lesions (39 of 41 farms, mean 28% of cows) and asymmetrical hooves (33 farms, mean 23% of cows). In WQ, there were 680 remarks out of 1083 possible (63%), the most frequent being that all 41 farms had cows with dirty legs (mean 68% of cows at herd level), dirty hindquarters (50%), hairless patches (61%) or lesions (32%).

Tie-stalls versus loose-housing system

In WQ, tie-stalls were rated significantly lower (mean ranking 25 ± 10) than loose-housing systems (mean ranking 13 ± 10) ($P = 0.001$), but there were no differences according to AC ($P = 0.5$, mean ranking 20 ± 13 for tie-stalls and 22 ± 11 for loose-housing systems) or OC ($P = 0.7$, mean number of negative remarks 1.4 ± 1.7 for tie-stalls and 1.3 ± 1.7 for loose-housing systems).

Correlations between protocols and NRM data

No correlations were found between the NRM data (kg ECM, calf mortality, mastitis, percentage of cows with >70 days between calving and first insemination) and the ranking obtained with the different protocols, with the exception of percentage of cows failing to become pregnant within 120 days of calving, which was correlated with the OC remarks ($r = 0.36$, $P = 0.02$; Supplementary Table S2). See Supplementary Table S3 for all mean values (± SD) of the three protocols and the NRM data.

Comparison between protocols on body condition

There were no remarks about lean or fat cows for any of the 41 farms when assessed with OC, while there were 18 farms with no lean cows according to AC and eight according to WQ (Figure 2). The 23 farms with lean cows according to AC had on average 6% lean cows (range 2–17%). The corresponding figure for WQ was 33 farms with on average 8% lean cows (range 2–56%). There were no differences in the proportion of lean cows in AC and WQ (3 ± 4% and 7 ± 10%, respectively; $P = 0.5$). There was a significant difference between cows scored as fat by AC and WQ (6 ± 8% and 4 ± 5%, respectively; $P = 0.02$). There were 15 farms with no fat cows according to AC and 18 according to WQ (Figure 2). The 26 farms with fat cows according to AC had on average 9% fat cows (range 3–40%). The corresponding figure for WQ was 23 farms with on average 7% fat cows (range 1–15%).

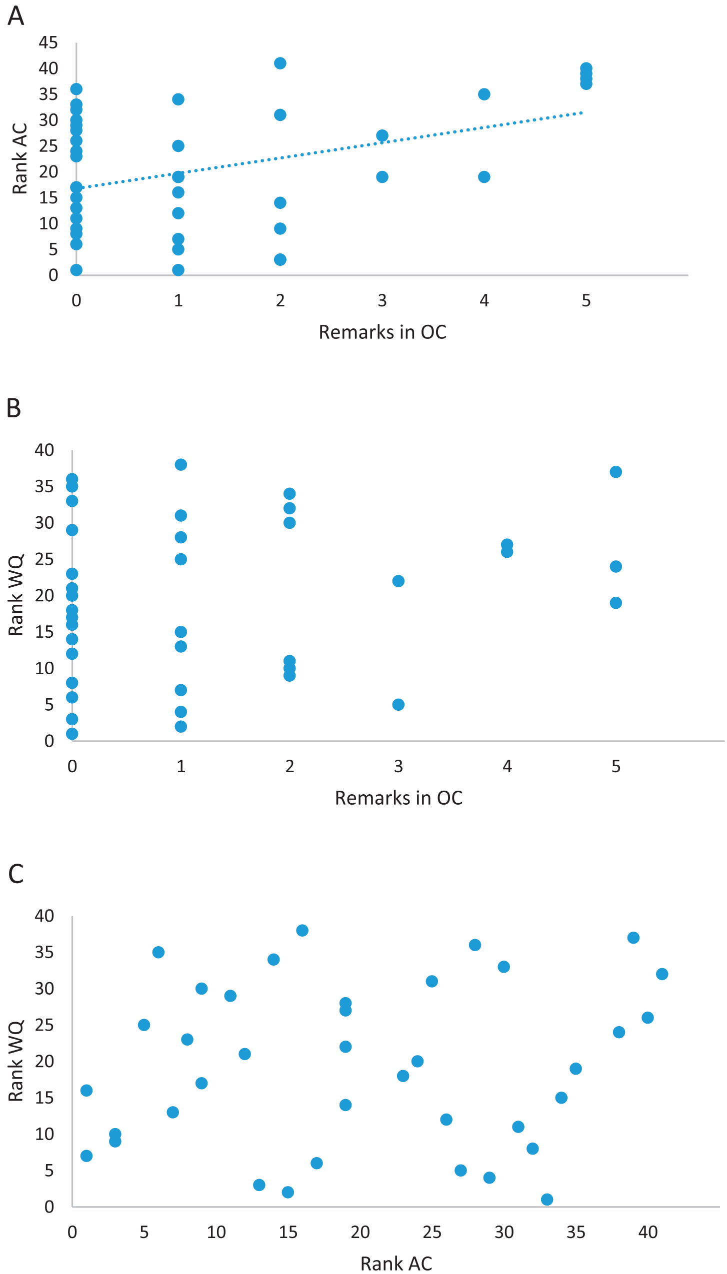

Figure 1. Correlation between (A) number of remarks in the Official Control (OC) and ranking of the 41 dairy farms using the 'Ask the Cow' (AC) protocol ($r = 0.40$, $P = 0.009$, $y = 2952x + 16{,}796$), (B) number of remarks in OC and ranking using the Welfare Quality (WQ) protocol ($P > 0.05$) and (C) the AC and WQ rankings. Correlations based on 41 farms, except for WQ ($n = 38$ farms).

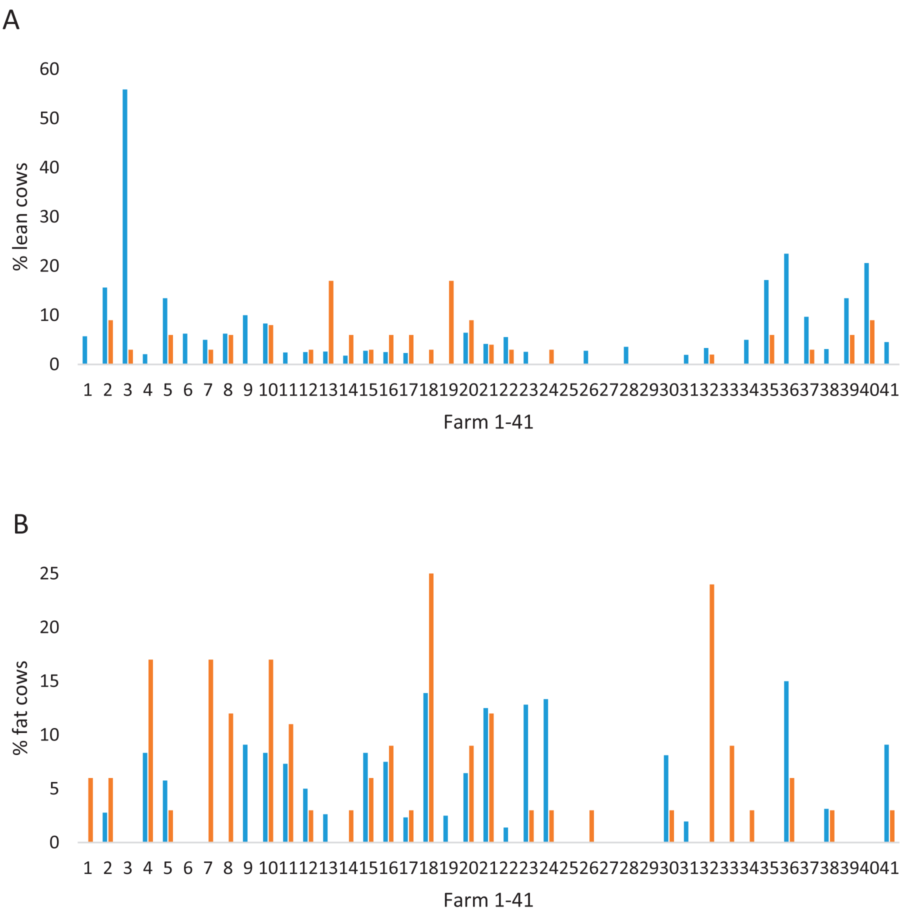

Figure 2. Proportion of lean cows (A) and fat cows (B) per herd in the Ask the Cow (AC) assessment (red bars) and Welfare Quality (WQ) assessment (blue bars). Farm number 1–41 is not related to the ranking.

Comparison between protocols on cleanliness

Only two herds received remarks about dirty cows in the OC assessment, but according to the AC assessment 31 herds out of 41 had dirty cows. In two herds, almost all cows (97%) were assessed as dirty in AC, and cows were assessed as severely dirty in seven herds. Ten herds did not have any dirty animals at all, according to the AC assessment (Figure 3). The 31 herds with dirty cows according to AC had on average 23% dirty cows (range 3–97%) and seven herds had severely dirty cows (mean 7%, range 3–20%).

According to the WQ assessments, all herds had cows with dirty legs and flanks (mean 68% and 50% respectively, range 8–100% and 4–97%, respectively), and all herds except one had cows with dirty udders (Figure 3). The 40 herds with dirty udders according to WQ had on average 44% cows with dirty udders (range 3–85%).

Comparison between protocols on skin lesions

The proportion of cows assessed as having skin lesions varied between 2% and 75% (32 ± 17%) in WQ and 0 and 68% (28 ± 17%) in AC (Figure 4). The proportion of cows with severe lesions according to AC was 5 ± 7% and varied between 0% and 26%. Cows on two farms had no lesions and cows on 16 farms had no severe lesions according to AC. The 39 herds with lesions had 6–68% (mean 30%) cows with lesions, and the 25 herds with severe lesions according to AC had 3–26% (mean 13%) cows with severe lesions. In WQ

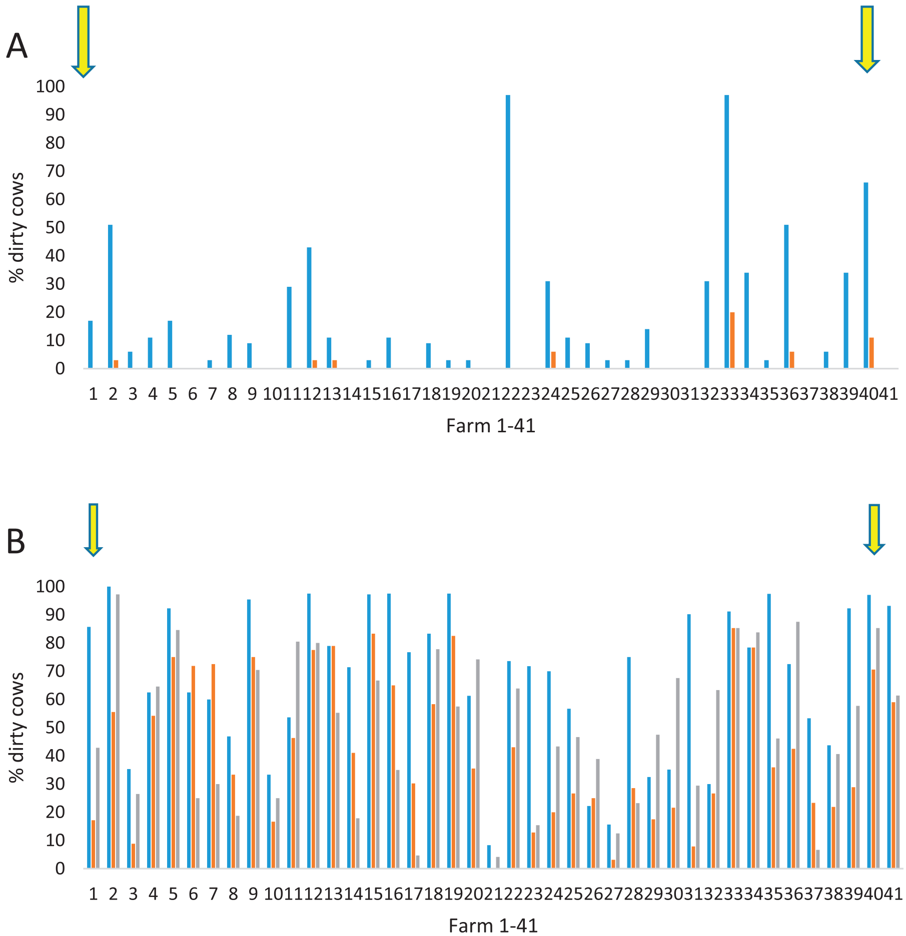

Figure 3. Proportion of (A) dirty cows (blue bars) and severely dirty cows (red bars) per herd in the Ask the Cow (AC) assessment and (B) proportion of cows per herd with dirty legs (blue bars), dirty udder (red bars) and dirty hindquarters (green bars) in the Welfare Quality (WQ) assessment. Farms that received negative remarks in Official Control (OC) are indicated with yellow arrows. Farm number 1–41 is not related to the ranking.

assessments, no farm was found to have any lesions in the herd. There was no correlation ($P > 0.05$) regarding lesions between AC and WQ. Skin lesions are not covered by the OC checklist.

Water supply and water quality

In OC, only one farm received a remark regarding water supply and/or the quality of the water. In AC assessments, seven farms received remarks on water hygiene and nine did not have a sufficient supply of water according to the AC protocol. In WQ assessments, three of the farms were considered not to have sufficient cleanliness of the water source and 31 farms were categorised as not having sufficient water flow.

Discussion

The aim of this study was to compare how 41 dairy farms were ranked using three animal welfare protocols, and to our knowledge, this is the first study in Sweden to do this. The starting hypothesis in this study, that the same farms would be identified as best and worse irrespective of the protocol used, was not supported by the results. The ranking of the farms differed substantially between the assessment systems, although a

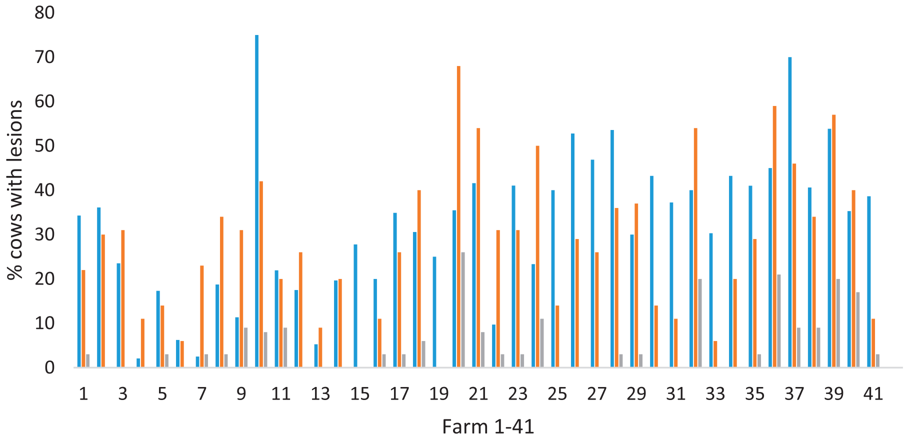

Figure 4. Proportion of cows with skin lesions in the Welfare Quality (WQ) assessment (blue bars) and Ask the Cow (AC) assessment (red bars = skin lesions, green bars = severe skin lesions). Farm number 1–41 is not related to the ranking.

moderate positive correlation was found between the outcomes of AC and OC. In contrast to our study, Otten et al. (2020) reported high agreement between two protocols, but both these protocols mainly used similar measurements (i.e. WQ or WQ based). The low agreement between the outcomes of the three protocols in our study is probably an effect of the use of different parameters, different scoring systems and scales, and different sample sizes. Stull et al. (2005) compared three assessment programs on 10 dairy farms in California, USA. The three assessment programs did not rank the farms similar, although all programs identified the same two farms with the lowest ranking.

However, the use of different assessors for the three protocols, might also have had impact on the agreement. A previous study by Knierim & Winckler (2009) raised the challenge of low reliability when assessing animal-based welfare parameters on farms and pointed out that robust agreement over time and/or between assessors is difficult to achieve (except for the 'avoidance distance' parameter in their study). However, the WQ inter-observer test in this study showed relatively high agreement between assessors. Another methodological strategy could have been to cross-train all observers on all three programs and randomly assign them to use a particular assessment type on a given farm. However, this was not practically feasible in the present study because the 41 farms were spread across Sweden. A strength of the present study is that experienced professional assessors carried out the assessments on the same day on each farm, which ensured professionalism and that conditions were the same although different protocols were used. As explained earlier, no joint training were performed within OC and AC, since their inspectors were experienced professional assessors. However, an inter-observer test would have been beneficial to investigate whether there were any differences in assessment outcomes between assessors.

In the OC system, the assessment of animal welfare is based to a low extent on animal-related parameters (e.g. cleanliness and body condition), with no documentation in the assessment protocol on number of animals affected and level of body condition. In OC, there were no remarks on body condition, whereas the AC and WQ outcomes showed that several cows were considered both lean and fat. There was no significant difference in the proportion of lean cows between the AC and WQ protocols (3% lean cows in AC and 7% in WQ), indicating that the two protocols work similarly for this parameter. The proportion of lean cows was similar to that in a previous study using WQ (Tremetsberger et al., 2019). In our study, one farm in the WQ protocol deviated by more than 3 SD from the mean regarding lean cows. However, that was the first WQ assessment by the evaluator and unfamiliarity with the protocol or the recording system might have influenced the rating. Another farm deviated by more than 3 SD from the mean regarding fat cows in the AC protocol, with 40% of the cows recorded as fat. However, in that case the assessor also made a comment in text about 'well-fed cows' and it can therefore be assumed that the high number of fat cows was according to that evaluator's opinion, while only 14% of the cows on the same farm were recorded as fat according to the WQ protocol/evaluator.

In both AC and WQ, substantial criticisms about cleanliness were made, i.e. in WQ all 41 farms had cows with dirty legs and dirty hindquarters and in AC 31 herds out of 41 had dirty cows. In OC, only two herds received comments about dirty cows. Both AC and WQ thoroughly evaluate cleanliness and the difference in outcome at herd level in this study was probably because the assessment criteria are different and the sample of animals within a herd was based on different individuals. It may also have been partly due to the dissimilar systems, with 12–35 cows assessed in AC and 12–73 in WQ, and to different assessors evaluating the cows differently. When assessing cleanliness, the entire body is checked in AC, whereas only one side of the body is checked in WQ. A study on Norwegian farms by Hauge et al. (2012) concluded that good housing, feeding and management are essential for keeping animals clean. In Austrian dairy herds, Tremetsberger et al. (2015) found that udder health and cleanliness improved significantly when implementing better husbandry practices, such as enhanced udder cleaning routines.

Lesions were registered in WQ and AC but there were no correlations regarding lesions between the two protocols. This was possibly due to different animals being measured or to the measuring procedure used. As previously mentioned, skin lesions, wounds and lameness are not included in the OC checklist. One lesion may occur by chance, but lesions on multiple cows suggest the presence of a systemic problem potentially leading to pain. By mainly focusing on resource-based measures and farmers' compliance with the legislation, OC excludes much information about individual animal welfare. In the OC checklist, the only question concerning sick animals or animals with lesions is whether the animals are given the necessary care and have access to a separate compartment. The checklist contains no items about the presence or proportion of sick animals or animals with lesions. Therefore, we suggest that including animal-based measurements of skin lesions, wounds and lameness could improve OC and the welfare of individual cows. No conclusions could be reached regarding the protocol with the best validity as regards lesions, so further studies on methodology are needed to determine which measures should be included in OC.

The most common remarks in OC were not animal-based but concerned with building ventilation. This is in agreement with the outcome of the official control in Sweden during the study year, where remarks about the ventilation system were the most common type of negative observation (K. Andersson, Swedish Board of Agriculture, pers. Comm. 17 January 2021). In contrast, neither AC nor WQ registered anything with respect to ventilation of the building, stable temperature, air quality or even animal-based indicators of poor ventilation or thermal comfort. We suggest a combination of resource and animal-based registrations to better monitor thermal comfort and air quality.

All three assessment systems studied include a resource-based measurement of water supply. As mentioned, water supply and water quality are checked in OC, while in AC and WQ the number of water bowls or length of water troughs available is recorded, as is the hygiene status of the water source. In addition, WQ measures water flow. The WQ system often gives tie-stalls a low ranking for the principle of good nutrition, because tied cows often have access to only one water bowl. In a study comparing WQ and the Danish Animal Welfare Index (DAWIN), Otten et al. (2020) concluded that insufficient water supply was the main area of concern according to the WQ protocol. On studying 92 dairy farms in England and Wales, Heath et al. (2014) found that the one resource-based parameter (absence from prolonged thirst) measured in WQ correctly classified 88% of the farms in the same way as the whole WQ assessment.

Tie-stalls were ranked worse than loose-housing systems with the WQ protocol, but no difference in ranking was observed with the AC or OC protocols. The reason for the poor ranking in WQ was that tied cows often had access to only one water bowl and were less able to express natural behaviours, since they had no access to daily exercise and were only kept loose when grazing during summer. Only one of the 22 herds with tie-stalls in this study offered the cows exercise all year around.

An animal welfare protocol used for official control must be based on validated methods to register animal welfare, have legal certainty but also be practically feasible to perform at the farm level. One challenge with the WQ assessment is that the behaviour tests are time-consuming (de Vries et al., 2013). Both the avoidance test evaluating the human–animal interaction, standing still looking at the animals' behaviour during the quality behaviour observation (20 min), and waiting for the cows for measuring and evaluating the lying down behaviour takes time. All assessors in our study reported that WQ requires too much time to be implemented fully in OC. Inclusion of new registrations to a protocol (e.g. OC) must therefore be made by weighing time consumption and the value and quality of the registrations for assessing animal welfare.

Conclusions

There were marked differences in assessment outcomes for the three protocols compared in this study, which is unsatisfactory from a legal perspective. Contrary to our hypothesis that the worst (with most remarks) and best (with fewest remarks) farms would be identified similarly by all three protocols, the results obtained showed the opposite, although there was a moderate correlation between the rankings in AC and OC. Substantial remarks on body condition and cleanliness were made according to AC and WQ, but not according to the OC system, with which remarks concerning the robustness of on-farm ventilation systems were instead most common. Skin lesions, wounds and lameness, which are not even measured in OC, received high numbers of remarks according to AC and WQ. We therefore suggest including animal-based measurements of skin lesions, wounds and lameness in order to improve OC and the welfare of individual cows. Including the amount of animals affected, and the level of dirtiness and body condition in the assessment would further improve the OC protocol. However, the outcome of animal-based assessments is highly dependent on the methods used and further studies are needed to develop and validate methods.

Ethics approval

Ethical review and approval from an animal ethics committee were not required for this study, since Swedish regulations explicitly state that animal experiments using privately owned animals kept in their normal environment and not subjected to any painful or stressful procedures may be conducted without an operating licence or ethical approval for animal research.

Data and model availability statement

The datasets considered used in this study contain sensitive information and cannot be made publicly accessible, but they are available (with participating farms anonymised) from the corresponding author upon reasonable request.

Acknowledgements

We are grateful to the farmers who participated in this study and to the animal welfare inspectors and all other participants for making the study possible. Thanks also to Christoph Winckler and Lukas Tremetsberger, University of Natural Resources and Life Sciences (BOKU) Austria, who performed the WQ training; Isabelle Veissier and Anne Lamadon, French National Institute for Agriculture, Food, and Environment (INRAE) France, who conducted the WQ calculations; Claudia von Brömssen at the Department of Energy and Technology, Swedish University of Agricultural Sciences, for statistical advice; and Mary McAfee for English editing. Part of this study has previously been published as the MSc thesis of the first author (Staaf Larsson, 2014).

Disclosure statement

Louise Winblad von Walter is employed at Växa Sverige, former Swedish Dairy Association, but she was not involved in the collection of data or analysis of the data. Birgitta Staaf Larsson, Anna Jansson, Mia Holmberg, Margareta Stéen and Kristina Dahlborn have no competing interests to declare.

Funding

This work was supported by the Swedish Government [grant number 1:23 anslagspost 7, punkt 13]; Svenska Djurskyddsföreningen [grant number 2014-05-06].

Financial support statement

This work was supported by the Swedish Government through the Swedish Centre for Animal Welfare, the Swedish University for Agricultural Sciences, and by the Swedish Association for the Protection of Animals (for preparation of this article).

ORCID

B. Staaf Larsson http://orcid.org/0000-0002-6748-6047
A. Jansson http://orcid.org/0000-0001-5612-9020
M. Holmberg http://orcid.org/0000-0002-2497-7647
L. Winblad von Walter http://orcid.org/0000-0002-7966-1842
M. Stéen http://orcid.org/0000-0001-8224-0224
K. Dahlborn http://orcid.org/0000-0002-1029-9013

References

Barry, C., Ellingsen-Dalskau, K., Garmo, R. T., Grønmo Kischel, S., Winckler, C. & Kielland, C. (2023). Obtaining an animal welfare status in Norwegian dairy herds – A mountain to climb. *Frontiers in Veterinary Science*, 10, 1125860. doi:10.3389/fvets.2023.1125860

Berthe, F., Vannier, P., Have, P., Serratosa, J., Bastino, E., Broom, D. M., Hartung, J. & Sharp, J. M. (2012). Special issue: The role of EFSA in assessing and promoting animal health and welfare. *EFSA Journal*, 10(10), s1002. (10 pp.). doi:10.2903/j.efsa.2012.s1002. Available online: www.efsa.europa.eu/efsajournal.

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.

Brscic, M., Otten, N. D., Contiero, B. & Kirchner, M. K. (2019). Investigation of a standardized qualitative behaviour assessment and exploration of potential influencing factors on the emotional state of dairy calves. *Animals*, 9(10), 757.

de Vries, M., Engel, B., den Uijl, I., van Schaik, G., Dijkstra, T., de Boer, I. J. M. & Bokkers, E. A. M. (2013). Assessment time of the Welfare Quality (protocol for dairy cattle). *Animal Welfare*, 22, 85–93.

European Commission. (1998). Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. *Official Journal of the European Communities*, L 221, 8.8.98, 23–27.

European Commission. (2008). Council Directive 2008/119/EC of 18 December 2008 laying down minimum standards for the protection of calves. *Official Journal of the European Communities*, L 10, 15.1.2009, 8.

Gieseke, D., Lambertz, C. & Gauly, M. (2022). Effects of housing and management factors on selected indicators of the welfare quality® protocol in loose-housed dairy cows. *Veterinary Sciences*, 9, 353. doi:10.3390/vetsci9070353

Hauge, S. J., Kielland, C., Ringdal, G., Skjerve, E. & Nafstad, O. (2012). Factors associated with cattle cleanliness on Norwegian dairy farms. *Journal of Dairy Science*, 7, 3213–3282.

Heath, C. A. E., Browne, W. J., Mullan, S. & Main, D. C. J. (2014). Navigating the iceberg: reducing the number of parameters within the Welfare Quality® assessment protocol for dairy cows. *Animal*, 8(12), 1978–1986.

Jensen, M. B., Pedersen, L. J. & Munksgaard, L. (2005). The effect of reward duration on demand functions for rest in dairy heifers and lying requirements as measured by demand functions. *Applied Animal Behaviour Science*, 90, 207–217. doi:10.1016/j.applanim.2004.08.006

Keeling, L. J. (2009). An analysis of animal-based versus resource-based comments in official animal welfare inspection reports from organic and conventional farms in Sweden. *Animal Welfare*, 18, 391–397.

KilBride, A. L., Mason, S. A., Honeyman, P. C., Pritchard, D. G., Hepple, S. & Green, L. E. (2012). Associations between membership of farm assurance and organic certification schemes and compliance with animal welfare legislation. *Veterinary Record*, 170, 152.

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(4), 451–458.

Lundmark, F., Röcklinsberg, H., Wahlberg, B. & Berg, C. (2016). Content and structure of Swedish animal welfare legislation and private standards for dairy cattle. *Acta Agriculturae Scandinavica Section A-Animal Science*, 66(1), 35–42.

Otten, N. D., Rousing, T., de Oliveira, V. H. S., Reiten, M., Michelsen, A. M., Hakansson, F., Lund, V. P., Houe, H., Denwood, M., Sorensen, J. T., Forkman, B. & Kirchner, M. K. (2020). A weak point analysis of welfare in Danish dairy herds using two different welfare assessment systems. *Animal Welfare*, 29(2), 197–207.

Radeski, M., Janevski, A. & Ilieski, V. (2015). Screening of selected indicators of dairy cattle welfare in Macedonia. *Macedonian Veterinary Review*, 38(1), 43–51.

Sandgren, C. H., Lindberg, A. & Keeling, L. J. (2009). Using a national dairy database to identify herds with poor welfare. *Animal Welfare*, 18, 523–532.

Staaf Larsson, B. (2014). *Comparison of three assessment programmes in Sweden; The official animal-based welfare control, Ask the Cow and Welfare Quality®*. MSc thesis, Gothenburg, Sweden: University of Gothenburg.

Stull, C. L., Reed, B. A. & Berry, S. L. (2005). A comparison of three animal welfare assessment programs on California dairies. *Journal of Dairy Science*, 88(4), 1595–1600. doi: 10.3168/jds.S0022-0302(05)72828-9. PMID: 15778329.

Swedish Board of Agriculture. (2019). Statens jordbruksverks föreskrifter och allmänna råd (SJVFS 2019:18) om nötkreaturshållning inom lantbruket m.m. [The Swedish Board of Agriculture's regulations and general recommendations (SJVFS 2019:18) on cattle husbandry in agriculture, etc.], Case No L 104.

Swedish Board of Agriculture. (2022). *Vägledning för offentlig djurskyddskontroll ['Guidelines for official animal welfare control']; Swedish Board of Agriculture: Jönköping* (Sweden).

Swedish Board of Agriculture. (2023). Antal djur och jordbruksföretag med djur efter Län, Kategori, Variabel, Tabelluppgift och År. PxWeb (sjv.se). [Number of animals and agricultural enterprises with animals by County, Category, Variable, Tabular information and Year.] Assessed 1 February 2023.

Swedish Government. (2018). Animal Welfare Act (2018:1192).

Swedish Government. (2019). Animal Welfare Ordinance (2019:66).

Tremetsberger, L., Leeb, C. & Winckler, C. (2015). Animal health and welfare planning improves udder health and cleanliness but not leg health in Austrian dairy herds. *Journal of Dairy Science*, 98(10), 6801–6811.

Tremetsberger, L., Winckler, C. & Kantelhardt, J. (2019). Animal health and welfare state and technical efficiency of dairy farms: possible synergies. *Animal Welfare*, 28(3), 345–352.

Van Os, J. M. C., Winckler, C., Trieb, J., Matarazzo, S. V., Lehenbauer, T. W., Champagne, J. D. & Tucker, C. B. (2018). Reliability of sampling strategies for measuring dairy cattle welfare on commercial farms. *Journal of Dairy Science*, 101, 1495–1504. doi:10.3168/jds.2017-13611

Welfare Quality®. (2009). *Welfare quality® assessment protocol for cattle. welfare quality®*. Consortium, Lelystad, Netherlands. http://www.welfarequality.net/media/1088/cattle_protocol_without_veal_calves.pdf.

II

ACTA AGRICULTURAE SCANDINAVICA, SECTION A — ANIMAL SCIENCE
2021, VOL. 70, NO. 1, 31–40
https://doi.org/10.1080/09064702.2020.1849380

Visual assessment of body condition and skin soiling in cattle by professionals and undergraduate students using photo slides

B. Staaf Larsson [a,b], E. Petersson [c], M. Stéen [a,b] and J. Hultgren [d]

[a]Swedish Centre for Animal Welfare, Swedish University of Agricultural Sciences, Uppsala, Sweden; [b]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden; [c]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm, Sweden; [d]Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Skara, Sweden

ABSTRACT

We aimed to study to what extent body condition and skin soiling in cattle are rated differently depending on the profession, education and professional experience of assessors. Data were collected at 15 group-training sessions in the years 2009–2016. Totally, 569 persons; Swedish animal welfare inspectors, veterinarians/advisers, animal welfare scientists, other animal professionals as well as undergraduate students in animal science rated a set of 6–40 photos with respect to animal body condition, animal skin soiling, and recommended corrective action in response to perceived animal soiling. The more extensive animal science education, the less prone animal welfare inspectors were to give higher body condition scores. Students had a higher overall probability of rating the degree of soiling higher than the animal welfare inspectors and veterinarians/advisers. Students also recommended stricter corrective actions than did welfare inspectors and veterinarians/advisers, and veterinarians/advisers recommended less strict actions than animal welfare inspectors.

ARTICLE HISTORY
Received 7 October 2020
Accepted 6 November 2020

KEYWORDS
Animal welfare; body condition; cattle; cleanliness; interactive audience response; skin soiling

Introduction

Good animal welfare (AW) is to some extent maintained through national and EU legislation and official control (European Parliament, 2017). Animal welfare legislation and official control are mainly preventive, focusing on potential risks in animal housing and management (Blokhuis et al., 2010; Broom, 2017). Resource-based measures are based on observations of the animal's environment and resources such as space allowance or air quality, while management-based measures include caretaking strategies and animal handling (Keeling, 2009). However, the European Commission (2012) considers the increased use of output-based measures, which reflect AW *per se*.

Complaints about body condition and skin soiling are common AW issues in cattle husbandry (Keeling, 2009). Extreme thinness, as well as obesity, increases the risk of diseases like milk fever, retained placenta, endometritis, ketosis, abomasal displacement and dystocia in cows (Gillund et al., 2001; Roche & Berry, 2006). Thinness has been associated with low milk production (Roche et al., 2007), low conception rate (Pryce et al., 2001) and an increased risk for sole ulcer and white line disease (Green et al., 2014). Green et al. (2014) showed that cows with a body condition score <2.5 (on a scale from 1 = thin to 5 = fat) are more likely to become lame. Over-conditioning may also cause reduced milk yield (Gillund et al., 2001).

There are several reasons for assessing skin soiling in farm animals. At official Swedish AW controls, the most prevalent recorded non-compliance is soiled animals (Keeling, 2009). Improved cattle cleanliness has many benefits, for example, strengthened food safety (Hughes, 2001), increased profits through intact hides at slaughter (Nafstad, 1999), reduced mastitis incidence (Hughes, 2001) and improved animal comfort. Known complications of chronic faecal soiling are etching of the skin, infections and irritation (Nafstad, 1999). Soiled cows have an increased risk of getting mastitis (Breen et al., 2009). It is painful to remove dry lumps of manure from the skin, faecal soiling causes skin lesions (Hauge et al., 2012), and damaged hides can be an economic setback for the farmer after slaughter (Nafstad, 1999). There are several risk factors for soiling, related to building design, management and stockmanship (Radeski et al., 2015). Hughes (2001) introduced a cleanliness

CONTACT B. Staaf Larsson birgitta.staaf.larsson@slu.se

scoring system, evaluating four separate skin areas on the cow to indicate the cause of soiling.

Interrater agreement (interrater reliability) is the degree of agreement between different assessors. It can be estimated in several ways (Gwet, 2014). Percentage agreement (joint probability of agreement) is the simplest measure in a nominal or ordinal rating system, but it does not account for random agreement. Other statistics that have been proposed for nominal or ordinal data with more than two levels, correcting for the fact that agreement may happen by chance, include Fleiss' kappa (Fleiss, 1971), generalised kappa (Abraira & Pérez de Vargas, 1999) and Gwet's agreement coefficient, AC1 (Gwet, 2008, 2014). Interpretation of agreement values has been suggested as 'none to slight' for 0.01–0.20, 'fair' for 0.21–0.40, 'moderate' for 0.41–0.60, 'substantial' for 0.61–0.80 and 'almost perfect' for 0.81–1.00 (Landis & Koch, 1977). Negative values indicate systematic disagreement. In contrast to kappa, Gwet's agreement coefficient is less sensitive to trait prevalence and marginal homogeneity and does not depend upon independence between assessors, which makes it versatile. The agreement between assessors regarding body condition and skin soiling in cattle, and regarding the perceived need for corrective control actions at an official AW control, has not been studied before. Nor have factors that influence these assessments.

A fair and secure legal treatment of farm-animal husbandry requires uniform and objective official AW control (Anneberg et al., 2012; Andrade & Anneberg, 2014). As Ruddat et al. (2014) pointed out, AW assessment is challenging, especially when no gold standard is available. Assessments of animal-based measures such as body condition and skin soiling require valid and reliable scoring systems and well-trained assessors. Further research is also needed to develop training programmes for AW inspectors. A common understanding of AW requirements is important (Sørensen & Fraser, 2010; Anneberg et al., 2012). Berg and Lundmark Hedman (2020) highlighted the importance of training programs, guidelines, and checklists for the inspectors, but also for the farmers to know and understand the requirements.

The objective of this study was to investigate how different categories of professionals and students assess important animal-based measures of cattle welfare and the need for corrective actions in response to a perceived violation of legislation at an imaginary official AW control visit. A second objective was to estimate the interrater agreement for these assessments. We hypothesised that previous animal-related education and professional experience with AW result in a more uniform and consistent assessment, compared to no such education and less experience, and that long professional AW practice improves interrater agreement.

Material and methods

Data were collected in 2009–2016 during 15 group sessions organised and chaired by the first author. Sessions were as uniform as possible and typically lasted for 1.5 h. The sessions aimed to improve the participants' understanding of AW science and skills for conducting official AW control inspections, and followed a standardised protocol, with an introductory lecture on the assessment of body condition and soiling in cattle, including an explanation of rating scales.

The number of participants per session varied between 12 and 80 (mean 38) (Table 1), with a total of 569 Swedish assessors. They were categorised with respect to profession as official 'AW inspectors' ($n = 281$), 'veterinarians/advisers' ($n = 95$), 'AW scientists' ($n = 27$), 'students' ($n = 88$) or 'other profession' ($n = 64$). Veterinarians/advisers included county veterinary officers, practising farm-animal veterinarians, and industry or government advisers. Students were animal science undergraduates, typically at a Master or Bachelor programme at the Swedish University of Agricultural Sciences. 'Other profession' included cattle farmers, veterinary nurses and some individuals not matching any of the other categories. The participants were asked to categorise themselves with respect to their level of education in animal science, as 'full' (university degree), 'part' (courses of more than 10 ECTS university credits) (European Credit Transfer System; European Commission, 2019), or 'none' (10 ECTS or less). They were also asked to state their work experience in their current profession as 'much' (at least 3 years) or 'little' (less than 3 years) as well as gender. Gender information was collected only at the last four sessions. Apart from the introductory lecture, the participants did not receive any special training in the study.

The participants were equipped with individual wireless interactive polling keypads (TurningPoint, version 5.4.1.2, Turning Technologies, Youngstown, Ohio, USA), shown a total of 6–40 photo slides projected on a lecture-hall screen, and asked to assess; (1) animal body condition (0–20 slides; Table 1), (2) the degree of skin soiling (0–11 slides), as well as (3) recommended control actions in response to soiled animals (0–10 slides) at an imaginary official control visit to a cattle farm, henceforth referred to as the three 'assessment domains'. The photo material was collected at regular Swedish AW inspections and study visits to cattle

Table 1. Number of ratings of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides; 15 assessment sessions 2009–2016.

		Number of participants						Number of slides			
Month, year	Profession categories	Little	Much	Full	Part	None	Total	Body condition	Soiling	Recom-mended action	Total
		Experience[a]		Animal-science education[b]							
December, 2009	AWI	27	13	26	11	3	40	18	6	10	34
March, 2009	STU	17	0	17	0	0	17	18	6	10	34
April, 2010	AWI	22	46	44	24	0	74	17	6	10	33
February, 2011	STU	22	2	23	0	1	24	17	6	10	33
May, 2011	AWI, VET, OTH	8	31	31	6	3	43	20	11	9	40
May, 2011	AWI, VET, OTH	13	39	42	12	0	57	20	11	9	40
May, 2011	AWI, VET, OTH	26	27	40	11	5	57	20	11	9	40
May, 2011	AWI, VET, OTH	27	49	55	18	6	80	19	11	8	38
June, 2011	AWI, VET, OTH	6	18	9	12	3	25	20	11	9	40
September, 2011	AWI, VET, OTH	16	18	25	7	1	34	20	11	9	40
October, 2011	VET, OTH	3	8	4	2	5	12	20	0	0	20
February, 2012	STU	17	0	16	2	0	18	20	11	8	39
September, 2013	VET, OTH	3	23	23	3	0	28	20	10	8	38
May, 2014	SCI, OTH	9	21	25	3	1	31	0	1	5	6
November, 2016	STU	20	7	21	5	1	29	19	11	9	39
	Sum:	236	302	401	116	29	569	268	123	123	514

Note: AWI, official AW inspectors; STU, undergraduate students; VET, veterinarians or advisers; SCI, scientists; OTH, other (farmers of dairy cows or cattle farmers, animal nurses or other professionals that do not belong to another professions).
[a]Little = less than 3 years of work experience; Much = at least 3 years of work experience.
[b]Full = university degree in animal science; Part = courses of more than 10 ECTS university credits in animal science; None = maximum 10 ECTS credits in animal science.

farms, and the majority taken by the first author. Since very thin, very obese, and extremely soiled animals are relatively rare, the photos were collected during several years with different cameras and showed one or several cattle of different types, viewed from different angles, in different housing conditions. Most photos displayed cows. Photos for assessing body condition mostly showed only one animal, which in some cases was a calf. Photos for scoring skin soiling were most often taken from behind, and some of the animals were young stock. The participants were present at only one session and assessed each photo only once. At each session, all participants were shown the same slides, for practical reasons the selection of slides varied between sessions.

The participants were given approximately 30 s per slide to respond by pressing the polling keypads, and they were explicitly instructed to make all assessments independently, without consulting fellow participants. Body condition was scored on a 5-level scale (1 = very poor, 2 = moderate, 3 = good, 4 = fat, 5 = very fat; Table 2; Figure 1(a–e)) (modified after Wildman et al., 1982; Svedberg, 2006). Skin soiling was rated on a 4-level scale (1 = clean, 2 = slightly soiled, 3 = moderately soiled, 4 = very soiled; Table 2; Figure 1(f–i)), according to Svedberg (2007). The recommended action was rated on a 4-level scale (1 = no action; 2 = remark; 3 = remark with a follow-up inspection; 4 = prohibition or order; Table 2), similar to different options when conducting a Swedish AW control, where a prohibition or order can be issued with or without a fine, to force the animal keeper to take action.

Statistical analysis

Analyses were made in Stata IC, version 15 (StataCorp, College Station, Texas, USA). To facilitate comparisons between slides, all recorded scores were transformed to standardised scores on a 3-level ordinal scale -1–0–1, calculated as the deviation of each value from the overall sample mode for the slide, as −1 for original scores below the mode, 0 for original scores equal to the mode and 1 for original scores above the mode. For example, if the median recorded body condition score in a slide was 4, a recorded score of 3 or lower was transformed to −1, 4 was transformed to 0 and 5 was transformed to 1. The standardised scores were then modelled statistically using generalised ordinal logistic regression. The participants were arranged in clusters identified by profession categories. The ratings were expected to be independent between profession clusters, but not necessarily within them, which affected the standard errors of coefficients. Estimated regression coefficients were converted to odds ratios (OR) and their 95% confidence intervals were calculated.

The fixed effect of slide number (categorical; 1 through 45) was forced into all models. Initially, independent variables that represented recording session (categorical; 1 through 15), profession (AWI, VET, SCI, STU or OTH), education (short, part or full) and work experience (much or little) were tested in simple models together with slide number. Three multivariable models were then constructed, one for each assessment domain. Variables representing profession and slide number were forced in, while the remaining independent variables were

Table 2. Overall distribution of scores of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides (modified for visual categorisation after Wildman et al., 1982).

Assessment domain	Level	Description of level	Number (percentage) of ratings
Body condition	Very poor	Individual spinous processes had limited flesh covering, were prominent, the ends are clearly visible, and together the processes formed a definite overhanging shelf effect to the loin region. Individual vertebrae of the chine, loin, and rump regions were prominent and distinct, hooks and pin bones were sharp with negligible flesh covering, and severe depressions between hooks and pin bones were noted. The area below the tail head and between the pin bones was severely depressed causing the bone structure of the area to appear extremely sharp.	737 (7.66%)
	Moderate	Individual spinous processes were visually discernible but were not prominent. Ends of processes were sharp although they had a greater flesh covering, and the processes did not have a distinct overhanging shelf effect. Individual vertebrae of chine, loin, and rump regions were not visually distinct. Hooks and pin bones were prominent, but the depression between them was less severe. The area below the tail head and between the pin bones was depressed, but the bone structure was not devoid of flesh covering.	2509 (26.1%)
	Good	Spinous processes were discernible. Together processes appeared smooth and the overhanging shelf effect was not noticeable. Vertebrae of the chine, loin, and rump regions appeared as a rounded ridge, and hooks and pin bones were rounded and smooth. The area between the tail head appeared smooth without a sign of fat deposition.	3761 (39.1%)
	Fat	Individual spinous processes could not be visual distinguished and, together, the processes appeared flat or rounded with no overhanging shelf effect. The ridge formed by the vertebral column of the chine region was rounded and smooth, but loin and rump regions appeared flat. Hooks were rounded, and the span between the hooks was flat. Area around tail head and pin bones was rounded, with evidence of subcutaneous fat deposition.	2226 (23.1%)
	Very fat	Bone structure of the vertebral column, spinous processes, hooks, and pin bone regions was not visually apparent, and evidence of subcutaneous fat deposition was prominent. The tail head appeared to be buried in fatty tissue.	392 (4.07%)
Soiling	Clean	The animals are clean from manure on the flanks, sides and legs.	572 (12.1%)
	Slightly soiled	The animals do not have manure or dirt on the whole body, including the back and sides. Belly, flanks and legs may have a reasonable amount of manure with dry, but not old manure. No layers of manure at all.	1795 (37.9%)
	Moderately soiled	Like 'slightly soiled', but with a certain part with layers of manure on belly, sides and flanks, but no large thick areas.	982 (20.7%)
	Very soiled	The animals are heavily covered with old layers of manure, with large areas on legs and/or parts of the belly, sometimes also on the sides of the body.	1385 (29.3%)
Recommended action	No action	No action taken	1426 (30.5%)
	Remark	Remark given, but no follow-up	1535 (32.8%)
	Remark with follow-up	Remark given and follow-up inspection scheduled	1244 (26.6%)
	Prohibition or order	Prohibition or order issued	470 (10.1%)

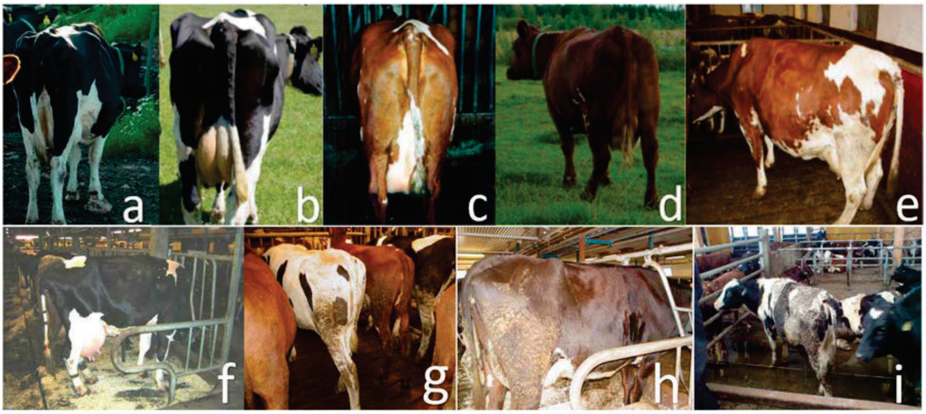

Figure 1. Photos illustrating different degrees of body condition [(a) very poor, (b) moderate, (c) good, (d) fat, (e) very fat] and skin soiling [(f) clean, (g) slightly soiled, (h) moderately soiled, (i) very soiled] in cattle, scored by study participants 2009–2016; Photographers: Jan Svedberg (a–c), Anne Larsen (d,e,g,h) and Birgitta Staaf Larsson (f,i).

tested and retained only if logically relevant. There was no indication of a confounding effect of participant gender, so this factor was excluded from further analysis. Plausible interactions were tested and included if significant at $p \leq 0.05$. Predictive margins were calculated and plotted against different predictor categories.

For each profession category and assessment domain, the agreement between participants was estimated by observed percent agreement (joint probability of agreement), a generalised kappa statistic adapted to ordinal data, multiple observers and incomplete designs (Abraira & Pérez de Vargas, 1999), and Gwet's AC1, treating slides as subjects and using the original scores. Associated 95% confidence intervals were calculated for kappa and Gwet's AC1.

Results

The slides were assessed 56–539 (mean 423) times each, generating totally 19,034 score records. The overall frequencies of the different scores are shown in Table 2 and the distribution between different levels of independent variables and outcome traits used in models are shown in Table 3.

Apart from profession and slide number, covariates included in the final multivariable models were education level and profession × education interaction in the model of body condition, and level of education, professional experience, profession × education interaction, profession × experience interaction in the model of soiling (Table 4). The effect of slide number was significant (joint $P < 0.0005$) for all three assessment domains.

The more extensive animal-related education, the less prone the AW inspectors were to give higher body condition scores (Figure 2). It was not possible to estimate the probabilities of different standardised scores of body condition for veterinarians/advisers with courses of more than 10 ECTS university credits. There were no major differences in the standardised scores of skin soiling between levels of education or professional experience. On the other hand, a statistically significant effect ($p \leq 0.05$) of profession was shown. Students had a higher overall probability of rating the degree of soiling higher than AW inspectors and veterinarians/advisers, regardless of professional experience (Figure 3). For recommended control action in response to soiling, participants with a high level of education or extensive professional experience did not differ significantly from less educated or unexperienced participants. Students recommended stricter AW control actions than did welfare inspectors and veterinarians/advisers, and veterinarians/advisers recommended less strict AW control actions than AW inspectors (Figure 4).

Table 3. Numbers and column percentages (%) of ratings for different participant categories and standardised scores from assessments of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides.

Variable	Level	Body condition ($n = 9625$)			Soiling ($n = 4734$)			Recommended action ($n = 4675$)		
		Below	Equal	Above	Below	Equal	Above	Below	Equal	Above
Profession	AWI	665 (56.0)	3361 (53.4)	942 (49.2)	533 (58.5)	1657 (53.2)	182 (30.9)	257 (53.4)	1549 (56.7)	659 (48.5)
	VET	179 (15.1)	1196 (19.0)	384 (20.1)	200 (22.1)	652 (20.9)	98 (16.6)	112 (23.3)	466 (17.1)	179 (13.2)
	SCI	0	0	0	0	0	0	3 (0.6)	71 (2.6)	53 (3.9)
	STU	217 (18.3)	997 (15.5)	343 (17.9)	43 (4.7)	453 (14.6)	246 (41.7)	53 (11.0)	364 (13.3)	371 (27.3)
	OTH	127 (10.7)	761 (12.1)	244 (12.8)	130 (14.3)	352 (11.3)	64 (10.8)	56 (11.6)	283 (10.4)	96 (7.1)
	Missing	30	152	47	26	89	9	15	65	23
Education level[a]	None	75 (6.4)	313 (5.0)	126 (6.7)	46 (5.1)	156 (5.1)	22 (3.8)	27 (5.7)	120 (4.5)	56 (4.2)
	Part	257 (22.1)	1349 (21.7)	426 (22.6)	212 (23.6)	677 (21.9)	112 (19.4)	93 (19.6)	624 (23.2)	264 (19.8)
	Full	831 (71.5)	4556 (73.3)	1332 (70.7)	639 (71.2)	2255 (73.0)	444 (76.8)	354 (74.7)	1949 (72.4)	1012 (76.0)
	Missing	55	229	76	35	115	21	22	105	49
Professional experience[b]	Little	549 (47.9)	2699 (44.3)	816 (44.0)	319 (36.1)	1305 (43.2)	310 (54.4)	171 (36.5)	1154 (43.5)	691 (52.6)
	Much	597 (52.1)	3395 (55.7)	1037 (56.0)	564 (63.9)	1714 (56.8)	260 (45.6)	297 (63.5)	1498 (56.5)	622 (47.4)
	Missing	72	353	107	49	184	29	28	146	68
Gender	Male	25 (15.0)	109 (12.9)	64 (20.3)	21 (30.9)	63 (13.3)	28 (13.3)	13 (22.8)	63 (19.3)	45 (13.8)
	Female	142 (85.0)	738 (87.1)	252 (79.7)	47 (69.1)	409 (86.7)	182 (86.7)	44 (77.2)	264 (80.7)	280 (86.2)
	Missing	1051	5600	1644	864	2731	389	439	2471	1056

Note: AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student; OTH, other profession.
[a]None = maximum 10 ECTS university credits in animal science; Part = courses of more than 10 ECTS university credits in animal science; Full = university degree in animal science.
[b]Little = less than 3 years of work experience; Much = at least 3 years of work experience.

Table 4. Effects of profession, education level and professional experience in three multivariable generalised[a] ordinal logistic models of standardised scores of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides.

			Body condition (n=9130)		Soiling (n=4610)		Recommended action (n=4375)	
Step	Variable	Level	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
−1 → 0	Intercept	–	4.56 (3.05, 6.83)	<0.0005	109 (49.7, 239)	<0.0005	1.48 (1.21, 1.81)	<0.0005
	Profession	AWI	Ref.		Ref.		Ref.	
		VET	0.94 (0.79, 1.12)	0.47	0.97 (0.68, 1.36)	0.85	0.75 (0.60, 0.95)	0.018
		SCI	–		-		1.53 (0.97, 2.42)	0.066
		STU	0.91 (0.54, 1.54)	0.73	24.1 (8.68, 67.2)	<0.0005	1.91 (1.35, 2.72)	0.0005
		OTH	0.82 (0.57, 1.17)	0.27	1.05 (0.53, 2.10)	0.89	0.75 (0.54, 1.04)	0.081
	Education level[5]	None	1.31 (1.05, 1.63)	0.016	1.08 (0.68, 1.72)	0.75	–	
		Part	Ref.		Ref.		–	
		Full	0.77 (0.66, 0.91)	0.002	1.01 (0.80, 1.27)	0.94	–	
	Profession × Education	STU × Short	0.52 (0.093, 2.91)	0.46	0.18 (0.02, 1.42)	0.10	–	
		STU × Full	1.43 (0.83, 2.48)	0.20	0.32 (0.11, 0.91)	0.033	–	
		VET × Full	1.54 (1.21, 1.95)	<0.0005	–		–	
		OTH × Short	0.90 (0.51, 1.59)	0.71	0.34 (0.11, 1.04)	0.058	–	
		OTH × Full	1.87 (1.23, 2.85)	0.004	0.43 (0.20, 0.94)	0.034	–	
	Experience[6]	Little	–		Ref.		–	
		Much	–		0.82 (0.66, 1.00)	0.055	–	
	Profession × Experience	STU × Much	–		1.29 (0.31, 5.28)	0.72	–	
		VET × Much	–		1.31 (0.85, 2.02)	0.22	–	
		OTH × Much	–		2.95 (1.72, 5.08)	<0.0005	–	
0 –> 1	Intercept	–	0.194 (0.127, 0.296)	<0.0005	0.251 (0.184, 0.342)	<0.0005	0.068 (0.053, 0.088)	<0.0005
	Profession	AWI	Ref.		Ref.		Ref.	
		VET	0.94 (0.79, 1.12)	0.47	0.97 (0.68, 1.36)	0.85	0.75 (0.60, 0.95)	0.018
		SCI	–		–		1.53 (0.97, 2.42)	0.066
		STU	0.91 (0.54, 1.54)	0.73	24.1 (8.68, 67.2)	<0.0005	2.68 (2.14, 3.36)	<0.0005
		OTH	0.82 (0.57, 1.17)	0.27	1.05 (0.53, 2.10)	0.89	0.75 (0.54, 1.04)	0.081
	Education level	Short	1.31 (1.05, 1.63)	0.016	1.08 (0.68, 1.72)	0.75	–	
		Part	Ref.		Ref.		–	
		Full	0.77 (0.66, 0.91)	0.002	1.01 (0.80, 1.27)	0.94	–	
	Profession × Education	STU × Short	0.52 (0.093, 2.91)	0.46	0.18 (0.02, 1.42)	0.10	–	
		STU × Full	1.43 (0.83, 2.48)	0.20	0.32 (0.11, 0.91)	0.033	–	
		VET × Full	1.54 (1.21, 1.95)	<0.0005	–	–	–	
		OTH × Short	0.90 (0.51, 1.59)	0.71	0.34 (0.11, 1.04)	0.058	–	
		OTH × Full	1.87 (1.23, 2.85)	0.004	0.43 (0.20, 0.94)	0.034	–	
	Experience	Little	–		Ref.		–	
		Much	–		0.82 (0.66, 1.00)	0.055	–	
	Profession × Experience	STU × Much	–		1.29 (0.31, 5.28)	0.72	–	
		VET × Much	–		1.31 (0.85, 2.02)	0.22	–	
		OTH × Much	–		2.95 (1.72, 5.08)	<0.0005	–	

Notes: OR, odds ratio; CI, confidence interval; AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student; OTH, other profession.
[a]Variables with identical values for the two steps meet the proportional odds assumption and thus did not require separate estimations.
[b]None = maximum 10 ECTS university credits in animal science; Part = courses of more than 10 ECTS university credits in animal science; Full = university degree in animal science.
[c]Little = less than 3 years of work experience; Much = at least 3 years of work experience.

As indicated by the generalised kappa and Gwet's AC1, agreement between participants within profession categories was fair to moderate, with kappa values between 0.293 and 0.372, and Gwet's AC1 between 0.308 and 0.479 (Table 5). In most cases, agreement regarding the recommended control action was slightly lower than for degree of soiling and body condition, and students had slightly lower values than AW inspectors and veterinarians/advisers, but the differences were small. For recommended AW control action, scientists had lower agreement and consistency values than the other professions. As expected, the values of percent agreement were generally somewhat higher than kappa and Gwet's AC1. Participants with at least 3 years of professional experience had a slightly better agreement for body condition than less experienced assessors, but the difference was not conclusive for soiling and control actions (Table 6).

Discussion

This study indicates that professionals and students from different backgrounds and experiences score the degree of cattle body condition and skin soiling differently, but

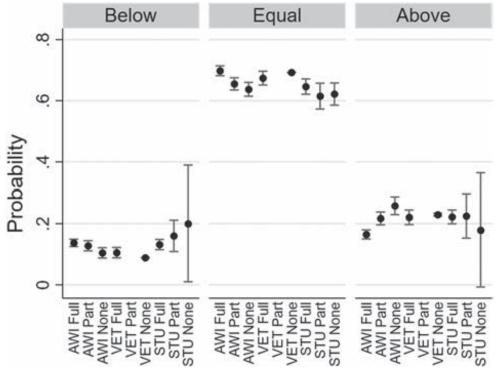

Figure 2. Predictive margins (with 95% confidence intervals) from ordinal logistic models of standardised scores of body condition, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different professional categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; STU = students) and different levels of animal science education (full = university degree; part = courses of more than 10 ECTS university credits; none = maximum 10 ECTS university credits), based on scorings of photo slides 2009–2016; probabilities for veterinarians/advisers with courses of more than 10 ECTS were not possible to estimate.

persons with the same background rate these measures rather similarly.

It is unclear to what extent education and training can help secure reliable assessment of AW. Vieira et al. (2015) found positive effects of training on the reliability of visual scoring of body condition in dairy goats using sketches. Shinozaki et al. (2019) found that training appears to affect the reliability of visual assessments of capillary refill time in humans. Practical experience from Swedish dairy adviser coordination meetings indicates that regular calibration exercises are important to secure assessor skills and ensure reliable assessments (pers. comm., L. Winblad von Walter, Växa Sverige, 5 December 2018). In this study, three or more years of professional experience with AW, compared to shorter experience, was only shown to influence ratings of soiling in the 'other profession' category. Agreement between participants regarding body condition was slightly better with 3 years or more of professional experience, compared to less experience, but the results were less conclusive for soiling and suggested actions. Especially ratings of body condition differed between participants, which underlines the need for an objective scoring scale, and shows the importance of proper calibration. Further research is needed to develop best practices for training inspectors in AW assessment.

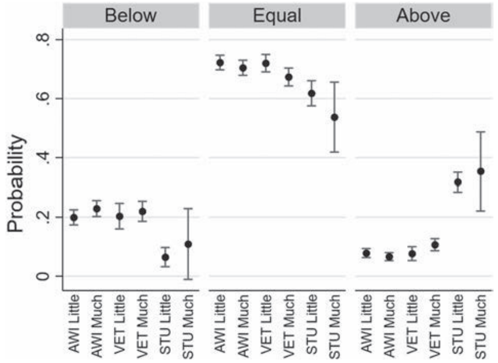

Figure 3. Predictive margins with 95% confidence intervals from ordinal logistic models of standardised scores of skin soiling, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different profession categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; STU = students) with a university degree and different levels of work experience (little = less than 3 years; much = at least 3 years), based on scorings of photo slides 2009–2016; confidence intervals for probabilities below and equal to the median for students with little work experience were not possible to estimate.

Found higher rates among students for soiling, compared to AW inspectors and veterinarians/advisers, indicate that students perceive animal soiling more severely. This may be due to differences in personal priorities because students do not encounter

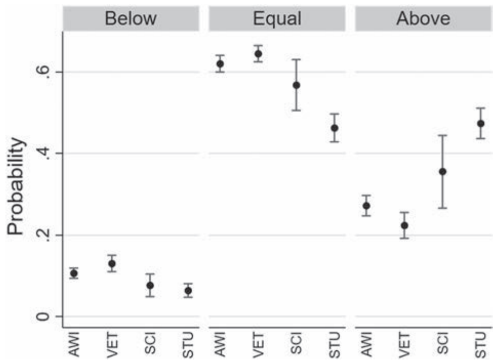

Figure 4. Predictive margins with 95% confidence intervals from ordinal logistic models of standardised scores of recommended action in response to perceived skin soiling at an imaginary official animal welfare, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different profession categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; SCI = animal welfare scientists; STU = students), based on scorings of photo slides 2009–2016.

Table 5. Percent agreement, generalised kappa and Gwet's agreement coefficient (AC1) (with 95% confidence intervals), as measures of agreement between participants of different profession categories, assessing cattle body condition, degree of skin soiling, and recommended corrective control action in response to soiling using photo slides; original scores.

Statistic	Profession category	Body condition[a]	Soiling[b]	Recommended action[c]
Percent agreement	AWI	0.564	0.582	0.502
	VET	0.568	0.578	0.530
	SCI	–	–	0.477
	STU	0.529	0.533	0.498
Kappa	AWI	0.348 (0.222, 0.583)	0.304 (−0.099, 0.879)	0.306 (0.225, 0.506)
	VET	0.372 (0.275, 0.578)	0.332 (−0.044, 0.870)	0.312 (0.177, 0.568)
	SCI	–	–	0.293 (0.162, 0.541)
	STU	0.317 (0.169, 0.569)	0.293 (−0.106, 0.822)	0.310 (0.191, 0.550)
Gwet's AC1	AWI	0.476 (0.281, 0.671)	0.477 (0.331, 0.624)	0.345 (0.226, 0.463)
	VET	0.479 (0.354, 0.604)	0.466 (0.369, 0.563)	0.391 (0.274, 0.508)
	SCI	–	–	0.308 (0.239, 0.376)
	STU	0.431 (0.304, 0.558)	0.402 (0.277, 0.527)	0.338 (0.175, 0.446)

Note: AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student.
[a]5-level scale (very poor–moderate–good–fat–very fat; Wildman et al., 1982; Svedberg, 2006).
[b]4-level scale (clean–slightly soiled–moderately soiled–very soiled; Svedberg, 2007).
[c]4-level scale (no action–remark–remark with a follow-up inspection–prohibition or order).

Table 6. Percent agreement, generalized kappa and Gwet's agreement coefficient (AC1) (with 95% confidence intervals), as measures of agreement between participants with different levels of professional experience, assessing cattle body condition, degree of skin soiling, and recommended corrective control action in response to soiling using photo slides; original scores.

Statistic	Experience[a]	Body condition[b]	Soiling[c]	Recommended action[d]
Percent agreement	Little	0.542	0.526	0.486
	Much	0.574	0.552	0.501
Kappa	Little	0.327 (0.172, 0.590)	0.322 (−0.149, 0.794)	0.296 (0.188, 0.520)
	Much	0.370 (0.287, 0.564)	0.294 (−0.099, 0.837)	0.294 (0.215, 0.489)
Gwet's AC1	Little	0.448 (0.298, 0.598)	0.400 (0.353, 0.446)	0.321 (0.252, 0.390)
	Much	0.487 (0.329, 0.645)	0.432 (0.321, 0.544)	0.348 (0.227, 0.469)

[a]Little = less than 3 years of work experience; Much = at least 3 years of work experience.
[b]5-level scale (very poor–moderate–good–fat–very fat; Wildman et al., 1982; Svedberg, 2006).
[c]4-level scale (clean–slightly soiled–moderately soiled–very soiled; Svedberg, 2007).
[d]4-level scale (no action–remark–remark with a follow-up inspection–prohibition or order).

soiled animals so often, to a change in moral standards with increasing age or to cultural relativism. It may also be due to social desirability bias (Kaminska & Foulsham, 2013) or a lack of agreement on or understanding of rating scales, despite efforts to standardise ratings. Higher scores in students for suggested control actions indicate that students are also less tolerant to animal soiling. It has been shown by Margoni et al. (2018) that young adults (21–39 years) rely more on intentions and less on outcomes in judging harmful actions, compared to older people (63–90 years), which may motivate them to take stricter measures against a particular AW infringement. A survey of European citizens (European Commission, 2016) showed that young people and students were more interested in the conditions for farm animals. A stricter view on AW in young persons may clash with the standards and beliefs of more experienced assessors, which requires careful consideration when e.g. official AW control is conducted. Participant age was not recorded in this study but students were likely generally younger than the other profession categories. Oliveira et al. (2017) found veterinarians to score foot pad dermatitis in chickens at the time of slaughter lower, compared to scientists, allegedly due to prior extensive experience of severe foot lesions at slaughterhouses.

AW inspectors and veterinarians/advisers perceived the animals as cleaner than students did. In a comparable study, Bracke et al. (2008) found that veterinarians overall give a higher welfare score for animal housing systems, compared to ethologists, probably reflecting differences in education. In this study, there were no significant differences regarding recommended control actions between AW inspectors and veterinarians/advisers, nor between scientists and other profession categories.

Various factors that were not possible to control or standardise, such as photo angle, image resolution, lighting conditions (contrast, shadows), disturbing objects and irrelevant housing conditions, may have influenced scorings and the agreement between assessors. For example, a dirty housing environment may have affected assessors unconsciously, thus biasing assessments of animal skin soiling towards higher scores. Because slightly different sets of images were used at different scoring sessions, such effects may also to some extent have influenced the estimated differences between different assessor categories. Poor image quality probably decreased agreement between assessors. Flowers et al. (2008) studied interrater agreement using a human medical wound imaging system, and reported on difficulties experienced by the assessors due to poor image quality, including low dark/light contrast and sharpness. Schmitt et al. (2008) found assessments of knee joint alignment from standardised photos of human subjects to be considerably influenced by the position of the subjects' legs.

Assessments of photo slides differ substantially from on-farm assessments of live animals. Examination in real life allows the assessor to apply different lighting conditions, perspectives and distances, which is likely to increase the reliability of the assessment. Zhu et al. (2017) showed that the visual information provided when viewing 3D images is not the same as when viewing 2D photos of human patients, which may change the clinical impression. However, photos similar to this study, are regularly used in education and training of students and professionals.

Conclusions

This study indicates that persons from different professional backgrounds, with different levels of experience, view and score skin soiling and body condition in cattle differently, while scorings by assessors with similar backgrounds agree fairly well. Undergraduate students in animal science rate cattle soiling stricter than professional assessors, and they seem prepared to take stricter corrective actions against animal soiling at an official control inspection.

Acknowledgements

We are grateful for the collaboration of the participants in the scoring sessions. The Swedish Association for the Protection of Animals supported this study by providing funding.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Swedish Association for the Protection of Animals.

ORCID

B. Staaf Larsson http://orcid.org/0000-0002-6748-6047
E. Petersson http://orcid.org/0000-0001-6443-4172
M. Stéen http://orcid.org/0000-0001-8224-0224
J. Hultgren http://orcid.org/0000-0001-9164-2321

References

Abraira, V. & Pérez de Vargas, A. (1999). Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs. *Qüestiió* 23, 561–571.

Andrade, S. B. & Anneberg, I. (2014). Farmers under pressure. Analysis of the social conditions of cases of animal neglect. *Journal of Agricultural and Environmental Ethics* 27, 103–126.

Anneberg, I., Vaarst, M. & Sørensen, J. T. (2012). The experience of animal welfare inspections as perceived by Danish livestock farmers: A qualitative research approach. *Livestock Science* 147(1–3), 49–58.

Berg, C. & Lundmark Hedman, F. (2020). Compliance with animal welfare regulations: Drivers and consequences. *CAB Reviews* 15(025), 1–10. doi:10/1079/PAVSNNR202015025. http://www.cabi.org/cabreviews.

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.

Bracke, M. B. M., Edwards, S. A., Engel, B., Buist, W. G. & Algers, B. (2008). Expert opinion as 'validation' of risk assessment applied to calf welfare. *Acta Veterinaria Scandinavica* 50, 29. doi:10.1186/1751-0147-50-29

Breen, J. E., Green, M. J. & Bradley, A. J. (2009). Quarter and cow risk factors associated with the occurrence of clinical mastitis in dairy cows in the United Kingdom. *Journal of Dairy Science* 92, 2551–2561.

Broom, D. M. (2017). Animal welfare in the European Union. Study for the PETI Committee. Brussels, Belgium: European Parliament. Accessed 30 September 2020, available at: https://op.europa.eu/sv/publication-detail/-/publication/74df7b49-ffe7-11e6-8a35-01aa75ed71a1/language-en.

European Commission. (2012). 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 2012–2015.

European Commission. (2016). Attitudes of Europeans towards Animal Welfare Special Eurobarometer 442. Accessed 15 September 2020, available at: http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/71348.

European Commission. (2019). ECTS - European Credit Transfer System. Accessed 14 August 2020, available at: https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en.

European Parliament. (2017). Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5), 378–382. doi:10.1037/h0031619

Flowers, C., Newall, N., Kapp, S., Lewin, G., Gliddon, T., Carville, K., Martinelli, D. & Santamaria, N. (2008). Clinician inter-rater reliability using a medical wound imaging system. *Wound Practice and Research* 16(1), 22–31.

Gillund, P., Reksen, O., Grohn, Y. T. & Karlberg, K. (2001). Body condition related to ketosis and reproductive performance in Norwegian dairy cows. *Journal of Dairy Science* 84(6), 1390–1396.

Green, L. E., Huxley, J. N., Banks, C. & Green, M. J. (2014). Temporal associations between low body condition, lameness and milk yield in a UK dairy herd. *Preventive Veterinary Medicine* 113(1), 63–71.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* 61, 29–48.

Gwet, K. L. (2014). *Handbook of Inter-rater Reliability* (Gaithersburg, MD, USA: *Advanced Analytics*, LLC).

Hauge, S. J., Kielland, C., Ringdal, G., Skjerve, E. & Nafstad, O. (2012). Factors associated with cattle cleanliness on Norwegian dairy farms. *Journal of Dairy Science* 95(5), 2485–2496.

Hughes, J. (2001). A system for assessing cow cleanliness. *In Practice* 23, 517–524.

Kaminska, O. & Foulsham, T. (2013). *Understanding Sources of Social Desirability Bias in Different Modes: Evidence from Eye-Tracking* (Colchester: Institute for Social and Economic Research (ISER), University of Essex). ISER Working Paper Series, No. 2013-04.

Keeling, L. J. (2009). An analysis of animal-based versus resource-based comments in official animal welfare inspection reports from organic and conventional farms in Sweden. *Animal Welfare* 18, 391–397.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Margoni, F., Geipel, J., Hadjichristidis, C. & Surian, L. (2018). Moral judgment in old age: Evidence for an intent-to-outcome shift. *Experimental Psychology* 65(2), 105–114. doi:10.1027/1618-3169/a000395

Nafstad, O. (1999). Skader og kvalitetsfeil på norske storfehuder [Hide damages and quality defects in cattle leather]. *Norsk Veterinærtidsskrift* 111(5), 311–319.

Oliveira, A. R. S., Lund, V. P., Christensen, J. P. & Nielsen, L. R. (2017). Inter-rater agreement in visual assessment of footpad dermatitis in Danish broiler chickens. *British Poultry Science* 58(3), 224–229.

Pryce, J. E., Coffey, M. P. & Simm, G. (2001). The relationship between body condition score and reproductive performance. *Journal of Dairy Science* 84(6), 1508–1515.

Radeski, M., Janevski, A. & Ilieski, V. (2015). Screening of selected indicators of dairy cattle welfare in Macedonia. *Macedonian Veterinary Review* 38(1), 43–51.

Roche, J. R. & Berry, D. P. (2006). Periparturient climatic, animal, and management factors influencing the incidence of milk fever in grazing systems. *Journal of Dairy Science* 89, 2775–2783.

Roche, J. R., Lee, J. M., Macdonald, K. A. & Berry, D. P. (2007). Relationships among body condition score, body weight, and milk production variables in pasture-based dairy cows. *Journal of Dairy Science* 90(8), 3802–3815.

Ruddat, I., Scholz, B., Bergmann, S., Buehring, A. L., Fischer, S., Manton, A., Prengel, D., Rauch, E., Steiner, S., Wiedmann, S., Kreienbrock, L. & Campe, A. (2014). Statistical tools to improve assessing agreement between several observers. *Animal* 8, 643–649.

Schmitt, H., Kappel, H., Moser, M. T., Cardenas-Montemayor, E., Engelleiter, K., Kuni, B. & Clarius, M. (2008). Determining knee joint alignment using digital photographs. *Knee Surgery, Sports Traumatology, Arthroscopy* 16, 776–780. doi:10.1007/s00167-008-0570-6

Shinozaki, K., Jacobson, L. S., Saeki, K., Kobayashi, N., Weisner, S., Falotico, J. M., Li, T., Kim, J., Lampe, J. W. & Becker, L. B. (2019). Does training level affect the accuracy of visual assessment of capillary refill time? *Critical Care* 23(1), 157. doi:10.1186/s13054-019-2444-3

Sørensen, J. T. & Fraser, D. (2010). On-farm welfare assessment for regulatory purposes: issues and possible solutions. *Livestock Science* 131, 1–7.

Svedberg, J. (2006). *Djurskyddstillsyn i Praktiken Nötkreatur Hull och Hullbedömningar [Animal Welfare Control in Practice Cattle Body Condition and Body Condition Scoring]* (Skara: SLU).

Svedberg, J. (2007). *Djurskyddstillsyn i Praktiken Nötkreatur Gödselförorenade Djur [Animal Welfare Control in Practice Cattle Defecated Animals]* (Skara: SLU).

Vieira, A., Brandão, S., Monteiro, A., Ajuda, I. & Stilwell, G. (2015). Development and validation of a visual body condition scoring system for dairy goats with picture-based training. *Journal of Dairy Science* 98, 6597–6608.

Wildman, E. E., Jones, G. M., Wagner, P. E. & Boman, R. L. (1982). A dairy cow body condition scoring system and its relationship to selected production characteristics. *Journal of Dairy Science* 65, 495–501.

Zhu, S., Yang, Y. & Khambay, B. (2017). A study to evaluate the reliability of using two-dimensional photographs, three-dimensional images, and stereoscopic projected three-dimensional images for patient assessment. *International Journal of Oral and Maxillofacial Surgery* 46, 394–400. doi:10.1016/j.ijom.2016.11.011

III

ACTA AGRICULTURAE SCANDINAVICA, SECTION A — ANIMAL SCIENCE
https://doi.org/10.1080/09064702.2024.2435339

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Methods for assessing skin temperature in two breeds of dairy cows and their correlation to indoor and rectal temperature

B. Staaf Larsson [a,b], K. Dahlborn [a] and A. Jansson [a]

[a]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden; [b]Swedish Centre for Animal Welfare, Swedish University of Agricultural Sciences, Uppsala, Sweden

ABSTRACT
Routines for assessing body temperature and thermal comfort are not included in official animal welfare controls although European legislation consider it important. This study investigated time consumption and feasibility of using handheld skin temperature sensors in a dairy farm and the correlation of the recordings with indoor and rectal temperature. Skin temperatures in 21 dairy cows of two breeds were recorded monthly during one year at the neck, hip and vulva, using two techniques (infrared radiation (IR) and conduction). Rectal and indoor ambient temperature were recorded on the same occasion. Time spent recording temperature was ~2 s/cow with IR and >1 min/cow with conduction technique. Skin temperatures did not correlate well with rectal temperature but correlated strongly with indoor temperature. Neck temperature recorded by IR best reflected indoor temperature, with no difference between breeds, and could be a tool for quick monitoring of ambient conditions in individual cows.

ARTICLE HISTORY
Received 19 August 2024
Accepted 22 November 2024

KEYWORDS
Ambient temperature; animal welfare; dairy cow; thermoneutral zone; rectal temperature; skin temperature

Introduction

European Union Council Directive (European Commission, 1998) on the protection of animals kept for farming purposes requires that ill animals must be treated without delay and that indoor temperature must be kept within limits that are not harmful to the animals. Corresponding Swedish regulations and general advice concerning cattle are similar and state 'In stables, animals must have a climate adapted to the type of animal and the type of animal husbandry (thermal comfort)' (Swedish Board of Agriculture, 2019). However, standardised routines for assessing body temperature (fever, hyper- and hypothermia) and thermal comfort are currently not included in official animal welfare controls, and in practice, it is not objectively evaluated. The range of ambient temperature conditions in which animals do not need to perform active strategies to maintain normal body temperature is called the thermoneutral zone, or comfort zone, and is defined by lower and upper critical ambient temperatures (Sjaastad et al., 2016). The lower critical ambient temperature can be identified by animals shivering and the upper temperature by cows sweating and panting (Sjaastad et al., 2016). Changes in skin surface temperature reflect changes in skin blood flow in response to alterations in environmental temperature (Scoley et al., 2019). Within the thermoneutral zone, the animal regulates body temperature by shifting blood flow to/from the skin, which causes alterations in skin temperature. In theory, skin temperature therefore has the potential to be an indicator of whether an animal is at the borders of its thermoneutral zone, i.e. near the initiation of active thermoregulation like shivering or panting/sweating. There are several options available for measuring skin temperature on farm animals (Nogami et al., 2014; Scoley et al., 2019; Furukawa et al., 2024). In this study, we evaluated the feasibility of methods based on conduction and infrared radiation (IR).

Body temperature measurements are also of interest for disease control, since fever is a common symptom of many infectious diseases of the cow (Smith & Risco, 2005). Easy identification of sick animals by farmers and animal welfare inspectors would enable early intervention and treatment. However, both farmers and animal welfare inspectors require quick, reliable and cost-effective methods. Measuring rectal temperature is the gold standard for assessment of body temperature and fever in animals (Sun et al., 2021). Tresoldi et al. (2020) concluded that the threshold of fever differs between researchers,

CONTACT B. Staaf Larsson Birgitta.Staaf.Larsson@slu.se
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09064702.2024.2435339.

from 38.9°C (Hillman et al., 2005) to 40°C (Burfeind et al., 2012; Pohl et al., 2014), and in the present study fever was set to >39.5°C. Measuring rectal temperature is time-consuming and requires physical contact with the animal and is accordingly not feasible neither for farmers nor animal welfare inspectors to perform at herd level. Therefore, there is a need for simpler and less invasive methods for accurate determination of body temperature in dairy cows under farm conditions.

Two breeds are dominating in Swedish dairy production, the Swedish Holstein (SH) and the Swedish Red and White Breed (SRB). Breed differences in body temperature have been observed (e.g. rectal temperature in Carvalho et al., 1995 and reticulorumen in Liang et al., 2013 and Stone et al., 2017), and Holstein cows have been shown to have lower heat tolerance than other dairy breeds (Legates et al., 1991). The SH and SRB breeds differ in terms of colour (black and brown, respectively) and fat accumulation pattern (Hjertén, 2006). Skin colour may affect temperature measurements made using methods based on detection of infrared radiation, since matt black surfaces (including cattle coats) are good emitters of infrared radiation (Hansen, 1990; Hellebrand et al., 2003), while fat accumulation pattern will determine the thickness of the insulating subcutaneous fat layer (Schröder & Staufenbiel, 2006). A recent study on data from Swedish dairy farms indicates, however, that none of the breeds (SH and SRB) have any advantage ameliorating high ambient temperatures in terms of milk production (Ahmed et al., 2022). It is not known if body temperatures differ between SH and SRB and this knowledge is needed if temperature registrations shall be implemented in a Swedish control system. It is well known that SH and SRB differ in terms of milk yield and that milk yield can be positively associated with temperature of the reticulorumen (Liang et al., 2013), and thereby the body temperature.

The aim of this study was to evaluate the feasibility of using two handheld skin temperature sensors during field conditions and the correlation of the values obtained with indoor temperature and the gold standard of body temperature, i.e. rectal temperature. Another aim was to investigate if rectal and skin temperatures differ between the two breeds. The hypothesis was that these types of recordings have potential as future tools both for farmers and inspectors in official animal welfare controls and that there might be breed differences.

Material and methods

Cows and management system

A total of 21 dairy cows (12 SRB, 9 SH) kept in an isolated and naturally ventilated loose-house system controlled based on indoor temperature through an adjustable open ridge in the ceiling, at the Swedish University of Agricultural Sciences research facility (Lövsta, Uppsala, Sweden), were used in the study. Breed differences have in earlier studies been observed between groups of e.g. eight (Gebremedhin et al., 2011) and 10 (Dikmen et al., 2008) animals of each breed or animal type. Therefore, the sample size was expected to be relevant. All cows were newly calved and were monitored over one year (February 2016 to January 2017), i.e. including the lactation period and in some cows also the following dry period. Information about the cows (breed, date of birth, parity, parturition date, start of dry period) is presented in Supplement A. The research facility, which can accommodate a total of 280 cows, was divided into four sections and lactating cows were moved between these sections depending on their energy requirement, lactation stage and health status. In the period May–August, all cows were kept in an outdoor enclosure at night and were indoors from morning milking until after afternoon milking. The cows were milked twice a day in an automatic milking rotary system (DeLaval AMR™, DeLaval, Sweden). Mean annual milk production was 10,282 kg energy-corrected milk per cow. Insemination was performed approximately two months into the lactation period and the dry period began 5–6 weeks before parturition. At the start of the dry period, cows were moved to a fifth section in the loose-house and kept together with replacement heifers. Two cows had been moved to the calving section by the last date of data collection and were therefore not included on that measurement occasion. Cow 1475 (SRB, see suppl. A) was removed from the study (slaughtered) in May due to disease and was replaced by cow 344 (SRB, see suppl. A) for the remainder of the study period. Three cows (90, 972 and 1475) were diagnosed with endometritis two days before the measurement occasion in March. One cow (972) had a cyst diagnosed at the same time. Cow 5357 was diagnosed with a cyst four days after the measurement occasion in May. Another cow (74) was treated for a sore teat at the measurement occasion in July. No data was removed due to these diagnoses and any findings related to this will be reported in the results section.

The study complied with ARRIVE guidelines and EU Directive 2010/63/EU on animal experiments.

Collection of data

Body temperature

Body temperature measurements were made monthly for 12 months (February 10, March 9, April 19, May

17, June 21, July 18, August 30, September 27, October 13, November 7, December 13, January 18). All measurements were performed by the same person, and all were made inside the barn during the afternoon and evening (12.40–20.10 h) except in October, when they were made during the morning (05.10–10.30 h). In the period May to August, when cows grazed outdoors during the night, the measurements were made in the afternoon, after the cows had been indoors for several hours. The cows were usually loose during the measurements (except for a few occasions when a cow would not stand still and had to be tied up). All cows were accustomed to being tied up occasionally and no adverse behaviour was observed to this.

Rectal temperature was measured using a digital rectal thermistor thermometer (MT20RA, Microlife AG, Widnau, Switzerland, precision 0.1°C), which was inserted 7–8 cm into the rectum and touched the intestinal wall. Skin temperature was recorded at three positions on the body, one in the cranial and two in the caudal direction (Figure 1), using a medical thermistor thermometer (MTT) (DM 852, Ellab, Hillerød, Denmark, range −1 to +50°C, precision 0.1°C) and an infrared thermometer (IRT) (TN1, ETI Ltd., AzoNetwork, UK Ltd, Manchester, UK, range −33 to 220°C, precision 0.1°C). The positions were: (1) neck (centred on a line between the withers and larynx, approximately 15 cm from the top of the neck), (2) hip (10 cm below tuber coxae) and (3) vulva (Figure 1). Areas least affected by the lying position and most exposed to ambient temperature was chosen as described in Scoley et al. (2019).

During measurement, MTT was placed under the hairs, to ensure contact with the skin, while IRT was placed on top of the hair (close but no contact). One recording was made at all positions with MTT, while two measurements were made at all position with IRT and the higher value was used in further analysis. According to Yan et al. (2021), maximum IR skin temperature is less sensitive to environmental parameters, but more correlated with core body temperature. In cow 961, a MTT measurement at the neck position was only available for one occasion (June 21), since the cow showed avoidance behaviour when the neck was approached. For practical reasons, MTT measurements were made on only one cow in May.

All data were collected by the first author and noted by an assistant who also registered time of day of registrations, the latter done by a stopwatch.

Indoor temperature and relative humidity

At the time of the cow measurements, indoor temperature and relative humidity (RH) were measured using a weather station (Nexus prologue, model: IW004/36-5136, Clas Ohlson, Insjön, Sweden) in the middle of the different loose-house sections, once each measurement day. Temperature and humidity index (THI) were calculated according to Tucker et al. (2008).

Statistical analysis

The data were analysed with a MIXED model (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA) with temperature as the dependent variable (y) and date and breed as fixed factors and individual as random factor. Correlations were analysed using Pearson correlation analysis (SAS, Version 9.4, SAS Institute Inc., Cary,

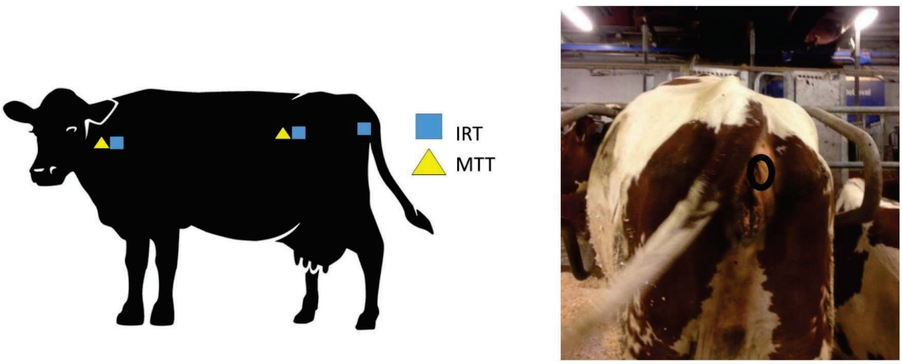

Figure 1. Positions on the cow's body at which skin temperature was measured using a medical thermistor thermometer (MTT) and an infrared thermometer (IRT): neck (centred on a line between the withers and larynx, approximately 15 cm from the top of the neck), hip (10 cm below tuber coxae) and vulva (IRT only, circle in image on the right).

NC, USA), with $P < 0.05$ considered significant. Values shown are least square (LS) Means ± standard error (SE) unless otherwise stated.

Results

Climate conditions

Indoor temperature ranged between 13.2 and 23.7°C (Figure 2) and RH was 67, 57, 47, 40, 53, 43, 51, 69, 68, 79, 51 and 72%, respectively from February to January. THI was 58, 57, 56, 64, 65, 69, 66, 63, 58, 55, 62 and 57 (February to January).

Findings in practical data collection

The time spent obtaining temperature measurements with MTT and IRT at each body position was approximately 70 and 2 s, respectively. However, if a cow did not stand still and had to be tied up, measurement with MTT could take several minutes. One cow also showed avoidance behaviour when the neck was approached, which resulted in missing values.

Rectal temperature and skin temperature

Individual minimum and maximum body temperature values varied by several degrees Celsius on all measurement occasions (Table 1). There was a significant effect of month on rectal temperature ($P = 0.02$), with the lowest mean values (38.3 ± 0.1°C) recorded in October and the highest (38.8 ± 0.1°C) in August (Figure 2). There was also a significant effect of month on skin temperature ($P < 0.0001$) (Figure 2).

There was no effect of breed on rectal temperature ($P = 0.932$) and generally no effect of breed on skin temperature. The only exception was temperature recorded at the hip, which was higher for cows of the SH breed than for SRB cows when measured with IRT (29.7 ± 0.2 vs. 29.1 ± 0.2°C, respectively, $P = 0.02$), and lower when measured with MTT (34.7 ± 0.2 vs. 35.1 ± 0.1°C, respectively, $P = 0.02$). We found no elevated rectal temperature on cows with diagnoses compared to their own mean value the other months.

Table 1. Minimum and maximum individual rectal temperatures (digital thermistor thermometer) and skin temperatures over one year of 21 dairy cows kept in an isolated loose-housing system.

	N =	Minimum °C	Maximum °C
Rectal	197	37.4	39.2
IRThip	205	21.2	35.7
IRTvul	201	22.6	35.6
IRTneck	190	22.8	33.9
MTThip	188	29.5	37.3
MTTneck	164	28.9	37.7

Note: Skin measurements were made 10 cm below tuber coxae (hip), at the lateral side of the vulva (vul) and at the neck, using an infrared thermometer (IRT) and a medical thermistor thermometer (MTT, conduction) at the hip and the neck.

Correlations

All skin temperature measurements showed very weak correlations ($r = 0.15–0.18$, $P < 0.04$) with rectal temperature (Table 2). Rectal temperature was not correlated

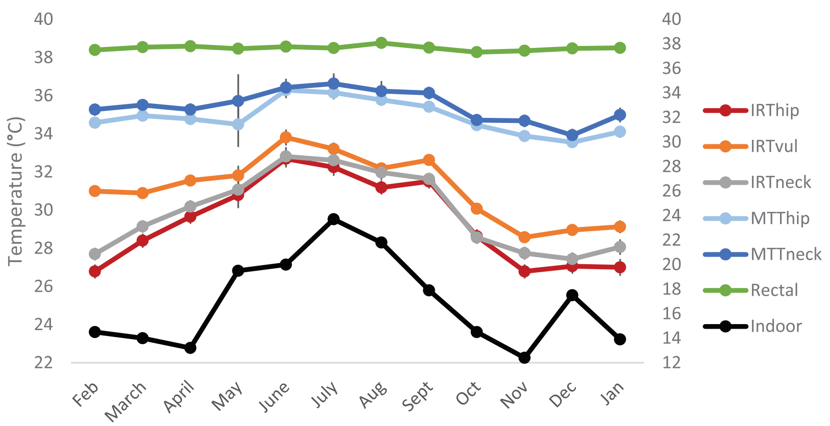

Figure 2. Monthly (February-January) rectal temperature (green) and skin temperatures (left axis) of 21 dairy cows recorded with an infrared thermometer 10 cm below tuber coxae (red, IRThip), at vulva (orange, IRTvul) and at the neck (grey, IRTneck) and with a medical thermistor thermometer 10 cm below tuber coxae (light blue, MTThip) and at the neck (dark blue, MTTneck). All cows calved between 22 December and 5 February. Black line shows indoor temperature (right axis).

Table 2. Correlations coefficients (r), *P* values and number of observations (N) for rectal temperature and skin temperatures registered 10 cm below tuber coxae (hip), at the caudal side of the vulva (vul) and at the neck using an infrared thermometer (IRT) or a medical thermistor thermometer (MTT).

		IRThip	IRTvul	IRTneck	MTThip	MTTneck
Rectal temperature	*r*	0.1545	0.1708	0.1489	0.1503	0.1788
	P	0.0301	0.0164	0.0425	0.0401	0.0233
	N	197	197	186	187	161

with indoor temperature, but all skin temperature values showed a strong correlation with indoor temperature (Table 3). The strongest correlation was obtained for measurements made with the IRT method at the neck.

Discussion

Rectal temperature values remained stable throughout most of the study period and no cow was observed with fever (>39.5°C, Suthar et al., 2012; Radostits et al., 2000) which limited our possibilities to link skin temperatures to fever or heat stress. However, a slight significant elevation was observed in August (+0.3°C compared with July), despite the fact that ambient temperature and THI peaked in July (23.7°C). Accordingly, the highest mean rectal temperature was not observed when ambient temperature and THI were highest (in July, 23.7°C and 69, respectively). The reason for the elevated temperature in August is unclear. No health problems were observed in the herd at that time, and it might be due to differences in physical activity. The few treatments in the herd during the study did not affect the measured temperatures. The study shows that cows (at peak and mid-lactation) were able to maintain their heat balance even when indoor temperature exceeded >19.5°C (on four occasions) and when it peaked at 23.7°C. However, active heat dissipation (sweating and elevated breathing frequency) might have occurred, but this was not observed during measurements. According to Li et al. (2020), heat stress is triggered at ambient temperature of around 25°C, rectal temperature of 38.6°C and respiration rate of 48 bpm. A study of Israeli Holstein cows found that rectal temperature increased with air temperatures of between 26°C and 36°C and concluded that the upper critical temperature is 25–26°C, irrespective of previous acclimatisation or milk production (Berman et al., 1985). Based on this, the air temperature in the present study did not reach the critical temperature to increase rectal temperature. However, milk production might be negatively affected at the ambient temperatures observed in this study. In a recent study by Ahmed et al. (2022) a sharp decrease in production was observed in Swedish dairy cows when the average maximum daily temperature of the past 7 days exceeded 22–23°C.

All skin temperature measurements showed a very weak correlation with rectal temperature, indicating that measurements at the positions evaluated (neck, hip, vulva) cannot be used to assess changes in actual body temperature (rectal temperature) at individual level in healthy cows (no fever) at the ambient temperature range prevailing in the study period (12.4–23.7°C). This finding is not surprising, since the mechanisms used by dairy cows to maintain constant body core temperature are both sensitive and fine-tuned, with thermoreceptors in skin and organs responding to temperature changes of less than 0.1°C and sending signals to the hypothalamus to adjust peripheral vasculature (Sjaastad et al., 2016). Blood is thereby directed to/away from the core and core temperature is maintained. The potential of the handheld temperature sensors to detect modified body temperature (e.g. elevated rectal temperatures in heat environment outside the thermoneutral zone) remains to be determined, since climate conditions were not extreme in the present study.

On the other hand, all skin temperature measurements showed a strong positive correlation with indoor temperature, reflecting effects of radiation from the surrounding and the effects of vasodilation or vasoconstriction. IRT measurements at the neck showed the strongest correlation ($r = 0.73$, $P < 0.0001$) with indoor temperature. In the period May–September, when indoor temperature was within the range 17.6–23.7°C, IRT values at the neck exceeded 31.5°C. During the rest of the year, when indoor temperatures were lower, the

Table 3. Correlations coefficients (r), *P* values and number of observations (*N*) for indoor temperature and rectal and skin temperatures registered 10 cm below tuber coxae (hip), at the caudal side of the vulva (vul) and at the neck using an infrared thermometer (IRT) or a medical thermistor thermometer (MTT).

		Rectal	IRThip	IRTvul	IRTneck	MTThip	MTTneck
Indoor temperature	*r*	0.0797	0.6804	0.6168	0.7262	0.5572	0.5045
	P	0.2657	<.0001	<.0001	<.0001	<.0001	<.0001
	N	197	205	201	190	188	164

IRT values at the neck were always below 30.2°C. The use of IRT at the neck gives a quick measurement with limited interaction with the cow and could therefore be investigated further as a tool for monitoring thermoregulation. IR cameras can also be permanently installed for temperature monitoring e.g. when approaching a water station (Schaefer et al., 2012) or a milking parlour.

There were variations in measured temperature values at all body sites. Mean rectal temperature ranged between 38.3 and 38.8°C, i.e. showed variation of <0.5°C, which can be taken as the normal within-individual variation in healthy cows. Previous studies (e.g. Liang et al., 2013) show that cows have a diurnal rectal temperature pattern with the lowest temperatures in the morning. At one occasion (October), temperatures were registered in the morning instead of the afternoon/evening in the present study but there was no significant difference compared to the other registrations.

Skin temperature measurements made using MTT showed variation of 2.7°C, with the highest temperatures obtained during the warmest period of the year (June–September) and not at the start of peak lactation (March to May), when metabolic rate can be expected to be highest. Skin temperature measurements made using IR showed the highest overall variation, with the largest variation (5.4°C) in temperature measured at the hip during June–September.

There was no difference in rectal temperature between the two dairy breeds but a difference between breeds was observed for temperature at the hip, where cows of the SH breed had higher temperature than SRB cows when measured using IR, but lower temperature when measured using MTT. The reason for this is unclear and might be of no biological relevance. However, it is possible that differences in colouring and subcutaneous fat layers played a role, as the IR and MTT values are in accordance with SRB being lighter (brown, not black) and having more insulation, i.e. subcutaneous fat (Hjertén, 2006). Arp et al. (1983) observed higher skin IR temperatures and respiratory rates in mostly black Holstein compared with mostly white Holstein at an ambient temperature of 33°C and concluded that black cattle are more subject to heat stress than red and white cattle. A correlation between total animal heat production and IR temperature at the flank has been observed previously in Holstein cows (Montanholi et al., 2008). Since SH generally have slightly higher milk (and heat) production than SRB, they can be expected to have higher IR temperatures at the flank, and perhaps also below the hip.

For temperature measurements to be included in official animal welfare controls, as well as in large-scale management systems, they must be possible to perform in a quick and safe way. Of the two handheld methods tested in the present study, the MTT sensor took much longer to obtain each measurement (more than one minute) and also required physical contact with the cow, a stationary cow and adjustment of the equipment by the operator, which was a challenge on some occasions. This technique is therefore not optimal as a routine tool for animal welfare control. The IRT device was quick (~2 s per measurement) and required no physical contact with the cow, and therefore has greater potential in this context. This type of equipment is also reasonably cheap (~200 Euro, Google search on 15 October 2023) and can be used without great financial risk in dirty indoor conditions.

Conclusions

In our study, there were few significant changes in rectal temperature over an indoor ambient temperature range of 12.7–23.7°C and no cow was observed with fever which limited our possibilities to link skin temperatures to fever and heat stress (defined as significantly elevated rectal temperatures). There was no difference between breeds in rectal temperature. Skin temperature values did not correlate well with rectal temperature under the prevailing conditions, but skin temperatures correlated well with indoor temperature, with IR temperature measurements at the neck best reflecting ambient temperature (and with no difference between breeds). The IR sensor was also quick to record and with limited interaction with the cow and could therefore be interesting for future studies investigating animals outside the thermoneutral zone.

Acknowledgements

We are grateful to the Swedish Association for the Protection of Animals for providing funding for the study. Thanks also to Jessica Isaksson for assistance with the measurements at Lövsta research facility. Part of this study has previously been published as the MSc thesis of Jessica Isaksson (Isaksson, 2017).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Swedish Association for the Protection of Animals [grant number 2014-05-06].

Ethics approval

All procedures involving cows were approved by Uppsala Animal Ethics Committee, approval number C 114/15 for Lövsta Research Centre.

Data and model availability statement

The datasets used in this study are available from the corresponding author upon reasonable request.

ORCID

B. Staaf Larsson http://orcid.org/0000-0002-6748-6047
K. Dahlborn http://orcid.org/0000-0002-1029-9013
A. Jansson http://orcid.org/0000-0001-5612-9020

References

Ahmed, H., Tamminen, L. M. & Emanuelson, U. (2022). Temperature, productivity, and heat tolerance: Evidence from Swedish dairy production. *Climatic Change*, 175(1-2), 10. doi:10.1007/s10584-022-03461-5

Arp, S. C., Owens, F. N., Armbruster, S. L. & Laudert, S. (1983). Relationships of coat color, body surface temperature and respiration rate in feedlot steers. *Animal Science Research Report*, 114, 82–86.

Berman, A., Folman, Y., Kaim, M., Mamen, M., Herz, Z., Wolfenson, D., Arieli, A. & Graber, Y. (1985). Upper critical temperatures and forced ventilation effects for high-yielding dairy cows in a subtropical climate. *Journal of Dairy Science*, 68(6), 1488–1495.

Burfeind, O., Suthar, V. S. & Heuwieser, W. (2012). Effect of heat stress on body temperature in healthy early postpartum dairy cows. *Theriogenology*, 78, 2031–2038. doi:10.1016/j.theriogenology.2012.07.024

Carvalho, F. A., Lammoglia, M. A., Simoes, M. J. & Randel, R. D. (1995). Breed affects thermoregulation and epithelial morphology in imported and native cattle subjected to heat stress. *Journal of Animal Science*, 73(12), 3570–3573. doi:10.2527/1995.73123570x

Dikmen, S., Alava, E., Pontes, E., Fear, J. M., Dikmen, B. Y., Olson, T. A. & Hansen, P. J. (2008). Differences in thermoregulatory ability between slick-haired and wild-type lactating Holstein cows in response to acute heat stress. *Journal of Dairy Science*, 91, 3395–3402. doi:10.3168/jds.2008-1072

European Commission Commission. (1998). Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. *Official Journal of the European Communities*, L 221, 8.8.98, 23–27.

Furukawa, E., Ozawa, T., Higaki, S., Suda, T., Sasaki, Y., Murayama, K., Noguchi, M. & Yoshioka, K. (2024). Changes in skin temperature and behaviors of preweaning Holstein calves in a hot environment monitored by a multimodal tail-attached device. *Journal of Dairy Science*, 5(4), 332–337. doi:10.3168/jdsc.2023-0515

Gebremedhin, K. G., Lee, C. N., Hillman, P. E. & Brown-Brandl, T. M. (2011). Body temperature and behavioral activities of four breeds of heifers in shade and full Sun. *Applied Engineering in Agriculture*, 27(6), 999–1006. doi:10.13031/2013.40620

Hansen, P. J. (1990). Effects of coat colour on physiological responses to solar radiation in Holsteins. *Veterinary Record*, 127, 333–334.

Hellebrand, H. J., Brehme, U., Beuche, H., Stollberg, U. & Jacobs, H. (2003). Application of thermal imaging for cattle management. Proceedings of the 1st European Conference on Precision Livestock Farming. Berlin, Germany), pp. 761–763).

Hillman, P. E., Lee, C. N. & Willard, S. T. (2005). Thermoregulatory responses associated with lying and standing in heat-stressed dairy cows. *Transactions of the ASAE*, 48, 795–801. doi:10.13031/2013.18322

Hjertén, J. (2006). Relationships between body condition score, subcutaneous fat, live weight and reproduction in Swedish Holstein and Swedish Red and White Cattle. Samband mellan hull, underhudsfett, levande vikt och fruktsamhet hos SRB och SLB - *Epsilon Archive for Student Projects* (slu.se).

Isaksson, J. (2017). *Changes in Dairy Cows' Temperature. Second Cycle, A2E* (Uppsala: SLU, Department of Anatomy, Physiology and Biochemistry). (until 231231).

Legates, J., Farthing, B., Casady, R. & Barrada, M. (1991). Body temperature and respiratory rate of lactating dairy cattle under field and chamber conditions. *Journal of Dairy Science*, 74(8), 2491–2500.

Li, G., Chen, S., Chen, J., Peng, D. & Gu, X. (2020). Predicting rectal temperature and respiration rate responses in lactating dairy cows exposed to heat stress. *Journal of Dairy Science*, 103, 5466–5484. doi:10.3168/jds.2019-16411

Liang, D., Wood, C. L., McQuerry, K. J., Ray, D. L., Clark, J. D. & Bewley, J. M. (2013). Influence of breed, milk production, season, and ambient temperature on dairy cow reticulorumen temperature. *Journal of Dairy Science*, 96, 5072–5081. doi:10.3168/jds.2012-6537

Montanholi, Y. R., Odongo, N. E., Swanson, K. C., Schenkel, F. S., McBride, B. W. & Miller, S. P. (2008). Application of infrared thermography as an indicator of heat and methane production and its use in the study of skin temperature in response to physiological events in dairy cattle (Bos taurus). *Journal of Thermal Biology*, 33(8), 468–475. doi:10.1016/j.jtherbio.2008.09.001

Nogami, H., Okada, H., Miyamoto, T., Maeda, R. & Itoh, T. (2014). Wearable wireless temperature sensor nodes appressed to base of a calf's tail. *Sensors and Materials*, 26(8), 539–545.

Pohl, A., Heuwieser, W. & Burfeind, O. (2014). Technical note: Assessment of milk temperature measured by automatic milking systems as an indicator of body temperature and fever in dairy cows. *Journal of Dairy Science*, 97, 4333–4339. doi:10.3168/jds.2014-7997

Radostits, O., Arundel, J. & Gay, C. (2000). *Veterinary Medicine: A Textbook of the Diseases of Cattle, Sheep, Pigs, Goats and Horses* (9th ed.). (Amsterdam, The Netherlands: Elsevier Health Sciences).

Schaefer, A. L., Cook, N. J., Bench, C., Chabot, J. B., Colyn, J., Liu, T., Okine, E. K., Stewart, M. & Webster, J. R. (2012). The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. *Research in Veterinary Science*, 93(2), 928–935. doi:10.1016/j.rvsc.2011.09.021

Schröder, U. J. & Staufenbiel, R. (2006). Invited review: Methods to determine body fat reserves in the dairy cow with special regard to ultrasonographic measurement of backfat thickness. *Journal of Dairy Science*, 89, 1–14. doi:10.3168/jds.S0022-0302(06)72064-1

Scoley, G., Gordon, A. & Morrison, S. J. (2019). The effect of calf jacket usage on performance, behaviour and physiological responses of group-housed dairy calves. *Animal*, 13(12), 2876–2884. doi:10.1017/S1751731119001071

Sjaastad, Ø. V, Sand, O. & Hove, K. (2016). *Physiology of Domestic Animals* (3rd ed. Oslo, Norway: Scandinavian Veterinary Press).

Smith, B. I. & Risco, C. A. (2005). Management of periparturient disorders in dairy cattle. *Veterinary Clinics of North America: Food Animal Practice*, 21, 503–521. doi:10.1016/j.cvfa.2005.02.007

Stone, A. E., Jones, B. W., Becker, C. A. & Bewley, J. M. (2017). Influence of breed, milk yield, and temperature-humidity index on dairy cow lying time, neck activity, reticulorumen temperature, and rumination behavior. *Journal of Dairy Science*, 100, 2395–2403. doi:10.3168/jds.2016-11607

Sun, D., Webb, L., van der Tol, P. P. J. & van Reenen, K. (2021). A systematic review of automatic health monitoring in calves: Glimpsing the future from current practice. *Frontiers in Veterinary Science*, 8, 761468. doi:10.3389/fvets.2021.761468

Suthar, V., Burfeind, O., Bonk, S., Voigtsberger, R., Keane, C. & Heuwieser, W. (2012). Factors associated with body temperature of healthy Holstein dairy cows during the first 10 days in milk. *Journal of Dairy Research*, 79, 135–142. doi:10.1017/S0022029911000896

Swedish Board of Agriculture. (2019). Statens jordbruksverks föreskrifter och allmänna råd (SJVFS 2019:18) om nötkreaturshållning inom lantbruket m.m. [The Swedish Board of Agriculture's regulations and general advice (SJVFS 2019:18) on cattle husbandry in agriculture, etc.], Chapter 4, Section 4.

Tresoldi, G., Schütz, K. E. & Tucker, C. B. (2020). Sampling strategy and measurement device affect vaginal temperature outcomes in lactating dairy cattle. *Journal of Dairy Science*, 103, 5414–5421. doi:10.3168/jds.2019-16667

Tucker, C. B., Rogers, A. R. & Schütz, K. E. (2008). Effect of solar radiation on dairy cattle behaviour, use of shade and body temperature in a pasture-based system. *Applied Animal Behaviour Science*, 109, 141–154. doi:10.1016/j.applanim.2007.03.015

Yan, G., Liu, K., Hao, Z., Li, H. & Shi, Z. (2021). Development and evaluation of thermal models for predicting skin temperature of dairy cattle. *Computers and Electronics in Agriculture*, 188, 106363. doi:10.1016/j.compag.2021.106363

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:49

The aim of this thesis was to investigate how the official animal welfare control can be improved by using more animal-based measurements. The suggestion is to include more detailed information on body condition and dirty animals, as well as presence of wounds, lame and sick cows. Inclusion of information on type of housing system, cow's rising behaviour and whether allogrooming occurs is also suggested, and possibly skin temperature and milk osmolality as objective measures of thermal comfort and fluid status.

Birgitta Staaf Larsson received her doctoral education at the Department of Animal Biosciences, Swedish University of Agricultural Sciences (SLU). She received her Degree of Bachelor of Science at the University of Gothenburg.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-484-0
ISBN (electronic version) 978-91-8046-534-2