Assessment, market potential and margins of improved dairy cattle welfare

Sophie de Graaf

Promotor: Prof. dr. Wim Verbeke

               Department of Agricultural Economics, Ghent University, Belgium

               Prof. dr. Frank A.M. Tuyttens

               Flanders research institute for agriculture, fisheries and food, Belgium

               Faculty of Veterinary Medicine, Ghent University, Belgium

               Prof. dr. Ludwig Lauwers

               Flanders research institute for agriculture, fisheries and food, Belgium

               Department of Agricultural Economics, Ghent University, Belgium

Dean: Prof. dr. Marc Van Meirvenne

Rector: Prof. dr. Anne De Paepe

Sophie de Graaf

Assessment, market potential and margins of improved dairy cattle welfare

Thesis submitted in fulfilment of the requirements

for the degree of Doctor (PhD) in Applied Biological Sciences

Dutch translation:

Beoordeling, marktpotentieel en marges van welzijnsverbetering bij melkvee

Way of citation: de Graaf, S. 2017. Assessment, market potential and margins of improved dairy cattle welfare. Doctoral thesis. Ghent University

ISBN-number: 978-90-5989-971-1

Members of the Jury

Prof. dr. Frank Devlieghere (Chairman)
Department of Food Safety and Food Quality, Ghent University, Belgium

Prof. dr. Stefaan De Smet (Secretary)
Department of Animal Production, Ghent University, Belgium

Dr. Hans De Steur
Department of Agricultural Economics, Ghent University, Belgium

Dr. Eddie Bokkers
Department of Animal Sciences, Wageningen University, the Netherlands

Prof. dr. Geert Opsomer
Department of Obstetrics, reproduction and herd health, Ghent University, Belgium

Dr. Dirk Lips

Table of contents

List of abbreviations

ADF	Avoidance distance at the Feed rack
ANI	Animal Needs Index
CAE	Cost Allocative Efficiency
CE	Cost Efficiency
CRM	Compensation-reduction methods
DA	Displaced abomasum
EC	European Council
EFSA	European Food Safety Authority
EU	European Union
GAIA	Global Action in the Interest of Animals
ICC	Intra-class correlation coefficient
IP	Intention to purchase
PCA	Principal Component Analyses
PS	Panting Score
QBA	Qualitative Behaviour Assessment
RO	Research Objectives
RQ	Research Questions
RR	Respiration Rate
SCC	Somatic Cell Count
SD	Standard Deviation
TE	Technical Efficiency
TPB	Theory of Planned Behaviour
VIF	Variance inflation factor
WI	Welfare index
WTP	Willingness to pay
WQ	Welfare Quality®
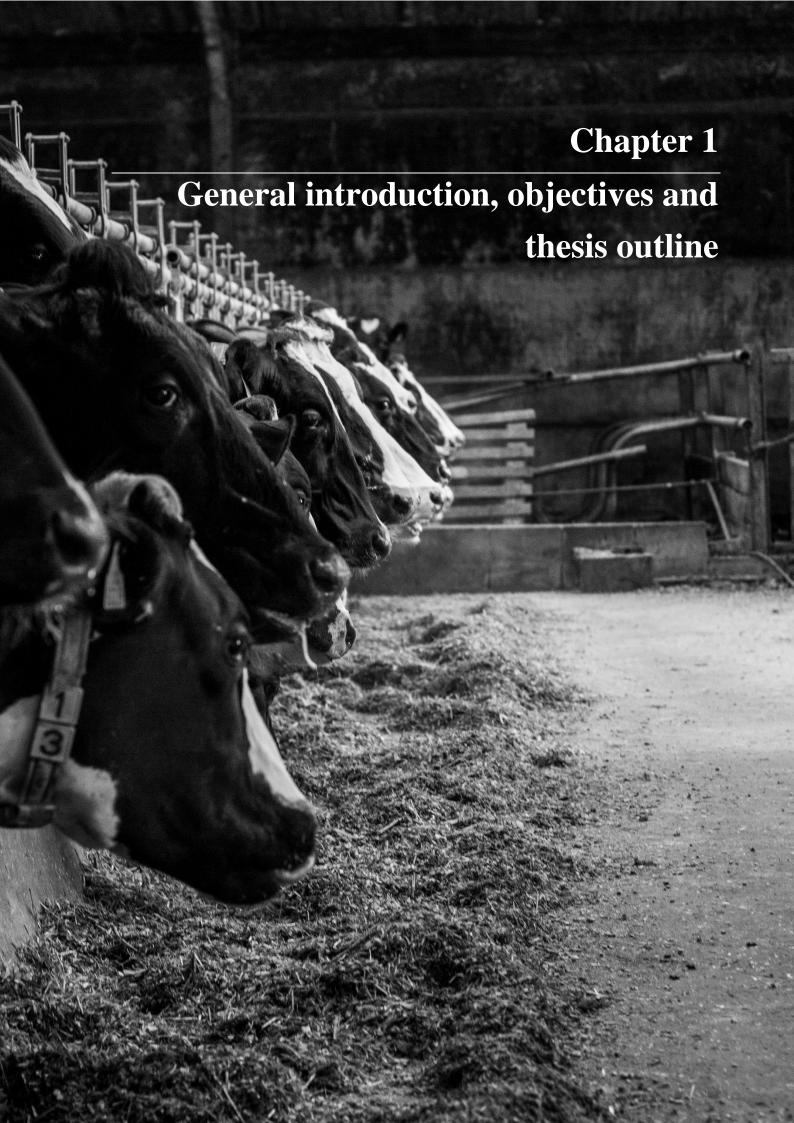
Chapter 1
General introduction, objectives and thesis outline

1.1   General introduction

The welfare of farm animals is of growing public interest and concern. Besides the widespread concern among the wider public (Mejdell, 2006), stakeholders in the production chain also recognise that animal welfare has become an important dimension in a responsible food chain (Pouta et al., 2010, Bracke et al., 2005). In the European Union (EU), legislation is currently the main method to remediate welfare problems in farm animals, primarily by mandating minimum requirements for housing and management (Bennett, 1997, Ingenbleek et al., 2012, Veissier et al., 2008). Besides, private initiatives originating from the sector are emerging and could be a promising avenue as well. Whereas animal welfare standards in legislation rely (partly) on information from scientific research, standards for private initiatives may in some cases be primarily based on needs of consumers (Robach, 2010). This is for instance reflected in various labels solely focused on outdoor access, which is consistently mentioned as highly important for animal welfare by consumers (e.g. Van Loo et al., 2014, Vanhonacker et al., 2008, Meuwissen et al., 2007). Retail may also play a role in mediating the consumers' demands, for example by positioning their retail brands based on an enhanced level of animal welfare (Burt, 2000). In some product categories both high quality retail brands and price fighting "generics" are combined under a single definition, and some retailers in the UK have used the retail brand strategically to launch initiatives championing "new" consumer values such as healthy eating, animal welfare or environmental issues. The supply side may respond to consumers' demands by adopting measures that allow for differentiated (niche) markets and create added value. This implies that private sector welfare standards are suitable for marketing strategies, but may sometimes lack scientific basis (Napolitano et al., 2010). In this, animal welfare science has a role to provide recommendations on accurate welfare assessment which might be used both in policy and private initiatives.

Animal welfare assurance schemes (WAS) aim to guarantee an enhanced level of farm animal welfare. Three, mutually linked, factors are of high importance for the success of a WAS in such a way that it provides added value for all stakeholders: credibility, implementation and costs. Credibility is important for acceptance of the WAS in society, by stakeholders like farmers, consumers, retailers and the government. Implementation refers to practical issues like communication of information acquired and monitoring welfare on farms. Costs have a major influence on the success of a WAS as well, because production animals are part of a chain that must make a profit to sustain.

An example of a WAS may be found in the Netherlands, where the animal protection organisation the 'Dierenbescherming' launched the 'Beter Leven Keurmerk' (literal translation: better life label), which is a label focused on animal welfare. A similar example can be found in the United Kingdom with the 'RSPCA assured' label, where the acronym stands for Royal Society for the Prevention of Cruelty to Animals. Mostly, these animal welfare assurance schemes focus on input-based (resources and management) measures of animal welfare. As input-based measures reflect risk factors or opportunities for impaired or good welfare rather than the direct animal-based measures (Blokhuis et al., 2003), their validity may be questioned (Johnsen et al., 2001). To assure an enhanced level of animal welfare on farms, welfare assurance schemes should adhere to several other criteria as well, e.g. cover the multidimensional nature of animal welfare and consist of reliable and feasible measures. Additionally, to differentiate between varying levels of farm animal welfare, such a scheme should be sensitive and discriminative. Not satisfying these criteria could have negative effects on consumer trust and would not provide farmers with valid welfare improvements.

Another challenge for a farm animal welfare assurance scheme is the balance between extra costs and revenues. Costs that are associated with an animal welfare assurance scheme are often expected to be paid by consumers. As consumers are the end user throughout the food chain, their demands should be accounted for. For this end, a more transparent food industry is needed, where animal welfare on-farm is connected to informed animal product consumption (Blokhuis et al., 2003).

Besides creating added value by satisfying certain demands of consumers, farmers may be able to invest in animal welfare when this directly influences farm economics by improving farm performance. However, a challenge which currently remains for science is investigating the relationship between both multidimensional concepts of farm efficiency and animal welfare.

Figure 1.1. illustrates in what way a WAS can lead to added value for the animal, the farmer and retailer, and the consumer, starting off from a valid welfare assessment. Animal welfare assessment may provide general information on overall welfare of production animals which can be used in communication with consumers. This may satisfy a consumer demand and allow for an informed product choice. The size of the market influences the price setting of such products and therewith farm performance. For this end, it is important to gain insight into the size of the market for animal-friendly products, and how to approach consumers. Besides

general information, welfare assessment may generate specific information via outcomes of welfare measures. This may motivate farmers to employ remedial actions for certain welfare problems. In addition, information from a welfare assessment may be linked directly to farm performance. For instance, clinical lameness has been found to negatively impact milk production (Green et al., 2002). Therefore, improvements in animal welfare problems like lameness may also lead to improvements in farm performance. This means that a welfare assessment which is to be used in a WAS should allow to clearly convey information to the consumer (one overall assessment of welfare on-farm), but should also allow to communicate specific information for the farmer. When this information leads to welfare-remedial changes in management and housing, this may lead to improved animal welfare which can be determined by repeating the welfare assessment.
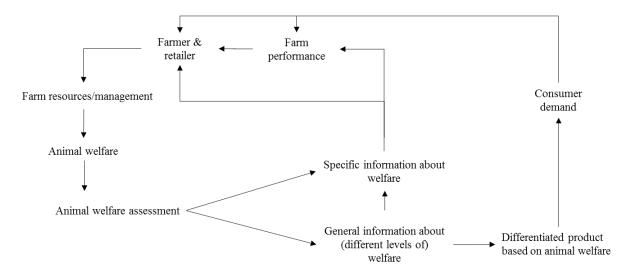
Figure 1.1.: Framework for pursuing added value of an animal welfare assurance scheme for farmers, consumers and animals (adapted from Blokhuis et al., 2003).

In this doctoral thesis, the possibilities of introducing a WAS in Flanders, Belgium are explored by means of a multidisciplinary approach in which three different groups which may benefit from a WAS are studied. The advantage of such an approach is that a more complete image can be formed from a problem (level of animal welfare) where these stakeholders interact in various ways (directly or indirectly). However, a disadvantage is that there is a need for focus through which necessary simplifications occur. In this thesis, the animal, the consumer and the farmer are the focus. This focus is illustrated in Figure 1.2., where the basis for the different research goals are displayed as the different bottlenecks which are identified in the positive feedback between these groups. These bottlenecks are first, assessing and monitoring animal welfare in

a valid and sensitive manner. This method should be credible to stakeholders and allow for communication with general information towards consumers (overall welfare) and at the same time allow for specific information about the welfare status to be communicated to the farmer (welfare measures). The second bottleneck concerns the existence of a consumer market for animal-friendly products, and how to communicate information on these products to consumers. This depends on (among other factors) price of the product, credibility, attractiveness of labels and therewith the welfare assessment method. The third and last bottleneck which will be investigated in this thesis is the existence and nature of a relationship between animal welfare and farm performance.
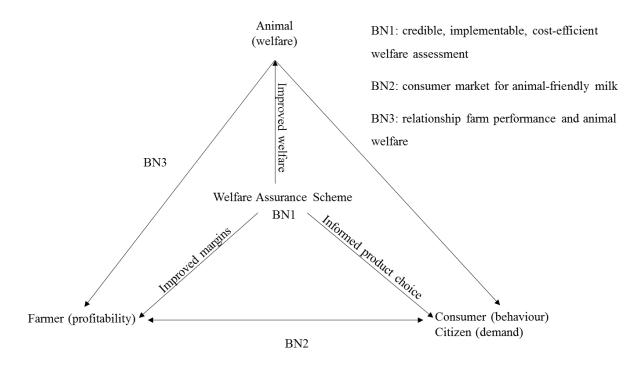

Figure 1.2: Thesis focus indicating bottlenecks (BN) in the added value of a Welfare Assurance Scheme for animals, consumers and farmers

In this thesis, we focus on dairy cattle welfare. The following reasons support this choice. First, as cows do not have to be killed to produce milk, consumers may perceive milk production as not related to meat production. This facilitates the provision of product information related to the animal and its welfare, more so than would be the case for meat, because many consumers do not want to be reminded that they are eating an animal which has been slaughtered for the purpose of their consumption when purchasing meat (Buller and Cesar, 2007, Te Velde et al., 2005, Kennedy et al., 2004, Holm and Møhl, 2000). Second, the dairy sector has a significant share (15%) of the EU agricultural output (European Commission, 2013). Third, market differentiation might be a welcome strategy for the dairy sector. With the recent abolishment of the milk quota in mind, market differentiation would be an alternative to the mere cost minimizing and milk production increasing strategy, which seems to be unfavourable for many farmers. Throughout this thesis, the term 'animal-friendly milk' is used in reference to milk from herds with an enhanced level of animal welfare when compared to other herds.

In this general introduction, dairy cattle welfare assessment will be discussed in more detail first, followed by the market for animal-friendly milk and by farm efficiency. Subsequently, the research objectives and research questions are presented, followed by a description of research design and sources of necessary data. Finally, the thesis outline is given.

1.1.1 Assessing dairy cattle welfare.

Animal welfare is a multidimensional concept, which is reflected in numerous definitions. For example, Fraser (2008) identified three views regarding animal welfare: 1) focusing on health and biological functioning (i.e. freedom of injuries and good health); 2) emphasizing affective states of animals, like pain, stress and pleasure; and 3) concentrating on the ability of animals to live natural lives, expressing natural behaviour (Fraser, 2008, Fraser et al., 1997). Even though overlap exists in these views, an increasing number of authors express the need to combine them in order to ensure that all aspects perceived to be part of animal welfare are included. Due to this multidimensional nature of animal welfare, a multi-criteria approach is needed to assess animal welfare (Blokhuis et al., 2003). This implies that various variables or measures should be used to assess overall welfare.

Animal welfare assurance schemes can be used to ensure a certain level of animal welfare on-farm. Such schemes can take on different forms; non-mandatory welfare guidelines by the industry or government, retailer assurance programmes (e.g. McDonald's animal welfare

standards) and product differentiation labelling programmes (e.g. Beter Leven Keurmerk) (Fraser, 2006). However, in order to actually ensure a proper level of welfare, such animal welfare assurance schemes should be based on accurate welfare assessments. Currently, this remains a challenge in animal welfare science.

As Fraser et al. (1995) argued, although specific aspects of welfare may be measured objectively (e.g. somatic cell count for clinical mastitis), integrating all these measures into one overall welfare assessment cannot be performed without some degree of subjectivity. The concept of animal welfare is therefore both science-based and values-based. This means that for the integration of various animal-based measures into one overall welfare index, weighting of the relative importance of these measures is subjective.

Welfare assessments traditionally mainly focused on input-based measures to assess housing conditions and management procedures. These assessments were based on what is deemed important for the animals' welfare (e.g. Bartussek et al., 2000, Bracke et al., 2000, 2002a,b, Sundrum et al., 1994), on pre-testing of housing equipment in experimental set-ups (Knierim et al., 2000) or on epidemiological studies (Ekesbo, 1992). Resource- and management-based data are used as input to generate a welfare score in these systems, because 1) this allows forming of recommendations for housing systems and management, 2) these aspects remain relatively constant and 3) they can often be measured objectively. Recently however, there is a growing consensus among animal welfare scientists that welfare should be assessed using output-based measures (animal-based measures), because these are thought to directly assess welfare as opposed to risks or chances for decreased or improved welfare. On a single farm, there are many interacting resources and management procedures which potentially affect animal welfare. Therefore, assessing welfare as a product of the interaction between the animal and its environment is highly complex, much more than measuring the actual effect of these interactions on the animals using animal-based measures.

In this doctoral thesis, we start from an existing and published assessment method, the Welfare Quality® (WQ) protocol. The WQ project was funded by the European Commission and involved researchers from 44 predominantly European institutes and universities (www.welfarequality.net). The project originated from public concern about farm animal welfare and a need for transparency in the food chain. The aims were to develop European standards for on-farm welfare assessment, product information systems and concrete strategies

for improving animal welfare. The WQ project resulted in protocols for on-farm welfare assessment of cattle (beef and dairy), pigs, and poultry (broiler chickens and laying hens). These protocols are based on mainly animal-based measures, which are aggregated using a three-step procedure into 12 criterion scores, four principle scores and an overall welfare category (Table 1.1, more details about the method are given in Chapters 2 and 3). There are in principle 31 measures in the protocol for dairy cow welfare assessment (Welfare Quality®, 2009, P. 69 – 70). However, throughout this thesis measures are either pooled or separated for the analyses of the integration method differently, resulting in various numbers of measures shown in the tables. When using multiple measures to assess welfare, it may be possible for a serious welfare problem to be compensated for by good scores (indicating good welfare) for other measures (Blokhuis et al., 2003, Spoolder et al., 2003, Capdeville & Veissier, 2001). The integration methods used in WQ are based on the intention to limit this compensation between welfare dimensions (Botreau et al., 2007b).

Table 1.1: All principles, the corresponding criteria and measures used in the Welfare Quality® assessment protocol for dairy cattle welfare

Principles	Criteria	Measures	Aggregation method measures
Good feeding	Absence of prolonged hunger	Body Condition Score (% very lean animals)	Spline curve fitting
	Absence of prolonged thirst	Availability & cleanliness water	Decision tree
Good housing	Comfort around resting	Time needed to lie down; collisions during lying down; on edge/outside of lying area; cleanliness	Converted to ordinal scores, combined in weighted sums and spline curve fitting
	Thermal comfort	No measure for dairy cattle	
	Ease of movement	Loose housing or tied housing and exercise	Decision tree
Good health	Absence of injuries	Lameness; integument alterations	Combined in weighted sums, spline curve fitting and Choquet integration
	Absence of diseases	Respiration/digestive diseases; mastitis; mortality; dystocia, downer cows	Converted to ordinal scores, combined in weighted sums and spline curve fitting
	Absence of pain induced by management procedures	Mutilations (dehorning; tail docking; use of anaesthetics/analgesics)	Decision tree
Appropriate behaviour	Expression of social behaviour	Prevalence agonistic interactions	Combined in weighted sums and spline curve fitting
	Expression of other behaviours	Access to pasture	Spline curve fitting
	Good human-animal relationship	Avoidance distance at feeding rack	Combined in weighted sums and spline curve fitting
	Positive emotional state	Qualitative Behavioural Assessment	Combined in weighted sums and spline curve fitting

Although the WQ protocol is currently the most comprehensive protocol which relies on predominantly animal-based measures for on-farm animal welfare assessment, concerns have been raised about various aspects of the protocol. Validity and reliability of certain measures is criticised (Tuyttens et al., 2010, Knierim and Winckler, 2009) along with the protocols' time-consuming nature and therefore costs that would be associated with implementation in practice (de Vries et al., 2013b). In addition, the integration methods are criticised for resulting in a focus on a limited set of (often resource-based) measures which colludes with the protocols' aim to be multidimensional (de Vries et al., 2013a, Heath et al., 2014). These concerns and causative factors are investigated further in this doctoral thesis with a specific focus on the integration methods. Such studies are required to determine whether and how the integration method of the WQ protocol should be adjusted. For this end, the sensitivity of the WQ integration method to extreme changes in individual measures is investigated along with the reasons for any differences in sensitivity (Chapter 2) and the degree of correspondence between integrated WQ scores and the opinion of experts (Chapter 3). In this thesis, because we focus on the WQ protocol, experts are defined as animal welfare scientists which are trained users of this protocol (more specific information in Chapter 3 and 4). Furthermore, as this critical evaluation strengthened concerns about the WQ integration method, an alternative welfare assessment protocol was developed designed to meet criticism on previous methods (Chapter 4). The aim of these specific chapters (2, 3 and 4) altogether is to work towards a (more) valid and reliable method to assess herd welfare which accurately reflects the welfare status on dairy farms.

1.1.2 Market for animal-friendly milk

Studies consistently show a high concern for farm animal welfare among citizens in Europe (European Commission, 2005 - 2016). In a recent study by the European Commission (2016), 94% of European citizens indicated to believe it is important to protect the welfare of farm animals and 82% stated that farm animals should be protected better than is currently the case. This citizen concern may not only relate to animal welfare, but also linked to a perceived relationship between poor animal welfare and food safety (Grunert, 2005, Harper and Henson, 2001) stimulated by disease-outbreaks like bluetongue, swine fever, BSE, and foot-and-mouth disease. Additionally, concerns are amplified by animal protection organisations which use campaigns to raise awareness about farm animal welfare issues (e.g. Global Action in the Interest of Animals (GAIA) in Belgium).

However, citizen concern is currently not fully reflected in a high market share for all animal-friendly products or for products that may be perceived as such by citizens. Studies shows that organic products are associated with enhanced animal welfare by citizens and that concerns about animal welfare are among the drivers to purchase organic products (Harper and Makatouni, 2002, Van Loo et al., 2014). However, the low market share for organic animal products (2.7% for organic milk in Belgium in 2015 (Samborski and Van Bellegem, 2016)) does not seem in agreement with citizens' stated concern for animal welfare. Previous studies argue this is due to a duality between citizen' attitudes and how they act as consumers when it comes to ethical decision making. This is referred to as the 'citizen-consumer duality' or 'attitude – behavioural intention gap' in different studies (Ajzen, 2001, Grunert, 2006, Kraus, 1995, Vermeir and Verbeke, 2006). Various explanations have been put forward for this duality, e.g. the habitual nature of food shopping (Minteer et al., 2004), and more individual factors like knowledge about, experience with and attitudes towards farm animals. Other, more practical, factors can be put forth as well like perceived lack of availability and higher prices of animal-friendly products (Harper and Henson, 2001). Recent studies tend to focus on heterogeneity in the consumer market for animal-friendly products, which is currently not fully met in the market supply (Weinrich and Spiller, 2016, de Jonge et al., 2015, van Herpen et al., 2015, Vanhonacker et al., 2007).

The supply market for animal-friendly milk products may be perceived as limited in Belgium. It is mainly a binary market, where the options are organic or conventional milk. Organic milk is perceived by citizens to be animal-friendlier and, besides health and safety reasons, animal welfare concerns are drivers to purchase organic products (Harper and Makatouni, 2002). However, although the regulations for organic milk do entail some welfare-related demands, their focus is not on animal welfare only. Organic products in Belgium follow Council Regulation (EC) No 834/2007, which states: 'Organic stock farming should respect high animal welfare standards and meet animals' species-specific behavioural needs' (L189/2). Yet, the only specific demands are that the animals should have permanent access to open air areas (preferably pasture) and that tethering for a prolonged period of time is prohibited. Therefore, demands do not incorporate the whole multidimensional nature of animal welfare. When comparing these demands to the welfare criteria as developed by WQ (Welfare Quality® Consortium, 2009), only two of twelve ('Expression of other normal behaviour' and 'Freedom of movement') are covered. High welfare standards for some of the most important welfare

criteria according to various stakeholders (e.g. 'Absence of disease'(Tuyttens et al., 2010)) are not included in this Council Regulation.

Previous studies have shown that the level and the type of societal concern for animal welfare varies among species of farm animals (Harper and Henson, 2001). In addition, the degree of 'feeling well-informed about welfare' is also influenced by farm animal species (European Commission, 2007). Also, variation between concern for farm animal welfare has been shown among (European) countries (Harper and Henson, 2001). Recent studies have provided evidence for the existence of a differentiated market in a country which could be described as one of the frontrunners when it comes to private initiatives and regulations on-farm animal welfare (the Netherlands (van Herpen et al., 2015, de Jonge and Van Trijp, 2014)). However, results may not be extrapolated to countries like Belgium, where fewer such initiatives exist. Therefore, to be able to stimulate the Flemish dairy industry towards processing and marketing animal-friendly milk, research is needed into whether a (heterogeneous) consumer market for animal-friendly milk actually exists in Flanders, and how these products should be positioned in the current milk market.

Positioning refers to the differentiation of a certain 'brand' from competitors by creating a specific image of the brand and therewith pinning it down for consumers (Fuchs and Diamantopoulos, 2010, Keller and Lehmann, 2006). However, specific consumer needs should be taken into account when positioning a product, if it ought to create added value to buying the specific product (Keller, 1993). The current study lays the basis for positioning of animal-friendly milk in the Flemish milk market by firstly investigating determinants of Flemish consumers' intention to purchase animal-friendly milk (Chapter 5). We focus on attitudinal determinants of the intention to purchase animal-friendly milk, based on a theory explaining human behaviour (Ajzen, 1985). In this theory, behaviour is assumed to arise from behavioural intention which is influenced by attitudes towards the behaviour, perceived social pressure to perform the behaviour and perceived behavioural control. All three these aspects are influenced by an individual's emotions and attitudes (Ajzen & Fishbein, 1980, Fishbein & Ajzen, 1975). Results from previous studies have suggested a link between intention to purchase animal-friendly products and attitudes towards animal welfare (e.g. Vanhonacker and Verbeke, 2009, Vermeir & Verbeke, 2006). Therefore, in Chapter 5, attitudinal determinants of the intention to purchase animal-friendly milk are the main focus, due to its influence on all three aspects of the TPB and its link with animal welfare. In addition, the role of product attribute importance and

associations between animal welfare and other product attributes will be investigated, because of their potential effect on purchasing decisions in the specific case of animal-friendly products. Furthermore, different market segments are distinguished, and these segments are profiled to examine how to communicate to these different groups of consumers (Chapter 6).

1.1.3 Farm performance

When legislation or market differentiation is needed to elevate the level of farm animal welfare, this implies that current farm resources and/or management practices impair animal welfare and that a shift towards a higher level of animal welfare might be costly. Within a cost increasing paradigm, market differentiation is one option to account for costs which may be associated with increasing levels of farm animal welfare. However, a question which arises is whether enhanced animal welfare might actually also be linked to higher productivity and revenues, and as such would be synonymous to good economic performance of the farm.

Like animal welfare, the economic performance of a farm is highly multifactorial as well. Performance of dairy farms is mostly measured and benchmarked with 'cost accounting' methods and partial productivity measures (Allendorf and Wettemann, 2015, Coelli et al., 2005). Cost accounting is a process that measures all costs of a company and compares input results to outputs (in dairy farms, input is e.g. feed costs and output is e.g. milk revenues). However, farm performance consists of more than only costs and must be balanced with any economic benefit of increased welfare. The same applies for partial productivity measures, which are (iceberg) measures of total productivity but only provide a limited view on both costs and benefits. A higher milk production per cow does not necessarily mean optimal economic results, as high production may be associated with reproductive loss and diseases (Lucy, 2001). Therefore, in this thesis, we use efficiency analysis as a more integrated method based on production theoretical principles such as the decreasing marginal returns. The concept of 'farm efficiency' is used to evaluate and compare economic success of farms incorporating both the physical non-monetary inputs and price information to decompose performance in a technical and economic compound (Barnes et al., 2011, Latruffe et al., 2004). The goal of efficiency analysis is to estimate farm efficiency relative to the sample and to label those farms with the highest output/input ratio (e.g. producing most litres of milk with as little feed as possible) as 'efficient' (Färe et al., 2013). More detailed information about farm efficiency is given in Chapter 7.

Currently, a very limited number of studies have focussed on farm efficiency in combination with animal welfare measures. Some recent studies on this subject either employ a limited set of welfare indicators focused on reproductive performance (Allendorf and Wettemann, 2015), or focus on difference in level of production costs between systems associated with different levels of animal welfare (Gocsik et al., 2016). To our knowledge no other study has used the multidimensional nature of animal welfare (i.e. using various animal-based welfare measures) in combination with farm efficiency. In this doctoral thesis, we link two such complex concepts to each other.

1.2   *Research objectives and research questions*

To promote transparency in the food chain and therewith stimulate the production and increase the market share of animal-friendly milk, this study lies the foundations for animal welfare assurance schemes. The aim of this thesis was therefore twofold. First, the aim was to identify a dairy cattle welfare assessment method which accurately reflects the welfare status on dairy farms. Furthermore the aim was to examine added value of dairy cattle welfare for milk through market potential and a positive relationship with farm efficiency. The research objectives with the corresponding research questions of this thesis are presented below.

The first research objective was to identify a (more) valid and sensitive method to assess herd welfare which accurately reflects the welfare status on dairy farms. As the WQ protocol currently is the most promising and renowned method for animal welfare assessment on-farm, this methods' sensitivity and correspondence with expert opinion was investigated using two research questions. Research question 1.1 was formulated as: 'is there variation in sensitivity of integrated WQ outcomes (criteria and principle scores and overall welfare category) to extremely low and high values of the various welfare measures and what are the reasons for this variation in sensitivity?'. Research question 1.2 was: 'to what extent do discrepancies exist between WQ integrated scores and expert opinion?'. As both these research questions lead to a critical evaluation of WQ integration methods which strengthens the growing criticism on the WQ protocols, a third research question (1.3) was employed: 'How can we develop an alternative integration protocol of dairy cattle welfare measures which uses only selected animal-based measures and is multidimensional, simple and transparent, discriminative and coherent with expert opinion?'.

Research objective 2 was formulated as 'examining the (existence of a) market for animal-friendly milk'. For this end, insight was gained into factors influencing the intention to purchase animal-friendly milk using research question 2.1: 'in what way do citizens' attitudes, perceptions of product-related factors, and milk consumption relate to their intention to purchase animal-friendly milk?'. In addition, the consumer market for animal-friendly milk was examined using research question 2.2: 'to what extent do market opportunities for animal-friendly milk exist in Flanders, Belgium?' and research question 2.3: 'what characterizes different segments within the market for animal-friendly milk?'.

The final research objective 3 was based on one research question (3.1): 'to what extent are animal welfare and cost efficiency related as investigated using integrated multi-criteria approaches?'.

1.3   Research design and data sources

For this thesis, data were collected to answer the various research questions through 1) farm visits performing the WQ protocol and collecting data on farm efficiency in Flanders, 2) collating several European databases of WQ data, 3) performing surveys among dairy cattle welfare experts and 4) performing an online consumer survey. Table 1.2 gives an overview of the data sources used for the six research chapters in this dissertation. More detailed descriptions of the study samples are included in each of the research chapters.

The data used in Chapter 2, 3 and 4 were collected in 2013 and 2014 during farms visits at 121 dairy farms in Flanders. During these farms visits, the WQ protocol for on-farm welfare assessment of dairy cattle welfare was used to collect data on 33 welfare measures. These data were used to examine the sensitivity of the WQ protocol to individual measures and to assess coherence of WQ integrated outcomes with expert opinion. However, both these purposes called for a substantial but realistic spread in observed WQ measure values. For this end, WQ data from nine different countries (Macedonia, the Netherlands, France, Scotland, Denmark, Romania, Northern Ireland, Spain and Austria) were collated into a database of 491 herds. Data on expert opinion used for Chapter 3 and Chapter 4 were collected using two online surveys which were performed in 2015 and 2016 respectively.

Data for Chapter 5 and 6 were collected in March 2014, through a consumer survey in the Northern Dutch speaking part of Belgium (Flanders). Total sample size was 787 respondents and participants were selected from a professional market research agency panel.

For Chapter 7, to assess farm efficiency, from some of the Flemish farms on which WQ assessments were performed (n = 41) farm economic data were collected as well.

Table 1.2 Overview of the data sources used for the various research objectives and corresponding research questions

Research objectives	Research question	Data source	Sample size
1 To identify a (more) valid and sensitive method to	1.1	WQ EU database (including Flemish WQ data)	491 herds
assess herd welfare which	1.2	WQ EU database and Expert Survey I	8 – 13 experts
accurately reflects the welfare status on dairy farms	1.3	WQ EU database and Expert Survey II	13 experts
2 Examining the (existence of a) market for animal-friendly milk	2.1 2.2 2.3	Online consumer survey	787 complete responses
3 Examining the relationship between farm efficiency and dairy cattle welfare	3.1	Farm economic data and Flemish WQ data	41 farms

1.4   Thesis outline

In this dissertation, six research chapters are compiled which contain (adjusted versions of) the six scientific manuscripts that have been published, accepted or submitted to international peer-reviewed journals.

In the second and third chapters, the Welfare Quality® protocol for on-farm dairy cattle welfare assessment is the main focus. Chapter 2 describes (reasons for) variations in sensitivity of the

integrated welfare scores to changes in individual welfare measures. In the same line, Chapter 3 focuses on correspondence of expert opinion with integrated Welfare Quality® outcomes of the dairy cattle welfare assessment protocol. As critical evaluation (in this thesis and in other studies (Buijs et al., 2016, Heath et al., 2014, de Vries et al., 2013a, Tuyttens et al., 2010) of the WQ protocol gave reason to question the integration method employed, in Chapter 4 the focus is on an alternative integration protocol of various dairy cattle welfare measures.

Chapters 5 and 6 focus on the consumer market for animal-friendly milk in Flanders. In Chapter 5 attitudinal determinants of consumers' intention to buy welfare friendly milk are examined. In Chapter 6, consumer segments are identified and profiled which differ in their intention to purchase animal-friendly milk and in their evaluation of the current state of dairy cattle welfare.

In Chapter 7, economic and technical performances of dairy farms in relation to animal welfare are examined.

In the final Chapter 8, results of the preceding research chapters are discussed and implications and limitations formulated

Chapter 2
Sensitivity of integrated Welfare Quality® outcomes

Abstract

The WQ protocol for on-farm dairy cattle welfare assessment describes 33 measures and a step-wise method to integrate the outcomes into 12 criteria scores, grouped further into 4 principle scores and finally into an overall welfare categorization with 4 possible levels. The relative contribution of various welfare measures to the integrated scores has been contested. Using a European dataset (491 herds), we investigated 1) the variation in sensitivity of integrated outcomes to extremely low and high values of measures, criteria and principles and 2) the reasons for this variation in sensitivity. Results indicate that, as intended by the WQ consortium, the sensitivity of integrated scores depends on 1) the observed value of the specific measures/criteria, 2) whether the change was positive or negative, and 3) the relative weight attributed to the measures. Additionally, two unintended factors with a considerable influence, appear to be side-effects of the complexity of the step-wise integration method. Namely 1) the number of measures that are integrated into criteria and principle scores, and 2) the aggregation method of the measures. As a result, resource-based measures related to drinkers, which have been subject to criticism regarding their validity in measuring absence of prolonged thirst, have a much larger influence on the integrated scores than health-related measures like 'mortality rate' and 'lameness score'. Hence, the integration method of the WQ protocol for dairy cattle should be revised to ensure that the relative contribution of the various welfare measures to the integrated scores more accurately reflect their relevance for dairy cattle welfare.

Introduction

Accurate welfare assessment is vital for improving herd animal welfare. In dairy cattle, measures have been developed and validated for a wide variety of both negative and positive aspects of welfare. However, only a few protocols exist that aggregate the scores of multiple welfare measures into one score or index reflecting the overall welfare status of a given herd. Such an overall welfare status score might be used for example in the communication with consumers (food labelling), as an incentive for on-farm welfare improvements and as regulative target (Blokhuis et al., 2010). Examples of schemes that calculate an overall welfare status of dairy cattle are a protocol by Whay et al (2003) based on the "Five Freedoms" (British Veterinary Association, 1992) which generates a ranking of herds' welfare status. The Animal Needs Index (ANI) produces an overall welfare score based on integrating mostly resource-based measures (measures of environmental aspects that affect welfare) (Bartussek et al., 2000). Finally, the Welfare Quality® (WQ) protocol categorizes overall welfare status of a herd as 'excellent', 'enhanced', 'acceptable' or 'not classified' based on a step-wise integration procedure (Welfare Quality® Consortium, 2009). The current study focuses on the WQ protocol, as this is the only protocol that predominantly uses animal-based measures to calculate an integrated welfare index. Such measures are generally preferred over resource-based measures as the latter tend to reflect risk factors for welfare impairments instead of directly measuring welfare (Blokhuis et al., 2010, Blokhuis et al., 2003).

In the EU project WQ, protocols for the welfare assessment of the main types of farm animals (cattle, pigs and chickens) were proposed. The protocol describes 31 welfare measures performed on-farm by means of behavioural observations, qualitative behaviour assessment, an avoidance distance test, a management questionnaire, a resource checklist and clinical scoring (Table 1.1). Subsequently, three steps are used to integrate separate measures into one overall welfare category. Measures are first integrated into criteria scores on a scale of $0 - 100$ which are in turn collated into four welfare principles principles ('good feeding', 'good housing', 'good health' and 'appropriate behaviour'). These principle scores are then used to determine herds' overall welfare category (Welfare Quality®, 2009). Integration methods are intended to limit compensation of poor scores with better scores on other welfare aspects (Veissier et al 2011). Expert opinion of social and animal scientists and stakeholders was used to determine weights for the integration method (Botreau et al 2007a). Additionally, the protocols were

designed with the intention of modifying and updating assessment methods according to advances in animal welfare science (www.welfarequalitynetwork.net/network/45848/7/0/40).

Discussion has arisen recently about WQ's integration methods. Some of the measures have been criticised for their poor or undocumented reliability, validity or feasibility (Tuyttens et al., 2015, de Jong et al., 2015, de Vries et al., 2013b, Knierim and Winckler, 2009). In addition, studies have indicated that a few, resource-based measures have a disproportionately large influence on the overall welfare category (de Vries et al 2013, Heath et al 2014). Both critical findings may harm the credibility and validity of the WQ protocol in assessing herd welfare. To further examine the functioning of the WQ protocol for dairy cattle, the aim of the current study was to examine 1) if there is variation in sensitivity of integrated outcomes (criteria and principle scores and overall welfare category) to extremely low and high values of measures, criteria and principles and 2) the reasons for this variation in sensitivity. More specifically, we aimed to critically evaluate whether differences in sensitivity appear to be deliberate and justifiable rather than unintentional side-effects of the complex integration method. To this end, we performed a sensitivity analysis by replacing individual observed values with both the theoretical and the actually observed worst and best values. The latter values were based on a large database of WQ data that reflect a wide range of herd types in Europe and thereby ensuring a substantial but realistic spread in observed values.

Materials and Methods

WQ Protocol

Only a brief description of the integration method of the WQ protocol for on-farm dairy cattle welfare assessment is given here. The full protocol can be found at http://www.welfarequalitynetwork.net/.

Step 1: from measures to criteria scores. Aggregation starts by combining 31 measures into 12 criteria (Table 1.1). Because the scales of measures differ, various aggregation methods are used. For categorical measures, decision trees are used resulting in a score between $0 - 100$ where 100 indicates the best possible score. Other measures are converted to ordinal scores where required (e.g. scores within 'comfort around resting' are converted into three categories: normal, moderate problem or serious problem using thresholds in seconds for 'time needed to lie down' and percentages of cows for the other measures) and then combined into index values

using weighted sums. Spline functions are used to re-weight these sums based on their severity according to expert opinion. Finally, when multiple spline functions were used, Choquet integrals are used to combine these functions into criteria scores on a scale of 0 – 100 (Botreau et al 2007b). These algorithmic operators calculate the criteria scores in such a way that a poor score cannot be fully compensated for by a better score in another measure (Botreau et al 2007b). Consequently, poor scores will have a bigger influence on the integrated scores than good scores. Using Choquet integrals, the weight given to each element (measures or criteria) depends on its value relative to the other elements, where the poorest score always gets the highest weight (Botreau et al., 2008, Welfare Quality 2009).

Step 2: from criterion scores to principle scores. To integrate criterion scores into principle scores, Choquet integrals are used (Welfare Quality® Consortium, 2009). The resulting principle scores range from 0 (worst) to 100 (best).

Step 3: from principle scores to overall welfare category. The third and final integration step is from principle scores to overall welfare category. Dairy cattle welfare in a herd is considered 'excellent' when it scores >50 for each principle and >75 on two of them. When a herd scores >15 on each principle and >50 on at least two of them, it is classified as 'enhanced'. 'Acceptable' herds score >5 for all principles and >15 for at least three principles. Herds that do not reach the thresholds for the category 'acceptable' are considered 'not classified' (Botreau et al., 2009).

Data Collection and Collation

To reflect the current range present in Europe across various herding systems, pre-existing research datasets of assessments using the WQ protocol for on-farm dairy cattle welfare were collated from seven European research institutes and included data from 10 countries. The collected samples were selected by the research institutes to be representative for 1) small scale dairy herds in Macedonia (n = 12); 2) non-organic and non-tie stall dairy herds in The Netherlands (n = 60) and France (n = 128); 3) random herds with individual Somatic Cell Count data available (SCC, to be able to calculate WQ scores) in Belgium (n = 140), Scotland (n = 16) and Denmark (n = 42); 4) typical herds for the regional low-input herding systems in Romania, Northern Ireland and Spain (n = 30); and 5) loose housed dairy herds with at least 20 cows in Austria (n = 65). The total number of herds in the collated database was 491. To ensure a homogenous integration method for all data, integrated WQ scores were calculated from raw

data using a custom-made integration procedure programmed in R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). The results were checked for coherence with the INRA WAFA webtool (http://www1.clermont.inra.fr/wq/), in which WQ measure values can be entered (for dairy cows, fattening pigs, growing pigs and broilers), and WQ criteria, principle and classification scores can be calculated.

Sensitivity Analysis

In order to investigate the extent to which values for separate measures affected the criteria and principle scores and the overall welfare category, we performed a sensitivity analysis. Such analyses are used to measure how estimates of values using a given model fluctuate when they are replaced with other values. In this study, each observed herd-level observation for each measure and each herd was replaced one by one with both the theoretical and the observed (of the entire dataset of 491 herds) worst and best values. This was repeated for individual criteria and principle scores to assess the impact of criteria and principle scores on the overall welfare category. For these calculations, farms that were already in the highest or lowest overall welfare category were excluded. Because these excluded farms were not able to shift categories, retaining them would give a distorted picture of the results. Subsequently, the median increase and decrease in criteria and principle scores and the percentage of herds that shifted to a lower or higher overall welfare category were quantified for each replacement by the theoretically and observed worst and best values.

For most measures, values that were altered were scored as either percentage of cows (e.g. % of severely lame cows) or 'yes' and 'no' (e.g. for drinker cleanliness). However, for some measures (ADF (Avoidance Distance at the feed rack), lameness and integument alterations) the aggregated measure indexes rather than individual percentages were replaced with worst and best scores. Because these measures together add up to 100% of animals, changing percentages within these could create an impossible situation (i.e. percentages would add up to over 100%). In addition, the theoretical best score for the measures 'length of drinking trough' and 'number of drinking bowls' depends on the average number of cows on the herd. Therefore, we replaced these with scores that would meet the requirements for all herds in the dataset (10,000 cm for drinking trough length and 100 for number of drinking bowls) as best scores. For the measures of dehorning and tail docking, we replaced the actual methods used at each herd with the methods which would generate the best (i.e. no dehorning, no tail docking

respectively) and the worst score (i.e. dehorning using surgery with no anaesthetics or analgesics, tail docking using a rubber band without anaesthetics and analgesics, respectively).

Results

None of the 491 herds were originally (i.e. before replacement with worst/best scores) in the 'excellent' category, 174 (35%) were in the 'enhanced' category, 308 (63%) in the 'acceptable' category and nine (2%) in the 'not classified' category. For eight of the nine 'not classified' herds, classification was due to a 'good feeding' principle score below 5 (the threshold for the not-classified category). The median, minimum, and maximum scores are given at the measure (Table 2.2) and principle and criterion level (Table 2.5). For several measures, the observed range spanned the entire theoretical range (i.e. $0 - 100$ for percentages, $0 - 24$ for hours and $0 - 365$ for days). However, for several other measures (18), criteria (6 out of 12) and principles (3 out of 4), the observed data range was narrower than was theoretically possible (Table 2.2 and Table 2.3). Only 5% of herds were not dehorned or disbudded, 18% were disbudded using caustic paste, 76% using thermocautery, and 1% was dehorned using surgery. Analgesics and/or anaesthetics were used during these procedures in 24% and 60% of the herds, respectively. Only 5 (ca. 1%) herds were tail docked (3 by rubber ring and 2 by surgery). Analgesics were never used during tail docking whilst anaesthetics were used in two herds.

Sensitivity analysis using observed values: measurement level

Sensitivity of the overall welfare category. When separate measure values were increased to the observed maximum value (i.e. to the level of the herd that scored best for that specific measure) fewer herds shifted between overall categories than when separate scores were decreased to the observed minimum value (Table 2.2). For most measures, the highest percentage of shifts occurred between the 'enhanced' and 'acceptable' category, although for increases in some measures ('% lean cows', 'number of water bowls', 'drinker cleanliness' and 'loose versus tied housing') highest % of shifts to a higher category were between 'not classified' and 'acceptable' (data not shown).

Replacements of measure values only rarely led to negative shifts of more than one category and never to positive shifts of more than one category (Table 2.2). The effects of replacing a measure often differed greatly, even between measures that belong to the same principle. 'Good

health' was the only principle for which changing the values of any of its underlying measures did not result in a substantial (>10%) effects on herd classification. All measures that were the only measure of a certain criterion caused a relatively high percentage of herds to shift category: '% lean cows', 'loose or tied housing' and the 'QBA index' when replaced with the theoretical worst score, with the exception of the 'ADF index'. Although seemingly combined with many other measures, most measures of the criterion 'absence of prolonged thirst' had a relatively large influence as well. Most upgrades to a higher overall welfare category were achieved by increasing (to the observed maximum levels) 'number of water bowls', 'trough length', and to a lesser extent '% of cows colliding'. Within the two criteria that contained most measures, either sensitivity was lowest for all measures ('absence of disease') or sensitivity was greater for those measures that were attributed the highest weight (i.e. within 'comfort around resting', the measures for resting behaviour are given a higher weight than cleanliness).

Table 2.2: Percentages of herds[1] (n = 491) that were downgraded or upgraded 1 or 2 overall welfare categories when individual values at measure level were replaced with observed worst and best values per measure (observed median, min. and max. score given between brackets)

Principles	Criteria, continuous measures	Observed median, min - max	Observed worst score		Observed best score
			% downgraded 1 category	% downgraded 2 categories	% upgraded 1 category
Good feeding	Absence of prolonged hunger				
	% lean cows[2]	4, 0 – 88	53	0	5
Good housing	Comfort around resting				
	Mean time needed to lie down (s)	6, 3 – 20	10	0	6
	% collisions during lying down movements	33, 0 – 100	5	0	12
	% of cows lying on edge/outside of lying area	0, 0 – 73	11	0	8
	% of cows with dirty flanks	64, 0 – 100	0	0	7
	% of cows with dirty lower legs	80, 0 – 100	2	0	7
	% cows with a dirty udder	37, 0 – 100	2	0	7
Good health	Absence of injuries				
	Lameness index	88, 37 – 100	6	0	5
	Integument alterations index	53, 0 – 100	2	0	4
	Absence of diseases				
	Range of all disease-measures[2]	-	1-2	0	0-1
Appropriate behaviour	Expression of social behaviour				
	Head butts/cow/15 min.	0.5, 0 – 7	13	0	1
	Displacements/cow/15 min.	0.4, 0 – 5	16	0	4
	Expression of other normal behaviour				
	Number of hours on pasture)	7.5, 0 - 24	9	0	1
	Number of days on pasture	175, 0 - 365	9	0	1
	Human-animal interaction				
	ADF index	67, 23 – 100	13	0	6

	Positive emotional state				
	QBA index	0.3, -11 – 5	24	1	7
	Criteria, binary measures	% farms with theoretical best score			
Good feeding	Absence of prolonged thirst				
	Water flow	82	22	3	3
	Trough length	18	26	1	19
	Number of water bowls		11	1	20
	Drinker cleanliness	76	23	0	8
	At least 2 drinkers/cow	84	9	0	1
Good housing	Ease of movement				
	Loose or tied housing	93	38	2	3
Good health	Absence of pain induced by management procedures				
	Dehorning method	5	9	0	3
	Tail docking method	95	8	0	0

[1]Percentages were based on the herds that were actually able to shift one or two categories. For downgrades of one category n = 482, for downgrades of two categories n = 174. For upgrades of one category n = 491.

[2]As absence of disease contains a very high number of measures with a very small range of shifts, we present only the range here. All separate measures can be found in the Appendix A.

Sensitivity of the principles and criteria scores. The sensitivity analysis of the effect of changes in separate measure values on the principles scores (Table 2.3) and on the criteria scores (Table 4) showed the same pattern as the sensitivity analysis of the overall welfare category. The decrease caused by changing a measure to the lowest observed value was usually greater than the increase caused by changing the same measure to its highest observed value. Exceptions to this trend often concerned measures of which the observed values were very poor. Furthermore, measures that caused the greatest difference tended to belong to criteria that contain few other measures. Exceptions to this trend once again concerned most measures within 'absence of prolonged thirst' and the measure '% of cows colliding with housing'. There was a difference in the sensitivity of the principles and the criteria in that measure values have a more direct influence on criteria scores, and therefore had a greater influence on criteria scores than on principle scores.

Sensitivity analysis using observed values: criteria and principle level. Of all principles, alteration of 'good feeding' led to the most negative shifts as well as positive shifts (Table 2.4). Moreover, replacing the 'good feeding' score to the lowest observed score in the database caused all 'enhanced' herds to be re-categorised as 'non-classified'. Alterations to the other principle scores never caused a change of more than one overall welfare category. Alteration of the 'good housing' principle caused the fewest positive shifts of all principles, as most farms already scored relatively high for this principle (median score of 54).

Of all criteria, replacement with the lowest observed score was most effective in generating negative shifts for 'absence of prolonged hunger' followed by 'absence of prolonged thirst'. Replacement with the highest observed score was most effective in generating a positive shift for 'absence of prolonged thirst'. Both criteria within the principle 'good housing' ('comfort around resting' and 'ease of movement') caused 27% of herds to be downgraded when replaced by the observed minimum. Effects of replacing criteria scores within the 'good health' and 'appropriate behaviour' principles varied considerably between criteria.

Table 2.3: Median (min – max) decrease and increase in principle and criteria scores when measure scores were replaced with worst and best observed measure scores were replaced with worst and best observed measure scores

Principles / Criteria	Measures	Changes in principles scores		Changes in criteria scores	
		Median decrease in worst scenario	Median increase in best scenario	Median decrease in worst scenario	Median increase in best scenario
Good feeding					
Absence of prolonged hunger	% lean cows	24 (0 – 71)	5 (0 – 69)	67 (0 - 98)	30 (0 - 98)
	Water flow	11 (0 – 85)	0 (0 – 85)	29 (0 - 97)	0 (0 - 0)
	Trough length	25 (0 – 85)	0 (0 – 85)	29 (0 - 97)	0 (0 - 97)
Absence of prolonged thirst	Number of water bowls	0 (0 – 85)	10 (0 – 85)	0 (0 - 97)	12 (0 - 97)
	Drinker cleanliness	12 (0 – 60)	0 (0 – 60)	40 (0 - 68)	0 (0 - 68)
	At least 2 drinkers/cow	0 (0 – 35)	0 (0 –35)	20 (0 – 97)	0 (0 - 40)
Good housing					
	Mean time to lie down	6 (0 – 20)	5 (0 – 20)	10 (0 – 32)	8 (0 - 31)
	% collisions during lying down movements	0 (0 – 19)	11 (0 – 17)	0 (0 - 32)	18 (0 - 27)
Comfort around resting	% of cows lying on edge/outside of lying area	10 (0 – 20)	0 (0 – 29)	16 (0 – 32)	0 (0 – 46)
	% cows with dirty flanks	0 (0 – 5)	4 (0 – 14)	0 (0 – 12)	6 (0 – 22)
	% cows with dirty lower legs	0 (0 – 9)	4 (0 – 12)	0 (0 – 15)	6 (0 – 18)
	% cows with a dirty udder	0 (0 - 9)	4 (0 – 8)	0 (0 – 15)	6 (0 – 18)
Ease of movement	Loose or tied housing	24 (0 – 37)	0 (0 – 40)	66 (0 – 66)	0 (0 – 85)
Good health					

Absence of injuries	Lameness index	13 (0 – 37)	5 (0 – 35)	27 (3 – 69)	33 (0 – 57)
	Integument alteration index	4 (0 – 24)	5 (0 – 26)	10 (0 – 44)	26 (0 – 42)
Absence of disease	Number of coughs/cow/minute	0 (0 – 0)	0 (0 – 0)	0 (0 – 0)	0 (0 – 0)
	% cows with nasal discharge	1 (0 – 12)	0 (0 – 10)	8 (0 – 35)	0 (0 – 21)
	% cows with ocular discharge	1 (0 – 12)	0 (0 – 8)	8 (0 – 35)	0 (0 – 35)
	% cows with hampered respiration	1 (0 – 5)	0 (0 – 1)	4 (0 – 14)	0 (0 – 14)
	% cows with diarrhoea	2 (0 – 12)	0 (0 – 10)	9 (0 – 35)	0 (0 – 35)
	% cows with vulvar discharge	3 (0 – 12)	0 (0 – 7)	10 (0 – 35)	0 (0 – 24)
	% cows with SCC >400.000	2 (0 – 12)	1 (0 – 12)	8 (0 – 35)	4 (0 – 35)
	% cows mortality	2 (0 – 11)	0 (0 – 12)	8 (0 – 35)	0 (0 – 35)
	% calvings with dystocia	1 (0 – 12)	0 (0 – 13)	7 (0 – 35)	4 (0 – 35)
	% downer cows	2 (0 – 12)	1 (0 – 13)	0 (0 – 35)	3 (0 – 35)
Absence of pain induced by management procedures	Dehorning method (none, surgery)	15 (0 – 35)	6 (0 – 40)	50 (0 – 89)	48 (0 – 98)
	Tail docking method (none, ring)	14 (0 – 34)	0 (0 – 6)	6 (0 – 89)	0 (0 – 0)
Appropriate behaviour					
Expression of social behaviour	Head butts/cow/15 min.	13 (0 – 37)	1 (0 – 16)	69 (0 – 100)	8 (0 – 49)
	Displacements/cow/15 min.	16 (0 – 44)	2 (0 – 30)	69 (0 – 100)	19 (0 – 93)
Expression of other behaviour	Number of hours on pasture	15 (0 – 38)	0 (0 – 34)	64 (1 – 100)	0 (0 – 85)
	Number of days on pasture	15 (0 – 38)	1 (0 – 24)	64 (1 – 100)	15 (0 – 86)
Good human-animal relationship	ADF index	10 (0 – 37)	9 (0 – 37)	31 (0 – 87)	56 (0 – 87)
Positive emotional state	QBA index	20 (0 – 50)	7 (0 – 44)	52 (0 – 93)	40 (0 – 93)

Table 2.4: Percentages of herds[1] (n = 491) that shifted into a different overall welfare category when individual scores were replaced with observed worst and best criteria or principle scores

Principles, Criteria	Original observed median, min - max	Observed worst score		Observed best score	
		% farms downgraded 1 category	% farms downgraded 2 categories	% farms upgraded 1 category	% farms upgraded 2 categories
Good feeding	40, 4 – 100	64	100	36	1
Absence of prolonged hunger	70, 3 – 100	59	0	6	0
Absence of prolonged thirst	60, 3 – 100	35	3	30	1
Good housing	54, 6 – 86	37	0	13	0
Comfort around resting	27, 0 – 80	27	0	13	0
Ease of movement	100, 15 – 100	27	0	0	0
Good health	34, 8 – 86	37	0	23	0
Absence of injuries	35, 4 – 100	21	0	8	0
Absence of diseases	40, 12 – 100	4	0	7	0
Absence of pain induced by management procedures	52, 2 – 100	9	0	3	0
Appropriate behaviour	35, 6 – 86	37	0	25	0
Expression of social behaviour	69, 0 – 100	16	0	5	0
Expression of other normal behaviour	64, 0 -100	9	0	8	0
Good human-animal relationship	44, 13 – 100	14	0	8	0
Positive emotional state	53, 0 – 93	24	1	7	0

[1]Percentages were based on the herds that were actually able to shift one or two categories. For downgrades of one category n = 482, for downgrades of 2 categories n = 174. For upgrades one category n = 491, for upgrades of two categories n = 317.

Differences between replacement with observed and theoretical scores. For several measures, criteria and principles, the observed range did not span the entire theoretical range. For three measures ('lameness index', 'head butts/cow/15 min' and 'ADF index'), four criteria ('absence of injuries', 'absence of diseases', and 'absence of pain induced by management procedures') and three principles ('good housing', 'good health' and 'appropriate behaviour'), replacement with the theoretical worst and best scores instead of the observed scores resulted in a higher % of herds shifting between overall welfare categories (Table 2.5). For four measures ('% lean cows', 'lameness index', 'number of coughs/cow/15 min.', '% cows with hampered respiration' and 'ADF index), this resulted in a higher median percentage of increase or decrease of the

principle and criteria scores than when worst or best observed scores were used (Table 2.5 and Table 2.6).

Table 2.5: Percentages of herds[1] (n = 491) that shifted into a different overall welfare category when scores at the measure, criterion, and principle level were replaced with theoretical[1] worst and best scores

	Worst score		Best score
	% downgraded 1 category	% downgraded 2 categories	% upgraded 1 category
Measures[1]			
Lameness index[3]	10	0	5
Head butts/cow/15 min.[3]	16	0	1
ADF index[3]	20	0	6
Criteria[1]			
Absence of injuries[4]	29	1	8
Absence of diseases[4]	36	1	7
Absence of pain induced by management procedures[4]	12	0	3
Good human-animal relationship[4]	23	0	8
Principles[1]			
Good housing[4]	64	100	13
Good health[4]	64	100	23
Appropriate behaviour[4]	64	100	25

[1]Percentages were based on the herds that were actually able to shift one or two categories. For downgrades of 1 category n = 482, for downgrades of 2 categories n = 174. For upgrades of 1 category n = 491.

[2]Scores shown are of those measures, criteria and principles where replacement with theoretical score generated different results than when replaced with observed score.

[3] Theoretical worst score was 100

[4] Theoretical worst score was 0

Table 2.6: Median (min – max) decrease and increase in principle and criterion scores when measures were replaced with worst and best theoretical values

Principles, criteria	Measures	Change in principle scores		Change in criteria scores	
		Median decrease in worst scenario	Median increase in best scenario	Median decrease in worst scenario	Median increase in best scenario
Good feeding[1]					
Absence of prolonged hunger	% lean cows	25 (2 – 73)	5 (0 – 69)	69 (2 – 100)	30 (0 – 98)
Good health[1]					
Absence of injuries	Lameness index	15 (2 – 39)	5 (0 – 35)	27 (3 – 69)	33 (0 – 57)
Absence of disease	Number of coughs/cow/15 min.	4 (0 – 12)	0 (0 – 0)	10 (5 – 35)	0 (0 – 0)
	% cows with hampered respiration	4 (1 – 12)	0 (0 – 1)	10 (6 – 35)	0 (0 – 14)
Appropriate behaviour[1]					
Good human-animal relationship	ADF index	46 (11 – 82)	9 (0 – 37)	44 (13 – 100)	55 (0 – 87)

[1] Scores shown are of those where replacement with theoretical score generated different results than when replaced with observed score

Discussion

This study investigated the sensitivity of the integrated scores of the WQ protocol for on-farm dairy cattle welfare assessment to extreme changes in individual measure, criterion and principle scores. The impact of one by one replacement of observed herd-level measure, criteria and principle scores by extremely low or high values had variable effects on the more highly integrated scores and on the overall welfare category. Investigation into what type of replacements have a large versus negligible impact suggests that a considerable part of this variation appears to be an unwanted side-effect of the complex step-wise integration method rather than being intentional or justifiable.

Sensitivity analysis using observed values: measurement level

Generally, the impact of a replacement with an extremely low score was bigger than replacement with an extremely high score. This reflects the intention of the WQ integration method to limit compensation of poor scores with better scores on other welfare aspects (Veissier et al., 2011). The effect of replacing observed measure scores with extreme values on more highly integrated scores (criteria and principles) and on the overall welfare category was very variable and seemed to depend on various aspects. Replacements of the measures '% lean cows', 'loose housing or tied housing', the 'QBA index', 'drinker trough length' and 'drinker cleanliness', had a bigger impact on overall classification compared to other measures (particularly when substituted by observed worst scores). The common feature shared by the first three measures is that they are the only measure of the criterion they belong to ('absence of prolonged hunger', 'ease of movement' and 'positive emotional state', respectively). One other criterion is also documented by a single measure, namely 'expression of other normal behaviour' measured with the ADF-test. This measure had less impact compared with the aforementioned three measures, presumably because the ADF index was already poor for most farms to begin with (so the change by replacing the actual score with the worst possible score was often very small).

The relatively large impact of drinker space and drinker cleanliness is in accordance with previous findings for both the dairy cattle protocol (de Vries et al 2013, Heath et al 2014) and the WQ broiler chicken protocol (Buijs et al 2016). This seems to be caused by a combination of factors. First, these measures both belong to the criterion of 'absence of prolonged thirst' which contains few measures that matter for calculating the criterion scores (in the decision tree only number/length of drinkers and cleanliness are taken into account). The other measures are either prerequisites for the required number/length of drinkers and therefore less directly influence criterion scores ('water flow'), or are related to the number of drinkers ('at least 2 drinkers/cow'). Second, the principle 'good feeding' contains only one other criterion apart from 'absence of prolonged thirst', whereas most other principles are composed of more criteria. It could be argued that the large impact of these measures is not necessarily problematic if they are valid measures of an important welfare problem. However, as resource-based measures, drinker space and cleanliness would appear to be potential risk factors rather than direct measures of thirst (Vanderhasselt et al 2014, Sprenger et al 2009). Moreover, to our knowledge, the validity of these measures of thirst has not yet been tested. Therefore, the

finding that these measures have a relatively large influence on integrated scores, can be considered problematic. Animal-based measures of thirst have been developed, such as voluntary water consumption (in broiler chickens; Vanderhasselt et al., 2014, Sprenger et al., 2009). It could be promising to develop voluntary water consumption tests further, and test for validity as a measure of prolonged thirst in dairy cattle.

Replacements of measures within the principle 'good health' with the best or worst scores had the lowest influence on principle and criterion scores and on overall classification, in accordance with previous results (de Vries et al., 2013, Buijs et al., 2016). This is remarkable because it includes measures which indicate important welfare problems in dairy cattle according to many experts, such as mortality, mastitis and lameness (Lievaart and Noordhuizen, 2011). In addition, Tuyttens et al (2010) reported that both consumers and farmers rank health aspects as the most important for farm animal welfare. The very limited effect of extreme changes in measures within the criterion 'absence of diseases' on integrated WQ scores seems at least partially caused by the aggregation method of this criterion. In this aggregation, prevalence of symptoms of diseases is compared to warning and alarm thresholds (e.g. warning threshold for nasal discharge is 5% of cows and alarm threshold 10% of cows). Subsequently, a weighted sum is calculated of warnings and alarms, with a weight of 1 for warnings and 3 for alarms, which is computed into the criterion score using a spline function. Because of this method, increasing prevalences that were already above the alarm threshold (or decreasing those that were already below the threshold) will not affect classification at all. Also, when the prevalence of one disease symptom changes, it has only a limited effect on the criterion scores because it is aggregated with many other disease symptoms.

Similarly to measures within 'absence of diseases', measures within 'absence of injuries' also had relatively low impact on the integrated scores. However, a different method is used to integrate the measures within 'absence of injuries' to one score. Partial scores for lameness and integument alterations are first calculated using weighted sums and i-spline curves, and are then combined using a Choquet integral. The lameness index had most influence, but still caused only 10% of herds to be downgraded when replaced with the theoretical worst score (i.e. 100% severely lame cows). This surprisingly lowest impact seems to be due to the large number of criteria within the principle 'good health', and to the observation that herds often score relatively low for these criteria. Therefore, changing another score within this principle to a low

score is likely to have a smaller effect than when it is done for a score in another principle with fewer criteria such as 'good feeding'. Due to the limited impact of good health measures on overall welfare categorisation, a situation could occur where farms categorised as 'acceptable' or better have 100% severely lame animals, while this may obviously be considered a major welfare problem.

Regarding positive shifts, the percentage of cows colliding with housing had a relatively large positive impact when replaced with best observed score. This is likely because a large proportion of farms (55%) were classified as having a serious problem for this measure to begin with, so for many farms a vast improvement was possible (compared to 37% for % of cows laying out and 28% which were above the threshold value of 6.3 seconds for mean time needed to lie down).

Sensitivity analysis using observed values: criteria and principle level

There are two, three, or four criteria per principle. This difference in the number of criteria is reflected in the results of the sensitivity analysis: replacement with the worst criteria scores within the principle ('good feeding') containing only two criteria ('absence of prolonged hunger' and 'absence of prolonged thirst') generated most shifts towards a different welfare category. The principle 'good housing' also consists of only two criteria for which measures have been developed (for its third criterion 'thermal comfort' no measure is available). The impact of both criteria are smaller compared to the two criteria of 'good feeding'. However, even though for 'thermal comfort' no data are collected, the missing criterion score is replaced with the best score among 'comfort around resting' and 'ease of movement'. This dilutes the effect of a very low score on either of these two criteria. Although some validated measures for thermal comfort exist for dairy cattle (e.g. respiration rate, Schutz et al., 2010), inclusion of such measures may complicate timing of farm visits, as these measures are highly dependent upon ambient temperature. Further research on how to deal with these complexities in the WQ protocol is necessary, or removal of 'thermal comfort' as a criterion for dairy cattle welfare should be considered.

In line with the criteria, of all principles, alteration of 'good feeding' led to the most negative and positive shifts when replaced with observed worst and best scores. For negative shifts this was because 'good feeding' was the only principle for which scores <5 were observed, which automatically categorizes a herd as 'not classified'. For positive shifts, this was because this

principle caused more 'not classified' and 'acceptable' categorizations than any other principle (as 131 farms originally had a score between 5 and 15 for this principle, as opposed to nine for housing, three for health and 23 for behaviour; data not shown). Therefore, more positive shifts could occur when 'good feeding' was altered than when the other principles were replaced with observed maximum scores.

Differences Between Replacement With Observed and Theoretical Scores

As the sample size in the current study was large and contained a wide variety of herds (given the different sampling aims), we can draw some conclusions about the observed scores in relation to theoretical scores. For most measures, observed scores spanned the entire theoretical range. This means that for the dairy cattle protocol, most limits set by WQ seem realistically attainable. For some measures however, observed scores were less extreme than the theoretical scores. In most cases, this did not affect criterion scores as these were within the criterion 'absence of diseases', where warning and alarm thresholds are used to integrate scores. For lameness index and ADF index however, fewer shifts of the overall welfare category were observed when replaced with the observed scores. This was also reflected in the corresponding criteria and principle scores, of which the worst possible score never occurred. This is one of the reasons that the principles 'good health' and 'appropriate behaviour' never caused herds to be categorized as 'not classified' when replaced by the observed minimum score.
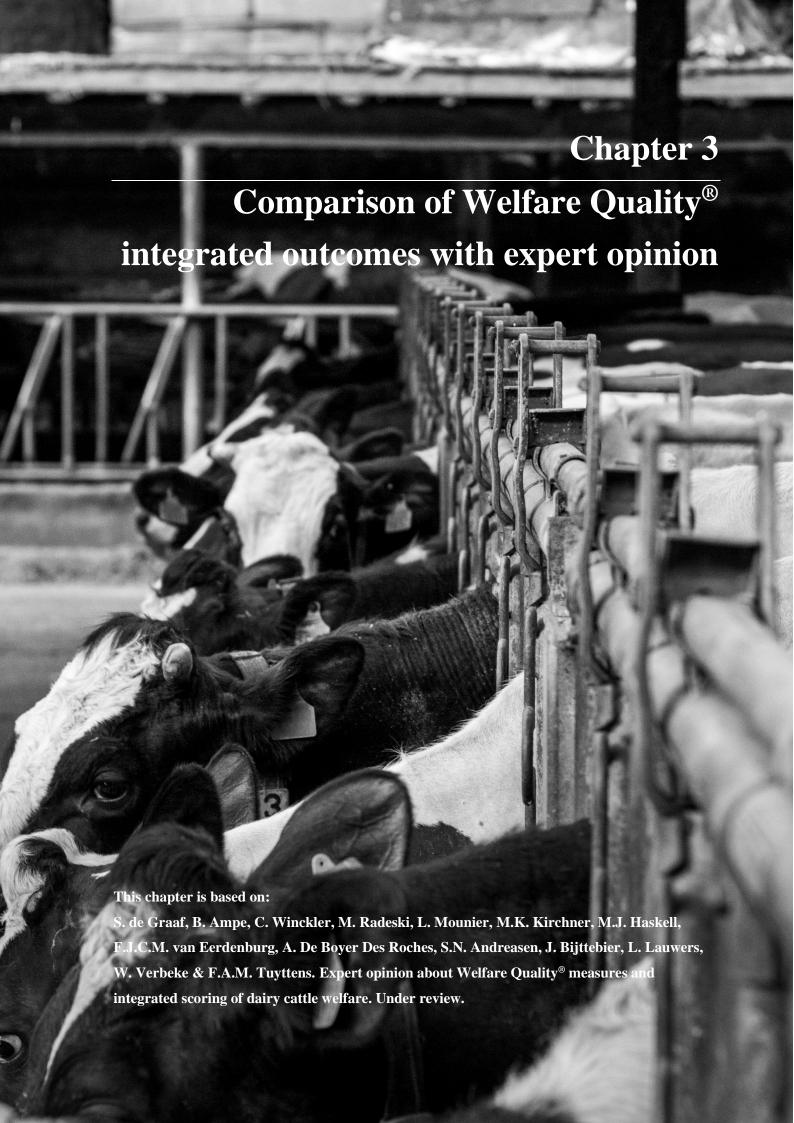
Conclusion

The results of the current study provide insight into the functioning of the integration methods for the dairy cattle WQ protocol. Findings indicate that the sensitivity of integrated scores to replacement of individual scores by extreme scores is dependent on a number of factors which were intended by the WQ protocol: 1) the observed value of the specific measure (or criterion), relative to the values of the other measure in the same criterion (or principle); 2) whether the values were replaced by an extremely low or an extremely high value (more impact of the former); 3) the relative weight WQ attributes to the measures. However, two other factors that were not intended and appear to be unwanted side-effects of the complexity of the step-wise integration method, also had considerable influence. These factors were: 1) the number of measures that are integrated into criteria and principle scores; and 2) the aggregation method of the measures. The effect of both integration method and grouping is problematic, as it should

be the severity of the welfare problem that affects the overall category. As a result, sensitivity is highest for changes in measures of the 'good feeding' principle, of which a large proportion of the measures are criticized for their validity (i.e. measures of 'absence of prolonged thirst'). On the contrary, measures within the principle 'good health' have the lowest impact while some of these measures are considered to most severely affect dairy cattle welfare. For instance, a farm in the 'acceptable' category or higher could theoretically have 100% severely lame animals.

Animal Welfare Implications

This study indicates that the WQ integration method does not adequately balance the relative importance of all welfare measures that are included in order to adhere to the multidimensional nature of animal welfare. Therefore, using the current integrated WQ scores could lead to a focus on a limited set of (often resource-based) measures which is hard to justify. As this harms the credibility of the assessment protocol, we recommend a revision of the integration method, so that the relative contribution of the various welfare measures to the integrated scores more correctly reflects their relevance for dairy cattle welfare.

Chapter 3

Comparison of Welfare Quality®
integrated outcomes with expert opinion

Abstract

The Welfare Quality® (WQ) protocol for on-farm dairy cattle welfare assessment describes 27 measures and a step-wise method to integrate values for these measures into 12 criteria scores, grouped further into four principle scores and finally into an overall welfare categorization with four levels. This WQ integration method requires investigation as it is increasingly being challenged, indicating discrepancies between the welfare assessment conducted by scientific dairy cattle welfare experts and the integrated scores. Therefore, we conducted an online survey to examine whether experts' (trained users of the WQ protocol for dairy cattle) opinions correspond with the integrated scores (criteria, principles and overall categorization) calculated according to the WQ protocol. First, the experts' scores ($n = 14 - 16$) for reliability, validity and their ranking of the importance of all measures for herd welfare were compared to the degree of actual impact of these measures on the WQ integrated scores. Logistic regression was applied to identify the measures that affected the WQ overall welfare categorization into the 'not classified' or 'enhanced' categories for a database of 491 European herds. The smallest multivariate model whilst maintaining the highest % of both sensitivity and specificity for the 'enhanced' category contained six measures, the model for not-classified contained four measures. Some of the measures that were ranked as least important by experts (e.g. measures relating to drinkers) had the highest influence on the WQ overall welfare categorization. Conversely, measures rated as most important by the experts (e.g. lameness and mortality) had a lower impact on the WQ overall category. In addition, experts were asked to allocate 'criterion' and 'overall' welfare scores to seven focal herds selected from the database ($n = 491$ herds). Data on all WQ measures for these focal herds relative to all other herds in the database were provided. The degree to which expert scores corresponded to each other, the systematic difference and the correspondence between median expert opinion and the WQ criterion scores were then tested. The level of correspondence between expert scoring vs. WQ scoring for 6 of the 12 criteria and for the overall welfare score was low. The integration methods of the WQ protocol for dairy cattle thus lacked correspondence with selected experts on the importance of several welfare measures.

Introduction

Assessing animal welfare is a highly complex task. Animal welfare is a multidimensional concept, which calls for a multi-criteria assessment using a multitude of welfare-measures (Fraser et al., 1997, Mason and Mendl 1993). To express the overall welfare status of a group of (farm) animals in one score or index, measure data should be integrated which requires interpretation and balancing. The lack of a 'gold standard' for animal welfare assessment (i.e. there is no standardized and commonly agreed-on method for assessing the overall welfare status of a group of farm animals) implies that some degree of subjectivity is inevitable when weighting different measures (Spoolder et al., 2003). To be widely accepted, an overall welfare index ought to correspond with society's concept of animal welfare and with the opinion of experts, i.e. people who are seen by society to have adequate knowledge and expertise about animal welfare. However, opinions on the concept of animal welfare may differ between experts and society, and even within society and within experts. For example, producers tend to highlight basic health and functioning of farm animals while non-producers tend to emphasize the need for a natural living environment of farm animals (reviewed by Sørensen and Fraser, 2010). It can be argued that for people without expertise in dairy cattle welfare and the specific welfare measures involved, it is too difficult to adequately balance the importance of different welfare measures. It has been shown that providing detailed information about on-farm collection methods of welfare measures, significantly influences the relative weights they are given by experts (Rodenburg et al., 2008). Therefore, the current study elicited experts on dairy cattle welfare only, and on the specific welfare measures involved.

To date, the WQ protocols are most likely the most renowned and comprehensive method for overall welfare assessment of different farm animal species (chickens, pigs and cattle) (Welfare Quality®, 2009). Unlike some other welfare assessment protocols, WQ relies predominantly on animal-based measures. Resource-based and management-based measures, in contrast, mostly reflect risk factors for welfare impairments instead of directly measuring welfare (Blokhuis et al., 2003, 2010). The WQ protocols are based on four main welfare principles ('good feeding', 'good housing', 'good health' and 'appropriate behavior') which are split into 12 independent welfare criteria (Table 1.1). Various welfare measures (n = 27 for dairy cows) were selected to assess these welfare criteria. The WQ protocol describes three steps to integrate these welfare measures into an overall final welfare category. Methods of integration are based upon expert

opinion of social and animal scientists and stakeholders (Botreau et al., 2007a), depending on the integration step. For aggregation of measures into criteria scores, animal scientists were consulted (n = 5-6) who were involved in the choice and development of the WQ measures (Botreau et al., 2008). They were asked to score several situations which could occur on-farm per criterion (e.g. for integument alterations within the criterion 'absence of injuries', experts were asked to score 11 hypothetical farms with varying prevalence of hairless patches, wounds and swellings). For aggregation from criteria to principle scores, social scientists were involved as well, using a similar approach. For the final step, several scenarios for reference profiles were developed to aggregate principle scores into an overall category. These scenarios were tested for 69 European dairy farms (Austrian, German and Italian) to firstly compare their ability to detect differences between farms. Secondly, stakeholders were consulted to assess which scenario was most appropriate and thirdly, the degree to which each scenario matched with the general impression of observers for 44/69 dairy farms was assessed. The four overall categories ('excellent', 'enhanced', 'acceptable' or 'not classified' (Welfare Quality®, 2009)) should reflect both the multi-dimensional nature of welfare and the relative importance of the various welfare measures (Botreau et al., 2009).

Recent critical evaluations of the WQ integration methods indicate that in the dairy cattle protocol a few resource-based measures appear to have a disproportionately large influence on integrated scores (de Vries et al., 2013a, Heath et al., 2014). For example, the measures for the criterion 'absence of prolonged thirst' (i.e. number, adequate functioning and drinker cleanliness) have a relatively large influence on integrated scores, although they are criticized for their low or undocumented validity (de Jong et al., 2016, Tuyttens et al., 2014, de Vries et al., 2013, Knierim and Winckler, 2009). In contrast, some of the most pressing welfare problems for dairy cattle as assessed by experts (i.e. mortality, lameness and mastitis, Nielsen et al., 2014, Lievaart and Noordhuizen, 2011, Whay et al., 2003), had a smaller influence on integrated scores. These findings point towards possible discrepancies between the welfare assessment by scientific dairy cattle welfare experts and the WQ integrated scores.

The WQ protocols were designed with the intention of modifying and updating assessment methods according to advances in animal welfare science. Currently, a large group of researchers has become familiar with the protocol to such a degree that they could be labelled as 'experts on the WQ Assessment protocol for dairy cows' (further referred to as 'experts'). In

addition, a large number of farm visits has been performed by these researchers, allowing for a thorough evaluation of the impact which measures have on overall welfare categorization. Therefore, analyzing the correspondence between WQ integrated scores and the opinion of such experts would now be feasible. Hence, the objective of the current study was to analyze correspondence between welfare assessment by experts and the WQ integration method. We performed this by examining whether measures which impact WQ categorization most, are also those which are deemed most important by experts. If there is a lack of correspondence here, the integrated WQ scores are not expected to compare well with how experts would score welfare. We tested this using a small sample of farms and selected experts.

Materials and Methods

WQ Protocol

A brief description of the WQ protocol for on-farm dairy cattle welfare assessment is presented below; the full protocol can be found at http://www.welfarequalitynetwork.net/. In short, the protocol describes 31 on-farm welfare measures (Table 1.1) that are subsequently integrated in a 3-step process to arrive at an overall welfare category. First, the welfare measures of various scales are combined into scores for 12 welfare criteria on a scale of 0 (worst) – 100 (best) (Table 1.1), using various aggregation methods (for details see Welfare Quality®, 2009) . Second, criteria are integrated into scores for four welfare principles using Choquet integrals, algorithmic operators which ensure that a poor score cannot be fully compensated by a better score in another criterion (Botreau et al., 2008). Principle scores can range from 0 (worst) to 100 (best). The third and final integration step is an outranking procedure from principle scores, arriving at an overall welfare category. Dairy welfare in a herd is considered 'excellent' when that herd scores >50 for each principle and >75 on two of them. When a herd scores >15 on each principle and >50 on at least two of them, it is classified as 'enhanced'. 'Acceptable' herds score >5 for all principles and >15 for at least three principles. Herds that do not reach the thresholds for the category 'acceptable' are considered 'not classified'. These reference profiles for overall welfare categorization were based on data from 69 herd assessments in the European Union (Botreau et al., 2009).

Collating WQ Data

Datasets of assessments using the WQ protocol for on-farm dairy cattle welfare were collated from seven European research institutes. Data from 10 countries (Macedonia, The Netherlands, France, Belgium, Scotland, Denmark, Romania, Northern Ireland, Spain and Austria) and 491 herds were used. The collected samples were selected to be representative for 1) small scale dairy herds in Macedonia (n = 12); 2) non-organic and non-tie stall dairy herds in The Netherlands (n = 60) and France (n = 128); 3) random herds with individual Somatic Cell Count data available (SCC, to be able to calculate WQ scores) in Belgium (n = 140), Scotland (n = 16) and Denmark (n = 42); 4) typical herds for the regional low-input herding systems in Romania, Northern Ireland and Spain (n = 30); and 5) loose housed dairy herds with at least 20 cows in Austria (n = 65). Integrated WQ scores were calculated from raw data using a custom-made integration procedure programmed in R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). The resulting welfare scores were in agreement with the INRA WAFA webtool (http://www1.clermont.inra.fr/wq/), in which WQ measure scores can be entered (for dairy cows, fattening cattle, growing pigs and broilers), and WQ criteria, principle and categorization scores are provided. None of the 491 herds were in the 'excellent' category, 174 (35%) were in the 'enhanced' category, 308 (63%) in the 'acceptable' category and nine (2%) in the 'not classified' category.

Survey

The survey was sent to 31 experts, partially completed by 14 - 15 (depending on the question) and totally completed by 8 experts. The survey was sent to experts which the co-authors knew to be experienced in the WQ assessment protocol for dairy cow welfare. These experts were in turn asked to provide contact details of any additional experts which would be suitable (trained to use the WQ protocol). All experts who filled out the survey, were not involved in creating the survey. All experts had experience with the WQ protocol for dairy cattle (i.e. were trained to perform the WQ protocol for dairy cattle and had performed on-farm WQ assessment of dairy herds), were animal scientists and had authored at least 1 peer-reviewed scientific paper about dairy cattle welfare involving the WQ protocol. Experts were all European and a total of 8 different nationalities was represented (British, Spanish, Macedonian, Dutch, Finnish, Austrian, German and French). Experts were surveyed on their judgement of the reliability, validity and importance of all WQ measures. In questions based on data from the WQ EU

database, they were asked to score the farms for each WQ criteria and to assign an overall welfare score.

Reliability, Validity and Ranking of all WQ Measures for Dairy Cattle. The experts were asked to indicate how acceptable they judged the reliability and validity of all measures using a tagged visual analogue scale from 0 to 100. For the purpose of this scoring, some measures were pooled so 27 measures remain, for example the measures pertaining to drinker space are pooled into one. Tags were 'not acceptable (<25)', 'just acceptable (25 – 50)', 'acceptable (50 – 75)', and 'very acceptable (75 – 100)'. 'Reliability' was defined in the survey as 'a combination of inter-observer, intra-observer and test-retest reliability'. 'Validity' was defined as 'the measure measures what it is supposed to'. Experts were then asked to rank all WQ measures according to importance for the overall welfare status of a herd of dairy cattle from 1 (most important) – 27 (least important). It was mentioned that for ranking, (inter alia) reliability, validity, perceived relevance and prevalence may be considered.

Comparing WQ Criteria Scores Using Expert Opinion. To assess the degree to which integrated WQ criteria scores correspond to expert opinion, we randomly selected five herds from the 'acceptable' welfare category and two herds from the 'enhanced' category out of the entire dataset. This reflects the distribution of the dataset in which 1.8% of the herds were categorized as 'not classified' (9 herds), 62.7% as 'acceptable' (308 herds), 35.4% as 'enhanced' (174 herds) and none as 'excellent'. The experts were shown graphs of all measures per criterion separately, showing the distribution of all herds in the database (example of one criterion: Figure 3.1, data shown in Table 3.2). The 'focus herds' were highlighted using triangles in different colors, and tables stated the data for each. Experts were asked to score the herds for all 11 criteria (excluding the criterion 'thermal comfort', as this is not measured on-farm for dairy cattle) on a 0-100 tagged visual analogue scale using the tags 'not classified (< 20)', 'acceptable (20 – 55)', 'enhanced (55- 80)' and 'excellent (>80)'.
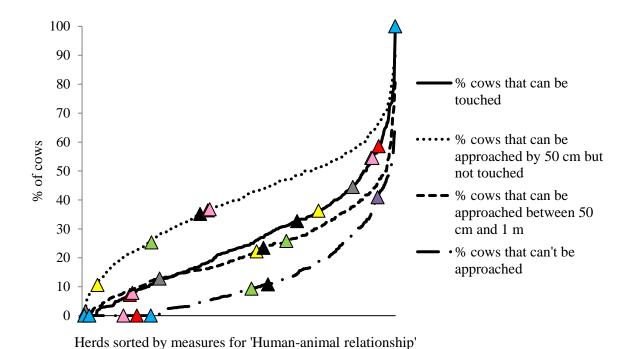
Figure 3.1: Example figure from the survey among experts, portraying the distribution of all herds in the database (n = 491) for the measures of the ADF, within the criterion 'Human-animal relationship'. Coloured triangles mark the seven focus herds.

Table 3.2: Measure values of each of the seven herds presented to experts in the survey

Criteria, measures Herd #	1	2	3	4	5	6	7
Absence of prolonged hunger							
% of lean cows	0	3	17	5	11	3	24
Absence of prolonged thirst							
Number of water bowls/cow	0.5	0.0	0.0	0.0	0.0	0.6	0.05
Trough length/cow (cm)	0.0	7.9	4.7	28.6	9.0	0.0	0.0
Drinker cleanliness	Yes	Yes	No	Yes	Yes	Yes	Yes
At least 2 drinkers/cow	No	Yes	No	No	Yes	No	Yes
Resting comfort							
Mean time needed to lie down (s)	4.6	4.6	7.5	4.1	6.6	5.4	6.8
% collisions during lying down movements	16	15	72	0	37	8	33
% of cows lying on edge/outside of lying area	50	11	0	0	0	35	0
% of cows with dirty flanks	34	55	81	14	67	79	70
% of cows with dirty lower legs	57	37	85	38	20	79	100
% cows with a dirty udder	18	21	77	10	42	48	95
Ease of movement							
Loose (L) or tied (T) housing	T	L	L	L	L	T	L
Absence of injuries							
% of moderately lame cows	0	13	88	0	23	0	84
% of severely lame cows	32	0	12	10	17	27	5
% of cows with at least one lesion	7	12	72	28	13	20	68
% of cows with no lesions but at least one hairless patch	98	18	28	38	21	100	32
Absence of disease							
Number of coughs/cow/minute	0.05	0.00	0.13	0.10	0.06	0.17	0.00
% cows with nasal discharge	59	0	0	0	5	18	0
% cows with ocular discharge	0	0	0	0	0	0	0
% cows with hampered respiration	0	0	0	5	0	0	0
% cows with diarrhea	5	0	0	0	0	0	16
% cows with vulvar discharge	0	0	0	0	0	0	3
% cows with SCC[1] > 400.000	8	21	25	0	14	8	12

% cows mortality	5	3	4	0	4	3	4
% calvings with dystocia	0	21	0	0	1	6	3
% downer cows	0	6	0	0	0	6	5
Absence of pain induced by management procedures							
Dehorning method, Thermal (T), Caustic paste (P) or None (N)	T	P	P	N	P	P	T
Use of Analgesics	No	Yes	Yes	No	Yes	No	No
Use of Anesthetics	No	Yes	Yes	No	Yes	No	Yes
Expression of social behavior							
Number of Head butts/cow/15 min.	0.8	4.0	0.7	0.0	0.4	0.4	1.0
Number of Displacements/cow/15 min.	0	1.2	0.1	0.0	0.2	0.0	0.8
Expression of other normal behavior							
Number of hours on pasture	214	180	0	0	0	214	195
Number of days on pasture	19	9	0	0	0	8	9
Human-animal relationship							
% of cows that could be touched	36	55	59	100	55	44	30
% closer than 50 cm but not touched	11	36	37	0	26	2	35
% between 50 and 1 m	23	9	9	0	11	14	24
% > 1 m	30	0	0	0	9	41	11
Positive emotional state							
QBA[2] score	43	40	8	91	77	66	54

[1]Somatic Cell Count

[2]Qualitative Behaviour Assessment

Expert Scoring Based on All WQ Measurements

The experts were then asked to score overall welfare based on all measures from the WQ protocol. They were shown one figure with box plots for all measures (part of the figure for one criterion: Figure 3.2). These showed the same herds as in the first figure using the same colored triangles. Experts were asked to score overall welfare of the seven focal herds us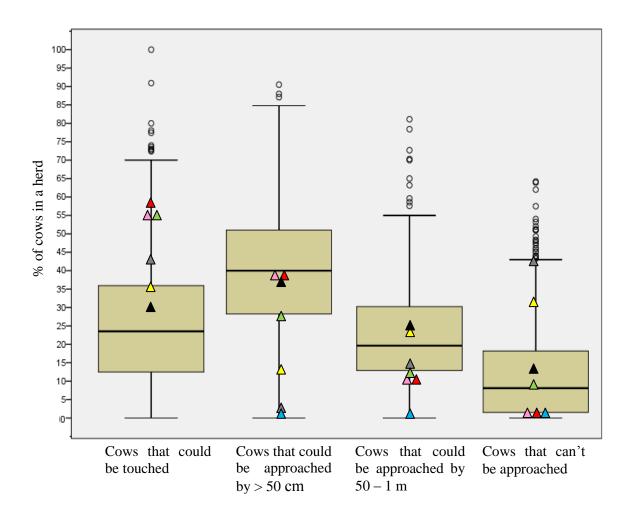ing a 0-100 tagged visual analogue scale, with the tags 'not classified' (< 20), 'acceptable' (20 – 55), 'enhanced' (55- 80) and 'excellent' (>80).

Figure 3.2: example boxplot figure from the survey among experts, portraying the distribution of all herds in the database (n = 491) for the measures of the ADF, within the criterion 'Human-animal relationship'. Coloured triangles mark the seven focus herds.

Statistical Analysis

The statistical analysis was performed in R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). The analysed data (except overall welfare categorization) were considered to be sufficiently normally distributed, based on the graphical evaluation (histogram and QQ-plot) of the residuals.

Reliability, Validity and Ranking of all WQ Measures for Dairy Cattle. To examine the influence of median reliability and validity scores and their interaction on median ranking of all measures, we used a linear mixed regression model with reliability and validity scores as independent variables, and importance rank as dependent variable. A random effect for expert was included in the model to account for the repeated measures.

Comparing WQ Criteria Scores with Expert Opinion. To assess the systematic difference between the median expert opinion score and the WQ criteria scores for each focal herd (n = 7), a paired t-test was performed. To model the correspondence of median expert opinion and the WQ criteria scores, a linear model was fitted and the coefficient of determination was calculated. Additionally, the intra-class correlation coefficient (ICC) was calculated to assess the degree of coherence between individual expert opinions.

Predicting Overall Welfare Categorization Using WQ Measures. To analyse which measures affected the WQ overall categorization both into the lowest (not classified) and the highest (enhanced, as no farms were categorized as excellent) categories, welfare categories of the entire European dataset (n = 491) were divided into two binary variables (1=enhanced, 0=other for variable 1; and 1=not classified, 0=other for variable 2). Logistic regression was used to identify measures that affected overall categorization both univariate and multivariate. For the latter, a model was built using step-wise forward selection, retaining measures with a $P < 0.05$ while maintaining the highest R^2. Collinearity was checked for measures used within the models. Model outcome was assessed by calculating specificity and sensitivity using the following formula:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad \text{Sensitivity} = \frac{TP}{TP + FN}$$

Where TN = true negatives, FP = false positives, TP = true positives and FN = false negatives. Negatives were those farms categorized as 'other' and positives were those farms categorized as either 'enhanced' for the first binary variable or 'not classified' for the second.

Results

Perceived Reliability, Validity and Ranking of WQ Measures

Median validity and reliability scores for all measures were 'acceptable' to 'very acceptable' (i.e. median scores > 50, Table 3.3). Nevertheless, there was variation in median scores for the various measures, ranging from 60 to 100 and from 50 to 90 for reliability and validity respectively. Highest median ranking was attached to 'lameness score' (rank 2), 'body condition score' (4), 'mortality rate' (7) and 'integument alterations' (7). 'Lameness score' and 'integument alternations' received the highest median validity scores (89 and 90, respectively), along with 'lying outside the lying area' (89) and 'tail docking method' (88). 'Tied versus loose housing' (100), measures of drinker space ('Centimeters of trough per cow (min. 6 cm), number of water bowls per cow (min. 0.10) and at least two drinkers available for each cow', 93) and 'water flow' (90) received the highest median reliability scores. The measure 'Qualitative Behaviour Assessment' (QBA) was given the worst median importance rank (22), the lowest median reliability score (60), and was among the lowest median validity scores (57). Measures of drinker space was given the lowest median validity score (50). 'Water flow' was among the lowest ranking measures in terms of importance (20) and among the lowest median validity scores (60) as well.

Table 3.3: Median (min – max) reliability and validity scores (scale 0 - 100) and rankings for each WQ measure by 16 dairy cattle welfare experts

	Reliability score	Validity score	Ranking
Body condition score (% of lean cows)	89 (61 – 95)	79 (63 – 100)	4 (1 – 17)
Centimetres of trough per cow (min. 6 cm), number of water bowls per cow (min. 0.10), at least two drinkers available for each cow	93 (31 – 100)	50 (25 – 99)	13 (2 – 20)
Water cleanliness, judged visually	80 (25 – 100)	70 (13 – 100)	19 (3 – 27)
Water flow	90 (13 – 100)	60 (25 – 100)	20 (1 – 27)
Time needed to lie down	75 (24 – 93)	78 (11 – 100)	9 (4 – 20)
Cows colliding with housing	70 (24 – 100)	82 (12 – 100)	16 (5 – 25)
Cows lying outside of lying area	85 (25 – 100)	89 (33 – 100)	16 (5 – 26)
Cleanliness of udders, flanks and lower legs	75 (24 – 100)	81 (50 – 100)	15 (4 – 27)
Tied versus loose housing	100 (54 – 100)	84 (25 – 100)	11 (2 – 27)
Lameness score	69 (39 – 95)	89 (38 – 100)	2 (1 – 21)
Integument alterations	75 (50 – 100)	90 (76 – 100)	7 (2 – 16)
Coughing	69 (11 – 100)	75 (25 – 100)	19 (8 – 26)
Nasal discharge	84 (11 – 100)	80 (36 – 100)	18 (5 – 24)
Ocular discharge	85 (38 – 100)	80 (40 – 100)	18 (5 – 24)
Hampered respiration	88 (38 – 100)	86 (15 – 100)	21 (3 – 27)
Diarrhoea	75 (37 – 100)	70 (50 – 100)	15 (8 – 27)
Vulvar discharge	77 (37 – 100)	86 (62 – 100)	18 (3 – 24)
Somatic cell count >400.000	83 (20 – 100)	81 (19 – 100)	13 (1 – 27)
Mortality	79 (20 – 100)	81 (36 – 100)	7 (4 – 17)
Dystocia	79 (28 – 100)	80 (39 – 100)	13 (7 – 27)
Downer cows	79 (12 – 100)	81 (38 – 100)	15 (4 – 25)
Dehorning method	90 (30 – 100)	86 (50 – 100)	11 (3 – 26)
Tail docking method	95 (30 – 100)	88 (70 – 100)	17 (4 – 27)
Head butts and displacements	70 (5 – 90)	75 (25 – 95)	14 (3 – 26)
Access to pasture (hours & days on pasture)	90 (56 – 100)	75 (11 – 100)	19 (3 – 22)
Avoidance distance test	66 (24 – 98)	76 (11 – 75)	17 (1 – 27)
Qualitative Behaviour Assessment	60 (3 – 94)	57 (63 – 100)	22 (1 – 27)

The importance rank of the measure was negatively associated with both the reliability and validity scores, although validity had a somewhat higher estimate (i.e. higher importance as

indicated by a lower ranking was associated with higher reliability and validity scores) (P = 0.03 for both, estimates -0.66 and -0.74, respectively, adjusted R^2 = 0.20). A very small but significant interaction was found between reliability and validity scores, where they did not strengthen each other's negative effect on ranking (P = 0.048, estimate = -0.009).

Predicting Overall Welfare Categorisation Using WQ Measures

When analysed univariately, 20 out of 42 measures significantly (P < 0.05) affected overall welfare categorization into the 'enhanced' category (Table 3.4), and 11 measures significantly affected categorization into the 'not classified' category for the entire European dataset (n = 491).

Table 3.4: P-values of the univariate logistic regression models examining predictability of single measures for a herd to be categorized as 'Enhanced' or 'Not classified' based on the collated European dataset (n = 491)

Criteria, Measures	Enhanced	Not classified
Absence of prolonged hunger		
% of lean cows	<0.001	<0.001
Absence of prolonged thirst		
Number of water bowls	0.070	0.863
Water flow	<0.001	0.505
Trough length/cow (cm)	0.001	0.008
At least 2 drinkers/cow	<0.001	0.006
Drinker cleanliness	<0.001	0.068
Resting comfort		
Mean time needed to lie down	<0.001	0.577
% collisions during lying down movements	<0.001	0.365
% of cows lying outside of lying area	<0.001	0.014
% of cows with dirty flanks	0.101	0.172
% of cows with dirty lower legs	0.023	0.110
% cows with a dirty udder	0.374	0.258
Ease of movement		
Loose or tied housing	<0.001	0.016
Absence of injuries		
% of moderately lame cows	0.002	0.392
% of severely lame cows	<0.001	0.096
% of cows with at least one lesion/swelling	<0.001	0.014
% of cows with at least one hairless patch	0.141	0.075
Absence of disease		
Number of coughs/cow/minute	0.168	0.350

% cows with nasal discharge	0.092	0.165
% cows with ocular discharge	0.044	0.426
% cows with hampered respiration	0.293	0.385
% cows with diarrhoea	0.386	0.546
% cows with vulvar discharge	0.588	0.936
% cows with SCC >400.000	0.130	0.014
% cows mortality	<0.001	0.189
% calvings with dystocia	0.619	0.841
% downer cows	0.742	0.423
Absence of pain induced by management procedures		
Method dehorning	0.130	0.021
Use of analgesics during/after dehorning	0.618	0.540
Use of anaesthesia during dehorning	0.759	0.110
Method tail docking	0.150	0.974
Use of analgesics during/after tail docking	0.011	0.008
Use of anaesthesia during tail docking	0.025	0.010
Expression of social behaviour		
Head butts/cow/15 min.	0.033	0.759
Displacements/cow/15 min.	0.615	0.159
Expression of other normal behaviour		
Number of hours on pasture	0.467	0.153
Number of days on pasture	0.810	0.454
Human-animal relationship		
% of cows that could be touched	0.711	0.188
% of cows that can be approached < 50 cm but not touched	0.012	0.379
% of cows that can be approached by 50 – 1 m	0.253	0.924
% of cows that can't be approached (> 1 m)	0.011	0.547
Positive emotional state		
QBA index score	0.079	<0.001

The multivariable model that had the fewest variables while maintaining the highest % of both sensitivity and specificity (67% and 85%, respectively) for the 'enhanced' category contained the following measures (from most to least influence): contained the following measures (from most to least influence): 'at least two drinkers/cow', 'water flow', '% of animals lying outside the lying area', 'mean time needed to lie down' 'drinker cleanliness', and '% of animals with at least one lesion/swelling' (Table 3.5). For not classified, the measures (from most to least influence) 'at least two drinkers/cow', 'number of lean cows', 'QBA index' and 'number of displacements/cow/h.' contributed to the model with fewest variables but the highest sensitivity (44%) and specificity (100%).

Table 3.5: P-values and model estimates of measures in the multivariate logistic regression models predicting a herd to be categorized as 'Enhanced' or 'Not classified' based on the collated European dataset (n = 491)

Outcome variables	Model for 'enhanced' categorisation		Model for 'not classified' categorisation	
	Estimate	P-value	Estimate	P-value
Number of lean cows	-	-	1.8	<0.001
Water flow	1.1	<0.001	-	-
At least 2 drinkers/cow	2.4	<0.001	-3.7	0.007
Drinker cleanliness	0.6	<0.001	-	-
Mean time needed to lie down	-0.7	<0.001	-	-
% of cows lying outside of lying area	-0.9	<0.001	-	-
% of cows with at least one lesion/swelling	-0.5	<0.001	-	-
Number of displacements/cow/h.	-	-	0.7	0.043
QBA index score	-	-	-1.6	0.002

Comparing WQ Overall Welfare Category and Criteria Scores with Expert Opinion

For 2 of 5 'acceptable' herds and for 1 of 2 'enhanced' herds, the majority of experts (n = 8) scored in accordance with WQ (Figure 3.3). Regarding scores that were not in accordance with WQ, the vast majority were a lower category than the WQ calculation (20 of 25 expert scores). Overall, ICC for overall welfare scores by experts was 0.5.
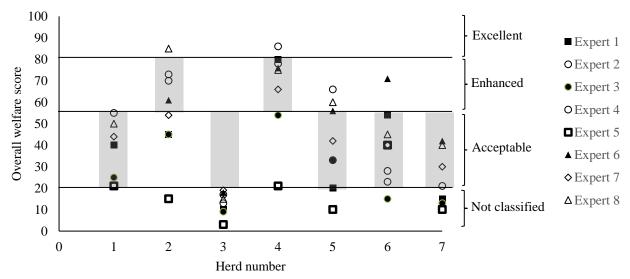
Figure 3.3: Overall welfare score for all seven focus herds by eight dairy cattle welfare experts, grey boxes indicate WQ overall welfare category

The criteria 'absence of injuries', 'absence of pain induced by management procedures', 'expression of social behaviour' and 'good human-animal relationship' were systematically scored lower by experts than the WQ score (Table 3.6). The expert and WQ scores were not significantly related for two criteria: 'absence of prolonged thirst' and 'absence of prolonged hunger' (Table 3.6). The correspondence between experts was insufficient (ICC < 0.6) for two criteria, namely 'absence of injuries' and 'absence of disease'. The number of measures within a criterion tended to be negatively related to ICC (P = 0.06, estimate = -0.04).

Table 3.6: Systematic t-test P-value, Linear Regression R^2 and ICC of WQ integrated scores and expert median scores (n =14) for the focus herds (n = 7) for each WQ criterion

Criteria	Median WQ score	Median expert score	Systematic t-test P-value	Regression R^2	ICC
Absence of prolonged hunger	67 (39)	50 (75)	0.475	0.237	0.6
Absence of prolonged thirst	20 (97)	50 (71)	0.737	0.007	0.7
Comfort around resting	27 (20)	25 (33)	0.181	0.880[**]	0.8
Freedom of movement	100 (33)	90 (90)	0.125	1.000[***]	1.0
Absence of injuries	28 (19)	18 (29)	0.006	0.926[***]	0.5
Absence of disease	40 (32)	42 (34)	0.296	0.903[**]	0.4
Absence of pain induced by management procedures	58 (18)	10 (50)	0.023	0.521[*]	0.8
Expression of social behavior	84 (24)	58 (50)	0.020	0.869[**]	0.6
Expression of other normal behavior	73 (78)	60 (78)	0.828	0.978[***]	0.9
Good human-animal relationship	54 (37)	52 (50)	0.023	0.984[***]	0.7
Positive emotional state	54 (32)	50 (37)	0.901	0.997[***]	0.8

[*]$P>0.05$, [**]$P>0.01$, [***]$P>0.001$

Discussion

This study gives insight into the relation of integrated scores of the WQ dairy cattle protocol with selected experts' opinion. The specific research design imposes some limitations, but also provides challenges for future research. For example, we chose to only select dairy cattle welfare experts who were trained users of the WQ dairy cattle protocol. This ensured that experts had a proper knowledge of the protocol and all measures, but limited the number of possible respondents which could be labelled as 'experts'. Despite these limitations, results show discrepancies between selected experts' opinion and WQ integration scores.

Expert opinion on Ranking, Reliability and Validity of Measures

The highest ranked measures in terms of perceived importance for the overall welfare status of a herd during the current study 'lameness score', 'body condition score', 'mortality rate' and 'integument alterations' are in agreement with earlier studies in which dairy cattle welfare experts were asked to score the importance of welfare measures (Nielsen et al., 2014, Lievaart and Noordhuizen, 2011, Whay et al., 2003). Reliability and validity scores both influenced ranking positively (based on the negative relationship between reliability and validity scores and ranking), but did not positively interact. This means that highest ranked measures in the current study did not necessarily receive the highest validity *and* reliability scores. In addition, although the set-up of this study was such that experts had to consider validity and reliability before ranking, other (unknown) factors appeared to influence the experts' opinion on the importance of the various measures for overall herd welfare as well (further supported by the models' low R^2 of 0.20). This was the case for lameness, for example, which was ranked highest for importance although its reliability was among the lowest.

Overall, QBA was the measure that was scored among the lowest by the experts with regard to reliability and validity (although still within the 'acceptable' range) and was ranked lowest on importance for dairy cattle welfare status. The QBA is a method that uses descriptors such as 'frustrated' or 'content', to interpret the behavior and body language of an animal (Rousing and Wemelsfelder, 2006). Inter observer reliability was tested and deemed acceptable for a QBA method using 'free' descriptors (i.e. not set but determined by observers themselves) and was validated by correlating results to behavioral observations (Rousing and Wemelsfelder, 2006; Napolitano et al., 2012). However, the method and specific set of descriptors used in the WQ protocol were tested for inter observer reliability in a study by Bokkers et al., (2012) and judged as not satisfactory by the authors involved (i.e. Kendall's coefficient of concordance < 0.7).

While some measures scored high for reliability, they scored low for validity, e.g. measures related to the criterion 'absence of prolonged thirst' ('centimeters of trough per cow'), or were ranked lowest on importance for dairy cattle welfare ('water flow'). Criticism expressed in earlier studies for these measures is related to their resource-based nature and the impact these specific measures have on the WQ integrated scores , while preference generally shall be given to animal-based measures (Buijs et al., 2016, Heath et al., 2014, de Vries et al., 2013).

Measuring functioning of water points, water provision and water cleanliness refers to assessing a risk for cows being in a certain welfare state and may therefore not be the most valid measure of an actual welfare state in dairy cattle, in this case due to prolonged thirst. Additionally, to our knowledge, no actual validity testing of the WQ drinker measures has occurred. This could explain the relatively low perceived validity score attached by the experts to these measures. Further testing of reliability and validity on certain measures is needed, based on the results of the current study and previous research (Knierim and Winckler, 2009). If from such studies it appears that measures are not sufficiently reliable or valid, then research should be performed to propose better measures.

There was variation between experts on the relative importance for the overall welfare status of dairy herds of different welfare measures (given the high variations in ranking and reliability and validity scores between experts). This possibly reflects diverging views in what experts find most important for dairy cattle welfare, as Fraser et al., (1997) showed in his study on animal welfare conceptualization among animal welfare scientists. This indicates that when using expert opinion to determine weights for various measures, such variation should be accounted for when selecting the expert panel. Therefore, it is not likely that an overall welfare score will always perfectly reflect an individual experts' opinion. Methods to achieve more consensus among experts exist. Examples are deliberative processes using a workshop like performed by Rodenburg et al. (2008), or more complex processes like a 'Delphi' method with multiple rounds of expert elicitation and feedback (Linstone and Turoff, 1975).

Comparison of the measures' impact on overall welfare categorization and expert opinion

Compared to previous studies (Buijs et al., 2016, Heath et al., 2014), more measures affected both the 'enhanced' and the 'not classified' categorization in the current study. This is likely due to a larger variation in data in the current study which used a much larger (and diverse, as data was collected in more than one country) database compared to both other studies. To specify, the current sample comprised of 491 herds, as opposed to 92 herds and 22 flocks for Heath et al., (2014) and Buijs et al., (2016) respectively. In accordance with Heath et al. (2014) drinker measures had the biggest influence for both the enhanced and not classified models, while these received some of the lowest ranks and/or validity scores by the selected experts. Additionally, the QBA score which scored lowest overall was among the best predictors for the

'not classified' categorization. By contrast, although there is often little agreement among experts on importance of various welfare measures, some measures which are often regarded as being highly important to cattle welfare did not have a great influence on the overall welfare status categorization. For example, although 'lameness score' and 'mortality rate' contributed to the 'enhanced' categorization in univariate models, they did not when combined into a multivariable model. These results show that the relative influence of measures on WQ integrated scores may not be in accordance with the selected experts' opinion. We tested this by comparing expert scoring of WQ criteria and overall welfare with calculated WQ scores.

Comparing WQ Integrated Scores with Expert Opinion

Overall welfare category. For only three out of the seven herds, the majority of experts scored in accordance with the WQ overall welfare categorization. The two herds that were scored as 'not classified' by at least half of the experts (Herds 3 and 7) both scored badly (i.e. relatively high prevalence on measures that were ranked as highly important by the experts, namely lesions/swellings and moderately lame cows.

Variation between experts was shown for the overall welfare scoring, given the relatively low ICCs. This was also shown for criteria scores, where ICCs tended to be lower for criteria which contain the most measures. This can indicate that 1) experts did not agree on their assessment of overall welfare caused by a different view of animal welfare (as mentioned above) and/or 2) some experts may have had difficulties in aggregating many welfare measures into one overall score. The latter explanation is supported by that fact that six of the 14 experts which completed the questions on criterion scores, did not complete the question on overall welfare scores.

Criteria scores. The criteria 'absence of injuries', 'absence of pain induced by management procedures', 'expression of social behaviour' and 'good human-animal relationship' were systematically scored lower by experts than the WQ integrated scores. In the WQ protocol, poor scores have more influence on integrated scores than good scores (Buijs et al., 2016). Therefore, lower scores on each of these criteria would have a major effect on principle scores and overall welfare category.

The correspondence between the expert and WQ score for the criterion 'absence of prolonged thirst' was extremely low. The finding that the experts considered some of these

measures of relatively poor validity may partly explain this lack of correspondence. It is a strong indication that experts of the present study did not agree with the way that the criterion 'absence of prolonged thirst' is evaluated in the WQ protocol.

Two complementary explanations can be put forward for the poor correspondence between experts scores and WQ integrated scores. First, except for the first step of the integration procedure, WQ consulted a much wider group of stakeholders (including animal scientists, social scientists, producers and retailers.) than we did in the current study. These stakeholders' views on the relative impact of the various measures on dairy cattle welfare may differ substantially from those of the experts in the current study. We opted to limit the current study to trained users only, because it could be argued that they are best qualified to assess overall dairy cattle welfare state and the relative importance of the various WQ measures. Third, WQ integration methods likely contribute to differences between expert opinion and WQ integrated scores. In Chapter 2, two factors were identified which influence the impact a measure has on the integrated WQ scores, but which seem unintended by the WQ consortium. Namely, 1) the number of integrated measures per criterion or principle, and 2) the various different aggregation methods of measures into criteria scores which influence the impact individual measures have on integrated scores. In this study a low level of correspondence between welfare measures which impact WQ categorization most and which were scored as most important by selected experts was found. Also, poor correspondence between expert opinion and some criterion scores indicated that this lack of correspondence already starts in the first step of integration. As regard to the translation of criteria scores into principle scores, and from principle scores into overall welfare scores, the current study does not provide insight. However, for other stakeholders (i.e. farmers and citizens), previous studies have indicated that the assumption of WQ that each principle should be attributed an equal weight might not be in correspondence with their opinion. In those studies, weight attributed to these four welfare principles differed between principle (Vanhonacker et al., 2016, Stadig et al., 2015, Tuyttens et al., 2010). However, further research is needed to determine whether this is the case for dairy cattle welfare experts as well.

Current results indicated a lack of correspondence between welfare scores and selected dairy cattle welfare experts' assessment of herd welfare. The opinion of these experts is the only 'silver standard' we have to validate animal welfare integrated scores, since they are

arguably best equipped to assess and quantify the welfare of a given herd. Moreover, these experts may be considered authorities for animal welfare assessment in society, and we feel it is important that scientists who use this method support it. Future research could focus on determining whether the way experts assess welfare is in correspondence with other stakeholders' assessment. An alternative approach might be to leave benchmarking of thresholds for farms to reach not-classified, acceptable, enhanced or excellent welfare categorization up to end-users of the protocol (e.g. producer organizations, animal welfare organizations or legislator bodies).

Conclusion

Expert opinion on the most and least important measures for the overall welfare status of a herd did not correspond well with the influence of these measures on the WQ overall welfare categorization. Some of the measures that were ranked as least important for herd welfare by selected experts (e.g. measures relating to drinkers) had the highest influence on the WQ overall welfare categorization. On the contrary, measures ranked as most important by the experts (e.g. lameness and mortality) had a lower impact on the WQ overall category. In addition, results indicate poor correspondence between selected experts' scoring and 6 of 11 WQ-criteria and the overall welfare category. In both cases, experts mostly allocated more negative scores. The integration method of the WQ protocol for assessing dairy cattle welfare on-farm should therefore be revised if it is to correspond better with the opinion of trained users of the protocol.

Chapter 4

Alternative integration protocol of dairy cattle welfare measures

Abstract

Aggregating data from various welfare measures into a 'multidimensional welfare index' (WI) which balances the scores across different welfare measures on a group of animals remains a challenge. Using on-farm dairy cattle welfare monitoring as a case, we aim to illustrate the five steps in the development of such an integrated index which uses only animal-based measures, is multidimensional, simple and transparent, discriminative and corresponds with expert opinion. This integrated index is based upon the average of a severity score (i.e. how severely a given welfare problem affects the welfare of an individual cow) multiplied with the herd prevalence for each welfare measure. Expert opinion was consulted using web surveys, selecting only trained users of the Welfare Quality® protocol for dairy cows. In Step 1, based upon the importance ranks of 14 experts complemented by literature review, six animal-based welfare measures were selected from the WQ protocol for dairy cattle: 'lameness score' (severe lameness), 'body condition score' (leanness), 'mortality rate', 'hairless patches', 'lesions/swellings' and 'Somatic Cell Count >400.000'. In Step 2, median expert severity scores for the welfare measures were determined. In Step 3, a dataset was collated of the selected welfare measures in 491 herds across Europe. Experts (n = 14) allocated a welfare score (from 0-100) to 12 fictitious herds whose properties represented the range in actual data for which the prevalence of each welfare measure was benchmarked against all 491 herds. Quadratic models indicated that the correspondence between these subjective scores and the proposed integrated index was high ($R^2 = 0.91$) and could not be improved by including an additional factor which reduces compensation of bad scores by good scores on the other measures. Step 4 comprised interpreting WI scores by assigning four welfare categories from poor ('not classified') to 'excellent' WI based upon expert scores for the 12 fictitious herds. In Step 5, the selected measures were found to be associated with all welfare problems which impair dairy cattle welfare most based on literature. We conclude that the proposed approach has resulted in a WI with a limited number of animal-based measures integrated in a transparent and simple way to produce a multidimensional and discriminative index based upon, and checked for, correspondence by a selective panel of experts. The index is not exhaustive and should be accompanied with a disclaimer for welfare aspects that cannot be detected. The proposed method is flexible though, enabling measures for additional welfare dimensions to be added by repeating the five-step process.

Introduction

One of the most important and complex challenges in animal welfare science, is aggregating data from various welfare measures into an integrated index reflecting the overall welfare status of a group of farm animals (Czycholl et al., 2015). As Fraser (2008) described, the conception of animal welfare is both science-based (in developing validated, objective welfare measure) and values-based (i.e. what people considers important or desirable for animals to have a good life). As there is no 'gold standard' for overall herd welfare, aggregating data on various welfare measures into an overall index requires some degree of subjectivity. Such a multidimensional welfare index (WI) could promote transparency in the food chain when used in the communication with consumers (labelling), and could function as an incentive for on-farm welfare improvements (Blokhuis et al., 2003). However, prerequisites for both aforementioned purposes of a WI are that it can discriminate between farms with different levels of welfare and acceptance of this index both by experts and by society at large as 'credible'.

Several methods for the assessment of herd welfare have been published in the last few decades (reviewed by Czycholl et al., 2015, Johnsen et al., 2001). Examples are the Animal Needs Index (Tiergerechtheitsindex, Bartussek et al., 2000), and the Welfare Quality® (WQ) protocols for welfare assessment of the main types of farm animals. However, concerns have been raised about current methods regarding the quality of the welfare measures included in the protocol and/or for the way data from these measures are aggregated into an overall WI. Criticisms on welfare measures often relate to their poor reliability, validity or feasibility (Tuyttens et al., 2015, de Jong et al., 2015, de Vries et al., 2013b, Knierim and Winckler, 2009). Many protocols also use resource-based or management-based measures which have been criticised for describing the potential or risk for good or bad welfare rather than directly measuring the welfare status itself. There is growing consensus that animal-based measures are to be preferred for directly assessing the outcome of the complex effects of the environment and management on the animal's actual state of welfare (Czycholl et al., 2016, European Food Safety Authority 2012, Blokhuis et al., 2003, Johnsen et al., 2001). The WQ protocols for welfare assessment of the main types of farm animals are probably the most renowned and comprehensive methods. They are based upon a hierarchical approach to integrate data on a multitude of predominantly animal-based welfare measures into an overall welfare categorisation. However, these protocols

have been subject to criticism not only concerning the choice of measures, but increasingly concerning the hierarchical integration method and the way it balances the relative importance of the various welfare measures (Buijs et al., 2016, Heath et al., 2014, de Vries et al., 2013a, Tuyttens et al., 2010). This appears to be an unwanted side-effect of the very complex integration method which was needed to aggregate measures of different scales using different thresholds. Having a large number of measures was seen as an advantage when WQ was created, as various aspects of welfare are detected when assessed using this system. However, the complex mathematical techniques have made the welfare assessment tool difficult to describe to end-users. In addition, using the current integrated WQ categories likely leads to a non-intended focus on a limited set of (often resource-based) measures that is hard to justify (Heath et al., 2014, de Vries et al., 2013a).

Revising WQ integration methods to be more simple, transparent and feasible without changing the protocols' structure seems difficult. In the current study, we therefore aim to illustrate a set of steps in the development of an alternative integration method which meets criticism on previous methods by using only selected animal-based measures and being multidimensional, simple and transparent, discriminative and coherent with expert opinion. This method may be used to integrate data from selected welfare measures in order to quantify the level of herd welfare, albeit without claiming to be exhaustive. Using dairy cattle welfare as a case, this integrated index is based upon a method by Burow et al. (2012). The WI is based on the average of a severity score (expert judgement of how severely a given welfare problem affects the welfare of an individual cow) multiplied by the herd prevalence for each welfare measure. However, the above integration method allows for compensation of poor scores with better scores. In some studies (e.g. Spoolder et al., 2013, Botreau et al., 2009), it is argued that such compensation should not be permitted as good results on one aspect cannot compensate for poor scores on other aspects (e.g. having a good body condition score cannot compensate for being severely lame). Other studies however, find indications that compensation between welfare aspects may be possible (reviewed by Leknes and Tracey, 2010). At present, there is little evidence that compensation-reduction is warranted, let alone what type of compensation-reduction method best corresponds with expert opinion. The latter is examined in one of the proposed steps in this study.

To strive for wide acceptance of a WI by society, the method of aggregation should ideally concord with society's concepts of animal welfare and fully address existing concerns, besides being validated using scientific experts' opinion (Tuyttens et al., 2010, Fraser et al., 2006). This could be achieved by eliciting input from society when developing the integration methods, as was done in the WQ project where social scientists and stakeholders (e.g. consumer and animal welfare organisations, producers and retailers) were involved in the last step of integration (i.e. determining thresholds to integrate welfare principles into one overall welfare category) to ensure wide acceptability of the protocol. However, it can be questioned whether people without in-depth knowledge and expertise in dairy cattle welfare and the measures involved can adequately balance the importance of different welfare measures. Indeed, the relative importance allocated to a given welfare measure could also depend on how exactly it is measured on-farm (e. g. selection of and size of the sample, to what extent confounding factors may influence the measures, objectivity of the measure). Moreover, it has been shown that detailed information on how data on welfare measures is collected on-farm can significantly influence the relative weights they are given by experts (Rodenburg et al., 2008). In response, the current study elicited exclusively dairy cattle welfare experts, i.e. people who have adequate knowledge and expertise about dairy cattle welfare and about the measures involved.

Material and Methods

Based upon Burow et al. (2012), the WI was constructed from perceived severity of welfare problems ('severity score') and observed prevalence of these welfare problems. The following formula forms the basis to integrate data on selected welfare measures into one score:

$$Welfare\ index\ score = \frac{1}{nm} \times \sum_{m=1}^{nm} Sm \times rPm$$

Here, n represents 'number', m refers to 'measure', S represents the 'severity score' which ranged from 0 - 1, and rP refers to 'relative prevalence' which is calculated as prevalence per herd/prevalence at 97.5th percentile of that measure among all herds in the EU database. In the proposed formula, rP rather than absolute prevalence was used so each herd covered the same possible spectrum for each measure. When including the 2.5 percent highest prevalence, extreme values may be included. When using relative prevalence, this would then mean that lower values (at the 90th percentile for instance) are given much less weight than when

excluding these extreme values. Therefore, prevalence of the 97.5th percentile was set as the maximum for each measure score, to prevent an extreme prevalence value of single measures from having a disproportionately large influence on the score. Therefore, herds with values equal to or higher than the 97.5th percentile were automatically given the maximum measure score. This allowed for a uniform method to determine thresholds for the different compensation-reduction methods (CRMs) which were tested. To achieve a score on a scale of 0 (very poor welfare) – 100 (excellent welfare) and to test various CRMs, the formula was complemented as follows:

$$Welfare\ index\ score = \mathbf{100} - \frac{1}{nm} \times \sum_{m=1}^{nm} Sm \times rPm \times \frac{\boldsymbol{C}}{\boldsymbol{Cmax}}$$

Here, C is a 'compensation-reduction factor' and Cmax is the maximum compensation-reduction factor. To gain input for this formula, we performed two independent online surveys among dairy cattle welfare experts. Because WQ measures were used in this study, experts were selected who had experience with the WQ protocol for dairy cattle (i.e. they were trained to perform the WQ protocol for dairy cattle and had performed on-farm WQ assessments of dairy herds), and had authored at least one peer-reviewed scientific paper about dairy cattle welfare. Experts were all European and eight different nationalities were represented. These are further referred to as 'selected experts'. The proposed method to develop an alternative WI consists of five steps (Figure 1). Step 1 entails selecting animal-based welfare measures to be included in the protocol. In Step 2, weights for the various welfare problems are determined by asking selected experts to score how severely each of the selected welfare problems impairs the welfare of an animal. In Step 3, WI is calculated and correspondence with expert opinion is analysed. Similarity between selected experts' welfare scores for several fictitious herds and integrated WI using the aforementioned formula with various CRMs is analysed. Step 4 consist of interpreting the WI (what score indicates poor/good welfare). Step 5 comprises of checking to what degree the selected welfare measures are associated with factors which have the most severe impact on dairy cattle welfare. The five steps are elaborated below.
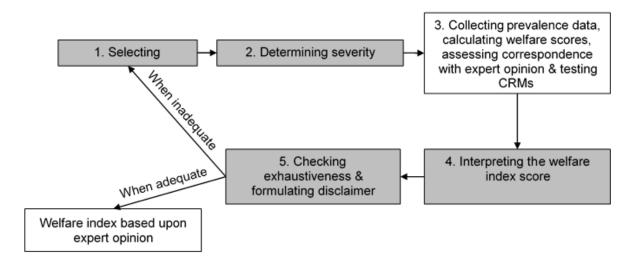
Figure 4.1: The different steps of the proposed method to integrate dairy cattle welfare measures. Grey boxes indicate steps where expert elicitation is used.

Step 1: Selecting Welfare Measures

A subset of welfare measures was selected from the WQ protocol (Welfare Quality® Consortium, 2009). We used three criteria for selecting measures: 1) they ought to be animal-based, 2) it must be possible to express them as a % and 3) they must be considered as important for dairy cattle welfare by selected experts. The importance of the measures was based upon the first online survey where 17 selected experts ranked all WQ measures (n = 27) on importance for the overall welfare status of a herd of dairy cattle. It was mentioned that for ranking, (inter alia) reliability, validity, perceived relevance and prevalence may be considered. Subsequently, we compared compliance of these selected measures with the outcomes of published studies in which selected experts ranked welfare measures on importance (Nielsen et al., 2014, Burow et al., 2012, Lievaart and Noordhuizen, 2011, Whay et al., 2003). When significant gaps were found between selected measures and literature, these measures were added.

Step 2: Determining Severity Scores

To determine the severity scores for the selected measures, input from 16 of the same selected experts which replied to the first survey was requested in a second survey, which 14 of the experts completed. In this second survey, selected experts were asked to score how severely the welfare of an individual cow would be affected by each of the six selected welfare impairments on a scale of 0 (totally not severe) – 100 (extremely severe). Selected experts were informed

that they may take the degree and duration of suffering into account. In the ensuing Step 3, median severity scores were used in calculating WI.

Step 3: Calculating WI and testing coherence with expert opinion

For checking correspondence between expert scores and aggregated WIs, in the second survey the 14 selected experts were presented with a graph showing the prevalence distribution of all selected welfare measures for 491 European herds (Figure 2). To reflect the current range present in Europe across various herding systems, datasets were collated from seven European research institutes and included data from 10 countries (more details in Chapter 2).

In the graph presented to the selected experts, six 'focus herds' were highlighted per expert (example: Figure 4.2; data shown in Table 4.1). Herds were fictitious, but were based upon real herd data from the European dataset. In total, twelve herds were created to fit the following descriptions: 1) two herds that scored high in prevalence (indicating poor welfare) on all measures; 2) two herds that scored low (indicating good welfare) on all measures; 3) two herds that scored medium on one half of the measures and high on the other half; 4) two herds that scored the other half of the measures medium and the other half high; 5) two herds that scored medium on all measures except for one (SCC > 400.000) and 6) two herds that scored medium on all measures but high for one (severe lameness). Highest prevalence belonged to the top 5% for all welfare measures, medium between 40% and 60% and lowest scores were from the lowest 5%. Each respondent was presented with six herds, one of the two for each category (Table 4.1). Selected experts were asked to allocate a welfare score to each herd they were presented with using a tagged visual analogue scale from 0 - 100. Tags were 'Not Classified (>20)', 'Acceptable (20 – 55)', 'Enhanced (55 – 80)' and 'Excellent (>80)', following WQ categorization (Welfare Quality, 2009).

Each of the twelve fictitious herds was thus scored by 6 to 8 selected experts. Subsequently, the degree of correspondence between expert scores and WI's were calculated with varying CRMs. One of the tested CRMs was 'veto', where thresholds are defined for each measure above which a value cannot be compensated for. This is achieved by automatically attributing the worst possible welfare score to a herd, independent of the prevalence of other welfare problems. The other tested CRMs use various formulas to allocate increasingly more weight to worse scores on a certain measure. Tested formula in the current study were 'Discrete', 'Linear', 'Broken

line' and 'Exponential' and are illustrated in Figure 3. In addition, scores were calculated without CRM ('no CRM').
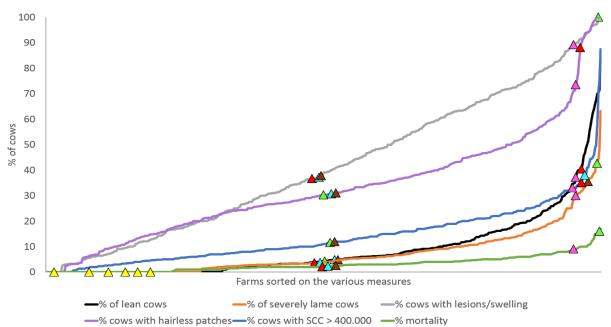

Figure 4.2: One of the graphs presented to experts in the second survey, showing the distribution of all herds in the database (n = 491) for the six selected measures. Coloured triangles mark six (of the 12) focus herds.

Table 4.1: Prevalence (% of cows) for the six selected dairy cattle welfare measures, for each of the 12 fictitious herds the experts (n=14) allocated an integrated index score

Herd #	Measure values	Very lean	Severely lame	Lesions and swellings	Hairless patches	SCC > 400.000	Mortality
1	All low[1]	0	0	3	3	0	0
2	All low[1]	0	0	0	0	0	0
3	All high[1]	46	33	92	92	38	10
4	All high[1]	37	30	90	74	33	9
5	Medium / high[1,2]	50	4	37	94	41	2
6	Medium / high[1,2]	41	4	37	88	35	2
7	High / medium[1,3]	4	32	92	30	11	10
8	High / medium[1,3]	4	44	100	30	11	16
9	Medium, high SCC[1]	4	4	37	30	35	2
10	Medium, high SCC[1]	4	4	37	30	38	2
11	Medium, high lameness[1]	5	34	39	30	11	2
12	Medium, high lameness[1]	4	33	37	30	11	2

[1]Highest scores belonged to the top 5% of herds in the European dataset (n = 491), medium between 40% - 60% and lowest to the lowest 5% of herds

[2]% of too lean cows', 'SCC >400.000' and 'nHP' were high

[3]% of cows with lesions', '% of cows with severe lameness' and '% of mortality' were high
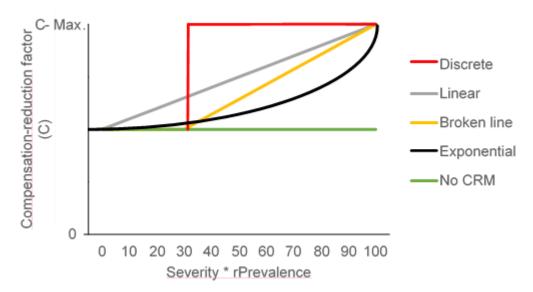
Figure 4.3: Illustration of the CRMs (except Veto) tested in this study. Discrete gives no compensation-reduction for measures up to a certain threshold of S*rP, above which the S*rP score is multiplied with the maximum C. For Linear CRM, C increases linearly with an increasing S*rP score of the welfare measures. The Broken line CRM increases C in a linear manner from a certain threshold onwards. Exponential CRM increases C exponentially with an increasing S*rP score of the welfare measures.

For discrete, broken line and Veto CRM, a threshold at which compensation-reduction starts needed to be determined. For all CRMs apart from veto, it also had to be determined what the maximum level of compensation-reduction (Cmax) was. We checked which threshold value of S*rP (ranging between 5 and 70 in increments of 5) and which value for Cmax (set at between 1.5, 2, 3, 5 and 10) corresponded best with expert opinion based on model R^2.

Statistical analyses were performed using the programme R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Both Linear and Quadratic models were used to test correspondence between expert scoring and the integrated scores to determine if adding a CRM to the WI formula generated better results (i.e. a higher R2) for varying thresholds and values of C.

Step 4: Interpreting the WI

To interpret the WI scores, we used expert scores for the 12 fictitious herds. Selected experts scored on a tagged visual analogue scale with labels for four welfare categories following WQ categorization ('not classified' from 0 - 20, 'acceptable' from 20 - 55, 'enhanced' from 55 – 80

and 'excellent' from 80 - 100). To determine thresholds of these welfare categories for the alternative WI, we (scatter) plotted the expert scores against the WI scores for the 12 fictitious herds and added the best fitting curve. We then identified the three points where the best fitting curve intersects with the thresholds of the scale on which the selected experts scored (expert scores 20, 55 and 80).
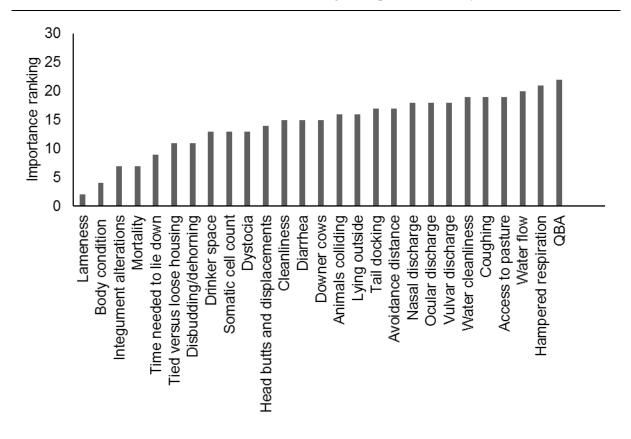
Step 5: Exhaustiveness Check

In Step 5, we assessed to what degree the selected measures are indicative of the 'worst adverse effects' (factors which have the most severe impact) on dairy cattle welfare. For this end, we compared the selection of welfare measures with a list of worst adverse effects on dairy cattle welfare and associated animal-based welfare measures in a European Food Safety Authority (EFSA) report by Nielsen et al. (2014). In this report, worst adverse effects were selected based upon several other EFSA reports (EFSA, 2009 a–e; EFSA 2012), Presi and Reist (2011), Brenninkmeyer and Winckler (2012) and expert opinion (Table 4.2).

Results

Step 1: Selecting Welfare Measures

Lowest median expert importance ranking for herd welfare, indicating highest perceived importance, was allocated to 'lameness', 'body condition score', 'integument alterations', and 'mortality rate', which were therefore selected to be included in the protocol (Figure 4). The other measures among the top ten ranked welfare measures were considered for inclusion as well: 'time needed to lie down', 'tied versus loose housing', 'disbudding/dehorning', 'drinker space', 'Somatic Cell Count (SCC)' and 'dystocia'. Only one of these measures (SCC > 400.000 as an indicator of mastitis) met all selection criteria. Lameness is measured using a gait score with categories 'not lame', 'moderately lame' and 'severely lame'. As moderately lame reflects lower welfare impairment, only severe lameness was used in the ensuing steps. Integument alterations consist of both hairless patches and lesions/swellings, as both may have different causes, we chose to separate the two in the ensuing steps of this study.

Figure 4.4: Median importance ranking for all WQ measures as judged by 14 dairy cattle welfare experts.

Step 2: Severity Scores

Median expert severity scores were highest for severe lameness (92, interquartile range = 90 – 97) and mortality (90, 69 - 100) followed by leanness (61, 50 - 71) and SCC > 400.000 (73, 43 - 80), and lowest for hairless patches (18, 12 - 34) and wounds/swellings (40, 39 - 58).

Step 3: calculating the WI and analysing coherence with expert scores

Welfare scores as indicated by the selected experts followed the patterns anticipated for the fictitious herds (Figure 5). Herds 1 & 2, with a low prevalence for all measures, received a good score while herds 3 & 4, with high prevalence for all measures (indicating poor welfare), received a bad score. Additionally, a high prevalence of the measure 'severe lameness' while all other prevalence were medium (herds 11 & 12), lead to a lower score than when only '% of cows with SCC > 400.000' was high (herds 9 & 10), in line with the higher severity scores for lameness than SCC.
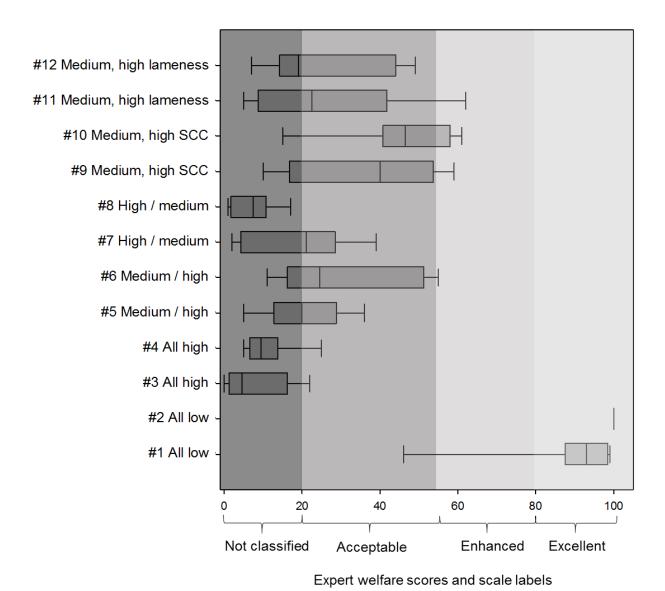
Figure 4.5: Medians and interquartile range (box) of the welfare scores allocated by experts (n = 14) to the 12 focus farms (confer Table 4.2) using a 0-100 tagged Visual Analogue Scale. Whiskers: data within 1.5× the interquartile range. Higher scores imply better welfare. Highest scores belonged to the top 5% of herds in the European dataset (n = 491), medium between 40% and 60% and lowest scores were from the lowest 5% of herds.

Quadratic models consistently achieved a higher R2 than linear models. The quadratic model with no CRM provided a good fit with the selected experts' scores ($R^2 = 0.91$, F = 401.4, Figure 6). None of the CRMs improved this fit (within the ranges of maximum C and cut-off values that were tested) and there were no indications that further increases or decreases of Cmax would improve the models. Models with Discrete, Linear, Broken Line and Exponential CRM

all gave a maximum R² of 0.90 (with F = 398, 369, 372, 368 and 355 respectively). Models with Veto CRM achieved a maximum R² of 0.74 (with F = 117). As there is no evidence that a method of compensation-reduction improves the fit with the expert scores, we can simplify the WI by removing C/Cmax from the formula.

Step 4: Interpreting the WI

Based upon expert scores in the different welfare categories, thresholds for the category 'Not classified' ranged from 0 to 46, for 'Acceptable' from 46 to 77, for 'Enhanced' from 77 to 93 and for 'Excellent' from 93 to 100 (Figure 6).
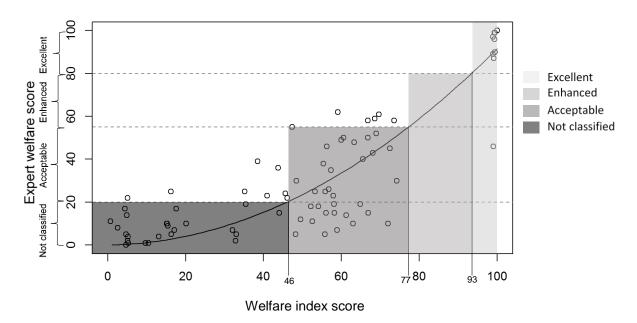

Figure 4.6: Expert (n = 14) welfare scores of the fictitious herds (n = 12) plotted against the calculated WI scores of these herds using no CRM, with best fitting quadratic curve ($R^2 = 0.91$). Higher scores indicate better welfare, category thresholds determined using the expert scores are indicated underneath the x-axis.

Step 5: Exhaustiveness Check

The welfare measures that were selected were all mentioned in Nielsen et al., (2014) as being associated with what they define as being the 'worst adverse effects' based on expert opinion and literature (Table 4.2).

Table 4.2: Summary of which of the 'worst adverse effects' for dairy cattle welfare are associated with the selection of welfare measures in the current study based upon Nielsen et al. (2014)

Adverse effects	Associated welfare measures
Foot disorders	Lameness, mortality, lesions/swellings
Leg injuries	Lameness, lesions/swellings
Mortality (unassisted)	Mortality
Mortality (euthanasia)	Mortality
Exhaustion (prolonged metabolic demand)	Leanness, mortality, lesions/swellings
Behavioural disruption – feeding (including social stress, pain, hunger, exhaustion, fear, frustration)	Leanness, lameness
Behavioural disruption - rest (including too little rest, pain, fear)	Lesions/swellings, lameness
Behavioural disruption - flooring/space (including fear, pain)	Lesions/swellings, lameness

Discussion

The proposed aggregation method for calculating an integrated (WI) for dairy cattle distinguishes itself from previous methods in four different ways. First, the use of exclusively animal-based measures implies direct assessment of dairy cattle welfare in contrast to the use of resource- or management-based measures. Second, the simple and transparent integration formula for calculating WI reduces the likelihood of unwanted side-effects that occur when using more complex aggregation procedures. The WQ protocol is an example of a complex integration method which was innovative in its use of methods, where welfare measures are first integrated into twelve criteria scores and subsequently into four principle scores which are then used to determine the overall welfare category (Welfare Quality®, 2009). The welfare principles were separated to reflect different dimensions of welfare, and the complex integration methods were necessary to cope with the large number of measures included. However, an unintended consequence of the large number of measures and the method used to integrate them is that sensitivity of the overall welfare category to changes in individual welfare measures

partly depends on the number of measures integrated into the criterion and principle scores (Buijs et al., 2016, Chapter 2 of this thesis). Third, the WI was tested and found to show high correspondence with expert opinion. Finally, WI being expressed on a continuous scale ensures a high degree of discrimination, which enables detection of relatively small differences between (or within) herds, at least if cows are affected by these conditions ('lameness', 'leanness', 'mortality rate' and 'integument alterations'). This implies that even small improvements in individual measures will lead to (slightly) higher integrated scores. Such a high degree of discrimination is perhaps more motivating for farmers to implement on-farm welfare improvements than a WI which changes only in response to very drastic improvements.

Models for none of the CRMs produced a better fit with the integrated welfare scores given to the fictitious herds by the selected experts when compared to applying no compensation-reduction. We suggest four possible explanations. First, other formulas than the ones we modelled could fit better with how experts perform compensation-reduction. (e.g. a synergetic formula, where a higher weight is attributed to a multitude of poor scores than for only one poor score) Second, selected experts may have been overtaxed by the survey task of allocating welfare scores to the fictitious herds so that they failed to apply compensation-reduction when giving scores, even though at hindsight they would have considered this desirable. In consumer studies, this problem is referred to as 'information overload'. This occurs when people are unable to distinguish relevant from irrelevant information when presented with too much information (Herbig and Kramer, 1994, Malhotra, 1982). Therefore, a focus on a limited set of welfare measures which may not be exhaustive but which do expose many different but important adverse effects may be preferred to using a large amount of measures. Third, the selected experts may not truly have considered compensation-reduction necessary. Apart from the study by Botreau at al. (2009), there appears to be hardly any evidence in the literature that supports the necessity for compensation-reduction. Fourth, enough compensation-reduction may have been achieved solely by using the different severity scores in combination with relative prevalence. Scores of more severe welfare problems thus automatically weigh more on the overall score and cannot be compensated fully by less severe welfare problems.

In the current study, the aim of the WI was not to detect all possible adverse effects on dairy cattle welfare. Instead we focused on an index which reflects the worst adverse effects on welfare (according to literature and selected experts). Incorporating an extensive list of welfare

measures would complicate Step 3 of the process (comparing expert scores with WIs) possibly leading to information overload. We therefore strongly advice to use a disclaimer with the proposed WI indicating which adverse effects may not be detected by the current selection of measures. This approach may be considered as more fair than claiming exhaustiveness (which in our view is currently near impossible). The proposed WI does enable detection of all the worst adverse effects on dairy cattle welfare according to the report by Nielsen et al., 2014). In the study by Nielsen et al. (2014), the worst adverse effects were selected from a list of adverse effects on dairy cattle welfare, based upon EFSA reports. Although this list was not assessed for comprehensiveness, this remains to be validated by future research, it was indicated that some additional (not-worst) adverse effects (i.e. reproductive disorders; thermal discomfort; pain, fear and frustration; abomasal displacement; respiratory distress/pain; other adverse effects related to diseases and other adverse effects related to injuries) are undetectable by the current selection of welfare measures. Measures of any omitted adverse effects could be added to the protocol when this is desired, by using the step-wise approach proposed in the current study (Figure 6). In addition, the inclusion of measures defining and differentiating the positive dimensions of welfare states rather than exclusively focusing on welfare impairments is desirable and should be promoted. The selected measures focus on welfare impairments and not on the positive spectrum of welfare state, which implies that the WI is not comprehensive as such. A growing number of animal scientists employ a hedonic approach to animal welfare, which means that not only physical health is involved in welfare states, especially if the aim is to assure enhanced welfare rather than excluding poor welfare state (Broom, 1991, Tannenbaum,1991).

Experts were stimulated in the survey to take validity and reliability of the WQ measures into account for ranking of the welfare measures. However, it still may be questioned whether validity and reliability were adequate. For example, mortality rate is based on herd records of which reliability has barely been documented. As is the case for any welfare assessment protocol, it is important to strive for high reliability of the measures by training observers to achieve high test-retest, inter- and intra-observer reliability and by unbiased sampling of animals.

Categorical differentiation between herds (i.e. welfare categorization), is useful to interpret the WI in terms of which scores indicate farms of poor or excellent welfare. In addition, such

welfare categories may be used for labelling purposes to identify farms of varying welfare levels. In the current study, we determined thresholds based upon expert scores in the different welfare categories for the 12 fictitious herds. These thresholds are only indicative, given the limited number of herds and selected experts which they are based on. In addition, as none of the fictitious farms were labelled as 'enhanced' using the alternative WI, model fit for this category needs to be further determined.

Two main inputs were used in the current study; expert opinion and a European database of selected welfare measures' prevalence. As expert opinion was vital in the current study, the use of stringent selection criteria of these selected experts was of high importance. While this limited the amount of selected experts, it also ensured adequate knowledge about dairy cattle welfare and the welfare measures concerned. Still, testing whether similar outcomes (in terms of selected welfare measures, severity scores and correspondence with expert opinion) would be achieved using a different group of dairy cattle welfare experts (including animal welfare scientists but also other stakeholders like end-users of the WI) would be relevant. Testing whether another setting (e.g. a workshop to achieve consensus like in Rodenburg et al., 2008) would influence the outcomes would be interesting as well. The second unique and important input used in the current study was the database containing prevalence data on the selected measures of 491 European dairy herds. This dataset allowed selected experts to benchmark results of the fictitious farms based on a wide range of data, which supported them in allocating welfare scores. Such a large database on other (non-WQ) measures where a uniform protocol was used may be harder to attain.

Conclusions

The step-wise approach employed in the current study led to an integration method for a selection of animal-based welfare measures which is simple, transparent and discriminative. In addition, the resulting WI was highly coherent with expert opinion. The current selection of all six welfare measures are associated with all the worst adverse effects for dairy cattle welfare as identified by Nielsen et al (2014). Nevertheless, the integrated welfare index should be accompanied with a disclaimer which lists adverse effects that cannot be detected adequately by the current selection of measures. However, the proposed method is flexible such that measures can be replaced or added as deemed desirable by repeating the proposed steps.

Chapter 5

Determinants of consumer intention to purchase animal-friendly milk

Abstract

Concern about the welfare of production animals is growing among various stakeholders, including the general public. Citizens can influence the market for premium welfare products by expressing public concerns, and consumers - the actors who actually purchase products - can do so through their purchasing behaviour. However, current market shares for premium welfare products are small in Europe. To better align purchase behaviour with public and individuals' concerns, insight is needed in determinants that influence the intention to purchase premium welfare products. A cross-sectional online survey of 787 Flemish milk consumers was conducted to investigate attitudes towards and intention to purchase animal-friendly milk. More than half of the sample (52.5%) expressed the intention to purchase animal-friendly milk. Linear regression modelling indicated that intention was positively influenced by 1) higher perceived product benefits from animal-friendly milk (milk with more health benefits and higher quality); 2) higher personal importance of extrinsic product attributes such as local production and country of origin; 3) higher personal importance of animal welfare; 4) a more natural living oriented attitude towards cows; and 5) a more positive general attitude towards milk. Intention was negatively influenced by 1) a stronger business-oriented attitude towards cows; and 2) by a higher personal importance attached to price. These insights in key components of purchase intention can assist producers, the dairy industry and retailers to position and market animal-friendly milk.

Introduction

A diverse group of stakeholders, including citizens, farmers, public authorities and the food industry, is increasingly confronted, interested and/or concerned with the welfare of production animals (Verbeke, 2009). Many studies related to this topic have focused on the end user of the chain, both in their role as citizen and consumer. The general public can influence the marketing and sale of premium welfare products by acting as citizens (influencing government policy by expressing existing public concerns) and as consumers (choosing to purchase certain products instead of others).

Results from both citizen and consumer studies illustrate a sense of public concern about farm animal welfare (e.g., Serpell, 2004) and a positive intention to purchase premium welfare products expressed in terms of self-reported willingness to pay (e.g. Taylor and Signal, 2009, Vanhonacker and Verbeke, 2009a, Olson and Jacoby, 1972). However, current market shares for premium welfare products are small in EU countries (e.g., a market share for organic milk of 2.7% in Belgium in 2015 (Samborski and Van Bellegem, 2016), which does not reflect citizens' stated valuation of animal welfare (European Commission, 2016). Several reasons are suggested for this citizen/consumer duality: limited (perceived) product availability (Vermeir and Verbeke, 2006); inadequate information provisioning (Vermeir and Verbeke, 2006, Harper and Henson, 2001), other product attributes (quality, food safety, price) that outweigh animal welfare in the purchasing decision (Vanhonacker and Verbeke, 2014, Ingenbleek et al., 2006, Heleski et al., 2004) and a lack of differentiation among products in terms of the level of animal welfare (de Jonge and van Trijp, 2013). The lack of differentiation refers to the products currently available, which are mainly either conventional or organic (Oosterkamp et al., 2011). Organic products in EU member states follow the rules set out in the EU Council Regulation (EC) No 834/2007 and include conditions which citizens generally associates with enhanced welfare such as permanent access to open air areas (preferably pasture). These organic products mainly target a niche market that is willing and able to pay (high) price premiums. Other consumer segments consider animal welfare as an important product attribute but not to a degree that they are willing to pay double or more for such products (Vanhonacker et al., 2007, Grunert et al., 2004). These segments are left with an unmet preference and purchasing desire. Better alignment of purchase behaviour with individuals' concerns and preferences can be supported by insight into determinants that influence intention to purchase premium welfare products.

According to the Theory of Planned Behaviour (TPB) (Ajzen, 1985), the intention to perform a certain behaviour precedes the actual behaviour. Ajzen suggested three factors that determine intention: attitudes, subjective norms and perceived behavioural control. This study focuses on the relationship between attitudes and intention. Fishbein and Ajzen (2011), p. 76) defined an attitude as 'a latent disposition or tendency to respond with some degree of favorableness or unfavorableness to a psychological object'.

Previous studies have mainly focused on attitudes in relation to general farm animal welfare (e.g. Vanhonacker et al., 2010, Austin et al., 2005). However, participants experience different degrees of 'feeling well-informed about the welfare' of various farm animal species (Harper and Henson, 2001) and experience different types and levels of concerns for the welfare of different species (European Commission, 2007). In this perspective, a species-specific approach is relevant and adds to the existing knowledge. The present study focuses on dairy cows for the following reasons. First, a large diversity of farm management practices can be found within the dairy sector that influence animal welfare (for instance, pasturing versus zero-grazing; tied housing, loose housing or deep litter (de Vries et al., 2015, Von Keyserlingk et al., 2012)). This diversity makes it possible to differentiate between products based on the characteristics of the production system. Second, milk is produced by cows without these having to be killed. This facilitates the provision of product information related to the animal and its welfare, more so than in meat as consumers do not want to be reminded of the living animal and that it had to be killed for the product when purchasing meat (e.g. Buller and Cesar, 2007, Kennedy et al., 2004, Te Velde et al., 2002, Holm and Møhl, 2000).

The aim of this study was to gain insight into the influence of attitudes, perceptions of product-related factors, and milk consumption on intention to purchase animal-friendly milk. Based on a theory explaining human behaviour (Ajzen, 1985), we hypothesize that the attitudinal determinants of intention to purchase animal-friendly milk are attitude towards milk, attitude towards the milk industry, and attitude towards dairy cows. Additionally, we will examine the role of product attribute importance and associations between animal welfare and other product attributes, because of their potential impact on purchasing decisions in the case of animal-friendly products.

Materials and Methods

Data Collection and Sample

Quantitative descriptive data were collected through a cross-sectional survey using a sample representative for age (20 – 65 years), gender, and region (n=787) from Flanders (the northern region of Belgium; Table 5.1). All participants were to some extent involved in food purchasing and purchased milk at least once a month (inclusion criteria). Participants were selected from a panel from a professional market research agency. All contact and questionnaire administration procedures were managed electronically. All data were collected during March - May 2014.

Table 5.1: Socio-demographic profile of the sample and the Flemish population

Characteristics		Sample (%)	Population[1] (%)
Age	21 - 29	17.5	18.3
	30 - 39	23.1	24.3
	40 - 49	24.7	25.8
	50 - 59	21.9	21.7
	60 - 65	12.9	10.0
Regional distribution	Antwerp	28.4	27.6
	East Flanders	23.9	22.9
	West Flanders	18.1	18.5
	Flemish Brabant	15.2	17.1
	Limburg	14.4	13.8
Gender	Female	52.4	49.5
	Male	47.6	50.5
Education	Elementary education (6 years of schooling)	4.6	15.7
	Secondary education (12 years of schooling)	51.8	54.9
	University college (Bachelor's degree)	33.1	14.8
	University degree (Master's degree)	10.5	14.5

[1] Based on Belgian Federal Government Statistics for the Flemish population, 2012

Questionnaire

The questionnaire assessed consumers' intention to purchase animal-friendly milk, socio-demographics (age, gender, education and region), milk consumption frequency, milk types bought, attitudes towards dairy cows, general attitudes towards milk and the milk industry, associations between animal welfare and other product attributes, and importance of product attributes when purchasing milk.

Intention to Purchase, Milk Consumption Frequency and Type. Consumers' intention to purchase animal-friendly milk was measured with the statement 'I am willing to purchase animal-friendly milk from now on, instead of the milk I usually purchase', using a 5-point Likert scale ranging from '1=fully disagree' to '5=fully agree'.

Participants were asked how often they consume cow's milk and which type of milk they usually purchase. Regarding frequency of milk consumption, participants could choose one of the following options: 'multiple times per day', 'every day', 'multiple times per week', 'once per week', 'less than once per week' or 'never'. For the type of milk, the main types present in the Belgian market were listed, including 'organic' (following Council Regulation (EC) No 834/2007), 'fair trade' (indicates that farmers received a fair and higher price for their milk), 'farm milk' (i.e., the whole process, from milking the cow to packaging the milk, is performed by the farmer), 'AA-milk' (a Belgian quality label with specific hygiene requirements for farmers and milk processors), and 'conventional milk' (defined as none of the above). Participants could also indicate 'our family does not purchase milk' (n=40), in which case they were excluded from the analysis (see inclusion criteria), resulting in 787 usable responses.

Attitudes Towards Dairy Cows. To measure attitudes towards dairy cows, participants were asked to score 13 items on 5-point Likert scales ('1=fully disagree' to '5=fully agree'). These items were based on Austin et al. (2005), who investigated the attitudes of dairy farmers and agriculture students towards farm animal welfare. They found five (in farmers) and six (in students) factors with two superordinate dimensions labelled as 'natural living orientation' and 'business orientation'. These two dimensions strongly resemble the 'affect' and 'utility' dimensions defined in a review on this matter by Serpell (2004). Serpell argued that determinants of human attitudes towards animals and their welfare can be divided in two primary dimensions: 'affect' (representing people's affective and/or emotional responses to animals) and 'utility' (representing people's responses to animals based on their instrumental value). We only report the survey items applicable to consumers and dairy cows (13 of 75 items

in total used in the original survey by Austin et al. (2005). This includes statements about the cow's welfare, the farmer, and production goals (e.g. "cows have to be kept in an environment that is as natural as possible" and "a farmer should think of his/her cows mainly in terms of the profit they will bring in"; see Appendix B). These items were deemed suitable for inclusion in the present study because they could be reduced to dimensions similar to those documented in previous studies (Kling-Eveillard et al., 2015, Kauppinen et al., 2012). Some of the statements used for the present study were rephrased from Austin et al. (2005) to make them applicable to consumers (Appendix B).

General Attitudes Towards Milk and the Milk Industry. General attitudes towards milk and the milk industry were scored on 5-point semantic differential scales based on Sparks and Guthrie (Sparks and Guthrie, 1998). For attitude towards milk, the bipolar adjectives were 'unhealthy'/'healthy', 'not essential'/'essential', 'not tasty'/'tasty', 'expensive'/'cheap', 'old fashioned'/'modern', and 'negative'/'positive'. For the milk industry, the bipolar adjectives were 'unjust'/'just', 'negative'/'positive', 'not sustainable'/'sustainable' and 'not transparent'/'transparent'.

Associations Between Animal Welfare and Other Product Attributes. To assess consumers' associations of animal-friendly milk with other product attributes, participants were probed for their level of agreement with the statement 'More attention for dairy cow welfare will lead to milk that is…' for 12 attributes in line with Vanhonacker et al. (2010). These attributes were: 'more traditional', 'safer', 'better quality', 'more hygienic', 'more authentic', 'easier to get', 'tastier', 'better for the environment', 'cheaper', 'healthier', 'more acceptable for me' and 'more profitable for the producer'. Each of these characteristics was evaluated on a 5-point Likert scale ranging from '1 = fully disagree' to '5 = fully agree'.

Product Attribute Importance. The product attribute importance was measured using the question: 'When purchasing milk, how important are the following product attributes for you?', and registered on a 5-point interval scale ranging from '1 = totally unimportant' to '5 = very important'. Attributes were: 'quality', 'food safety', 'freshness', 'healthiness', 'price', 'packaging', 'taste', 'sustainability', 'production method', 'animal welfare', 'local production', 'country of origin', 'environmental friendliness' and 'fair trade' (based on (Vanhonacker et al., 2010)).

Statistical Analysis

Data were analysed using SPSS 22 (Chicago, IL, USA). Factor analysis using principal components was performed for each of the following five sets of measures: 1) attitudes towards dairy cows, 2) general attitudes towards milk, 3) general attitudes towards the milk industry, 4) associations between animal welfare and other product attributes and 5) product attribute importance. Items were excluded using backward selection when factor scores were lower than 0.6 and when the item loaded high (> half of the highest) on two or more factors, as recommended by Field (2009). Cronbach's alphas were checked and deemed acceptable when higher than 0.7.

Paired T-tests were performed to determine differences between the perceived importance of product attributes using $P < 0.05$ as a cut-off value for significance. Linear regression analysis was performed with intention to purchase animal-friendly milk as the dependent variable. Initial explanatory variables were socio-demographics variables (age, gender, education and region), frequency of milk consumption, type of milk consumed, and factors of the attitudes towards dairy cows, general attitudes towards milk and the milk industry, associations with other product attributes and product attribute importance.

The data complied with all assumptions for linear regression (i.e., normality of the residuals, heteroscedasticity, linearity and multicollinearity using Spearman correlation index and Variance inflation factor (VIF) values). In line with Field (2009), outliers were detected using standardized residuals values (>3.29), covariance statistics (based on sample size, >1.026 and <0.973) and Cook's distance (>1). Outliers were removed (n=46) and components that had resulted from Principal Component Analyses (PCA) were checked for consistency after removal of the outliers. The final model was chosen using backward selection, excluding explanatory variables expressing a P-value >0.10. Although differences between segments are often significant, only relevant (e.g. >0.5 point difference) are discussed and used to profile segments.

Results

Milk Consumption. Most participants consumed milk at least once per day (54.1%), or multiple times per week (27.3%). The majority of the participants consumed conventional milk (65.1%) or AA-milk (22.6%). A small percentage of people consumed farm milk (5.0%), fair trade milk (4.2%) or organic milk (3.2%). On average, the most important milk attributes were

freshness and safety (mean importance scores of 4.19 and 4.17, respectively) (Figure 5.1). Generally, the participants allocated about as much importance to animal welfare as to price (both 'rather important'). More than half of the study participants expressed a positive intention to purchase animal-friendly milk (52.5% agree or fully agree), while a small group answered negatively (7.3% disagree or fully agree).
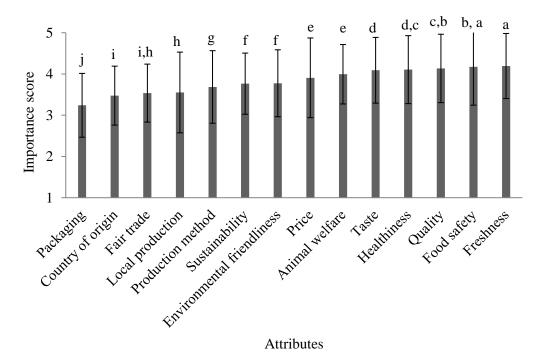

Figure 5.1: Product attribute importance of milk, mean attribute scores ± SD for the question 'When purchasing cow's milk, how important are the following product attributes for you?,' measured using a 5-point interval scale ranging from 1 = totally unimportant to 5 = very important. Different letters indicate significant differences (Paired T-tests, $P < 0.05$).

Attitudes Towards Dairy Cows. Factor analysis on the 13 items revealed three factors (Table 5.2), which we labelled as 'business orientation', 'natural living orientation' and 'functioning orientation'. The factor 'business orientation' (Cronbach's alpha, $\alpha = 0.83$) incorporated items concentrating on the economic function of cows for a farmer. 'Natural living orientation' ($\alpha = 0.73$) contained items about natural behaviour and the cows' living environment. 'Functioning orientation' consisted of statements in which good animal welfare is closely associated with a cow's health status and performance ($\alpha = 0.81$). Mean scores (± Standard Deviation, SD) were 2.60 (± 0.63) for the factor business orientation, 4.10 (± 0.76) for natural living orientation and 3.35 (± 0.74) for functioning orientation.

Table 5.2: Rotated factor loadings of the Principal Factor Analysis of the attitudes towards cows and cow welfare, with three factors (F1-3) labelled 'Business orientation', 'Natural living orientation', and 'Functioning orientation'

Items[2]	F1: Business orientation	F2: Functioning orientation	F3: Natural living orientation
A farmer should think of his/her stock mainly in terms of their market value or cost	**0.86**[1]	0.06	-0.24
A farmer should mainly think of his/her cows in terms of the profit they will bring in	**0.83**	0.14	-0.07
A farmer should view his/her cows as useful objects with which he can gain personal benefit	**0.78**	0.07	-0.11
Production efficiency should be first priority of the farmer	**0.70**	0.11	-0.24
A cow that is physically fit experiences good welfare by definition	0.07	**0.88**	0.09
If a cow is reproducing efficiently, her welfare standards must be good	0.03	**0.88**	0.08
If a cow is growing well she must be experiencing good welfare	0.22	**0.78**	-0.00
Cows need a natural habitat just as much as wildlife does	-0.17	-0.00	**0.82**
Cows have to be kept in an environment that is as natural as possible	-0.10	0.04	**0.81**
It is important for cows to be able to express natural behaviour	-0.30	0.16	**0.73**

[1]Boldface type indicates items and their loading that are included in each factor

[2]Total variance explained: 69.1%

General Attitude Towards Milk and the Milk Industry. Factor analysis revealed that all descriptors for attitude towards milk loaded on one factor (total variance explained: 70.2%, $\alpha=0.85$), except for the items 'expensive versus cheap' and 'old fashioned versus modern', which were excluded because factor loadings were too low (0.43 and 0.57). The items of attitude towards the milk industry could all be fit into one factor (total variance explained: 73.2%, $\alpha=0.88$). Mean scores ($\pm$ SD) were 4.03 ($\pm$ 0.77) for general attitude towards milk and 2.97 ($\pm$ 0.82) for general attitude towards the milk industry.

Associations between Animal Welfare and Other Product Attributes. Factor analysis identified two factors, explaining 62.4% of the total variance in associations between animal welfare and other product attributes. The first factor, 'perceived benefits', included items related to health, quality, taste, hygiene, acceptability, safety, authenticity, environmental friendliness and traditionality ($\alpha=0.91$). The second factor, 'perceived disadvantages', included items related to price, availability and profitability for the farmer ($\alpha=0.69$). Mean scores ($\pm$ SD) were for perceived benefits 3.79 ($\pm$ 0.61) and for perceived disadvantages 3.08 ($\pm$ 0.69).

Product Attribute Importance. Factor analysis identified two factors explaining 66% of the total variance in perceived product attribute importance. The first factor, 'intrinsic product attributes', included the attributes quality, freshness, food safety, healthiness and taste ($\alpha=0.91$). The second factor, 'extrinsic product attributes', included local production, country of origin, packaging and fair trade ($\alpha=0.71$). This grouping of product attributes was based on a study by Olson and Jacoby (1972), who stated that intrinsic attributes are specific to each product and are no longer present when it is consumed and cannot be changed without changing the product itself. Extrinsic attributes are not physically a part of the product but are strongly associated with it (e.g., production method). Mean importance scores ($\pm$ SD) were 4.14 ($\pm$ 0.63) for intrinsic product attributes and 3.45 ($\pm$ 0.71) for extrinsic product attributes.

The attributes sustainability, price and animal welfare could not be fit into either of the factors due to a high factor loading on both factors. However, in this study price and animal welfare were very important. We therefore chose to include these in the linear regression model as single-item explanatory variables.

Linear Regression Results

A significant correlation (correlation coefficient > 0.4) was found between the factor of attitude towards dairy cows 'functioning orientation' and the factor 'intrinsic attribute importance',

therefore the latter was excluded from analysis. Descriptive statistics of the variables are presented in Table 5.3.

Table 5.3: Descriptive statistics (mean and SD) of dependent and explanatory variables (linear regression model, n=741)

Variables	Mean	SD
Dependent variable		
Intention to purchase animal-friendly milk	3.61	0.85
Explanatory variables		
Attitude towards milk	4.05	0.79
Attitudes towards dairy cows - Business orientation	2.60	0.63
Attitudes towards dairy cows - Natural living orientation	4.10	0.76
Associations with other product attributes - Benefits	3.84	0.61
Product attribute importance - Price	3.91	0.82
Product attribute importance - Animal Welfare	4.00	0.84
Product attribute importance factor - Extrinsic product attributes	3.53	0.72

The final linear regression model (Table 5.4) shows that the intention to purchase animal-friendly milk was positively influenced by 1) higher perceived benefits from animal-friendly milk in terms of the product being perceived as healthier, of better quality, tastier, more hygienic, more acceptable, safer, more authentic, environmentally friendlier and more traditional; 2) a higher personal importance attached to extrinsic product attributes (local production, country of origin, packaging and fair trade); 3) a higher personal importance attached to the product attribute 'animal welfare', 4) a higher score for the attitude towards dairy cows of 'natural living orientation' and 5) a more positive attitude towards milk. Intention was negatively influenced by 1) a higher score for the attitude towards dairy cows of 'business orientation' and 2) a higher personal importance of the attribute price. Neither the socio-demographic characteristics (i.e. gender, age, education, region), nor the behavioural characteristics such as milk consumption frequency or type of milk consumed, expressed a P-

value < 0.10 and these were therefore not included in the final model. The final model explained 42.4% (R^2) of the variation in the intention to purchase animal-friendly milk.

Table 5.4: Determinants of the intention to purchase animal-friendly milk (linear regression model, n=741)

Variables	Standardized Beta	P-value	95.0% Confidence Interval for B	
			Lower Bound	Upper Bound
Product attribute importance factor – Extrinsic product attributes	0.22	<0.001	0.18	0.34
Associations with other product attributes - Benefits	0.21	<0.001	0.18	0.39
Attitudes towards dairy cows - natural living orientation	0.19	<0.001	0.10	0.22
Product attribute importance - animal welfare	0.18	<0.001	0.10	0.25
Attitudes towards dairy cows - business orientation	-0.17	<0.001	-0.19	-0.09
Product attribute importance - price	-0.17	<0.001	-0.24	-0.12
Attitude towards milk	0.06	0.049	0.00	0.13

Discussion

The goals of this study were to gain insight into consumers' intention to purchase animal-friendly milk, and to explore attitudes and perceptions that influence this purchase intention. The study complements previous studies by focusing on the welfare of a specific species (i.e. dairy cows), and on a particular food product category (i.e. cow's milk). These insights may therefore be valuable in formulating recommendations for producers and/or retailers aiming to market animal-friendly milk.

A heterogeneous intention to purchase animal-friendly milk was found in our study, where a majority (52.5%) expressed a positive intention. Although these percentages are based on self-reporting, and therefore potentially overestimated, they suggest a substantial market interest and preference for animal-friendly milk. The heterogeneity in intention to purchase animal-friendly products has also been reported in earlier studies (e.g. Vanhonacker and Verbeke, 2009b). However, the market for milk products is at present poorly differentiated in terms of animal welfare in Belgium. Therefore, a deeper understanding of determinants that influence intention to purchase animal-friendly milk can be relevant.

Corresponding with the TPB, our results showed a significant relationship between attitudes and intention. Intention to purchase was positively influenced by a more natural living oriented attitude towards cows, in line with Vanhonacker et al. (2010), and a more positive general attitude towards milk. By contrast, we found intention to be negatively influenced by a stronger business-oriented attitude towards cows. This corroborates with findings by Vanhonacker et al. (2007) in which a segment of consumers with a business-oriented view on farm animals was identified which also had a lower intention to purchase animal-friendly products.

Similar to attitudes, we found different product attribute perceptions to significantly influence the intention to purchase animal-friendly milk. The evaluation of food products is a multi-attribute decision process (Fishbein, 1976). Previous studies reported that a higher priority is generally given by consumers to intrinsic product attributes like quality, health and safety than to aspects related to moral and/or sustainability issues like fair trade, local production and animal welfare (Ingenbleek et al., 2006). This was confirmed for the pooled sample in the current study, where on average intrinsic rather than extrinsic product attributes were given the highest importance score. However, looking beyond the pooled sample averages, results suggested some degree of heterogeneity in perceived product attribute importance. Namely, participants who attached a higher importance to extrinsic product attributes, were found to have a stronger intention to purchase animal-friendly milk; while participants who indicated a higher personal importance attached to price showed a weaker intention to purchase animal-friendly milk. These results are consistent with findings by Vanhonacker and Verbeke (2009b) on the choice for animal-friendly poultry products.

Furthermore, we found price and animal welfare as product attributes to be rated equally important in our sample. This is in contrast with many earlier studies where price was considered more important to consumers than animal welfare (e.g. Vanhonacker et al., 2010,

Harper and Henson, 2001). This might be an effect of the growing concern and interest in animal welfare among European consumers, as illustrated by European Commission studies from different years (European Commision, 2005 - 2016). However, in other survey-based consumer studies, some social desirability bias may have led to an under-estimation of the importance of price versus an over-estimation of the importance of animal welfare. This was illustrated by Wolf et al. (2016) where participants generally indicated to find animal welfare more important than low milk prices but also indicated that they thought the 'average American' would not necessarily agree. These authors argue that due to social desirability bias responses projected on the average American are probably closer to the actual attitudes of the respondents.

Even though, generally, price and animal welfare were considered to be of equal importance in our study sample, not all consumers will be attracted to the highest level of animal welfare (Vanhonacker et al., 2007, Vermeir and Verbeke, 2006), but intermediate levels of price and animal welfare might suit the Flemish market for animal-friendly milk. De Jonge et al. (2015) labelled such intermediate products (broiler filets in their study) as 'compromise products' and Van Herpen et al. (2015) referred to these (meat products) as 'mildly sustainable' products. De Jonge et al. (2015) reported that the consumption of animal-friendly meat products increased when such compromise products are offered in the assortment. Van Herpen et al. (2015) reported similar results but nuanced their findings by adding that in-store display and price level influence product choice. Further studies could focus on market segmentation for animal-friendly milk specifically, and on identifying different segments' characteristics.

Results of the current study give some insight on how to communicate about animal-friendly milk. Opportunities for communication stem from the finding that the intention to purchase animal-friendly milk was related to other product attributes in the factor 'benefits' (in terms of the animal-friendly product being perceived as healthier, of higher quality, tastier, more hygienic, more acceptable, safer, more authentic, environmentally friendlier and more traditional). This indicates that the participants who were more likely to purchase animal-friendly milk also associated other benefits more strongly with animal welfare. This is in accordance with earlier studies by Vanhonacker et al. (2007, 2009b) and Cardoso et al. (2016) who found that people related higher welfare standards with product attributes like health, taste, and quality. We conclude from these findings that communication to promote animal-friendly milk does not necessarily have to focus on highlighting the welfare of the cows only, but could also highlight the associated benefits of the resulting milk and dairy products. This could attract a wider group of consumers. In order to use such benefit claims, further research might be

needed to evidence the link between animal welfare and other product benefits, such as for example indications of a positive relationship between animal welfare and milk quality (Müller-Lindenlauf et al., 2010) or a higher nutritional quality in products which originated from organic dairy farms (Palupi et al., 2012).

In contrast to attitudes and perceptions, intention to purchase animal-friendly milk was neither significantly related to socio-demographic factors (such as gender and age) nor to behavioural characteristics (such as frequency and type of milk consumed) – when analysed in a multivariate setting together with attitudinal factors. In fact, earlier studies also questioned the usefulness of socio-demographic and/or behavioural factors in food-related decisions involving ethical considerations (Diamantopoulos et al., 2003).

Limitations and Future Research

Every study is based on a specific research design and applies certain methodologies and analyses. This choice imposes some   limitations, providing at the same time challenges for future research. This study was developed to gain first insights in determinants that influence consumers' intention to purchase animal-friendly milk. We mainly focused on attitudes, applying Ajzen's TPB to measure attitudes towards dairy cows based on Austin et al. (2005). These questions were mainly gauging specific elements of welfare instead of the entirety of dairy cattle welfare, particularly towards natural living, but paid less attention to other elements of animal welfare like biological functioning and affective state (Fraser et al., 1997). Further research could study more and/or other possible attitudinal determinants, or other constructs such as beliefs, life-style variables or values.

Further, we have used a structured questionnaire and survey method, which has the advantage to collect data from a large number of participants, allowing quantitative analyses. Its disadvantage is potential bias induced by the hypothetical nature of the survey and a possible social desirability bias (Fisher and Katz, 2008). Given the debated relationship between what an individual states as a citizen and how he/she actually behaves as a consumer (Vermeir and Verbeke, 2006), further research is recommended to validate the findings of this study in non-hypothetical research settings. Finally, further research could study whether the findings are applicable for animal products from other animal species and in other geographical regions or specific (consumer) population groups.

Conclusions

This study showed that Flemish (Belgian) adult milk consumers varied in their self-reported intention to purchase animal-friendly milk, with a small majority indicating a positive intention. This variation was (partly) explained by the perceived benefits from animal-friendly milk, the importance attached to extrinsic product attributes, the importance of animal welfare and price as a product attribute, the natural living oriented and business-oriented attitudes towards dairy cows, and the attitude towards milk. These results can be used to position animal-friendly milk on the market. The use of various determinants found to positively influence intention to purchase, such as highlighting the welfare of the cows along with the associated benefits of such a product, can be used in communication about animal-friendly milk, in order to address specific consumers' beliefs and desires.

Chapter 6
Market opportunities for animal-friendly milk in different consumer segments

Abstract

Consumers have increasing, but highly variable, interest in sustainability attributes of food, including animal welfare. We explored market opportunities for animal-friendly cow's milk based on segmentation (cluster) analysis. Flemish survey participants (n=787) were clustered (n=6) based on their intention to purchase (IP) animal-friendly milk, and their evaluation of cows' welfare state (EV). Three market opportunity segments were derived from clusters and labelled as 'high', 'moderate' and 'limited'. Only 8% of the participants belong to the 'high market opportunities' segment, characterized by a high IP and a low EV. The 'limited' segment (44%) indicated a neutral to low IP and a positive EV. The 'moderate' segment (48%) had a moderately positive IP and positive/negative EV. Reported willingness to pay, interest in information about the state of animal welfare and importance of the product attribute 'animal welfare' differed among segments and were strongly related to IP. Most promising selling propositions about animal-friendly milk were related to pasture access. The high degree of differentiation within the Flemish milk market reveals market opportunities for animal-friendly milk, but for an effective market share increase supply of animal-friendly products needs to get more aligned with the heterogeneous demand.

Introduction

The public is increasingly interested in sustainability aspects of food consumption, including the various characteristics of the production process and of the final food products themselves. For the livestock production, these concerns also relate to animal welfare (Briggeman and Lusk, 2010). However, although European citizens report high levels of concern in relation to farm animal welfare (Krystallis et al., 2009), the market for animal-friendly products – or of products that are commonly perceived as such - is small. For example, the market share of organic milk, which is associated with higher levels of animal welfare by consumers (Van Loo et al., 2014, Harper and Makatouni, 2002), was 2.7% in Belgium in 2015 (Samborski and Van Bellegem, 2016). This may illustrate that concerns do not always translate into purchase behaviour (Harvey and Hubbard, 2013, Toma et al., 2011), called the 'consumer-citizen-duality'. While citizens are driven by moral concerns, consumers are driven by the product characteristics involved in their food purchasing decision process (Dagevos and Sterrenberg, 2003). To account for this duality, it is important to take both consumer-related and citizen-oriented measures into consideration when studying this topic.

The European food market for premium animal welfare products is currently a niche market where differentiation levels in the products vary according to the farm animal species. Eggs, for example, have moderate differentiation (i.e., cage, barn eggs, free range, organic), whereas cow's milk shows little differentiation in terms of animal welfare with the exception of welfare-related options such as 'organic' (following EU Council Regulation No 834/2007) and 'access to pasture'. The current market for animal-friendly products mostly attracts consumers with only one specific profile, while previous studies (Liljenstolpe, 2008, Bourlakis et al., 2007, Hall and Sandilands, 2007) demonstrated a more heterogeneous interest in farm animal welfare. While some consumer clusters are highly interested in animal-friendly products, other clusters may consider animal welfare to be an important product attribute but not dominant to other product attributes such as price, taste or quality. For these consumers, products that compromise between animal welfare and price, termed 'compromise products', 'mildly sustainable products' or 'medium welfare products' in different papers, may be an attractive alternative (European Commission, 2007). However, knowledge about market opportunities in different consumer clusters for specific types of animal(-derived) products is currently lacking.

Societal concern about animal welfare varies among species of farm animals along with the degree of 'feeling well-informed about welfare' (Harper and Henson, 2001). In response, the current study focuses on dairy cattle with cow's milk as product, for several reasons. First, consumers do not want to be reminded of the living animal when it had to be killed for the product while purchasing meat (so-called 'strategic ignorance' (Onwezen and van der Weele, 2016)). Hence, it is expectedly easier to communicate about the animal and its welfare for products, such as milk, that do not require the animal to be killed. A second advantage is that dairy farms strongly differ in the housing and management factors affecting cow welfare (de Vries et al., 2015, Von Keyserlingk et al., 2012). This makes it possible to differentiate the milk based on animal welfare status. Third, the dairy sector represents a significant proportion (15%) of the EU agricultural output (Europea, 2013), and market differentiation might be a welcome strategy for this sector which regularly faces economic crises and searches for overall sustainability.

The overall objective of this study is therefore, to explore market opportunities for animal-friendly milk by clustering individuals based on a consumer- and a citizen-oriented measure. The overall objective is further specified into three research goals. First, this study will identify consumer segments for animal-friendly milk. Second, it will be examined to what extent market opportunities exist within different consumer segments for animal-friendly milk in Flanders, Belgium. The specific focus on the region Flanders is motivated by the significance of the dairy sector in the study area itself and for milk production in Belgium. Third, the resulting segments will be profiled in terms of perceived attribute importance and interest in information as basis for the development of marketing and communication strategies for animal-friendly milk. These insights can support a targeted marketing approach, relevant to seize market opportunities.

Materials and Methods

An online questionnaire was completed in May 2014 by 786 consumers living in Flanders, the northern region of Belgium. Participants were recruited by a subcontracted professional market research agency. The sample was representative for the Flemish population in terms of age, regional distribution and gender (Table 5.1). As 44% of the participants had an university college or university degree, the sample was slightly biased towards higher educated people

(which was ±30% in the Flemish population in 2014), which is common in online surveys (Granello and Wheaton, 2004). All respondents were involved in food purchasing (not necessarily the main person responsible). The survey consisted of four tiers of variables: 1) segmentation variables; 2) variables aimed to form a socio-demographic profile of the clusters; 3) variables used to examine market opportunities of the different clusters; and 4) measures aimed to support the market positioning of animal-friendly milk, mainly in terms of communication.

Questionnaire

Segmentation Variables. The consumer-related measures used as segmentation variables were the intention to try and the intention to effectively purchase animal-friendly milk. These were measured with the statements 'I am willing to buy animal-friendly milk from now on, instead of the milk I usually buy', and 'I am willing to buy animal-friendly milk to try it' using a 5-point Likert scale ranging from 'totally disagree' to 'totally agree'. The citizen-oriented measure was the perceived evaluation of the current state of dairy cattle welfare (EV), which was measured using the question 'How would you rate the welfare of dairy cattle in Flanders?' on a 7-point scale ranging from 'very bad' to 'very good'. The term 'dairy cattle welfare' was not defined in the survey because there are varying definitions of animal welfare, among consumers and even among animal scientists (Fraser, 2008). We wanted respondents to use their own perception of animal welfare, as they would do so as a consumer too.

Socio-demographics. Age, regional distribution (Flemish provinces), gender and education were probed using categorical scales. Rural or urban living conditions, familiarity with agriculture and livestock were scored on 5-point semantic differential scales. The bipolar adjectives were 'rural'/'urban', 'not familiar at all with agriculture'/'very familiar with agriculture' and 'not familiar at all with livestock'/'very familiar with livestock'.

Market Opportunities. To determine market opportunities in Flanders, Belgium for animal-friendly milk among different consumer segments, we examined the different segments' milk consumption pattern, general attitudes towards milk and the milk processing industry, and willingness to pay for animal-friendly milk.

Milk consumption. Participants were asked about their milk consumption frequency and about the type of milk they usually buy. Options for frequency of milk consumption were: 'multiple times per day', 'every day', 'multiple times per week', 'once per week' or 'less than once per week'. For the type of milk, the main types present in the Belgian market were listed, including 'organic' (following Council Regulation (EC) No 834/2007), 'fair trade' (indicates that farmers received a fair and higher price for their milk), 'farm milk' (i.e., the whole process, from milking the cow to packaging the milk, is performed by the farmer), 'AA-milk' (a Belgian quality label with specific hygiene requirements for farmers and milk processors), and 'conventional milk' (defined as none of the above). Participants could also indicate 'our family does not purchase milk', in which case they (n=40) were excluded from the analysis (see inclusion criteria), resulting in 787 usable responses.

General attitudes towards milk and milk processing industry. General attitudes towards milk and the milk industry were scored on 5-point semantic differential scales. Specific information on scales can be found in Chapter 5.

Willingness to pay (WTP). WTP was measured using contingent valuation (Vossler and Kerkvliet, 2003) using the question: 'suppose that a label came onto the market that you trust which guarantees animal-friendly production of the milk. To which degree would you be willing to buy this product instead of a product without such a label?'. This is a widely used method in consumer studies that is easily understood by participants. Despite limitations, e.g. relating to hypothetical bias, the contingent valuation method can yield valuable insight into eventual differences between consumer segments regarding WTP. Respondents were provided with a reference price of the current average conventional milk price in Belgian supermarkets (based on an online consultation of five Belgian supermarkets, €1.10 per litre). WTP for animal-friendly label milk was probed for different price points, including no change in price (€1.10 per litre) and for price increases of 5% (€1.16 per litre), 10% (€1.21 per litre), 20% (€1.32 per litre), 50% (€1.65 per litre) and 100% (€2.20 per litre). Answers were on a 5-point scale ranging from 'very unlikely' to 'very likely'.

Communication About Animal-friendly Milk. To form the basis for development of marketing and communication strategies about animal-friendly milk, segments were profiled in terms of perceived attribute importance, trust in information sources, preferred format of information on dairy cattle welfare, and interest in information about dairy cattle welfare.

Perceived attribute importance. The perceived attribute importance was measured as described in Chapter 5, using the question: 'When buying milk, how important are the following product attributes for you?' on a 5-point interval scale ranging from '1 = totally unimportant' to '5 = very important'.

Trust in information sources and preferred format of information on dairy cattle welfare. Respondents were asked to indicate their level of trust in different information sources in relation to dairy cattle welfare on a 5-point scale ranging from 'a complete lack of trust' to 'a great deal of trust'. Sources were 'government', 'animal welfare organisation', 'supermarket', 'milk processing sector', 'farmer', 'consumer organisation', 'science' and 'veterinarian'. Additionally, respondents were asked to indicate their preference towards five possible welfare labels on a 5-point scale ranging from 'absolutely no preference' to 'strong preference'. Options were chosen to allow comparison between 1) a general animal welfare label versus a specific animal welfare label, 2) a tangible animal welfare issue versus an intangible animal welfare issue (i.e. more difficult to imagine without having specific knowledge on dairy cattle) from the consumer point-of-view and 3) a product label with a broader scope including animal welfare versus an animal welfare label. Therefore, options were 'a general cow welfare label' (general animal welfare label), 'a cow welfare label with a focus on access to pasture' (specific animal welfare label, tangible issue), 'a cow welfare label with a focus on welfare problems such as udder infection and lameness' (specific animal welfare label, intangible issue), 'a label with focus on sustainability, where higher animal welfare requirements are one aspect'(product label with a broader scope, including animal welfare) and 'a label with a focus on local production, where higher animal welfare requirements is only one of the aspects (product label with a broader scope, including animal welfare).

Interest in information about dairy cattle welfare. Respondents were probed about their interest in information about dairy cattle welfare using measures and criteria derived from the Welfare

Quality® (WQ) protocol for dairy cattle . This protocol describes methods to integrate dairy cattle welfare measures into 12 criteria, four principles and an overall welfare category. All 26 welfare measures and 11 of the 12 criteria were used in the current study (Appendix C). In dairy cattle, the criterion 'Thermal comfort' is not measured and was thus not included in the survey. The full WQ protocol for dairy cattle can be found at http://www.welfarequalitynetwork.net/.

To measure interest in information, the following question was used for each WQ measure and criterion: 'Several aspects that can influence the welfare of dairy cows are listed below. To what degree do you wish to be informed about each of these aspects in order to make the right choice when buying cow's milk?'. Answers were on a 100-point scale ranging from 'I do not want to be informed about this at all when I am buying cow's milk' to 'This is very interesting information for me when I am buying cow's milk'. Measures and criteria were shown separately and in random order to prohibit the respondents from scoring the measures purely based on the criteria they belong to.

Statistical Analysis

Factor Analysis using principal components was performed for both attitude measures. Items were excluded using backward selection when factor scores were lower than 0.6 and when the item loaded high (> half of the highest) on two or more factors, as recommended by Field (Field, 2009). Internal consistency (using Cronbach's alpha's) was checked and deemed acceptable when higher than 0.7 following Nunnally (1967). Factor Analysis revealed that all descriptors for attitude towards milk loaded on one factor (total variance explained: 70.2%, α=0.85), except for the items 'expensive versus cheap' and 'old fashioned versus modern', which were removed because factor loadings were not high enough (0.43 and 0.57). The items of attitude towards the milk industry could all be fit into one factor (total variance explained: 73.2%, α=0.88). Because of the high Cronbach's alphas, a mean attitude score towards milk and the milk processing industry was calculated, by summating and averaging the respondents scores for the individual adjectives, and used in further analyses (bivariate analysis using one-way ANOVA). Data were analysed using SPSS 22 (Chicago, IL, USA). For the cluster analysis, first hierarchical clusters were developed using Ward's method. Subsequently, K-means cluster analysis were performed to obtain clusters (following (Kuo et al., 2002)). Z-scores were used for all segmentation variables to get a better view of the relative position of the clusters.

Bivariate analysis using one-way ANOVA with Dunnett's T3 post hoc comparison of mean scores and cross-tabulation with $\chi2$ –statistics were used to profile the clusters.

Results and Discussion

Cluster Analysis

The evaluation of the current state of dairy cattle welfare (EV) was neutral on average (not poor and not good, mean 4.1±1.3 on a scale of 1 - 7). EV was perceived as very negative – rather negative by 29% of respondents, as neutral by 31% and as rather good – very good by 40%. With regard to the questions dealing with intention, mean intention to try was rather positive (3.9±0.8 on a scale of 1 - 5) and intention to purchase from now on neutral – rather positive (3.6±0.9). Most respondents (72%) (strongly) stated to be willing to try animal-friendly milk, while only 4% (strongly) disagreed. For more than half of the respondents this resulted in a positive intention to purchase animal-friendly milk from now on (53%). As IT and IP were highly correlated (r = 0.73, P < 0.001), we will discuss them together under the term 'intention to purchase' (IP) from here on.

The clustering procedure resulted in six clusters (Table 6.2, Figure 6.1). The existence of multiple consumer segments is in accordance with previous studies showing heterogeneity in interest in farm animal welfare . We divided the clusters into three segments based on market opportunities. The first segment, with 'high market opportunities', consists of cluster 2 (CL2) which is characterized by a high IP and a low EV. The second segment, referred to as 'limited market opportunities', consist of three clusters (CL4, CL5 and CL6) and is characterized by a more positive perception of the current state of dairy cattle welfare and a neutral to low IP. The third segment, referred to as 'moderate market opportunities', consists of two clusters (CL1 & CL3). Whereas CL1 does not seem concerned about dairy cattle welfare, they do have a positive IP, while CL3 does seem to be concerned about animal welfare (negative EV) with a neutral – rather positive IP. As some important nuances did exist between clusters in these three segments, further analysis was performed using all clusters, but results are mainly discussed using the three segments. However, where differences between clusters within segments are present, these will be discussed.

Table 6.2: Segmentation variables, percentages of the sample and market opportunities for each cluster (CL) (Mean (± SD), n = 787)

Cluster code	CL1	CL2	CL3	CL4	CL5	CL6
Market opportunities[1]	Moderate	High	Moderate	Limited	Limited	Limited
% of sample	24.4	8.1	23.8	3.6	16.0	24.1
Evaluation dairy cattle welfare (EV, z-score)	0.8	-1.7	-0.8	0.4	0.5	0.2
Evaluation dairy cattle welfare[2]	5.1 ± 0.7	1.8 ± 0.8	3.0 ± 0.9	4.6 ± 1.3	4.7 ± 0.9	4.3 ± 0.9
Intention to try (IT, z-score)	0.7	1.3	0.2	-2.7	0.2	-1.1
Intention to try[3]	4.5 ± 0.5	5.0 ± 0.0	4.1 ± 0.4	1.6 ± 0.6	4.1 ± 0.9	3.0 ± 0.9
Intention to purchase from now on (IP, z-score)	0.8	1.5	0.2	-2.2	-0.9	-0.7
Intention to purchase from now on[2]	4.3 ± 0.5	4.9 ± 0.3	3.8 ± 0.4	1.7 ± 0.5	2.8 ± 0.4	3.0 ± 0.3

[1] As derived from the segmentation variables

[2] 7-point scale from 'very bad' to 'very good'

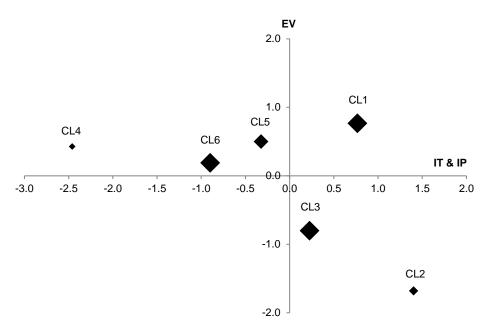[3] 5-point scale from 'strongly disagree' to 'fully agree'

Figure 6.1: Mapping of the cluster centers (CL1-CL6) according to the evaluation of welfare (EV z-score) and intention to purchase animal-friendly milk (IT and IP z-score). The size of the marks reflects the relative size of the clusters (total n = 787).

Socio-demographics

Fewer male respondents were found in the high market opportunities segment (20% versus 49 – 59% for moderate market opportunities and limited market opportunities, P < 0.001). This is in accordance with previous studies where females were more interested/concerned about animal welfare (Phillips et al., 2010, Kendall et al., 2006). The high market opportunities segment also contained more respondents with a higher education (54.7%), while there were fewer in the limited market opportunities segment (specifically CL4 and CL6 with 32.1% and 34.9%, CL1: 45.8%, CL3: 44.9% and CL5: 50.0%, P < 0.001). The profile of the high market opportunities segment corresponds to the 'ethical consumers' that Vermeir and Verbeke (2009) characterize as being more educated and better informed than average.

In CL4, more respondents indicated that they were familiar with livestock compared to the other clusters (39.3% versus 15 - 25% for the high market opportunities and moderate market opportunities segments, P < 0.001). This is in line with a study by Boogaard et al. (2010) among Dutch citizens who found that respondents with more stated familiarity with farming were more content with dairy farming practices. Vanhonacker et al. (2007) found a similar cluster in their

study and argued that the respondents were involved with animal production more on a socio-economical level than on a moral and ethical level.

As segments did not differ with respect to province, rural versus urban living conditions, familiarity with agriculture and age of the respondents, EV and IP appeared to be influenced by the respondents' attitudes, knowledge and opinions rather than by their socio-demographics. This is in accordance with earlier studies that question the usefulness of socio-demographic factors in studies involving ethical concerns (e.g. Verbeke and Vackier, 2004).

Market Opportunities

Milk consumption. Self-reported frequency of milk consumption did not differ between segments, which is in contrast with earlier studies concerning meat, where different clusters were found to indicate a different consumption pattern . Regarding type of milk, respondents in the high market opportunities segment purchased significantly more organic milk compared to consumers of the other clusters (P < 0.001, 10.9% for CL2 and 7.1% for CL4, versus 3.1%, 4.3%, 1.6% and 0.0% for CL1, CL3, CL5 and CL6). This is in line with expectations for this segment given their very positive intention to purchase animal-friendly milk.

General attitudes towards milk and the milk processing industry. Both attitudes differed among clusters (P < 0.001). Attitude towards the milk processing industry was more negative among the high market opportunities segment (2.1 ± 1.0) than for all other clusters. This is very likely related to their negative EV, which reflects poorly on the milk processing industry.

Willingness to pay. The WTP differed for each price increase category between clusters (P < 0.001 for all). As measuring WTP using contingent valuation may be subject to hypothetical bias, we here use the results of this measure only to compare between segments, not to determine thresholds for maximum prices of animal-friendly milk. The pattern followed the IP indicated by the three segments. The high market opportunities segment indicated a higher WTP for all price increases. Most clusters with limited market opportunities (CL4 and CL6) indicated a lower WTP for 0% increase, 5% and 10% (neutral – rather unlikely, Figure 6.2), and no majority indicated a positive WTP at any price level (i.e. maximum of 35% and 39% for a 0% price increase). The majority of respondents indicated a positive (likely or very likely) WTP up to a price increase of 20% for CL2, 10% for CL1 and CL3 and 5% for CL5.
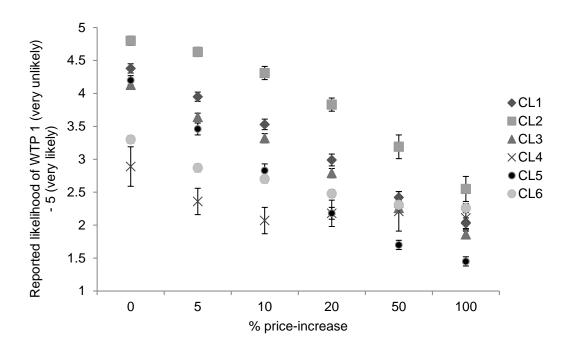
Figure 6.2: Mean willingness to pay scores ± SEM for CL1-CL6, per % of price increase

Considerations - market opportunities. The most pronounced differences between segments were found for the WTP, where the segment that indicated a higher IT and IP (high market opportunities) also showed a higher WTP. This segment would likely purchase products with the highest level of animal welfare, with a matching price (currently organic milk in the Flemish market). Attitudes towards milk as a product were positive, with generally neutral attitudes towards the dairy industry. Along with the finding that milk consumption frequency did not differ between clusters, these findings indicate the existence of market opportunities for animal-friendly milk, both for the highest level of animal welfare and for more intermediate products.

Communication About Animal-friendly Milk

Perceived importance of product attributes. The perceived importance of all milk attributes differed among clusters ($P < 0.001$, Table 6.3), in a somewhat different pattern than the segments. Generally, CL1 and CL2 indicated the highest importance for most attributes (all rather important to very important), CL3 and CL5 mostly indicated 'neutral' – 'rather interested' and the CL4 and CL6 most often indicated 'neutral'. An exception was the attribute 'price', which CL1 and CL5 scored highest (both 'very important'). These findings indicate

that although CL1 might be a more conscious consumer than CL3, they do value price a lot, which causes lower market opportunities than for CL2.

The absolute importance scores for the product attribute 'animal welfare' ranged from 'neutral' for CL4 and CL6, to 'rather important' for CL3 and CL5 clusters and 'very important' for CL1 and CL2. However, when viewing the product attribute 'animal welfare' relative to other product attributes, the high market opportunities and moderate market opportunities segments scored animal welfare among the more important product attributes, and the limited market opportunities segment among the less important product attributes (paired samples t-test, P < 0.001).

Table 6.3: Perceived importance of product attributes (Mean ± SD, n = 787)

	Sample	CL1	CL2	CL3	CL4	CL5	CL6	P-value
Packaging	3.2±1.0	3.5±1.0[b]	3.2±1.3[ab]	3.2±1.0[ab]	3.1±1.1[a]	3.1±1.0[a]	3.1±0.8[a]	<0.001
Country of origin	3.5±1.0	3.9±0.9[b]	4.0±1.1[b]	3.4±1.0[a]	3.0±1.1[a]	3.3±1.0[a]	3.2±0.8[a]	<0.001
Fair trade	3.5±1.0	3.9±0.9[c]	4.2±1.1[c]	3.6±0.8[b]	3.0±1.2[ab]	3.2±1.0[a]	3.2±0.7[a]	<0.001
Local production	3.6±0.9	4.0±0.8[b]	4.1±1.0[b]	3.5±0.9[a]	3.0±1.1[a]	3.3±1.0[a]	3.2±0.7[a]	<0.001
Method of production	3.7±0.9	4.0±0.7[c]	4.6±0.6[d]	3.8±0.8[b]	3.0±1.0[a]	3.4±0.9[a]	3.3±0.7[a]	<0.001
Sustainability	3.8±0.8	4.1±0.7[d]	4.3±0.7[d]	3.8±0.7[c]	3.1±1.1[ab]	3.6±0.9[bc]	3.4±0.7[a]	<0.001
Environmental friendliness	3.8±0.8	4.1±0.7[c]	4.5±0.8[d]	3.9±0.6[b]	3.1±0.9[a]	3.5±0.9[a]	3.3±0.7[a]	<0.001
Price	3.9±0.8	4.1±0.8[bc]	3.7±1.0[ab]	3.9±0.8[b]	3.8±1.1[abc]	4.2±0.7[c]	3.6±0.8[a]	<0.001
Animal welfare	4.0±0.9	4.3±0.7[c]	4.9±0.3[d]	4.2±0.6[c]	3.3±1.0[ab]	3.7±0.8[b]	3.4±0.8[a]	<0.001
Taste	4.1±0.7	4.4±0.6[c]	4.6±0.5[d]	4.1±0.6[b]	3.5±0.7[a]	4.1±0.6[b]	3.6±0.8[a]	<0.001
Health	4.1±0.8	4.4±0.6[d]	4.8±0.4[e]	4.2±0.7[c]	3.5±0.9[abc]	4.0±0.7[b]	3.6±0.8[a]	<0.001
Quality	4.1±0.7	4.5±0.6[c]	4.7±0.5[c]	4.2±0.6[ab]	3.6±1.1[a]	4.2±0.6[b]	3.6±0.8[a]	<0.001
Food safety	4.2±0.8	4.5±0.6[c]	4.8±0.4[d]	4.2±0.6[b]	3.5±1.1[a]	4.3±0.7[b]	3.6±0.8[a]	<0.001

| Freshness | 4.2±0.7 | 4.5±0.6[c] | 4.7±0.7[c] | 4.2±0.6[b] | 3.9±0.8[ab] | 4.3±0.7[bc] | 3.7±0.8[a] | <0.001 |

[a-e] Means ± SD within a row without common superscript differ significantly ($P < 0.05$, one-way ANOVA with Dunnett's T3 post hoc comparison of mean scores)

Information sources and preferred format of information on dairy cattle welfare. All clusters indicated a neutral to rather negative score for 'the dairy processing industry' regarding trust as a source of animal welfare information (Table 6.4). Possibly, this is caused by the industries' vested interests in milk, which potentially biases towards creating a positive image. Besides a negative score for 'the dairy processing industry', the high market opportunities segment indicated rather negative scores for trust in 'government', 'supermarket', and 'farmer' as well. This is likely related to their low EV-score which reflects poorly on these stakeholders. The most trusted sources of an animal-friendly milk label were veterinarians and animal welfare organisations, as all clusters indicated a neutral to positive trust in these segments.

Content of the label for all categories (i.e., 'general cow welfare', 'focus on access to pasture', 'focus on specific welfare problems', 'sustainability' and 'local production') was least preferred by CL4 and CL6 (scored neutral to rather uninteresting, Table 6.3). The high market opportunities segment showed the highest preference for all suggested label content (scored rather to very interesting). The same goes for the type of label. None of the segments indicated a clear preference for any of the label contents.

Regarding the type of dairy cattle welfare label, preference was given to a label with a star rating system by the high market opportunities and moderate market opportunities segments. The limited segments' scores did not differ significantly between the options. These results are in line with the 'extremeness aversion' theory (Simonson and Tversky, 1992), meaning that when a wider range of products is provided, consumers are less likely to choose the cheapest product. In animal welfare products, a similar construct has been examined studies , who found a smaller share of respondents choosing for the product lowest in price and animal welfare when products at intermediate levels of animal welfare and price were offered. An example of a similar and rather successful labelling scheme (representing approximately 22% of the egg market share) can be found in the Netherlands, where an animal welfare organisation (Dierenbescherming) developed a label with star rating system to indicate different levels of animal welfare for various farm animal species. Also the recently established state-controlled

animal welfare label in Denmark (Bedre Dyrevelfærd) that will enter the market in 2017 distinguishes between three levels of improved welfare following similar principles, though using hearts instead of stars.

Table 6.4: Preferred source, content of a label and type of label (Mean ± SD, n = 787, between segments P-value <0.001 for all)

		Sample	CL1	CL2	CL3	CL4	CL5	CL6
Source	Veterinarian	3.7±0.8	4.0±0.8[c]	3.7±1.1[bc]	3.7±0.8[b]	3.8±1.0[bc]	3.8±0.7[bc]	3.2±0.6[a]
	Animal welfare organisation	3.7±0.8	4.0±0.8[c]	4.2±0.9[c]	3.7±0.7[b]	3.3±1.1[ab]	3.6±0.8[b]	3.2±0.7[a]
	Science	3.5±0.8	3.8±0.8[b]	3.7±0.8[b]	3.6±0.8[b]	3.1±1.1[a]	3.7±0.7[b]	3.3±0.6[a]
	Consumer organisation	3.5±0.8	3.9±0.7[c]	3.5±1.0[b]	3.5±0.8[b]	3.4±0.9[ab]	3.6±0.7[cb]	3.2±0.5[a]
	Farmer	3.1±0.8	3.3±0.9[c]	2.3±1.1[a]	3.0±0.9[b]	3.3±0.9[bc]	3.5±0.8[c]	3.1±0.6[b]
	Supermarket	3.1±0.7	3.2±0.8[b]	2.6±0.9[a]	3.0±0.7[b]	3.2±0.8[b]	3.3±0.7[b]	3.2±0.5[b]
	Government	3.1±0.9	3.3±1.0[b]	2.5±1.0[a]	3.0±0.9[b]	3.2±1.1[b]	3.1±0.8[b]	3.1±0.6[b]
	Dairy processing sector	3.0±0.9	3.3±0.9[cd]	2.4±1.1[a]	3.0±0.9[b]	3.1±1.0[bcd]	3.3±0.9[d]	3.0±0.5[bc]
Content	General cow welfare	3.7±0.9	4.3±0.9[c]	4.4±0.7[c]	3.9±0.7[b]	2.9±1.1[a]	3.8±0.8[b]	3.1±0.7[a]
	Focus on access to pasture	3.6±0.9	4.2±0.8[c]	4.3±0.7[c]	3.8±0.7[b]	2.8±0.8[a]	3.6±0.8[b]	3.1±0.7[a]
	Focus on specific welfare problems	3.6±0.9	4.2±0.9[d]	4.4±0.8[d]	3.7±0.7[c]	2.6±1.0[a]	3.7±0.9[c]	3.1±0.7[b]
	Sustainability	3.6±0.9	4.2±0.8[c]	4.0±0.7[c]	3.7±0.7[b]	2.7±1.2[a]	3.6±0.8[b]	3.1±0.7[a]
	Local production	3.5±0.9	4.0±0.8[c]	4.3±1.0[c]	3.6±0.8[b]	2.7±1.1[a]	3.3±0.8[b]	3.0±0.7[a]
Type	A label with a star rating system	3.7±1.0	4.4±0.9[d]	4.3±1.0[d]	3.8±1.0[c]	2.5±1.0[a]	3.6±1.0[c]	3.2±0.8[b]
	A label without star rating system	3.2±1.0	3.4±1.1[b]	3.3±1.3[ab]	3.2±1.0[ab]	2.7±1.1[a]	3.3±1.0[b]	3.1±0.8[ab]

[a-d] Means within a row without common superscript differ significantly (P < 0.05, one-way ANOVA with Dunnett's T3 post hoc comparison of mean scores)

Interest in information about WQ measures and criteria. Two findings were most apparent from the respondents' interest in information about WQ measures and criteria. First, for all measures and criteria, interest in information between clusters differed significantly (P < 0.001 for all; interest scores for all measures and criteria can be found in the Appendix C). Post hoc tests showed a clear pattern: interest in information in measures and criteria was lowest for the 'limited market opportunities' segment (mean scores ranged from 27.1 – 35.9 for CL4, 27.9 - 58.6 for CL5, and 32.8 - 45.2 for CL6), followed by the 'moderate' segment (CL1: 44.3 - 78.4 and CL3: 43.9 -72.9) and highest for the segment with 'high market opportunities' (CL2: 66.1 - 93.6) (P < 0.05 for all differences). Second, even though interest in all measures differed, generally the same criteria and measures received the highest importance score for all clusters. This is not in accordance with previous studies where heterogeneity in perceived important factors for farm animal welfare was found (e.g. for broiler chickens). However, in these studies respondents were not given a list of (WQ) welfare impairments, but were asked to think of them themselves. When a predetermined list was given to respondents, (as performed by Tuyttens et al., (2010) using WQ criteria), the same criteria were found among the most and least important scores.

Both the high and moderate segments indicated most interest in 'freedom of movement', whereas the limited opportunities segment attached most interest to 'absence of diseases' (all p-values < 0.05). The moderate market opportunities segment and CL5 indicated the second-to most interest in absence of diseases, although moderate clusters indicated as much interest for 'absence of injuries' (CL5), 'comfort around resting' (CL3), 'good human-animal relationship' (CL3) and 'positive emotional state' (CL3). The high market opportunities and moderate market opportunities segments, and CL6, generally indicated least interest for the criteria 'expression of other normal behaviour' and 'expression of social behaviour'. Both CL4 and CL5 did not score any of the criteria as lower than most other criteria. Regarding measures, highest scores (all P < 0.05) by the moderate market opportunities segment and CL5 and CL6 were attributed to 'access to pasture', but some scored 'expression of positive behaviour' (CL1, CL3, CL5, CL6), 'expression of negative behaviour' (CL1, CL3, CL5, CL6), dehorning method (CL3), tail docking method, 'number of drinking points, cleanliness and functioning of the drinking points'(CL3) and 'tied housing' (CL3, CL5, CL6) equally high. CL2 scored all the aforementioned measures equally high, and CL4 scored all equally low.

While most segments attributed lowest scores to the criterion which is measured using 'access to pasture' (criterion 'expression of normal behaviour'), this measure was given the highest importance score by the majority of segments. Possibly, freedom of movement was associated with access to pasture by the respondents (which actually is a measure of the criterion 'expression of normal behaviour') which explains this discrepancy. Access to pasture is consistently mentioned in literature as being highly important to citizens (Schuppli et al., 2014, Boogaard et al., 2008). Additionally, although the criterion 'absence of diseases' was scored high by many clusters, none of the measures for diseases were among the highest scores for most clusters. It is likely that a high level of knowledge or familiarity is needed to understand the importance of these specific welfare measures, while a general lack of knowledge about farming practices and specific welfare problems among consumers has been shown (Miele, 2010). Besides a lack of knowledge, respondents also seemed to not prefer specific information, given that the least preferred content for an animal-friendly label was 'specific welfare problems'. The latter is in accordance with previous studies which found that too detailed information (or information overload) could have a detrimental effect on the intention to purchase (Kolodinsky, 2012, Van Kleef et al., 2008).

Considerations – communication targeted to the various segments. Generally, animal welfare was scored as a neutral to a very important product attribute, but less important than taste, health, quality, food safety and freshness. This is in accordance with earlier studies, which report that consumers generally give higher priority to primary product attributes like quality, health and safety than to aspects related to moral issues and/or sustainability such as fair trade, local production and animal welfare (Ingenbleek et al., 2006). By underlining the product attributes that are considered most important in milk, consumers who are negative or neutral in their intention to purchase might see benefits of the animal-friendly products. Perceived benefits from animal-friendly milk (e.g., healthier and better quality) have previously been found to be related to intention to purchase (Chapter 5) and could change consumers' purchase behaviour. Naturally, these benefits would have to actually exist in order to make such claims about animal-friendly milk. A study has already shown a positive relationship between milk quality and animal welfare, but more research on this subject is still needed.

The most trusted sources for information on cow welfare were veterinarians and animal welfare organisations. Regarding type of dairy cattle welfare label, general preference was given to a

label with a star rating. Regarding interest in information about the current state of animal welfare, the measure 'access to pasture' and criterion 'freedom of movement' were considered to be most interesting. Although the WQ criterion 'absence of diseases' was among the highest scores as well, using individual diseases or injuries like lameness or mastitis as attributes to communicate does not seem a promising route, because no preference was given to this type of welfare label, nor to the WQ measures of disease.

Differences between clusters within segments

Although results were mostly discussed for the segments, some profound differences between clusters within segments were found. Within the moderate market opportunities segment, clusters differed regarding their EV (CL1 rather positive, CL3 rather low, Table 6.5). Also, a majority of respondents in CL1 were willing to pay a higher percentage of price increase than in CL3 (10% versus 5%). Furthermore, product attribute importance of all attributes was scored somewhat higher by CL1 than CL3 (rather important to very important versus neutral to rather important). Within the limited market opportunities segment, CL5 differed from CL4 and CL6 in a higher WTP (majority willing to pay a 5% price-increase versus no majority willing to pay any price increase), a somewhat higher importance attached to all product attributes, including animal welfare (neutral to rather important versus neutral) and a somewhat higher interest in information about dairy cattle welfare. This indicates that CL1 and CL5 may have higher market opportunities than the other clusters in the segments they were grouped in. Therefore, there may be opportunities (e.g. using informational approaches) to transfer respondents from these clusters to higher market opportunities segments.

Table 6.5: Summarizing differences between clusters within segments

Segment:	Moderate	High	Moderate	Limited	Limited	Limited
	CL1	CL2	CL3	CL4	CL5	CL6
IP[1], EV[2]	Positive IP, rather positive EV	Highly positive IP and low EV	Rather positive IP, rather low EV	Rather negative IP, neutral EV	Neutral IP, neutral EV	Neutral IP, neutral EV
Positive Willingness To Pay	Majority up to 10% price-increase	Majority up to 20% price-increase	Majority up to 5% price-increase	No majority at any price level	Majority up to 5% price-increase	No majority at any price level
Product attribute importance – all attributes	Rather important / very important	Rather important / very important	Neutral / rather important	Neutral	Neutral / rather important, price very important	Neutral
Product attribute importance – animal welfare	Very important	Very important	Rather important	Neutral	Rather important	Neutral
Interest in information about the current state of animal welfare	Moderate	Highest	Moderate	Lowest	Moderate	Lowest

[1]Combination of intention to try and intention to purchase animal-friendly milk

[2]Evaluation of the current state of dairy cattle welfare in Flanders, Belgium

[3]Throughout the table, the term 'more' indicates a higher percentage of respondents than other clusters, and the term 'fewer' indicates a lower percentage than other clusters ($P < 0.05$)

Conclusions

The objective of this paper was to explore market opportunities for sustainable products, specifically animal-friendly milk, by identifying and profiling market segments based on a consumer- and a citizen-oriented segmentation variable (IP and EV). A high degree of differentiation within the Flemish market for milk was shown (six consumer clusters), with identification of market opportunities for animal-friendly milk, while the market supply does not show a similar differentiation. Intention to purchase, and to a lesser degree WTP, were strongly linked to interest in information about animal welfare. The consumer segment with the

highest IP and WTP is believed to be most interested in the products with the highest level of animal welfare (e.g., organic; high market opportunities segment). In addition, this segment was most negative regarding the state of animal welfare and the dairy industry. Milk with the highest level of both price and animal welfare could be positioned in the market by primarily focusing on enhanced welfare as this appears to be a dominant selection criterion for the segment with high market opportunities.

Although in this paper, we collated clusters based on market opportunities to form three segments, some crucial differences between clusters within segments were apparent. While CL5 was identified as having 'limited' market opportunities, this cluster did seem somewhat more interested in animal-friendly milk than CL4 and CL6 within the same segment. Since age was not associated with the clusters, a transition of individuals from one segment to another, e.g. from limited to moderate, or from moderate to high, is unlikely to happen automatically over time as the simple result of people's ageing. Specific efforts that aim at stimulating such a transition will be needed, and the fact that gender is associated with clusters indicates that extra efforts targeting male consumers might be needed. The use of informational approaches emerges as a potential strategy. The finding that education is associated with clusters suggests that better and more easily accessible information may be effective in moving people from one to another segment. Such informational approaches could focus on the state of animal welfare or on the precise characteristics of animal-friendly products.

This study lays the foundation for future research, which could focus on examining consumer segments with moderate and/or limited market opportunities, where communication could highlight specific benefits (e.g., healthier and better quality) at an intermediate level of animal welfare and price to change these consumers' purchase behaviour. Promising selling propositions in future communication about animal-friendly milk were found to be 'access to pasture' and the related criterion (according to respondents) 'freedom of movement' and the criterion 'absence of diseases'. Our results reveal an information need that confirms actual practices as foreseen in operational labelling systems, such as a star rating system differentiating between diverse levels of animal welfare at different price levels. Marketing efforts which underline access to pasture and the healthiness of cows would play into the publics' interests regarding dairy cattle welfare. To position products with the highest level of price and animal welfare, enhanced welfare could be the main focus, whereas intermediate products could be

positioned based on benefits of increased cow welfare for milk. When the market supply would be better aligned with the heterogeneous demand, for instance by using a star (or equivalent) rating system, the market share of animal-friendly products could increase. Additionally, such a rating system could encourage farmers to invest in improved cow welfare as a higher rating could translate into a higher price for their dairy products.

The results obtained in the present study are specific for milk in Flanders, Belgium. Nevertheless, it is likely that the finding of a high degree of differentiation in the market based on consumers' intention to purchase and their evaluation of the state of animal welfare as an attribute of livestock products, as well as the existence of related market opportunities, can be extrapolated to other livestock products, animal species, and regions. Further research into this domain is recommended.

The relationship between animal welfare and farm performance

Abstract

The relationship between animal welfare and cost efficiency is investigated on 41 dairy farms in Flanders, Belgium. On a sample of 263 farms, farm-specific cost efficiency (CE), decomposed in technical (TE) and cost allocative (CAE) efficiency, is estimated with data envelopment analysis (DEA). On 41 of these farms, animal welfare was assessed with the Welfare Quality® (WQ) protocol and an Alternative Welfare Index. Average scores were 0.66, 0.87, and 0.57 for TE, CAE and CE respectively, with a negative correlation between TE and CAE. Technically efficient farms used proportionally more concentrates compared to grassland and roughage to produce milk, whereas cost allocative efficient farms used proportionally less concentrates compared to grassland and roughage. Efficiency scores did not differ according to the farms' WQ classification as 'enhanced' (n=10) and 'acceptable'(n=31). None of the efficiency scores was correlated with the Alternative Welfare Index, but some were correlated with single welfare measures. These correlations show either win-win, e.g. TE and prevalence of mastitis, CE and prevalence of lesions, or trade-offs, e.g. TE and prevalence of hairless patches, CAE and share of cows lying outside the lying area and CAE and cows with dirty flanks and upper legs. Our results indicate that pursuing both cost efficient milk production and improved animal welfare is feasible, but reaching one goal not necessarily implies accomplishment of the other goal.

Introduction

Welfare of farm animals is of increasing interest and concern to consumers, farmers, public authorities and the food industry (Verbeke, 2009). Farmers may be stimulated by enforcement, e.g. legislation mandating minimum requirements for housing and management factors that influence animal welfare (Ingenbleek et al., 2012) or by price premiums, e.g. quality assurance schemes which allow both producers and retailers to differentiate themselves from others (Manning et al., 2006). Dependence on stimuli reinforces the perception that welfare efforts come at the expense of farms economic performance, usually measured as net farm profit, but is this really the case? The interrelationship between a farms' economic performance and animal welfare has been the subject of ongoing debate (Lusk and Norwood, 2011). Efforts to improve animal welfare or health, often imply extra managerial interventions, treatment costs or investments in animal housing conditions (Özkan et al., 2016). The other way around, improvements in animal welfare, may reduce the prevalence of animal diseases, increase productivity and reduce possible economic losses.

Efforts to deal with animal welfare and farm economic performance are diverse and mainly concern cost-benefit analyses, often in a normative budgeting exercise, or empirical analyses of welfare problems like mastitis and lameness in relation to productivity (Huxley, 2013, Green et al., 2002, Fregonesi & Leaver, 2001, Warnick et al., 2001). Productivity analysis has the advantage of directly measuring the amount of input needed for a unit of output, but it remains only a partial and indirect measure for economic performance. Other authors went a step further and included both the costs of managerial interventions and the gains resulting from improved welfare of the dairy herd (Ettema et al., 2006, Seegers et al., 2003). One popular economic measure is gross margin, being the difference between revenues (outputs multiplied by output price) and direct costs necessary to come to these outputs. These studies, however, only focus on single welfare measures. But as both animal welfare and a farms' economic performance are multidimensional concepts (Vanhonacker et al., 2012, Wilson, 2011, Fraser, 2008, Coelli et al., 2005, Fraser et al., 1997, Mason & Mendl, 1993), they require multi-criteria evaluation approaches (Burow et al., 2013, Blokhuis et al., 2003). Maximizing farms' profitability relies on managerial competencies to optimise milk yield and composition, while minimizing operational and fixed costs of production. Furthermore, one animal welfare measure can be negatively correlated with farm performance, while for another measure this correlation may be the other way round. For example, Coignard et al. (2004) found positive associations

between high-yielding dairy herds and less aggressive behaviour between cows, but negative associations with the prevalence of diseases and injuries. Health disorders can result in a decrease in milk production, troubling the cause and consequence distinction and making things even more complex. Integral approaches are therefore needed to reveal the link between animal welfare and farms' economic performances.

Current research efforts on both animal welfare and farm performance evolved to capture their multi-criteria complexity, and may help to analyse the economic-welfare link in a more integrated way. On the economic side, production efficiency models (Coelli et al., 2005) became widely used to assess farm-specific technical and cost efficiency, also in dairy farming, and even in particular to analyse possible links with specific animal diseases (van der Voort et al., 2016, Barnes et al., 2011, Lawson et al., 2004). In contrast to single economic and partial productivity measures, efficiency analysis, originating from the work by Farrell (1957), enables to decompose cost efficiency into technical and cost allocative components. Technical efficiency is an integrated measure of the more physical process of input-output transformation, whereas Cost Allocative Efficiency refers to the ability to use inputs in a cost minimizing proportion given the input prices. On the animal welfare side, several published methods integrate the scores of multiple animal welfare measures into one score or index reflecting the overall welfare status of a given herd (see reviews by Czycholl et al., 2015, Johnsen et al., 2001). The Welfare Quality® (WQ) protocol is probably the most renowned method and one of the few protocols using mainly animal-based measures (Welfare Quality®, 2009). A growing consensus exists among animal welfare scientists that animal-based measures are to be preferred for directly assessing the animal's actual state of welfare as opposed to input-based measures like management or resources (Czycholl et al., 2016, EFSA, 2012, Blokhuis et al., 2003, Johnsen et al., 2001). However, recently these protocols have been criticised for their complex integration methods which does not seem to adequately balance the relative importance of the various welfare measures (Chapter 2 in this thesis, Buijs et al., 2016, Heath et al., 2014, de Vries et al., 2013b, Tuyttens et al., 2010). Therefore, in Chapter 4 of this thesis, an Alternative Welfare Index was proposed which is simple and transparent, multidimensional, discriminatory and highly coherent with expert opinion. In this study, we use both production efficiency models and integrated welfare measures. Details are given in the methods and materials section.

The objective of this paper is 1) to explore the relation between overall animal welfare and cost efficiency using integrated multi-criteria approaches and 2) to explain this relation by decomposing the overall scores into their components and analyse links between these components. Animal welfare is assessed by the WQ protocol and an Alternative Welfare Index as integrated scores, whereas cost efficiency is assessed with efficiency analysis that integrates productivity with input allocative aspects. The integrated animal welfare approaches are decomposed into single dairy cattle welfare measures, while cost efficiency is decomposed into its technical and cost allocative component parts. The research is done on a set of 41 dairy farms, with detailed data on both dairy cattle welfare and farm accountancy available. To position the sample in a broader context, the efficiency analysis is also done on a wider sample within Flanders, Belgium.

Materials and Methods

Data gathering

During the winters of 2013 and 2014, animal welfare was assessed on 43 dairy farms in Flanders with an average herd size of 68.7 cows (SD = 30.4, range = 28 − 150). Farms were selected based on collection of individual SCC data (which is needed to calculate WQ scores) and farmer consent. Assessments commenced shortly after the morning milking, and were always performed by one out of two experienced assessors. Animal welfare was measured using the Welfare Quality® protocol (2009). Farm accountancy data were collected with farmers' consent directly from the accounting firms to analyse farm characteristics and to be used for efficiency analysis.

To position the farms in a wider, representative sample for dairy farms in Flanders with respect to technical and economic performance, data of 254 extra farms from the regional farm accountancy data network (Department of Agricultural and Fisheries in Flanders) were used. To position the farms regarding their values for dairy cattle welfare measures, they are compared to benchmarks derived from WQ (Welfare Quality®, 2009). If not available in the protocol, these benchmarks were based on the 25% worst scoring herds out of a larger sample of 121 Flemish dairy herds (containing the 43 abovementioned dairy farms). All 121 farms were visited in the same time period.

The WQ Protocol for Dairy Cow Welfare Assessment

The WQ protocol for on-farm dairy cattle welfare assessment comprises 31 measures and a step-wise method to integrate these measures into 12 criteria scores, grouped further into four principle scores and finally into an overall welfare categorization with four possible levels: 'excellent', 'enhanced', 'acceptable' or 'not classified'. Data for the underlying welfare measures are collected on-farm by means of behavioural observations, an avoidance distance test, a management questionnaire, a resource checklist, clinical scoring and a QBA. Behavioural observations take approximately two hours and are performed to gain data on lying down behaviour (time needed to lie down, collisions during lying down, on edge/outside of lying area), respiration disease (coughing) and prevalence of agonistic interactions. The avoidance distance test (ADF) is performed at the feed rack, and is used as a measure of human-animal relationship. The animals are approached by an observer while their head is through the feed rack, and the distance at which the animals show avoidance behaviour is noted. Clinical scoring of the animals is used to gather data on diseases, injuries, cleanliness and body condition. The measures are mostly animal-based, with the exception of the resource-based and management-based measures regarding the drinkers, loose or tied housing, access to pasture and mutilations. The QBA is a method where the herd is observed for a given period of time (20 minutes in the WQ protocol) after which expressive and emotional descriptors (e.g. 'frustrated' or 'content') are scored to interpret the behaviour and body language of the herd (Rousing and Wemelsfelder, 2006).

Alternative Welfare Index Score

The Alternative Welfare Index described in Chapter 4 comprises six of the 31 WQ measures, selected to detect those problems with the most severe impact on dairy cattle welfare (based on expert opinion and literature). The six welfare measures are 'severe lameness', 'mortality', 'leanness', 'hairless patches', 'lesions/swellings' and 'SCC > 400.000'. Based on Burow et al. (2013), the Welfare Index Score is constructed from severity of welfare problems ('severity score' as determined by dairy cattle welfare experts with experience in the WQ protocol for dairy cattle (n = 14)) and prevalence of these welfare problems. The following formula is used to integrate data on selected welfare measures into one score:

$$Welfare\ index\ score = \mathbf{100} - \frac{1}{6} \times \sum_{m=1}^{6} Sm \times rPm$$

Here, S is 'severity score' ranging from 0 to 1, m is 'measure' and rP is 'relative prevalence'. The resulting Welfare Index Score is on a scale of 0 (very poor welfare) to 100 (excellent welfare). In the formula, relative prevalence (rP, prevalence per herd/prevalence at 95th percentile) rather than absolute prevalence is used so each herd covers the same possible spectrum for each measure.

Data Envelopment Analysis

Data envelopment analysis (DEA), a frequently used method for efficiency analysis, is used to assess the aggregate technical and economic performance of the farms in the sample. DEA involves the use of linear programming to construct a non-parametric frontier, which envelops the data points by piecewise connecting the best performing farms in the dataset (Coelli et al., 2005). In our study, we approach efficiency as the efficient transformation of different feed inputs into milk. The reason why our study focuses on only feed inputs will be described in next paragraph. Technical efficiency (TE) then measures the farmers' ability to use a minimal amount of feed inputs to produce milk as an output. The piecewise frontier links technically efficient farms that constitute the angular points of the frontier. Cost allocative efficiency (CAE) measures the ability to use inputs in an optimal mix given the input prices. Cost efficiency (CE) combines TE and CAE. Fully cost efficient farms are situated on the frontier (they are fully technical efficient) and use an allocation of inputs that minimizes input costs (they are fully cost allocative efficient). Inefficiency is identified by comparing the current performance levels of farms with their potential optimal performance level. The efficiency score of a farm lies between 0 and 1, 0 indicating fully inefficient and 1 indicating fully efficient (Coelli et al., 2005).

To assess TE, CAE and CE scores, milk production (l) is considered as single output variable, because the main objective of specialised dairy farms is to produce milk. Milk production is adjusted for fat and protein content of the milk to the same percentage (4% fat and 3.3% protein) to enable comparing milk production across farms. Our model includes three input variables related to feed intake of the dairy cows: 1) purchased feeds (mainly concentrates and by products intake (€)), 2) roughage intake (ha) excluding grass and 3) pasture (ha), being the sum of grazed grassland and grassland used for roughage. By-products are residual products from beer, sugar and potato processing. The choice of three feed inputs is justified because feed is a dominant cost, the main fixed costs are attributed to roughage production and feed mix is pivotal

in operational management. Costs of roughage (both grassland and other roughage) production are based on the variable and fixed costs as derived from the accountancy data. Costs of purchased feed correspond to the purchase prices. For the farms where animal welfare was assessed, these data were collected using the annual accounts from 2012, which was the last common year in which accountancy data were available for the comparative analysis with welfare data.

DEA is sensitive to outliers, since the piecewise frontier in DEA connects real farms within the sample (Coelli et al., 2005). Outliers are removed based on partial productivities of the inputs used and on dairy enterprise size: farms with total milk production below 100 000 litre per year are removed from the sample. In total 34 farms are removed (of which two farms where animal welfare was assessed) which leaves us with 263 farms for efficiency analysis and 41 for the comparative analysis.

Statistical Analysis

Data preparation and analysis are performed using SPSS Statistics 23. Differences in efficiency scores and descriptive statistics of farms appertained to different classes of animal welfare ('acceptable' and 'enhanced') according to the WQ protocol, are assessed using independent samples t-tests, as all farms are classified in only 2 of 4 classes of animal welfare. A Spearman's rank test is used to assess the correlation between animal welfare measures and efficiency scores, including the correlation between the Alternative Welfare Index, efficiency scores and farm statistics.

Results

Animal Welfare

Assessment of animal welfare by using the Welfare Quality protocol reveals 31 farms to be classified as 'acceptable', 10 farms are classified as 'enhanced'. None of the farms within the sample are classified as 'not classified' or 'excellent'. The Alternative Welfare Index reveals an average score of 78.1, ranging from 57.9 to 92.2. The results of the welfare measurements according to the WQ protocol and the Alternative Welfare Index are given in Table 7.1. Besides the average scores of each measure included in both WQ and the WI, also threshold values and the share of herds above the threshold are shown in Table 7.1. Values above these thresholds are signalling welfare problems for the individual measures. The majority of farms in the

sample scores above the threshold value for measures within the criterion 'comfort around resting'. For example, the thresholds "at least 19% of the cows have dirty udders", "at least 19% of the cows have dirty flanks and upper legs", "at least 50% of the cows have dirty lower legs" and "at least 5% of the cows lie partly outside lying area" are exceeded by more than 70% of the farms.

Table 7.1: Values for animal welfare measures included in the WQ protocol and the Alternative Welfare Index* (mean ± SD). Threshold values[1] for welfare problems in individual herds and observed prevalence (%) of these welfare problems within herds.

Criteria Measures	Threshold value welfare problem	Overall (n=41)	
		Mean ± SD	% herds above threshold
Absence of prolonged hunger			
% of lean cows*	5%	3.49 ± 4.44	27
Absence of prolonged thirst			
Per 10 cows: < 1 drinker or < 60 cm trough	No		24
Drinker cleanliness	No		51
At least 2 drinkers/cow	No		15
Comfort around resting			
Mean time needed to lie down	6.3 (s)	4.92 ± 0.81	2
% lying cows which lie partly outside lying area	5%	12.98 ± 11.58	76
% collisions during lying down movements	30%	40.40 ± 29.06	63
% cows with dirty lower legs	50%	63.52 ± 22.49	73
% cows with a dirty udder	19%	35.77 ± 19.69	78
% cows with dirty flanks and upper legs	19%	73.51 ± 15.96	100
Ease of movement			
Tied or loose housing	Tied		7
Absence of injuries			
% cows with at least one hairless patch and no lesion*	46%	34.54 ± 14.87	20
% cows with at least one lesion*	70%	50.02 ± 25.42	24
Moderate lameness	28%	21.05 ± 9.17	20
Severe lameness*	8%	5.63 ± 5.16	32
Absence of disease			
% cows with nasal discharge	10%	11.77 ± 10.34	39
% cows with hampered respiration	6.5%	0.00 ± 0.00	0
% cows with ocular discharge	6%	9.80 ± 7.22	61
% cows with diarrhoea	6.5%	0.81 ± 2.11	2
% cows with vulvar discharge	4.5%	2.24 ± 2.93	12
Frequency of coughs/cow/15 min	6	0.35 ± 0.23	0
% mastitis (milk Somatic Cell Count >400 000)*	17.5%	6.72 ± 4.88	12
% calvings with dystocia	5.5%	15.11 ± 20.24	63

% downer cows	5.5%	6.12 ± 5.58	44
% cows mortality*	4.5%	2.55 ± 2.24	15
Absence of pain due to management procedures			
Disbud/dehorned cattle	Yes		100
Method	Caustic paste		7
Use of analgesics	No		90
Use of anesthetics	No		22
Expression of social behaviour			
Frequency of head butts per cow per hour	0.7	0.60 ± 0.39	39
Frequency of displacements per cow per hour	0.4	0.32 ± 0.24	32
Expression of other normal behaviour			
Number of days on pasture per year	105	134.02 ± 83.05	27
Number of hours on pasture per day	< 6	8.71 ± 7.50	37
Good human-animal relationship			
ADF index[2]	81	73.12 ± 13.40	24
Positive emotional state			
QBA index	-0.3	0.71 ± 1.74	27

[1]Thresholds were either derived from Welfare Quality® (2009), or if not available in the protocol, threshold values were based on the 25% worst scoring herds out of 121 Flemish dairy herds

[2]Higher values indicate poorer welfare

*Measures of the Welfare Quality protocol that are used to calculate the Alternative Welfare Index

Animal Welfare and Cost Efficiency: Link Between Integrated Measures

Efficiency analysis reveals an average cost efficiency score of 0.57 for the extended sample (n=263) and 0.59 and 0.60 for farms classified as 'enhanced' and 'accepted' respectively (Table 7.2). These scores do not differ between both groups of farms. Furthermore, cost efficiency scores are not correlated with the Alternative Welfare Index (Table 7.2). When using gross margin, which is total revenues minus direct costs, we observe on average of 200€ less gross margin per cow for farms classified as 'enhanced', however, this difference is not significant (P=0.119) (Table 7.2). The absence of a correlation between animal welfare, measured by the AWI scores, and gross margin per cow confirms this result. Contrary to the results in gross margins between 'enhanced' and 'acceptable' farms, which indicates a possible trade-off, the link between gross margin and Alternative Welfare Index suggests a win-win. This is confirmed by a similar possible win-win (but again not significant) between CE and AWI.

Table 7.2: Means and SD for efficiency scores and farm statistics for the extended sample (n=263) and for farms classified as 'acceptable' and 'enhanced' according to the WQ protocol.

	All (n=263)		Acceptable (n=31)		Enhanced (n=10)	
	Mean	SD	Mean	SD	Mean	SD
TE	0.66	0.13	0.70	0.11	0.70	0.10
CAE	0.87	0.08	0.85	0.07	0.84	0.06
CE	0.57	0.10	0.60	0.09	0.59	0.09
N° of dairy cows	56.39	28.70	70.91	33.30	58.30	16.31
Milk production (l FPCM[1])	443030.3	257277.2	626177.9[a]	299056.7	471362.2[b]	120169.0
Milk production/cow (l FPCM/cow)	7714.8	1386.7	8758.8	1088.4	8174.3	761.9
Milk/ha roughage (l FPCM/ha)	11256.9	4033.6	13330.2	3803.3	12549.7	3393.3
N° of dairy cows/ha	1.45	0.43	1.51	0.40	1.55	0.42
Concentrates (kg/ 100 l FPCM)	195.2	63.6	283.2	70.4	239.5	53.9
Concentrates (k/cow)	1536.3	641.0	2274.1[a]	1007.6	1535.8[b]	908.2
Grassland (ha/100 l FPCM)	3.80	1.42	2.33	0.78	2.82	0.90
Roughage (excl grassland) (ha/100 l FPCM)	2.29	0.75	2.18	0.50	2.29	1.08
Purchased feed (€/100 l FPCM)	75.37	17.95	84.94	13.72	79.53	17.05
Roughage – cost (€/ha)	1530.26	566.51	1514.99	259.41	1603.61	376.86
Grassland – cost (€/ha)	943.94	280.24	1191.60	308.09	1077.47	194.72
Gross margin (€/cow)	955.56	374.97	1287.59	319.43	1087.43	358.52
Grassland/purchased feed ratio	5.60	3.22	2.83	1.13	3.95	2.33
Roughage/purchased feed ratio	3.30	1.60	2.66	0.82	3.16	2.00
Grassland/roughage ratio	1.86	1.08	1.11	0.42	1.47	0.82

[a-b]means within a row with different superscripts are significantly different (P < 0.05)

[1]FPCM= fat and protein corrected milk

Animal welfare and Cost efficiency: CE decomposed and the correlation with AWI

All partial productivity measures, expressed in litres of milk per unit input, show significant correlations with TE (Table 7.3). This sounds reasonable as TE analysis must be considered as an approach to integrate partial input productivity measures. A remarkable observation is the negative correlation between TE and CAE (ρ=-0.538). More technically efficient farms use proportionally more concentrates (expressed as purchased feed) per ha grassland and per ha roughage. These farms produce more milk per cow (ρ=0.147) and more milk per ha grassland (ρ=0.427) and ha roughage (ρ=0.500). Cost allocative efficient farms use proportionally less concentrates per ha grassland and per ha roughage. This results in lower productivity per ha

grassland (ρ=-0.192) and per ha other roughage (ρ=-0.173), resulting in the negative correlation between TE and CAE. As CE and TE are positively correlated, TE should be improved to increase cost efficiency. As a consequence, also partial productivities, expressed as milk production per ha grassland, per ha roughage and per unit of concentrate use, need to be improved. Technical efficient farms produce more milk both per kg of concentrates and per cow. Cost efficient farms produce more milk per kg of concentrates by reducing the amount of concentrates fed to the cow while maintaining milk production per cow. Interesting now is that the partial productivity measure, milk production per kg concentrates used, is also positively correlated with the Alternative Welfare Index score. This might indicate that economizing on concentrates may act as a leverage for a win-win between cost efficiency and animal welfare. AWI is negatively correlated with herd size (ρ= -0.336; P=0.032), and there is a tendency towards lower on farm milk production with increasing AWI score (ρ=-0.311; P=0.051). Although there is no correlation between milk production per cow and AWI, there is a positive correlation between this welfare measure and milk produced per kg of concentrates. Thus, farms with high AWI scores use less concentrates to produce milk, while the ratio of grassland/purchased feed and other roughage/purchased feed does not differ compared to farms with lower AWI scores. However, this does not result in better TE scores on farms with high AWI. So, although there is a correlation between AWI and this partial productivity measure, milk production per concentrate use, this win-win is not reflected in a positive correlation with the integrated TE, which combines multiple partial productivities.

Table 7.3: Overview of correlations (Spearman's ρ) between efficiency scores and farm descriptives, and between the Alternative Welfare Index and efficiency scores and farm descriptives.

	Overall (n=263)			n=41
	TE	CAE	CE	Alternative Welfare Index
TE	1.000	-0.538**	0.864**	0.120
CE	0.864**	-0.117	1.000	0.182
N° of dairy cows	-0.042	0.015	0.012	-0.336*
Milk production (l FPCM)	0.015	0.027	0.078	-0.311†
Milk production/cow (l FPCM/cow)	0.147*	0.053	0.196**	0.106
N° of dairy cows/ha	-0.051	0.137*	0.053	-0.173
Milk/ha grassland (l FPCM/ha)	0.427**	-0.192**	0.456**	-0.033
Milk/ha other roughage (l FPCM/ha)	0.500**	-0.173**	0.473**	0.233
Milk/kg concentrates (l FPCM/kg)	0.192**	0.109	0.291**	0.302†
Milk/purchased feed (l FPCM/€)	0.342**	0.088	0.475**	0.232
Milk/ha roughage (l FPCM/ha)	0.545**	-0.194**	0.573**	0.030
Concentrates (kg/cow)	-0.086	-0.079	-0.135*	-0.270†
Roughage – cost (€/ha)	0.155*	0.11	0.227**	-0.105
Grassland – cost (€/ha)	0.230**	0.111	0.330**	-0.017
Gross margin (€/cow)	0.414**	-0.043	0.482**	0.189
Grassland/purchased feed ratio	-0.154*	0.175**	-0.114	0.157
Roughage/purchased feed ratio	-0.174**	0.166**	-0.090	-0.052
Grassland/roughage ratio	-0.020	0.099	-0.021	0.184

*$P<0.05$; **$P<0.01$; †$P<0.1$

Animal Welfare and Cost efficiency: trade-offs and win-wins between component parts of integrated measures

The results of the correlation analysis between efficiency scores and single welfare measures are summarized in Table 7.4. Farms with higher technical efficiency scores show a lower prevalence of mastitis ($\rho= -0.344$; $P=0.028$)), measured as the % of cows with a SCC of above 400 000 cells/mL of milk, and a higher prevalence of hairless patches ($\rho= 0.318$; $P=0.043$). Moreover, they tend to have a lower prevalence of lesions ($\rho= -0.387$; $P=0.012$) resulting in a positive correlation between cost efficiency and the absence of lesions in the dairy herd ($\rho=-

0.312; P=0.047). On cost allocative efficient farms, more cows are lying outside the lying area (ρ= 0.311; P=0.048), and these farms tend to have more problems with cows with dirty flanks and upper legs (ρ= 0.305; P=0.053).

Table 7.4: Overview of correlation (Spearman's ρ) between efficiency scores and welfare measures

WQ criteria, measures	Overall (n=41)		
	TE	CAE	CE
Absence of prolonged hunger			
% too lean cows	0.077	0.034	0.048
Absence of prolonged thirst			
Per 10 cows: <1 drinker or <60 cm through	-	-	-
Clean drinkers	-	-	-
At least 2 drinkers/cow	-	-	-
Comfort around resting			
Duration of lying down movements	-0.159	0.039	-0.181
% lying cows which lie partly outside lying area	-0.093	0.311*	-0.001
% collisions during lying down movements	-0.203	0.086	-0.219
% cows with dirty lower legs	-0.070	-0.085	-0.106
% cows with dirty udder	0.041	-0.05	0.016
% cows with dirty flank and upper legs	-0.258	0.305†	-0.165
Ease of movement			
Tied/loose housing	-	-	-
Absence of injuries			
% cows with at least one hairless patch and no lesion	0.318*	-0.254	0.165
% cows with at least one lesion	-0.387*	0.118	-0.312*
% moderate lameness	-0.066	-0.078	-0.153
% severe lameness	-0.146	-0.096	-0.173
Absence of diseases			
% cows with nasal discharge	0.015	-0.056	0.02
% cows with increased respiratory rate	-	-	-
% cows with ocular discharge	-0.121	0.016	-0.092
% cows with diarrhoea	-0.021	0.048	-0.026
% cows with vulvar discharge	0.030	-0.113	-0.052

Frequency of coughing per cow per 15 min	0.203	-0.256	0.136
% mastitis (milk Somatic Cell Count > 400 000)	-0.344*	0.223	-0.239
% dystocia	0.092	0.095	0.129
% downer cows	-0.142	0.058	-0.102
% mortality during the last 12 months	0.078	-0.159	-0.056
Absence of pain due to management procedures			
Method	-	-	-
Analgesics	-	-	-
Anesthetics	-	-	-
Expression of social behaviour			
Frequency of butts per cow per hour	0.124	-0.191	0.025
Frequency of displacements per cow per hour	0.025	-0.101	-0.067
Expression of other normal behaviour			
Number of days on pasture per year	-0.075	0.131	-0.031
Number of hours on pasture per day	-0.199	0.044	-0.208
Good human-animal relationship			
ADF index	0.023	-0.171	-0.07
Positive emotional state			
QBA index	-0.112	0.096	-0.015

†P<0.1; *P<0.05; **P<0.01

Discussion

Our study did not reveal a relationship between overall animal welfare and economic performance on dairy farms. Farms with 'enhanced' animal welfare according to WQ, did not have higher nor lower gross margins or cost efficiency scores, compared to farms with a lower level of animal welfare ('acceptable'). These results were confirmed by the absence of a correlation between the Alternative Welfare Index and both cost efficiency scores and gross margin per cow. Up to this point, it is not clear whether the absence of a correlation might be influenced due to the way economics and welfare are measured. To improve our understanding on the presence or absence of a correlation between both goals, animal welfare was decomposed into single welfare measures instead of using an integrated index. Moreover, cost efficiency was decomposed into technical and Cost Allocative Efficiency, and underlying partial productivities and input ratios were considered. One reason for the absence of a correlation may

be the endogenous compensation or counterbalancing effect between driving factors. The observed trade-off between Cost Allocative Efficiency and technical efficiency is one possible reason. Technical efficient farms used proportionally more concentrates (expressed as purchased feed) compared to grassland and roughage to produce milk, whereas cost allocative efficient farms used proportionally fewer concentrates compared to grassland and roughage. This negative correlation between CAE and TE might question whether the potential of farms to simultaneously improve TE and CAE is achievable with the objective to maximize cost efficiency. Economizing on concentrates is one avenue, at least when expressed per cow, to both improve cost efficiency and the Alternative Welfare Index, and this will in first instance also give rise to increased milk produced per kg concentrates used, as this partial productivity measure is positively correlated with both CE and AWI.

A second clarification can be found in the counterbalancing effects of different measures of welfare and efficiency scores. Our results revealed a negative correlation between technical efficiency and both the prevalence of lesions/swellings and mastitis, but farms with high technical efficiency scores also had a higher prevalence of hairless patches. Our results confirm the findings of Hansson et al. (2011) that fully technical efficient farms have lower prevalence of mastitis. Hansson et al. attributed this to some preventive measures against mastitis on the efficient farms. Other authors, however, described an increased SCC on farms with higher scores for technical efficiency (Allendorf and Wettemann, 2015), although this effect was not linear. Until a specific threshold (159000 cells/mL), efficiency decreased with increasing cell count. Above that value, efficiency increased. For lameness, contrasting results have been published as well. While Barnes (2011) found that low prevalence of lameness (<10%) was associated with higher levels of technical efficiency, Lawson et al. (2004) showed that dairy farms with higher levels of lameness tended to register higher levels of technical efficiency. These results confirm a relationship between single measures of dairy cattle welfare and technical efficiency, although the nature of these relationships varies among studies. In our study, we consider both technical efficiency and the cost allocation of inputs, whereas other studies presumably focus on technical efficiency as a proxy for economic success. Although there is a positive correlation between technical and cost efficiency, the negative correlation between cost allocative and technical efficiency observed in this study emphasizes the need to have more detailed insights in the counterbalancing impact.

The discussion above confirms the complex relationship between animal welfare of production animals and farms' economic performance. Besides counterbalancing effects of different measures of both animal welfare and efficiency scores, methodological choices might (partially) explain contradictory results between studies as well. In this study, animal welfare was assessed by the WQ protocol as these protocols use mainly animal-based, validated welfare measures to assess animal welfare on-farm. As the procedures for integration of single measures into one index in the WQ protocol have been contested (Buijs et al., 2016, Heath et al., 2014, de Vries et al., 2013b, Tuyttens et al., 2010, Chapter 2 and Chapter 3 of this thesis), we additionally used an Alternative Welfare Index, based on six of the 31 measures from the WQ protocol. Similar to the analysis with WQ, no significant correlations were observed between efficiency scores and welfare. The alternative index shows some additional strengths compared to the WQ protocol, in particular because of its continuous and more discriminatory nature. Observations on, for example, the link between concentrates use per cow and animal welfare, help to explore avenues for improving both welfare and economics.
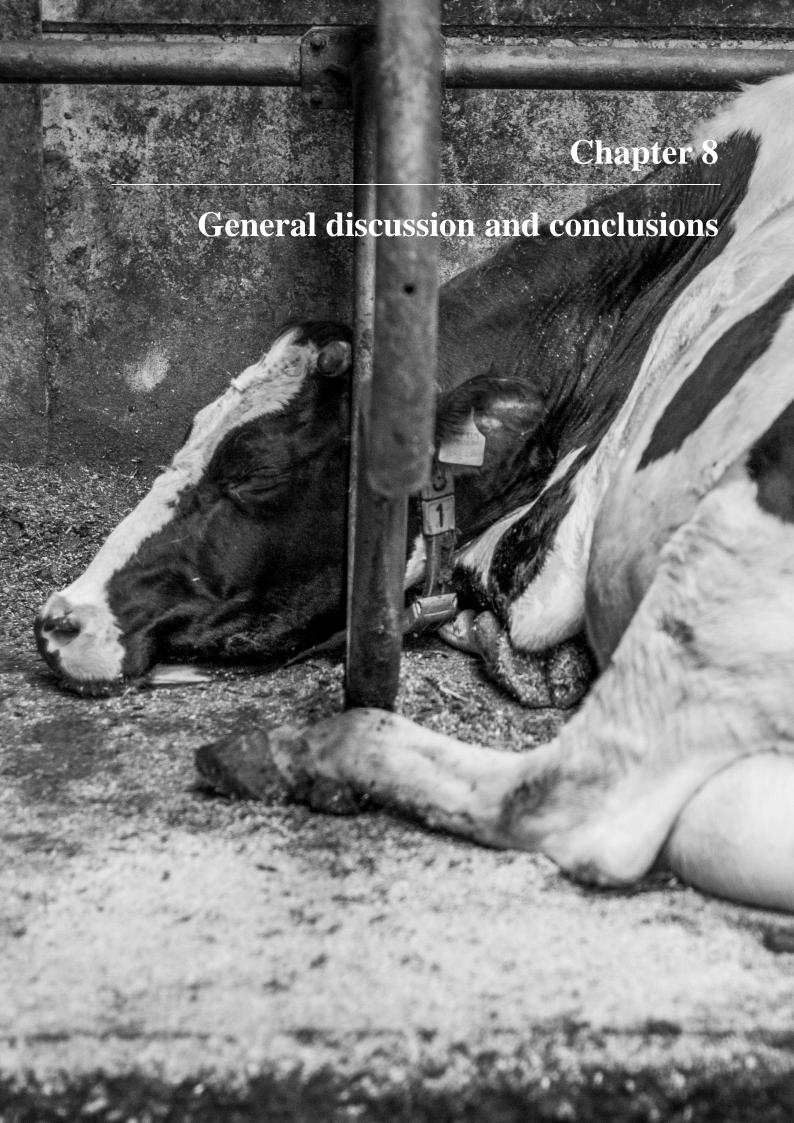
Farm technical and economic performance was evaluated using efficiency analysis. We chose DEA since it does not require 'ex ante' specification of the functional form of the production efficiency to be estimated. Efficiency scores are highly depending on the variables used in the model. Whereas earlier studies consider all inputs and outputs at the farm or enterprise level, inputs in this study were confined to feed and only milk as output was taken into account. This study included only feed inputs, as efficiency of nutrient use is a major factor affecting farm profitability on modern dairy farms, and feed accounts for about half of total expenses (Vandehaar, 1998). The significant positive correlation between gross margin (€ per cow) and cost efficiency indicates that the cost efficiency scores in this study can be a proxy for the economic performance of the farm. As the inputs 'pasture' and 'other roughage' were expressed in hectares, technical efficiency does not take into account differences in yields nor in the quality of roughage produced among farms. We might assume, however, that by working with farm-specific prices for the calculation of Cost Allocative Efficiency, differences in yield and quality of roughage are at least partially covered in the calculation of cost efficiency.

We additionally should emphasize that the current analysis of the relationship between animal welfare and efficiency scores remains exploratory. It is based on Spearman correlation analysis, revealing monotonous correlations between variables. Previous studies showed that the effect of some welfare measures on efficiency might be of a non-monotonous character (Allendorf and Wettemann, 2015, Roche et al., 2009). Roche et al. (2009) described the non-monotonous association between body condition score and milk yield, as they observed a positive association between body condition score and milk yield until an optimum body condition score, but a negative association thereafter. A similar non-monotonous association was observed between technical efficiency and cow losses, and between technical efficiency and SCC (Allendorf and Wettemann, 2015).

Nonetheless, besides the methodological considerations as described above, the relationship between animal welfare and farm economics, is presumably largely dependent on a number of confounding factors, as both animal welfare and farm economics are functions of multiple factors. A farms' performance depends on structural factors as farm size and milking systems, on production practices, but also on factors linked to the farmers' managerial capacities and genetic potential of the animals (Tauer and Mishra, 2006, Gloy et al., 2002, Mishra and Morehart, 2001). These factors may influence animal welfare as well, but possibly in a different way.

Conclusions

Exploring possible trade-offs or win-wins between animal welfare and farm economic performance suffers from counterbalancing effects of underlying factors. The final categorization issuing from the WQ protocol does not sufficiently and significantly differentiate between economic measures. A more continuous alternative for WQ may help, and in our case showed that win-win between economic and welfare is not to be excluded. In order to explore possible avenues for economic-animal welfare win-wins, a decomposition of these integrated scores into their underlying components appears necessary. This study already unravelled some mechanisms as challenges for further research, namely an endogenous counterbalancing of technical and Cost Allocative Efficiency, the two main determinants of cost efficiency, and the need for optimising concentrate use as a potential win-win between animal welfare and cost efficiency.

Chapter 8

General discussion and conclusions

In this final chapter, first, the relevance of the results for the various research objectives are discussed along with their limitations and future research opportunities (8.1., 8.2. and 8.3.). Second, implications of the findings across the six papers are discussed for a dairy cattle welfare assurance scheme (8.4). Finally, general conclusions are given (8.5).

8.1. RO1: Identifying a (more) valid and sensitive method to assess herd welfare that accurately reflects the welfare status on dairy farms

8.1.1. Revisiting research questions

Several methods for assessing herd welfare have been developed and published, of which the WQ protocols for welfare assessment of the main types of farm animals are probably the most renowned and most comprehensive. However, concerns have been raised about this protocol (discussed in Chapter 1), some of which have been investigated further in this doctoral thesis with a specific focus on the integration methods. During development of the WQ protocol, thresholds for WQ categorisation were determined based on 69 herds assessments (Botreau et al., 2009). By comparing intuitive welfare scores of observers for 44 of these herds with WQ categorisation, internal validity was checked. A significant relationship between these intuitive welfare scores and the WQ category was found. However, whether validity can also be assumed for other herds should be tested as well. In Chapter 2, the variation in sensitivity of integrated WQ outcomes (criteria and principle scores and overall welfare category) to extremely low and high values of the various welfare measures in a large European dataset of WQ assessments was examined, along with the reasons for this variation in sensitivity.

Results presented in Chapter 2 showed that extreme changes of some measures had a much larger impact on integrated scores than similar changes of other measures. Furthermore, some explanations for this variation in sensitivity have been suggested. Although most of these explanations were intended by the WQ protocol, two unintended side-effects of the integration method had a considerable influence on sensitivity too. Namely 1) the number of measures that are integrated into criteria and principle scores, and 2) the aggregation method of the measures. The complexity of the WQ integration methods (many measures, multiple steps, various different aggregation methods) probably causes these unintended side-effects. A simpler and uniform integration method for all measures would likely be more transparent and more sensitive to changes in individual measure values.

In Chapter 3, we investigated the balancing of welfare measures in the WQ protocol by examining whether integrated WQ scores were in accordance with expert opinion. Results show that the way in which WQ categorization reflects the relative importance of some welfare measures, is not in correspondence with selected experts' opinion. This lack of correspondence is probably because a different expert panel was employed than during the WQ project (i.e. no stakeholders like retailers and consumers) and likely because of the experience with the various measures which the consulted experts gained since introduction of the WQ protocol (discussed in detail in Chapter 3). One could argue that dairy cattle welfare experts, and to an even higher extent trained users of the WQ protocol, are most qualified to balance the relative importance of various welfare measures to assess herd welfare. Therefore, this poor correspondence between some WQ integrated outcomes and our specific group of experts' opinions questions whether these outcomes can be used to accurately reflect dairy cattle welfare on-farm. Valid welfare assessment is essential to an animal welfare assurance scheme because this will guide improvements that should positively affect dairy cattle welfare.

Findings of both Chapter 2 and Chapter 3 thus strengthen the growing evidence that the WQ integration method should be revised in order to accurately reflect herd welfare in a way that corresponds with the opinions of trained users of the WQ protocol for dairy cattle welfare. As results suggested that the highly complex integration procedure causes some unwanted side-effects we opted to develop an alternative and more simple integration protocol. Results are presented in Chapter 4, where a five-step procedure was presented that leads to a protocol for integration of measures of welfare problems which most severely affect dairy cattle welfare into a WI which is transparent and simple, multidimensional, discriminative and based upon, and checked for correspondence by, a selective panel of experts.

8.1.2. Limitations and future research opportunities

The two main inputs for all three research chapters were the European dataset of 491 dairy herd welfare assessments and the surveys of a selected group of experts. The European dataset can be considered as unique in that data on welfare measures collected using the same protocol in 10 different countries, from seven different research institutes were collated. This ensured a wide range for all WQ measures, which is beneficial for testing changes in measure values in a sensitivity analysis. In addition, this dataset allowed selected experts to benchmark results of

the focal farms in Chapter 3 and the fictitious farms in Chapter 4 based on a wide range of data, which supported them in allocating welfare scores. However, using a database from various different research institutes may also entail risks regarding inter-observer reliability. Although, the use of the same protocol to perform these measures may have limited the influence of different observers. In addition, the various different housing systems represented in the dataset, specifically tied versus loose housing, may have influenced results of the measures as well. However, the only measure which differs between these housing systems is the lameness score, which is integrated in the same manner. The selected group of experts can be considered unique as well, as all experts were experienced in performing the WQ protocol for dairy cow welfare. Therefore, they knew about any practical implications of the WQ measures which provides valuable information for balancing importance of these measures for herd welfare.

However, the selection of experts also forms a limitation. The criteria to which experts had to adhere to guarantee adequate knowledge about the WQ measures, strongly restricted the number of 'suitable' experts. In addition, to generate wide acceptance for an animal welfare assurance scheme, the integration protocol should be tested for correspondence with the opinion of other stakeholders as well. In WQ, the following stakeholders were involved in developing the integration methods (more specifically integration from principles into overall welfare categorisation): "animal scientists (recruited for their knowledge of animals and of the measures used for scoring), social scientists (recruited for their knowledge of expectations of societal groups), and stakeholders who would possibly be using the scoring system in the future (here: the Advisory committee of WQ on which sat representatives of farmers, breeders, retailers, veterinarians, animal protectors, and of institutions — EU Commission and OIE)" (Botreau et al., 2009, p.265). Acceptance by such stakeholders of an animal assessment scheme is important because they are the end-users of the scheme. Therefore, major dissimilarities between selected welfare measures and how these are balanced could decrease the likelihood that a welfare assurance scheme will be widely implemented. In future studies, compliance of the alternative integration protocol with a similar group of stakeholders could be examined. For this end, stakeholders' input could be elicited in multiple points of the five-step procedure to develop the integration method. First, stakeholders may be asked to perform the same exercise as the selected experts in Chapter 3 and 4, ranking WQ measures for importance. This may provide feedback based on which additional measures may be added to the current selection, and the relevance of these measures for stakeholders. Another option is to determine thresholds for the

various welfare categories, (partly) based on the opinions of stakeholders. For this end, the same approach as used for experts in this doctoral thesis could be employed where focus farms are scored by stakeholders and welfare scores compared with WI scores.

A limitation specific to Chapter 4 relates to the selection of welfare measures. First, only WQ measures were considered for inclusion in the alternative protocol. Some of these measures were previously criticised for their validity, reliability and/or feasibility (Tuyttens et al., 2015, Bokkers et al., 2012, de Vries et al., 2013b, Knierim and Winckler, 2009). The lowest ranked measures in Chapter 4, were also those criticised in earlier studies, namely measures of thirst and the QBA. Therefore, future studies could focus on replacing those measures which were considered not reliable, valid and/or feasible with alternatives and then repeating the exercise of ranking welfare measures. Second, in Chapter 4, it was determined that the selection of welfare measures (severe lameness, leanness, mortality rate, hairless patches, lesions/swellings and Somatic Cell Count > 400.000) would allow to detect the welfare problems which most severely impact dairy cattle welfare on-farm according to selected experts and literature (Nielsen et al., 2014). For specific purposes of welfare assessment (e.g. labelling for consumers) however, it could be desirable to detect additional welfare problems. Following Nielsen et al (2014) and EFSA (2012) important welfare problems which would not be detected specifically with the selected welfare measures are: reproductive disorders; thermal discomfort; pain, fear and frustration due to handling, milking, dehorning, tail docking and downer cows; abomasal displacement; respiratory distress; social stress; and prolonged thirst. In addition, positive welfare measures could be added to incorporate positive emotions, besides welfare impairments. Although some links between selected welfare measures and the abovementioned welfare problems may exist (e.g. abomasal displacement could influence a cows' body condition score), the selected measures would not be specific and sensitive enough to fully detect these welfare problems.

WQ has selected or developed animal-based measures for some of the welfare problems not detected by the selected welfare measures in Chapter 4. Reproductive disorders which directly influence dairy cattle welfare are detected using the WQ measures of dystocia and vulvar discharge. The effect of human-animal interactions on fearfulness in dairy cattle is measured in the WQ protocol using the avoidance distance test at the feed rack (ADF). Pain related to dehorning and tail docking is detected using a management-based measure, where the farmer

is asked about the methods used for these procedures. Using animal-based measures for these specific welfare problems is less obvious, as pain caused by these procedures is limited to the period in which they are performed, and this period rarely coincides with the moment of welfare assessment. Respiratory distress is measured in WQ by counting the amount of coughs per animal per 15 minutes during behavioural observations, and by assessing whether animals express hampered respiration during clinical scoring. Social stress is detected in WQ by quantifying the amount of head butts and displacements performed by cows during behavioural observations (Welfare Quality®, 2009). Below, the welfare problems for which no animal-based measure is described in WQ are discussed.

Thermal discomfort. Especially in warm climates, thermal discomfort may pose a threat to dairy cattle welfare. Previous studies have identified a Panting Score (PS) and Respiration Rate (RR) as feasible and valid methods to assess thermal discomfort in dairy cattle (Samborski and Van Bellegem, 2015, Schütz et al., 2014, Schütz et al., 2010, Van Laer et al., 2015). Such measures could be added to a dairy cattle welfare monitoring scheme. However, these measures are highly dependent upon outside temperature, which decreases the test-retest reliability of such measures. Increasing test-retest reliability of measures of thermal comfort, or finding manners how to deal with this remains a challenge for future research.

Abomasal displacement. Especially in high-producing dairy herds, displaced abomasum (DA) is a common disease (Cameron et al., 1998, Grymer, 1980). DA entails the abomasum (the fourth stomach of the cow) filling with gas and floating in the dorsal part of the abdomen. This condition may result in reduced feed intake (or even anorexia), drop in milk yield, discomfort, depression and sometimes even death (Van Winden and Kuiper, 2003). In dairy cattle welfare assurance schemes, occurrences of DA may be determined for a given herd using veterinary records of treatments of or death due to DA. However, whether these veterinary records are sufficiently reliable currently remains to be validated.

Prolonged thirst. Animal-based measures of thirst have been developed for poultry, such as blood sodium concentrations, plasma osmolality (Vanderhasselt et al., 2013, Reece, 2009,) and voluntary water consumption (Vanderhasselt et al., 2014, Sprenger et al., 2009,). Whereas blood parameters are too invasive and laborious to perform on-farm, it could be promising to develop voluntary water consumption tests further, and test for validity as a measure of

prolonged thirst in dairy cattle. During this test, as described by Sprenger et al., (2009), several groups of animals are enclosed within predefined locations and presented with an unfamiliar open drinker. The amount of water consumed from these drinkers is then measured after a given period of time (120 min. in Sprenger et al., 2009). When validating such a measure, the procedure as described by Vanderhasselt (2013) for broilers may be used, where water consumption was measured before and after the animals were subjected to water deprivation for a given period of time.

Positive welfare measures. When assessing overall welfare, a component lacking in the current proposed set of welfare measures is the assessment of positive emotions. Some argue that welfare not only comprises of the absence of negative experiences, but also of the presence of positive experiences (Boissy et al., 2007, Spruijt et al., 2001, Duncan, 1996, Fraser, 1995,). Moreover, one of the main symptoms of depression is anhedonia (i.e. the inability to experience pleasure) (Willner, 1997, Fawcett et al., 1983). In this train of thought, an overall welfare assessment scheme should entail assessment of positive welfare. In a review by Boissy et al. (2007), behavioural measures of positive emotions which would be feasible to apply on-farm are suggested to be play behaviour, affiliative behaviour and vocalisations. As play behaviour occurs mostly in juvenile animals, this measure is less applicable to dairy cows. Affiliative behaviour however, is expressed by dairy cattle, more specifically allogrooming (social grooming between members of the same species). Allogrooming has been found to strengthen social bonds in herds of dairy cattle and to reduce heart rate in receivers (Sato et al., 1993). However, allogrooming in other species (more specifically horses) was found to be influenced by confounding factors such as season (Kimura, 1998) and reproductive state (Van Dierendonck et al., 2004). Additionally, in pigs allogrooming was found to increase in barren situations (de Jong et al., 1998) and positive effects may differ between receiver and performer (Laister et al., 2011). Therefore, the validity of social grooming as an measure of positive emotional state in dairy cattle should be studied further. Vocalisations have previously been used primarily to indicate negative emotions, although they hold promise for assessing positive emotions in farm animals as well. In sheep some work already shows a specific low-pitched bleating to be associated with positive situations like mothers nursing their lambs (Fisher and Matthews, 2001). As vocalisations can be registered using recording devices and assessed using sound analysis like performed for stress vocalisations in pigs by Schön et al. (2004), positive vocalisations hold potential in measuring positive emotional states on-farm.

8.2.   RO2: Examining the (existence of) the market for animal-friendly milk

8.2.1. Revisiting research questions

The Flemish market for animal-friendly milk was studied in two chapters. Chapter 5 aimed at investigating factors that determine consumer intention to purchase animal-friendly milk. In particular, the influence of citizens' attitudes, perceptions of product-related factors, and milk consumption on intention to purchase animal-friendly milk. Chapter 6 aimed at identifying market opportunities and at profiling consumers as a basis for marketing and communication strategies about animal-friendly milk. As Chapter 5 (in accordance with literature) identified a high degree of variation between consumers in the intention to purchase animal-friendly milk, different consumer segments were studied in the ensuing Chapter 6.

In both chapters, a high degree of differentiation was shown within the consumer market for animal-friendly milk. In Chapter 5, the variation in intention to purchase animal-friendly milk was found to be (partly) explained by the attitude of consumers towards dairy cattle welfare (positively influenced by a more natural living oriented attitude and negatively by a more business-oriented attitude) and by their attitude towards milk. While attitudes can be changed (Malotte et al., 2000), this would not be the aim of an animal welfare label in the short-term. Rather, such labels would use existing consumer attitudes towards dairy cattle welfare to position animal-friendly products in the market.

From findings of Chapter 5 and Chapter 6, we conclude that marketing animal-friendly milk at several levels of animal welfare and price could reach a wide group of consumers. Although possibly somewhat overstated due to a hypothetical bias and a social desirability bias (further discussed under 8.2.2), the cluster with high market opportunities formed 8% and the cluster with moderate opportunities 48% of milk consumers involved in the purchase decision. This indicates the theoretical existence of a large potential consumer market for animal-friendly milk. Markets for milk perceived by consumers as of enhanced welfare (organic milk) may be higher than the supply in Europe and Belgium. Bioforum Vlaanderen (2015, the Flemish sector organisation for organic farming and food) states that within Belgium, sufficient market demand exists for a doubling of the Belgian organic milk supply. It is possible that part of this market demand could also be filled by farms with high levels of welfare, as one of the incentives

for purchasing organic products is an increased level of farm animal welfare (Harper and Makatouni, 2002). Consumers who expressed a higher intention to purchase animal-friendly milk (8% of the study sample), were found to perceive animal welfare as a very important product attribute, were highly interested in information about dairy cattle welfare and had a more natural living oriented attitude towards dairy cattle welfare. For these consumers, animal-friendly milk of the highest level of animal welfare and price could be positioned by purely focusing on enhanced welfare.

Other consumer segments with moderate or even low market opportunities (48% and 40% of the sample respectively) may consider animal welfare to be an important product attribute but not dominant to other product attributes such as price, taste or quality. Such consumer segments were found to be less interested in animal welfare and more sensitive to price, given their lower reported WTP and the finding that intention to purchase animal-friendly milk was negatively influenced by a higher personal importance attached to price. Consumers in these segments would therefore likely be more willing to purchase intermediate products. Intermediate products would compromise between animal welfare and price where a higher level of animal welfare is associated with a higher price (de Jonge et al., 2015, van Herpen et al., 2015). Following results of Chapter 5, consumer segments with lower market opportunities might be more willing to purchase animal-friendly milk when benefits like health benefits or premium quality of such milk would be underlined. Such benefit-based positioning has been shown previously to more effectively increase consumer choice for these products than positioning strategies which focus on a single concrete feature of a product (Fuchs and Diamantopoulos, 2010). However, when claiming that such benefits exists, scientific research should support these statements. It remains to be tested whether milk of farms with a welfare label would be of higher quality, or would be perceived as tasting better by consumers. Van Herpen et al., (2015) studied how to increase the choice for intermediate meat products by means of product positioning and product display in the supermarket. They found that products with the highest level of price and welfare could best be displayed in a separate section to increasing the purchase intention for these products. For intermediate products, a mixed display would more effectively increasing product choice, allowing consumers to compare between conventional and intermediate products. Future studies could focus on whether these results can be applied to differentiated animal-friendly milk products as well.

8.2.2. Limitations and future research opportunities

In both Chapter 5 and 6, online surveys were used. The main limitation of such a method is that choices which subjects make, remain purely hypothetical. This means that purchase intention might be overstated due to a social desirability bias, and not represent actual consumer behaviour (Fisher, 1993). To examine whether results of the current study are robust in the situation of an actual purchase, non-hypothetical choice experiments could be used. Such methods place the participants in a more realistic setting and force them to make evaluative choices and trade-offs between multiple product attributes (de Jonge and van Trijp, 2013, Olynk et al., 2010). An example is the 'Vickrey auction' (Vickrey, 1961), where individuals are asked to submit a bid of the maximum price they would pay for a given product. Subsequently, the winner is the highest bidder who has to pay the second highest price for the product. This method has previously been used in studies on the influence of information about animal welfare and WTP for products (Liljenstolpe, 2008, Napolitano et al . 2008, Lagerkvist et al., 2006). Building further on findings of this doctoral thesis, future studies could test the different positioning strategies and influence on WTP of different consumer segments.

Another limitation of the research methodology employed in this doctoral thesis, pertains to the study sample. The sample was representative for age, region and gender, but there was a slight sample bias towards higher educated consumers and only Flemish consumers were surveyed. Belgian, and therefore Flemish, milk farmers export milk as well, mostly to other EU countries namely Germany (41% of export in 2015), France (27%) and the Netherlands (20%) (VLAM, 2016). Therefore, consumer needs in those countries are important in determining Flemish farmers' strategies as well. For Dutch, German and French consumers, previous studies have also identified a differentiated market for animal-friendly products (de Jonge and Van Trijp, 2014, Nocella et al., 2010, Krystallis et al., 2009). However, extrapolation of research results like e.g. interest in dairy cattle welfare measures to these countries, remains to be validated.

Chapter 5 is based on the Theory of Planned Behaviour (TPB) which identifies attitudes, subjective norms and perceived behavioural control as determinants of purchase intention (Ajzen, 1985). Although in Chapter 5, attitudinal determinants of the purchase intention were the main focus, future studies could include subjective norms and perceived behavioural control as well.

8.3. RO3: Examining the relationship between farm economic performance and dairy cattle welfare

8.3.1. Revisiting the research question

For farmers to create value from animal welfare improvements on their farm, two main options are possible. The first option follows from the results of the previous research objective and searches for added value from product differentiation based on enhanced animal welfare. The second option is to explore possibilities of a link between an enhanced level of animal welfare and farm performance. The latter was the focus of Chapter 7 where the relationship between dairy cattle welfare and farm performance was investigated using integrated productive efficiency techniques. The efficiency analysis integrates physical productivities (litres of milk) expressed relative to another input with optimisation of input substitution, measured by technical and Cost Allocative Efficiency. This gives a more comprehensive diagnosis of farm performance, additional to the mere gross margin and profitability analysis.

Previous studies have indicated that relationships between welfare measures and farm efficiency may differ depending on the measures involved. For instance, in a study by Allendorf and Wettemann (2015), % of cow losses was negatively associated with efficiency up to a given point, while a high SCC was positively associated with efficiency. Moreover, in the same study a non-linear relationship between several individual welfare measures and efficiency was found. For example, up to a certain point, higher % of cow losses was associated with higher efficiency scores, while above this point increasing mortality had an adverse effect on efficiency. These literature findings imply some warnings for the research objective. Due to the complex relationships, it is possible that the overall link between an integrated welfare score and integrated cost efficiency scores ends up neutral.

In Chapter 7, for a sample of 41 Flemish dairy farms, the two components of Cost Efficiency (CE), Technical Efficiency (TE) and Cost Allocative Efficiency (CAE) were compared with WQ categorisation and the alternative WI, along with their underlying measures. CE was significantly correlated with gross margin, but the decomposition revealed a negative correlation between both TE and CAE. Although TE was positively correlated with CE, which CAE was not, the negative correlation between both components of CE indicate that improving TE can also lead to a decreased CAE. More, correlation between these CE measures and

integrated welfare scores were absent, and indication of either trade-offs or win-wins were only found on the individual measures. Whereas positive relationships were found between some specific welfare measures and TE, only negative relationships were found between CAE and some welfare measures of comfort around resting, which appears to be somewhat lower on farms with a higher CAE.

Results indicate that a high level of CE and an enhanced level of animal welfare could be combined in one farm. Stimulated by an increased welfare on-farm and access to a differentiated market which would allow farms to distinguish themselves from others, farmers might therefore be able to and willing to employ welfare improvements on-farm.

8.3.2. Limitations and future research opportunities

One major limitation to the current research design pertains to the usage of only feed as input for the cost efficiency analysis. Although feed accounts for about half of total costs on dairy farms (Vandehaar, 1998), future studies should include additional inputs such as labour and capital to gain a more complete image of farm efficiency. Both inputs are, however, difficult to measure and to compare within farms e.g. due to non-standardized labour registration systems and/or depreciation schemes. Taking those inputs into account that are easily adaptable in the short run by farmers, like feed in the current study, is already a first step in unravelling possible management factors suitable for stimulating farmers to invest in animal welfare. As welfare improvements on-farm do not always have to be costly (Appleby, 2005), possibilities lie in welfare improvements which do not imply costly changes to farm resources or management.

From Chapter 7, we concluded that linking both aggregated approaches of animal welfare and farm economics concealed some underlying relationships, so it was not possible to detect a clear win-win nor a clear trade-off. Because of the limited study sample (n = 41 farms), however, the existence of win-wins or trade-offs cannot be ruled out. When measured with WQ, the best performing farms on animal welfare ('enhanced' welfare category) showed a somewhat lower total milk production and no differences in efficiency scores compared to farms of the 'acceptable' welfare category. This result indicates a possible trade-off between animal welfare and farm performance. When measured with the alternative WI, some positive links were found, which indicates a possible win-win situation. This highlights the importance of the animal

welfare integration method and the need to have insights in the links of separate measures with economics.

Besides the intrinsic economic evaluation of market differentiation and of influence of animal welfare on farm performance, one must also take into account extra costs associated with an animal welfare assurance scheme. The European Commission (D.G. SANCO, 2009) identified two different types of costs associated with food labelling: certification costs and production costs. Certification costs are paid by farms to receive a certificate, and consists of e.g. time spent on documentation, external advisory service and operating costs like auditing and member fees. Production costs are those costs incurred to meet the requirements of a welfare assurance scheme. For farms to reach an excellent welfare category, investments may be needed in improved housing conditions (cubicles, flooring, lighting, water supply, etc.) or by managerial changes which might imply training of staff and/or additional time costs. Variable production costs may entail additional sampling (e.g. individual SCC testing), more veterinary supervision, more labour and possible influences on farm performance. For the specific situation of a label for animal-friendly milk with different levels of welfare and price, future studies could focus on mapping all associated costs to evaluate the balance between costs and benefits (Belletti et al., 2007).

8.4. Implications of the current results for a dairy cattle welfare assurance scheme

Results on some specific topics of a WAS were found across research chapters in this thesis, these specific findings are discussed in the paragraph underneath. Furthermore, the general implications of research findings for the introduction of a WAS are discussed in the ensuing paragraph.

8.4.1. Implications of specific findings across research chapters

Using an animal welfare assurance scheme, added value can be achieved for all stakeholders when the welfare assessment method allows for an accurate assessment of welfare on farms, but communication about welfare status on these farms should also address consumer concerns about welfare. In this doctoral thesis, there was poor correspondence between which WQ measures were ranked as most important for dairy cattle herd welfare by experts (Chapter 3 and Chapter 4) and what measures consumers indicated to find most interesting to make an informed

purchase decision (Chapter 6). Whereas the measures 'access to pasture', 'positive emotional state' and 'negative emotional state' were scored highest of interest by most consumer segments, the WQ measures lameness, mortality, integument alterations and leanness were ranked highest by experts. So, measures which selected experts indicated to most severely affect herd welfare, are not necessarily most interesting to society. The WQ criteria which were allocated the highest interest score by consumers were 'absence of diseases' and 'freedom of movement' (the latter of which may have been associated with access to pasture). Therefore, incorporating measures of health in a labelling scheme might be interesting to consumers but communicating about the underlying measures specifically is probably not. A label underlining the 'healthiness' of cows may be interesting for a large part of consumer segments. This would also be consistent with studies involving farmers, where health (absence of diseases) was indicated among the most important WQ criteria by Flemish farmers (Tuyttens et al., 2010) and by a broader group of European farmers (Averäs et al., 2013) to most effectively increase animal welfare.

Outdoor access is consistently mentioned by citizens as one of the most important factors for farm animal welfare. This has been shown across species (Vanhonacker et al., 2008), and for specific farm animal species like chickens (Van Loo et al., 2014, Vanhonacker and Verbeke, 2009) and pigs (Meuwissen et al., 2007, Dransfield et al., 2005). This high importance of outdoor access might be because citizens associate this with a more natural living environment. In addition, outdoor access is a very tangible and visible aspect. Therefore, even consumers with very limited knowledge on livestock can relate to the picture of cows on pasture. However, access to pasture is a resource-based measure and was not mentioned by experts among the most important welfare measures. In studies where lack of access to pasture is treated as a risk factor for impaired welfare, predominantly positive effects of access to pasture on (aspects of) welfare are found. For instance, a lower prevalence of lameness (de Vries et al., 2015, Onyiro and Brotherstone, 2008), a lower incidence of clinical mastitis (Washburn et al., 2002), and a lower mortality rate (Burow et al., 2011) were observed in cows with access to pasture. In addition, Burow et al. (2013) used an aggregated welfare index to assess the welfare of loose housed Danish herds (all with access to pasture in summer) during the indoor and outdoor periods. They concluded that welfare was better during summer than during winter. However, it is possible that herds housed in systems which take into account the animals' needs but where the cows are not granted access to pasture, could score high on the WI score proposed in this

thesis. Consumers, however, might feel that without outdoor access welfare cannot be enhanced. This is a contraposition between citizens and animal welfare experts which could be solved in different ways. First, access to pasture could be added as a measure (possibly expressed as the % of days/year the animals are not granted outdoor access), but this measure may receive a low severity score by the experts. Moreover, factors such as pasture quality (e.g. poorly drained pasture may negatively influence claw health) are confounding to the influence of access to pasture on dairy cattle welfare. Therefore, it might be possible that zero-grazing herds are categorized as having enhanced or even excellent welfare and that farms with access to pasture reach the not classified category. The second solution is that access to pasture may be used as a prerequisite to a labelling scheme. However, this implies the exclusion of farms where welfare may be excellent, regardless of their zero-grazing policy. Given the trend of increased zero-grazing among Belgian dairy farms (Schellekens and Van de Ven, 2008), such a prerequisite may strongly limit the number of farms included in the animal welfare assurance scheme and thus the amount of farmers which are stimulated to improve welfare on-farm. Both scenarios, and the option of not including access to pasture as a measure, should be discussed among stakeholders and animal welfare scientists when developing a dairy cattle welfare assurance scheme.

Findings that access to pasture influences dairy cattle welfare, have implications for timing of welfare assessments. Moreover, differences in individual WQ measures were observed in a study where dairy cattle welfare was assessed at the beginning and at the end of the indoor period (de Graaf et al., in press). Therefore, test-retest reliability of the proposed welfare index should be determined to assess whether fluctuations in welfare score occur during the year. When such fluctuation of the WI score during the year exists, the time period in which the welfare assessments have to be performed in order to accurately compare between and within farms might be very limited. This influences the feasibility of an animal welfare monitoring scheme.

In terms of associated costs with an animal welfare assurance scheme, feasibility is one of the key components for a welfare assessment method. One of the main restraints of performing animal-based measures on-farm, however, is their time-consuming nature (Blokhuis et al., 2010). Consequently, WQ is criticised for its time-consuming nature (de Vries et al., 2013, Knierim and Winckler, 2009). De Vries et al. (2013) found that mean total assessment time of

the WQ protocol was 381 minutes. When the selection of welfare measures of the proposed Alternative Welfare Index is used, only clinical observations are needed to measure lameness, leanness and integument alterations. Mortality is routinely collected on all Flemish dairy farms and can therefore be collected off-farm. Individual cow SCC data is collected on about 45% of Flemish dairy farms (G. Haesebeyt, CRV, personal communication). Knierim and Winckler (2009) stated that the entire clinical observation (including other health-measures like hampered respiration and ocular and nasal discharge) takes on average 3 min/cow. This means that for a herd of 50 cows, sampling 33 cows (according to WQ guidelines) would take 99 minutes. For a herd of 100 cows, 49 would be sampled which would take 147 minutes and for a herd of 200 cows, sampling 65 of them would take 195 minutes. This would mean a considerable reduction of assessment time, and would imply that multiple farms could be visited in one day.

The often time-consuming nature of animal-based measures should also be taken into account when adding measures to the proposed WI. Example given, behavioural observations have been found as time-consuming as clinical observations by de Vries et al. (2013), taking 150 min. on average. Therefore, collecting data on fearful behaviour or positive welfare measures like allogrooming, would mean a vast increase in assessment time of the protocol. This should be a considered when determining if and which measures should be added to the protocol.

The WI of the proposed assessment protocol was currently interpreted using thresholds for four welfare categories following WQ. These categories may be used in communication with consumers for intermediate milk ('enhanced welfare') and milk of the highest level of welfare and price ('excellent welfare'). When applying these thresholds to the WQ European dataset (n = 491 herd assessments), 41.6% of farms would be categorized as 'Enhanced' and 1.2% as 'Excellent' (Figure 8.1). In particular, for the Flemish dataset, 58.7% of herds would be categorized as 'Enhanced' and none of the farms as 'Excellent'.
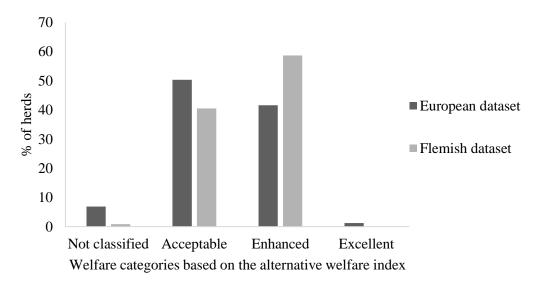
Figure 8.1: Percentage of herds per welfare category based on the Alternative Welfare Index for the European dataset (n = 491) and the Flemish dataset (n = 121)

These Flemish Enhanced and Excellent dairy farms would be able to differentiate themselves based on the proposed WI score, without further investing in welfare improvements. As the current sample may not be representative for the entire population, actual distribution of Flemish dairy farms across the welfare categories may differ. However, current results indicate that dairy farms which could differentiate themselves based on animal welfare exist among the Flemish population. Future studies could focus on using a representative sample of Flemish dairy farms to assess distribution across welfare categories.

In Chapter 6, across consumer segments the most trusted sources of an animal-friendly milk label were veterinarians and (non-governmental) animal welfare organisations. This means that a similar label as the Dutch 'Beter Leven Keurmerk', and the United Kingdoms' 'RSPCA assured' label may be successful in Flanders as well. Most similar animal welfare organisation compared to the Dierenbescherming and the RSPCA in Flanders would be 'GAIA' (http://www.gaia.be/), although they may be perceived as somewhat more extreme in their views regarding farm animal welfare. Alternatively, a government-managed label is an option as well. An example of such a label managed by the government has recently been launched in Denmark by the Ministry of Environment and Food and is entitled 'Bedre Dyrevelfaerd' (literally translated: 'better animal welfare'). This voluntary welfare label is currently only focused on pigs, has three welfare levels and states that requirements for attaining the label are

stricter than the current legislator requirements in Denmark and in the EU (more information at www.foedevarestyrelsen.dk). Welfare labels which employ welfare assessments using mainly based animal-based measures are rare. However some concrete initiatives exist, for example a label which plans to use the WQ protocol is 'Free Cow's Milk' from the Finnish dairy company 'Juustoportti Foods Oy'.

8.4.2. Implications for credibility, implementation and costs

This doctoral thesis aims to contribute to the greater goal of improving farm animal welfare, by examining possibilities to implement a WAS in Flanders, Belgium. Animal welfare cannot be viewed separately from needs of society for affordable, healthy and sustainable food production, and from the need of farmers for a profitable business. Therefore, a WAS will only be successful when it meets (or forms a compromise between) the needs of all stakeholders. In the general introduction, three mutually linked factors were identified for the success of a WAS, namely credibility, implementation and costs. Results of all six research chapters are discussed underneath to highlight these factors, what the results of this thesis have contributed and priorities for further research.

Credibility. The welfare assessment which is used to monitor animal welfare in a WAS should be valid for such a scheme to be credible, and to guarantee a certain level of animal welfare. In this thesis, it was found that welfare assessment using the current WQ protocol does not accurately reflect the level of dairy cattle welfare in a way that corresponds to the opinion of trained users of the same protocol. This was found to be at least partly due to the highly complex WQ integration method. Therefore, an alternative welfare assessment protocol was proposed in this thesis which is transparent and simple, multidimensional, sensitive to individual measures and which corresponds with the opinion of selected experts. This welfare index allows the provision of general information about dairy cattle welfare for consumers (overall welfare index, categorized), and specific information for farmers (welfare measures). However, there are aspects of welfare which are not measured with the limited selection of welfare measures in the alternative welfare index. To provide information about these missing welfare aspects welfare measures may be added welfare, as mentioned in paragraph 8.1.2.

In addition to the need for additional welfare measures, the welfare assessment as proposed in Chapter 4 needs to be tested for credibility as perceived by consumers, its users and other stakeholders. When the overall welfare assessment based on the selected experts does not match with the assessment of these stakeholders, one solution may be to leave the thresholds for different levels of welfare up to these stakeholders. This could mean that for different uses of the protocol, different thresholds are employed adjusted to the demands of for example consumers, retailers or the government for the level of animal welfare. Besides overall welfare assessment, the selection of welfare measures (e.g. omission of the measure 'access to pasture' as discussed in paragraph 8.4.1) may not be in agreement with the opinions of stakeholders. In this case, strategic decisions need to be made in order to balance credibility, implementation and costs. These decisions may depend on the reason for performing a welfare assessment (e.g. for a consumer label or legislation), but should never undermine the validity of the assessment.

Results presented in Chapter 6 of this thesis show that the dairy industry was trusted least as source of information about dairy cattle welfare by consumers, which indicates that such a label would have credibility issues. Animal welfare organisation and veterinarians were trusted most. Although veterinarians are likely not in the position to initiate an animal welfare label, animal welfare organisation would be able to, like the RPSCA in the UK and the Dierenbescherming in the Netherlands have previously.

Implementation. An important prerequisite for implementation of a WAS when aiming for continuous increased welfare on-farm, is that farmers are motivated to cooperate. In Figure 1.1. three possible motivating factors (added value) were illustrated for farmers. The first is an improved level of animal welfare on-farm. The general welfare assessment should therefore be sensitive for improvement or deterioration in individual welfare measures. Farmers will then see influence of the changes they made on their farms on the welfare score. The second motivation identified in Figure 1.1. was improved farm performance. Results from Chapter 7 showed that a high level of animal welfare may be combined with a high level of cost-efficiency. Investigating the links between welfare-remedial actions and farm performance could show where win-win situations may occur. The third motivator for farmers was access to a differentiated consumer market. Results from Chapter 5 and 6 indicated the existence of such a segmented consumer market for animal-friendly milk in Flanders, Belgium.

Another important implementation aspect of a WAS pertains to welfare monitoring. Factors which influence credibility and price are the frequency and timing of the assessments as well as the training of auditors. Regarding timing, previous studies identified that access to pasture influences outcomes of welfare measures (de Graaf et al., in press, Burow et al., 2013, Corrazin et al., 2010), which complicates determining the time-period in which welfare assessment may be performed. Besides timing, frequency of farm visits to assure a certain level of welfare on-farm remains to be determined. Blokhuis et al. (2010) stated that risk factor analysis would allow identification of problem areas and determine the required frequency of farm visits. In addition, longitudinal studies (e.g. using routinely collected data) would allow to investigate the degree of fluctuation of welfare state. As auditors are of critical importance for the reliability of welfare assessments, their competency and credibility is of high importance. There is a need for an uniform and robust training followed by an evaluation of their performance. Besides, re-evaluation to ensure a sustained level of inter- and intra-observer reliability is necessary (Blokhuis et al., 2010, Butterworth, 2009). Auditor drift (i.e. changes in the way which measures are used over time) may occur which could decrease the reliability within and between auditors. However, similar to the welfare assessments themselves, the necessary frequency of auditor evaluations is not determined. Therefore, further studies on frequency and timing of farm visits, and training of auditors is warranted to assure that welfare monitoring is reliable and allows for comparisons within and between farms.

Communication of general information about overall welfare to consumers is a third important aspect for implementation of a WAS. In chapters 5 and 6, animal health, access to pasture and perceived advantages were found to be promising for communication towards consumers. Future market research can build from these finding, in creating marketing and positioning strategies for animal-friendly milk. A prerequisite to communicating perceived advantages (e.g. healthier and better quality) of animal-friendly milk is determining whether these actually exist. For this end, taste tests and quality assessments may be performed using milk originating from herds with different levels of overall welfare.

Besides farm economics and animal welfare, environmental sustainability is an important topic of discussion in dairy farming and in society as well (Belgische Confederatie van de Zuivelindustrie, 2016, Capper et al., 2008). In Chapter 5, consumers' intention to purchase animal-friendly milk was positively associated with perceived importance of extrinsic product

attributes like sustainability and local production. This indicates that Flemish consumers may be interested in a combination of animal welfare and environmental sustainability. To achieve a sustainable dairy sector, it needs to be determined whether a high level of farm performance can co-exist with a high level of animal welfare and environmental sustainability on farms. For this end, animal welfare may be incorporated as one of the measures (or set of measures) into sustainability assessment tools. An example of such a tool is the 'RISE' tool, which is being used in Denmark to assess sustainability performance of organic farms (de Olde et al., 2016). However, as trade-offs exist between different measures in such a sustainability assessment tool (Schader et al., 2014), future research is needed on how to position animal welfare herein.

Costs. Regarding costs of welfare monitoring, animal-based measures generally consume more time and are therefore more costly as compared to input-based measures (Blokhuis et al., 2010). This is of importance when considering adding welfare measures to the welfare index which was proposed in Chapter 4. Previous WAS therefore opted for input-based measures of housing and management, while the validity of these types of measures cannot always be guaranteed. There are two possible solutions for this problem. The first is a protocol consisting of two versions: an elaborate protocol for which measures may be added to the current selection and the protocol as proposed in Chapter 4. This set-up would allow for specific information to be communicated to the farmer about welfare problems on-farm, but only when assessment using the reduced, and therefore more cost-efficient, protocol warrants this. The second solution may be the use of routinely collected data. The proposed set of welfare measures in Chapter 4 contains some measures that are already routinely collected in Flanders (mortality and somatic cell count). However, data on lameness and integument alterations is currently not routinely collected. Such data would be highly valuable, both for cost-efficient welfare monitoring and epidemiological studies on these welfare problems.

It is often expected that extra costs which are associated with a WAS are, at least partially, paid by consumers. Therefore, the size of the consumer market along with their WTP is of importance for the success of a WAS. In chapter 6, a segmented consumer market was identified which would allow for product differentiation. Milk originating from herds with different levels of animal welfare may therefore be marketed at different price levels. As an increasing level of price would likely attract a lower number of consumers, price setting of such milk should be investigated in a non-hypothetical setting. This would allow for an estimation of which part of

the costs associated with a WAS may be compensated by consumers, and which part needs to be compensated for otherwise.

In Chapter 7, no clear relationship between overall welfare score and cost efficiency as a measure of farm performance was found. This means that there are no indications that improving animal welfare will lead to better farm performance, but it also means that high levels of both may occur on one farm. In addition, findings do not rule out that there are procedures to remediate welfare problems which may also increase the level of welfare. Furthermore, the added value for the farmer may be found in the improved level of animal welfare on his farm and access to a differentiated market.

8.5. General conclusions

Results presented in this doctoral thesis have shown that animal welfare has the potential to provide dairy farms with options to differentiate products or to stay within a cost minimizing approach and therewith increase farm efficiency. In addition, a WI which is simple, intuitive, transparent and which detects the most severe welfare problems according to dairy cattle welfare experts was developed. Future research could focus on identifying to what degree the proposed welfare index would be widely accepted by several stakeholders, among which consumers. In addition, future research needs to determine whether the proposed selection of welfare measures should be complemented with additional welfare measures.

A heterogeneous market for animal-friendly milk was identified, for which healthiness of the cows and outdoor access are promising selling propositions from the consumer point-of-view. Product differentiation could be achieved by means of, example given, a rating system differentiating between different levels of animal welfare (enhanced or excellent) at different price levels. Products from these different levels of price and welfare could be positioned differently in the market, where products with excellent levels of welfare could focus primarily on animal welfare, and lower levels (enhanced) could focus on benefits of animal-friendly milk. Access to such a differentiated market, along with an improved welfare of cows in the herd, could allow for and stimulate farmers to implement welfare improvements.

The general conclusion from this doctoral thesis is that the introduction of a WAS for dairy cattle is possible in Flanders. However, the success of a WAS will depend on the strategic decisions which will determine the balance between credibility, implementation and costs. The approach in this thesis could be extended to investigate other species of farm animals, regarding heterogeneity in the consumer market, development of a WI using the proposed five-step methodology and examining the relationship between animal welfare and farm efficiency.

Appendices

Appendix A: Sensitivity of overall welfare category to replacement of individual measures of the criterion 'absence of disease' (Chapter 2)

Percentages of herds[1] (n = 491) that were downgraded and upgraded by 1 or 2 overall welfare categories when individual values at measure level within the criterion 'absence of disease' were replaced with theoretical worst and best values per measure

Measures	Observed worst score	Observed best score
	% downgraded 1 category	% upgraded 1 category
Number of coughs/cow/minute	2	0
% cows with nasal discharge	2	0
% cows with ocular discharge	2	0
% cows with hampered respiration	1	0
% cows with diarrhoea	2	0
% cows with vulvar discharge	2	0
% cows with SCC >400.000	2	1
% cows mortality	2	1
% calvings with dystocia	1	0
% downer cows	1	1

[1]Percentages were based on the herds that were actually able to shift one or two categories. For downgrades of one category n = 482, for downgrades of two categories n = 174. For upgrades of one category n = 491.

Appendix B: The items for attitudes towards dairy cows as used in the current study, adjusted from Austin et al. (2005) (Chapter 5)

Items
Cows have to be kept in an environment that is as natural as possible
It is important for cows to be able to express natural behaviour
Seeing a neglected cow doesn't affect me as much as it would affect most people
The idea of a 'natural environment' applies to farm animals as well as wild animals
Production efficiency should be first priority of the farmer
A farmer should think of his/her cows mainly in terms of the profit they will bring in
A farmer should think of his/her cows mainly in terms of their market value or cost
A farmer should view his/her animals as useful objects with which he/she could gain profit
I tend to think of cows as being very similar to machines
A cow that is physically fit experiences good welfare by definition
If a cow is reproducing efficiently her welfare standards must be good
If a cow is growing well she must be experiencing good welfare
A cow that is physically healthy cannot be suffering

Appendix C: Mean scores of interest in information about WQ measures and criteria for the entire sample and per cluster (Chapter 6)

WQ Criteria and WQ measures	Sample	CL1	CL2	CL3	CL4	CL5	CL6
Absence of prolonged hunger	57.58	68.70[c]	91.73[d]	63.83[c]	27.21[a]	46.02[b]	40.74[ab]
The number of cows on a farm that are too thin	50.95	59.71[b]	84.23[c]	57.50[b]	25.64[a]	36.38[a]	37.74[a]
Absence of prolonged thirst	57.80	68.82[b]	90.23[c]	64.89[b]	29.82[a]	46.33[a]	40.40[a]
Number of drinking points, cleanliness and functioning of the drinking points	56.40	66.51[b]	89.41[c]	62.68[b]	27.64[a]	44.84[a]	40.69[a]
Comfort around resting	58.56	70.19[c]	90.53[d]	66.10[c]	25.82[a]	45.94[b]	41.66[b]
Number of cows on a farm colliding with housing equipment when they lie down	43.22	46.96[b]	79.53[c]	48.57[b]	23.64[a]	31.17[a]	32.77[a]
The number of cows on a farm that lie down on places that are not meant for lying (indication of a dirty or unsuitable lying area)	50.90	58.76[b]	88.28[c]	56.43[b]	25.14[a]	38.44[a]	36.90[a]
The number of cows on the farm with dirty legs, udder or flanks	48.32	54.91[b]	78.61[c]	55.13[b]	23.64[a]	34.03[a]	37.83[a]
The time that it takes for a cow to lie down as indicator of pain or fear of colliding with housing equipment	47.77	54.11[b]	83.59[c]	53.41[b]	27.61[a]	33.85[a]	35.87[a]
Freedom of movement	65.22	78.43[c]	93.44[d]	72.89[c]	33.25[a]	58.62[b]	43.78[a]
Housing where the cows are tied instead of able to walk around freely	57.93	67.11[b]	92.23[c]	64.72[b]	32.32[a]	47.31[a]	41.13[a]
Absence of injuries	57.88	68.27[c]	86.97[d]	64.89[c]	27.43[a]	47.25[b]	42.29[ab]
The number of lame cows on the farm	53.11	60.05[b]	88.67[c]	59.37[b]	32.43[a]	39.90[a]	39.69[a]
The number of cows on a farm with hairless patches. cuts and/or swellings	54.23	62.09[b]	88.59[c]	61.14[b]	32.71[a]	41.59[a]	39.37[a]

176

Absence of diseases	63.00	73.99[c]	89.36[d]	68.64[c]	35.93[ab]	57.17[b]	45.24[a]
The number of cows on a farm that cough and sneeze	51.99	59.80[b]	84.69[c]	57.58[b]	29.50[a]	41.96[a]	37.49[a]
The number of cows on a farm with a runny nose or runny eye (symptom of an infection)	53.97	62.85[b]	87.50[c]	58.08[b]	33.39[a]	42.35[a]	40.33[a]
The number of cows on a farm that have difficulty breathing	54.19	62.10[b]	88.86[c]	59.70[b]	32.00[a]	44.22[a]	38.90[a]
The number of cows on a farm with diarrhoea	54.24	61.46[b]	87.77[c]	59.23[b]	33.86[a]	42.34[a]	41.56[a]
The number of cows on a farm with a vaginal or uterine infection	53.39	62.75[b]	88.83[c]	57.81[b]	27.54[a]	40.94[a]	39.63[a]
The number of cows on a farm with an udder infection	57.77	66.86[b]	89.33[c]	63.07[b]	34.71[a]	48.05[a]	42.52[a]
The number of cows on a farm that have died	50.98	59.22[b]	84.67[c]	54.64[b]	31.46[a]	40.75[a]	37.29[a]
The number of difficult calvings	39.86	44.32[b]	66.08[c]	43.88[b]	24.32[a]	27.86[a]	32.78[a]
The number of cows on a farm that could not stand upright anymore	52.09	60.35[b]	87.31[c]	57.93[b]	25.43[a]	41.45[a]	37.02[a]
Absence of pain induced by management procedures	56.99	67.53[c]	89.50[d]	63.98[c]	27.11[a]	45.64[b]	40.35[ab]
The number of cows on a farm where the horns are removed using a caustic paste in comparison to a soldering iron and whether anaesthesia or pain medication is used during this procedure	54.17	61.44[b]	90.81[c]	62.05[b]	28.07[a]	43.25[a]	37.72[a]
The number of cows on a farm whose tail has been removed using a rubber band instead of an operation and whether anaesthesia or pain medication was used during this procedure	54.18	61.55[b]	90.30[c]	61.23[b]	26.71[a]	43.16[a]	38.92[a]
Expression of social behaviour	55.44	64.72[b]	83.75[c]	63.04[b]	32.68[a]	45.98[a]	38.58[a]
Cows butting heads or showing other aggressive behaviour among themselves	43.77	48.79[b]	76.56[c]	49.34[b]	27.21[a]	29.75[a]	33.86[a]
	55.03	65.39[c]	84.89[d]	62.11[c]	25.18[a]	45.00[b]	38.51[ab]
Expression of other normal behaviour The number of days per year that the cows have access to a pasture for more than 6 hours per day	60.31	72.18[c]	89.27[d]	65.78[c]	29.46[a]	51.79[b]	43.30[ab]
Good human-animal relationship	60.60	71.86[b]	91.17[c]	68.64[b]	33.14[a]	48.37[a]	43.07[a]
Number of cows that are not approachable by people (indicates cows' fear of people)	50.25	57.84[b]	83.66[a]	56.71[b]	25.75[a]	37.76[a]	36.78[a]

Positive emotional state	57.34	68.18[b]	89.25[c]	66.39[b]	26.57[a]	44.67[a]	39.57[a]
Showing positive behaviour (active, relaxed, calm, content, friendly, playful, positively occupied, lively, curious, social and happy)	60.10	72.82[b]	92.45[c]	67.50[b]	30.32[a]	48.45[a]	41.08[a]
Showing negative behaviour (fearful, irritated, indifferent, frustrated, bored, agitated, nervous, apathetic and stressed)	60.92	72.48[b]	93.59[c]	68.48[b]	35.71[a]	49.41[a]	42.03[a]

[a-d] Means within a row without common superscript differ significantly ($P < 0.05$, one-way ANOVA with Dunnett's T3 post hoc comparison of mean scores)

Summary

To promote transparency in the milk chain and therewith stimulate the production and increase the market share of animal-friendly milk, this study lies foundations for dairy cattle welfare assurance schemes. In this doctoral thesis, the possibilities of introducing a WAS in Flanders, Belgium are explored by means of a multidisciplinary approach in which three different groups which may benefit from a WAS are studied, namely the animal, the consumer and the farmer. For consumers, a valid welfare assurance scheme could be translated into a more informed product choice regarding animal welfare. This may be achieved through marketing and positioning of the products, and may influence farm performance. Access to a differentiated market and an improved farm economic performance may stimulate farmers to implement changes in farm resources/management which improve dairy cattle welfare. When a valid method is used to assess animal welfare, such an animal welfare assurance scheme could enhance farm animal welfare. This doctoral thesis had two main aims, focused on bottlenecks for added value of a WAS. First, the aim was to identify a dairy cattle welfare assessment method which accurately reflects the welfare status on dairy farms. Second, the aim was to examine added value of dairy cattle welfare for milk through market potential and a positive relationship with farm performance.

Dairy Cattle Welfare Assessment

Several methods for assessing herd welfare have been developed and published. To accurately reflect herd welfare, an assessment method should be multidimensional, contain reliable and feasible measures, be sensitive and discriminative and should cohere with expert assessment of herd welfare. The Welfare Quality® (WQ) protocols are the most renowned and comprehensive. The WQ protocol for on-farm dairy cattle welfare assessment describes 31 measures and a step-wise method to integrate values for these measures into 12 criteria scores, grouped further into four principle scores and finally into an overall welfare categorization with four possible levels. However, some concerns have been raised about this protocol, which are investigated further in this doctoral thesis with a specific focus on the integration method.

In Chapter 2, the research question was: is there variation in sensitivity of integrated WQ outcomes (criteria and principle scores and overall welfare category) to extremely low and high values of the welfare measures and what are the reasons for this variation in sensitivity? This

was investigated by replacing each observed individual value for each measure and each herd one by one with both the theoretically possible and the observed worst and best values for a dataset of WQ assessments for 491 European herds (further referred to as 'EU dataset'). Results showed a varying degree of sensitivity of integrated WQ outcomes to extreme changes in individual measures and provided explanations for this variation in sensitivity. Some of these explanations for variation in sensitivity were intended by the WQ consortium (e.g. a negative change (indicating poor welfare) influenced integrated outcomes more strongly than a positive change). However, two unintended side-effects of the step-wise integration method had a considerable influence too. Namely 1) the number of measures that are integrated into criteria and principle scores, and 2) the aggregation method of the measures. A simpler integration method which is uniform for all measures could potentially solve this problem of unintended side-effects.

The research question of Chapter 3 was: to what extent do discrepancies exist between WQ integrated scores and expert opinion? It can be argued that animal scientists and, even more so, people with experience in the measures involved, would be most qualified to balance the various measures of animal welfare and therewith judge the state of animal welfare of a given herd. We therefore conducted an online survey to examine to what extent opinions of trained users of the WQ protocol (experts) correspond with the integrated scores (i.e. criteria, principles and overall categorization) calculated according to the WQ dairy cattle protocol. Experts were asked to allocate criterion and overall welfare scores to seven focal herds from the EU dataset. In addition, logistic regression was applied to identify the measures that affected the WQ overall welfare categorization for the EU dataset. Results indicated that the correspondence between selected experts' opinion and WQ scores for six of the 12 criteria and for overall welfare was low. Moreover, some of the measures that were ranked as least important (e.g. measures of drinkers) by experts had the highest influence on the WQ overall welfare categorization. Conversely, measures rated as most important by the experts (e.g. lameness and mortality) had a lower impact on the WQ overall category. The integration methods of the WQ protocol for dairy cattle thus lacked correspondence with selected experts on the importance of several welfare measures.

Findings of both Chapter 2 and Chapter 3 strengthen the growing evidence that the WQ integration method should be revised in order to accurately reflect herd welfare in a way that

corresponds with the opinions of trained users of the WQ protocol for dairy cattle welfare. As the sensitivity analysis showed that the highly complex integration procedure causes some unwanted side-effects we opted to develop an alternative integration protocol (Chapter 4). Using a five-step procedure, a protocol for integration of dairy welfare measures was developed into a Welfare Index (WI) which is transparent and simple, multidimensional, discriminative and based upon, and checked for, correspondence by a panel of experts. This WI is based upon the average of a severity score (i.e. how severely a given welfare problem affects the welfare of a cow) multiplied with the herd prevalence for each welfare measure. In Step 1, based upon the importance ranks by 14 experts complemented with literature review, six animal-based welfare measures were selected from the WQ protocol for dairy cattle: 'lameness score' (severe lameness), 'body condition score' (leanness), 'mortality rate', 'hairless patches', 'lesions/swellings' and 'SCC >400.000'. In Step 2, median expert severity scores for these welfare measures were determined. In Step 3, a dataset was collated of the selected welfare measures in 491 herds across Europe. Experts (n = 14) allocated a welfare score (from 0-100) to 12 fictitious herds for which the prevalence of each welfare measure was benchmarked against all 491 herds. Quadratic models indicated that the correspondence between these expert scores and the proposed WI was high (R^2= 0.91). Step 4 comprised interpreting WI scores by determining four welfare categories from poor ('not classified') to 'excellent' welfare based upon expert scores for the 12 fictitious herds. In Step 5, the selected measures were found to be associated with all welfare problems which impair dairy cattle welfare most, based on literature. The proposed WI is not exhaustive, and should be accompanied with a disclaimer for welfare aspects that cannot be detected. However, the proposed method is flexible in such a way that measures can be replaced or added as deemed desirable by repeating the proposed steps.

The Flemish Consumer Market for Animal-friendly Milk

The Flemish consumer market for animal-friendly milk was studied in Chapter 5 and 6. The research question of Chapter 5 was: 'in what way do citizens' attitudes, perceptions of product-related factors, and milk consumption relate to their intention to purchase animal-friendly milk?'. Research questions for Chapter 6 were 1) to what extent do market opportunities for animal-friendly milk exist in Flanders, Belgium?; and 2) what characterizes different segments within the market for animal-friendly milk? Data were collected using an online survey among 787 Flemish dairy consumers who were involved in the purchase decision.

Findings of Chapter 5 and Chapter 6 showed that the Flemish consumer market for animal-friendly milk is highly differentiated. Marketing animal-friendly milk at several levels of animal welfare and price could reach a wide group of consumers. Consumers who have a higher intention to purchase animal-friendly milk (8% of the sample) were found to perceive animal welfare as a very important product attribute, were highly interested in information about dairy cattle welfare and, had a more natural living oriented attitude towards dairy cattle welfare. For these consumers, focusing positioning of animal-friendly milk on enhanced welfare could be sufficient to stimulate purchase of animal-friendly milk of the highest level of animal welfare and price. Other consumer segments may consider animal welfare to be an important product attribute but not dominant to other product attributes such as price, taste or quality, and would therefore likely be more willing to purchase intermediate products regarding level of welfare and price. Consumer segments with moderate or low market opportunities formed 48% and 40% of the study sample respectively. Most promising selling propositions in future communication about animal-friendly milk were found to be 'access to pasture' and the related WQ criterion (according to respondents) 'freedom of movement' and the WQ criterion 'absence of diseases'.

Relationship Between Farm Performance and Dairy Cattle Welfare

For farmers to invest in animal welfare improvements on-farm when no return on investment can be guaranteed, is unlikely or even impossible. Therefore, it is important to assess the trade-offs existing in the animal-friendly milk supply chain. In this thesis, added value of farm differentiation based on enhanced animal welfare was pursued by means of 1) access to a differentiated consumer market for animal-friendly products (Chapter 5 and 6) and 2) the link between an enhanced level of animal welfare and farm performance (Chapter 7). For the latter, the relationship between dairy cattle welfare and farm performance was examined, as measured using cost efficiency. Research question was: to what extent are animal welfare and cost efficiency related as investigated using integrated multi-criteria approaches? In Chapter 7, for a sample of 41 Flemish dairy farms, Cost Efficiency (CE) and its underlying components, Technical Efficiency (TE), and Cost Allocative Efficiency (CAE) were compared with WQ categorisation and the alternative WI score. TE measures the farmers' ability to use a minimal amount of feed inputs to produce milk as an output, whereas CAE measures the ability to use

inputs in an optimal combination given the input prices. Fully cost efficient farms are both fully technically efficient and fully cost allocative efficient. CE was not significantly correlated with both WQ categorisation and WI scores. Evidence of trade-off or win-wins between enhanced animal welfare and farm performance were only found in the individual components (TE and CAE and individual welfare measures). Results indicate that a high level of CE and an enhanced level of animal welfare could be combined in one farm. Stimulated by an increased welfare on-farm and access to a differentiated market which would allow farms to distinguish themselves from others, farmers might therefore be able and willing to employ welfare improvements on-farm.

General Conclusions

This doctoral thesis has shown that animal welfare has the potential to provide dairy farms with options to increase farm competitiveness in such a way that cost efficiency is not influenced. In addition, a WI which is simple, intuitive, transparent and which detects the most severe welfare problems according to dairy cattle welfare experts was developed. Future research could focus on identifying to what degree the proposed welfare index would be widely accepted by several stakeholders, among which society.

A heterogeneous market for animal-friendly milk was identified, for which healthiness of the cows and outdoor access are promising selling propositions. Product differentiation could be achieved by means of e.g. a star rating system differentiating between different levels of animal welfare at different price levels. Products from these different levels of price and welfare could be positioned in the market differently, where higher levels could focus primarily on enhanced welfare, and lower levels could focus on other benefits of animal-friendly milk such as healthier or off better quality. Access to such a differentiated market, along with an improved welfare of cows in the herd, could allow for and stimulate farmers to implement welfare improvements.

The general conclusion of this doctoral thesis that the introduction of a WAS for dairy cattle is possible in Flanders, but its success will depend on strategic decisions which will determine the balance between credibility, implementation and costs.

Samenvatting

Deze studie legt de basis voor garantiesystemen voor het welzijn van melkvee (WAS) die transparantie in de melkketen stimuleren, en daarmee de productie en het marktaandeel van diervriendelijke melk vergroten. In deze thesis werden de mogelijkheden voor het introduceren van een WAS onderzocht door een multidisciplinaire aanpak waarbij drie groepen die baat kunnen hebben bij een dergelijk systeem bestudeerd werden: het dier, de consument en de veehouder. Voor consumenten zou een dergelijk systeem kunnen leiden tot een beter geïnformeerde keuze wat betreft dierenwelzijn van melkvee door middel van product-marketing en positionering. Toegang tot een gedifferentieerde markt en verbeterde bedrijfseconomische prestaties kan veehouders stimuleren om aanpassingen in hun bedrijf door te voeren die het dierenwelzijnsniveau verhogen. Wanneer een valide methode gebruikt wordt om welzijn te meten, kunnen deze aanpassingen tot een verhoogd dierenwelzijnsniveau leiden. Deze doctorale thesis had twee doelen gericht op knelpunten voor toegevoegde waarde van een WAS. Ten eerste het identificeren van een methode die de welzijnsstatus op een bedrijf nauwkeurig weergeeft. Het tweede doel was om toegevoegde waarde van diervriendelijke melk te onderzoeken op basis van marktpotentieel en bedrijfsprestaties.

Welzijn Beoordelen van Melkvee

Om het welzijn van een kudde dieren accuraat weer te geven, dient de evaluatiemethode multidimensionaal te zijn, betrouwbare metingen te bevatten, gevoelig en discriminerend te zijn en moeten uitkomsten overeenkomen met de beoordeling van experts. De Welfare Quality® (WQ) protocollen voor het beoordelen van welzijn omvatten de meest bekende en meest extensieve methode. Het WQ protocol voor de welzijnsbeoordeling van melkvee beschrijft 31 metingen en een stapsgewijze methode om uitkomsten van deze metingen te integreren in 12 criteriascores. Deze worden verder gegroepeerd in vier principescores en uiteindelijk in een algehele welzijnscategorisering met vier mogelijke niveaus. In eerdere studies zijn bezorgdheden geuit over dit protocol, die verder worden onderzocht in deze thesis met als focus de integratiemethode.

In Hoofdstuk 2 was de onderzoeksvraag: is er variatie in de gevoeligheid van de geïntegreerde WQ uitkomsten (criteria- en principescores en de algehele welzijnscategorie) voor extreem lage en hoge waarden van de verschillende metingen, en wat zijn de redenen voor deze variatie in

gevoeligheid? Dit werd onderzocht met behulp van een sensitiviteitsanalyse, voor een dataset van WQ beoordelingen uitgevoerd bij 491 Europese melkveekuddes ('EU dataset'). Hierbij werd elke waargenomen individuele waarde voor elke meting en van elke kudde één voor één met zowel de theoretisch als de waargenomen slechtste en beste waarden vervangen. De resultaten toonden een variërende mate van gevoeligheid van de geïntegreerde WQ uitkomsten voor extreme veranderingen in de afzonderlijke welzijnsmetingen. Enkele verklaringen voor deze variatie in gevoeligheid werden uiteengezet. Een aantal van deze verklaringen zijn vooropgesteld door het WQ consortium (bijvoorbeeld, een negatieve verandering (slechter welzijn) beïnvloedde de geïntegreerde resultaten sterker dan een positieve verandering). Echter hadden twee kenmerken van de complexe stapsgewijze integratiemethode ook een ongewenste invloed. Variatie in gevoeligheid wordt mede bepaalt door 1) het aantal metingen die zijn geïntegreerd in de criteria- en principescores, en 2) de verschillende integratiemethodes van de metingen. Een eenvoudigere en uniforme integratiemethode voor alle metingen is een potentiële oplossing.

Hoofdstuk 3 was gebaseerd op de onderzoeksvraag: in welke mate verschillen WQ geïntegreerde scores van de welzijnsbeoordeling door experts? Gesteld kan worden dat dierwetenschappers, en in het bijzonder mensen die ervaring hebben met de specifieke welzijnsmetingen, het meest gekwalificeerd zijn om de dierenwelzijnsstatus van een kudde melkvee te beoordelen op basis van data van WQ metingen. Om deze reden is een online enquête afgenomen onder getrainde gebruikers van het WQ protocol ('experts') om te onderzoeken in hoeverre hun mening over de staat van welzijn op melkveebedrijven overeen komt met de geïntegreerde scores berekend volgens het WQ protocol. Experts werden gevraagd om een rangorde en een betrouwbaarheids- en validiteitsscore te geven voor alle WQ welzijnsmetingen. Daarnaast werd ze gevraagd criterium- en algehele welzijnsscores toe te wijzen aan zeven kuddes uit de EU dataset. Ook werd logistische regressie toegepast om de metingen te identificeren die de WQ algemene welzijnscategorisatie het meest beïnvloeden voor de EU dataset. Metingen die door experts als minst belangrijk werden gerangschikt (bijvoorbeeld die met betrekking tot watervoorziening) hadden de meeste invloed op de WQ welzijnscategorisering. Daarentegen hadden metingen beoordeeld als belangrijkst door de experts (bijvoorbeeld kreupelheid en mortaliteit) een lagere impact op WQ welzijnscategorisering. Bovendien was de mate van overeenkomst tussen de mening van de experts en WQ scores voor zes van de 12 criteria en voor de algemene welzijnscategorie laag.

Dit illustreert een gebrek aan overeenkomst tussen geïntegreerde uitkomsten van het WQ protocol voor melkvee en geselecteerde experts over het belang van een aantal welzijnsmetingen.

Bevindingen van hoofdstukken 2 en 3 versterken de groeiende bewijslast dat de WQ integratiemethode herzien moet worden om een accurate weergave van het welzijn van melkvee te bieden die overeenkomt met getrainde gebruikers van ditzelfde protocol. Uit de gevoeligheidsanalyse bleek dat de zeer complexe integratiemethode enige ongewenste neveneffecten veroorzaakt. Het oplossen van deze problemen, zonder afbreuk te doen aan de stapsgewijze opbouw van het WQ protocol, is zeer moeilijk. Derhalve is een alternatief protocol ontwikkeld door middel van een vijf-staps-procedure voor integratie van welzijnsmetingen tot een welzijnsindex (WI, Hoofdstuk 4). Deze WI is gebaseerd op het gemiddelde van ernstigheidsscores (in welke mate tast een welzijnsprobleem het algemene welzijn van een koe aan) en de prevalentie van dat welzijnsprobleem in de kudde voor elke welzijnsmeting. In Stap 1 werden op basis van de mening van 13 experts aangevuld met literatuurstudie, zes welzijnsmetingen geselecteerd uit het WQ protocol voor melkvee: 'kreupelheidsscore' (ernstige kreupelheid), 'conditiescore' (te mager), 'sterfte', 'haarloze plekken', 'laesies / zwellingen' en 'celgetal > 400.000'. In Stap 2 werd de mediaan van ernstigheidsscores voor de verschillende metingen door de experts bepaald. In Stap 3 werd experts gevraagd het welzijn van in totaal 12 fictieve kuddes te scoren waarvoor de prevalentie van elke meting werd vergeleken met die van de EU-dataset. Kwadratische modellen gaven aan dat de mate van overeenkomst tussen deze expertscores en de WI hoog was ($R^2 = 0,91$). Stap 4 bestond uit het interpreteren van WI scores door het bepalen van vier welzijnscategorieën van slecht tot excellent welzijn op basis van expertscores voor de 12 fictieve kuddes. In Stap 5 werd vastgesteld dat de gekozen metingen gerelateerd zijn aan alle belangrijkste welzijnsproblemen die afbreuk doen aan het welzijn van melkvee, op basis van literatuur. De voorgestelde WI is echter niet exhaustief en moet daarom gepaard gaan met een disclaimer voor de aspecten van welzijn die niet gemeten kunnen worden. Omdat de werkwijze flexibel is, kunnen metingen vervangen of toegevoegd worden door de vijf stappen te herhalen.

De Vlaamse Consumentenmarkt voor Diervriendelijke Melk

De Vlaamse consumentenmarkt voor diervriendelijke melk werd bestudeerd met behulp van drie onderzoeksvragen. De onderzoeksvraag van Hoofdstuk 5 was op welke manier

beïnvloeden attitudes over, en perceptie van product-gerelateerde factoren en de consumptie van melk de intentie tot het kopen diervriendelijke melk? Onderzoeksvragen in Hoofdstuk 6 waren 1) in welke mate bestaan er marktkansen binnen verschillende consumentensegmenten voor diervriendelijke melk in Vlaanderen?, en 2) wat karakteriseert verschillende segmenten binnen de markt voor diervriendelijke melk? Data werd verzameld door middel van een online enquête afgenomen bij 787 Vlaamse consumenten van melk die betrokken zijn bij de aankoopbeslissing.

De bevindingen van Hoofdstuk 5 en Hoofdstuk 6 tonen aan dat de consumentenmarkt voor diervriendelijke melk in hoge mate gedifferentieerd is. Hierop kan ingespeeld worden door producten van verschillende dierenwelzijns- en prijsniveaus aan te bieden. Consumenten die een hogere intentie hebben tot het kopen van diervriendelijke melk (8% van de onderzoekspopulatie), zagen dierenwelzijn als een zeer belangrijk productattribuut, waren zeer geïnteresseerd in informatie over het welzijn van melkvee en hadden een attitude ten opzichte van het welzijn van melkvee die het natuurlijk leven van deze dieren benadrukt. Voor deze consumenten kan men bij de positionering van diervriendelijke melk in de markt, de nadruk leggen op verbeterd dierenwelzijn. Andere consumentensegmenten beschouwden dierenwelzijn als een belangrijk productkenmerk, maar niet dominant over andere producteigenschappen zoals prijs, smaak en kwaliteit. Consumentensegmenten met middelmatig of weinig marktkansen vormden respectievelijk 48% en 40% van de onderzoekspopulatie. Deze segmenten zijn waarschijnlijk eerder bereid om intermediaire producten te kopen, tussen conventionele melk en het hoogste niveau van dierenwelzijn in. De WQ metingen 'toegang tot de weide' en het WQ criteria 'bewegingsvrijheid' en 'afwezigheid van ziekten' bleken over segmenten heen meest veelbelovend voor communicatie over diervriendelijke melk.

De Relatie Tussen Bedrijfsprestaties en Dierenwelzijn

Het is onwaarschijnlijk dat veehouders kunnen investeren in welzijnsverbeteringen op hun bedrijf wanneer er geen rendement op deze investering kan worden behaald. Daarom is het belangrijk om de wisselwerking tussen dierenwelzijn en bedrijfsprestaties te analyseren. Ten eerste door toegang tot een gedifferentieerde consumentenmarkt voor diervriendelijke producten (Hoofdstuk 5 en 6) en ten tweede door het verband tussen dierenwelzijn en bedrijfsprestaties (Hoofdstuk 7). Voor dit tweede punt werd de relatie tussen het welzijn van

melkvee en bedrijfsprestaties onderzocht, gemeten door middel van Kostenefficiëntie (KE). De onderzoeksvraag was: in welke mate zijn dierenwelzijn en KE gerelateerd als onderzocht met behulp van geïntegreerde metingen. In Hoofdstuk 7 werden voor een steekproef van 41 Vlaamse melkveebedrijven de KE en diens onderliggende componenten, technische efficiëntie (TE) en kosten allocatieve efficiëntie (KAE) vergeleken met hun WQ categorisatie en de alternatieve WI score. De input voor het berekenen van de KE was voeding, en de output was liters melk. TE meet in hoeverre de veehouders in staat zijn om een minimale hoeveelheid voer als input te gebruiken om melk te produceren als output. KAE meet of veehouders in staat zijn om input (krachtvoeder, ruwvoeder en weide) te gebruiken in een optimale mix, gezien de prijzen voor deze input. Volledig kostenefficiënte bedrijven zijn zowel volledig technisch efficiënt als volledig kosten allocatief efficiënt. KE score was niet gecorreleerd met zowel WQ categorisatie als WI scores. Bewijs van positieve of negatieve wisselwerkingen tussen verbeterd dierenwelzijn en bedrijfsprestaties waren alleen te vinden in de afzonderlijke componenten (relaties tussen TE en KAE en individuele welzijnsmetingen). Deze resultaten geven aan dat een hoog niveau van KE en een hoger niveau van dierenwelzijn kunnen worden gecombineerd in één melkveebedrijf. Gestimuleerd door een toegenomen welzijnsniveau en door toegang tot een gedifferentieerde markt die het mogelijk maakt voor melkveebedrijven om zich te onderscheiden van anderen, kunnen de veehouders in staat én bereid zijn om welzijnsverbeteringen op hun bedrijf toe te passen.

Algemene Conclusies

Uit dit proefschrift blijkt dat dierenwelzijn mogelijkheden biedt voor melkveehouderijen om zich te onderscheiden zonder de focus op kostenefficiëntie te verliezen. Daarnaast is een welzijnsindex ontwikkeld die eenvoudig, discriminerend, sensitief en transparant is en de meest ernstige welzijnsproblemen detecteert. Toekomstig onderzoek zou zich kunnen richten op de mate waarin de voorgestelde welzijnsindex door verschillende stakeholders, waaronder de samenleving, zou worden aanvaard en in hoeverre er metingen aan de huidige selectie moeten worden toegevoegd.

Een heterogene markt voor diervriendelijke melk werd geïdentificeerd, waarbij gezondheid van de koeien, toegang tot de weide en gepercipieerde voordelen van diervriendelijke melk (indien deze daadwerkelijk bestaan) veelbelovend waren voor de communicatie over diervriendelijke

melk. Productdifferentiatie kan toegepast worden door onderscheid te maken tussen verschillende dierenwelzijnsniveaus en deze aan te bieden voor verschillende prijsniveaus. Producten van deze niveaus kunnen op een verschillende manier worden gepositioneerd in de markt, waarbij hogere niveaus zich puur kunnen richten op een beter welzijn en lagere niveaus op de voordelen van diervriendelijke melk voor de consument. De toegang tot een dergelijke gedifferentieerde markt, samen met een verbeterd welzijnsniveau, kan veehouders stimuleren om welzijnsverbeteringen op hun bedrijf uit te voeren. De algehele conclusie van deze thesis is dat de introductie van een WAS voor melkvee in Vlaanderen mogelijk is, maar haar slagen hangt af van strategische beslissingen die de balans tussen geloofwaardigheid, implementatie en kosten zullen bepalen.

References

Ajzen I 1985. From intentions to actions: A theory of planned behavior. In Action control, p. 11-39. Springer-Verlag, Berlin, Germany.

Ajzen I 2001. Nature and operation of attitudes. Annual Review of Psychology 52, 27-58.

Allendorf JJ and Wettemann PJC 2015. Does animal welfare influence dairy farm efficiency? A two-stage approach. Journal of Dairy Science 98, 7730-7740.

Austin EJ, Deary IJ, Edwards-Jones G and Arey D 2005. Attitudes to farm animal welfare: factor structure and personality correlates in farmers and agriculture students. Journal of Individual Differences 26, 107.

Averós X, Aparicio MA, Ferrari P, Guy JH, Hubbard C, Schmid O, Ilieski V and Spoolder HA 2013. The effect of steps to promote higher levels of farm animal welfare across the EU. Societal versus animal scientists' perceptions of animal welfare. Animals 3, 786-807.

Barnes AP, Rutherford KMD, Langford FM and Haskell MJ 2011. The effect of lameness prevalence on technical efficiency at the dairy farm level: An adjusted data envelopment analysis approach. Journal of Dairy Science 94, 5449-5457.

Bartussek H, Leeb C and Held S 2000. Animal Needs Index for Cattle (Ani 35 L/2000-cattle). Federal Research Institute for Agriculture in Alpine Regions BAL Gumpenstein, Irdning, Austria.

Bennett RM 1997. Farm animal welfare and food policy. Food Policy 22, 281-288.

Belletti G, Burgassi T, Marescotti A and Scaramuzzi S 2007. The effects of certification costs on the success of a PDO/PGI. Quality Management in Food Chains, Wageningen, The Netherlands.

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. Animal Welfare 12, 445-455.

Blokhuis HJ, Veissier I, Miele M and Jones B 2010. The Welfare Quality® project and beyond: Safeguarding farm animal well-being. Acta Agriculturae Scand Section A 60, 129-140.

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I and Langbein J 2007. Assessment of positive emotions in animals to improve their welfare. Physiology & Behavior 92, 375-397.

Boogaard BK, Oosting SJ and Bock BB 2008. Defining sustainability as a socio-cultural concept: Citizen panels visiting dairy farms in the Netherlands. Livestock Science 117, 24-33.

Boogaard BK, Bock BB, Oosting SJ and Krogh E 2010. Visiting a farm: An exploratory study of the social construction of animal farming in Norway and the Netherlands based on sensory perception. International Journal of Sociology of Agriculture and Food 17, 24-50.

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007a. Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, 225-228. Deliverable 2.8b, subtask 2.3.1.2, Welfare Quality® (EU Food-CT-2004-506508)

Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, Van Reenen CG and Veissier I 2007b. Aggregation of measures to produce an overall assessment of animal welfare. Part 2: Analysis of constraints.

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

Botreau R, Perny P and Veissier I 2008. Reports on the construction of welfare criteria for different livestock species Part 3 – Criteria construction for all animal types on farm.

Bokkers EAM, de Vries M, Antonissen ICMA and de Boer IJM 2012. Inter-and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. Animal Welfare 21, 307-318.

Bourlakis M, Bock BB and van Huik MM 2007. Animal welfare: the attitudes and behaviour of European pig farmers. British Food Journal 109, 931-944.

Bracke MBM, Mets JHM and Spruijt BM 2001. Development of a Decision Support System to Assess Farm Animal Welfare. Acta Agriculturae Scandinavica, section A: Animal Science supplement 30, 17 – 20

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a. Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. Journal of Animal Science 80, 1819 – 1834.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002b. Decision support system for overall welfare assessment in pregnant sows B: Validation by expert opinion. Journal of Animal Science 80:,1835 – 1845.

Bracke MBM, De Greef KH and Hopster H 2005. Qualitative stakeholder analysis for the development of sustainable monitoring systems for farm animal welfare. Journal of Agricultural and Environmental Ethics 18, 27-56.

Brenninkmeyer C and Winckler C 1-3-2012. Relationships between animal welfare hazards and animal-based welfare measures. EFSA Supporting Publications 9, 253E-2n/a.

Briggeman BC and Lusk JL 2010. Preferences for fairness and equity in the food system. European Review of Agricultural Economics 38, 1-29.

British Veterinary Association 1992. FAWC updates the five freedoms. Veterinary Record 131, 357.

Buijs S, Ampe B and Tuyttens FAM 2016. Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? Animal 15, 1-10.

Buller H and Cesar C 2007. Eating well, eating fare: farm animal welfare in France. International Journal of Sociology of Food and Agriculture 15, 45-58.

Burow E, Rousing T, Thomsen PT, Otten ND and Sørensen JT 2013. Effect of grazing on the cow welfare of dairy herds evaluated by a multidimensional welfare index. Animal 7, 834-842.

Burt S. 2000. The strategic role of retail brands in British grocery retailing. European Journal of marketing 34: 875-890.

Cameron REB, Dyk PB, Herdt TH, Kaneene JB, Miller R, Bucholtz HF, Liesman JS, Vandehaar MJ and Emery RS 1998. Dry cow diet, management, and energy balance as risk factors for displaced abomasum in high producing dairy herds. Journal of Dairy Science 81, 132-139.

Capdeville J and Veissier I 2001. A method for assessing dairy cows welfare in a loose housing herd focussing on animal observations. Acta Agriculturae Scandinavia 30: 62-68

Cardoso CS, Hötzel MJ, Weary DM, Robbins JA and von Keyserlingk MA 2016. Imagining the ideal dairy farm. Journal of Dairy Science 99, 1663-1671.

Coelli TJ, Rao DSP, O'Donnell CJ, and Battese GE 2005. An introduction to efficiency and productive analysis. Springer, New York, USA.

Coignard M, Guatteo R, Veissier I, Lehébel A, Hoogveld C, Mounier L, and Bareille N. 2014. Does milk yield reflect the level of welfare in dairy herds? The Veterinary Journal, 199: 184-187.

Cook NB 2002. The influence of barn design on dairy cow hygiene, lameness and udder health. Proc.of the 35 Th Ann.Conv.Amer.Assoc.Bov.Pract., Madison, WI.Amer.Assoc.Bov.Pract., Rome, GA

Corazzin M, Piasentier E, Dovier S and Bovolenta S 2010. Effect of summer grazing on welfare of dairy cows reared in mountain tie-stall barns. Italian Journal of Animal Science 9, 97-103.

Czycholl I, Büttner K, grosse Beilage E and Krieter J 2015. Review of the assessment of animal welfare with special emphasis on the" Welfare Quality® animal welfare assessment protocol for growing pigs". Archiv Fuer Tierzucht 58, 237.

Dagevos H and Sterrenberg L 2003. Burgers en consumenten: tussen tweedeling en twee-eenheid. Wageningen, the Netherlands.

de Graaf S, Van Loo EJ, Bijttebier J, Vanhonacker F, Lauwers L, Tuyttens FA and Verbeke W 2016. Determinants of consumer intention to purchase animal-friendly milk. Journal of Dairy Science 99, 8304-8313.

de Graaf S, Ampe B and Tuyttens FAM in press. Assessing dairy cow welfare at the beginning and end of the indoor period using the Welfare Quality® protocol. Animal Welfare.

de Graaf S, Vanhonacker F, Van Loo EJ, Bijttebier J, Lauwers L, Tuyttens FAM and Verbeke W 2016. Market Opportunities for Animal-Friendly Milk in Different Consumer Segments. Sustainability 8, 1302.

de Jonge J, van der Lans IA and van Trijp HC 2015. Different shades of grey: Compromise products to encourage animal friendly consumption. Food Quality and Preference 45, 87-99.

de Jonge J and Van Trijp H 2014. Heterogeneity in consumer perceptions of the animal friendliness of broiler production systems. Food Policy 49, 174-185.

de Jonge J and van Trijp H 2013. Meeting heterogeneity in consumer demand for animal welfare: A reflection on existing knowledge and implications for the meat sector. Journal of Agricultural and Environmental Ethics 26, 629-661.

de Jong IC, Ekkel ED, van de Burgwal JA, Lambooij E, Korte SM, Ruis MA, Koolhaas JM and Blokhuis HJ 1998. Effects of strawbedding on physiological responses to stressors and behaviour in growing pigs. Physiology & Behaviour 64, 303-310.

de Jong IC, Hindle VA, Butterworth A, Engel B, Ferrari P, Gunnink H, Moya TP, Tuyttens FAM and Van Reenen CG 2016. Simplifying the Welfare Quality® assessment protocol for broiler chicken welfare. Animal 10, 117-127.

de Vries M, Bokkers EAM, van Schaik G, Botreau Rl, Engel B, Dijkstra T and de Boer IJM 2013a. Evaluating results of the Welfare Quality multi-criteria evaluation model for classification of dairy cattle welfare at the herd level. Journal of Dairy Science 96, 6264-6273.

de Vries M, Engel B, den Uijl I, van Schaik G, Dijkstra T, de Boer IJM and Bokkers EAM 2013b. Assessment time of the Welfare Quality-protocol for dairy cattle. Animal Welfare 22, 85-93.

de Vries M, Bokkers EAM, Van Reenen CG, Engel B, van Schaik G, Dijkstra T and de Boer IJM 2015. Housing and management factors associated with indicators of dairy cattle welfare. Preventive Veterinary Medicine 118, 80-92.

D.G. SANCO 2009. Feasibility study on animal welfare labelling and establishing a Community Reference Centre for Animal Protection and Welfare. pp.

Diamantopoulos A, Schlegelmilch BB, Sinkovics RR and Bohlen GM 2003. Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. Journal of Business Research 56, 465-480.

Dransfield E, Ngapo TM, Nielsen NA, Bredahl L, Sjöden PO, Magnusson M, Campo MM and Nute GR 2005. Consumer choice and suggested price for pork as influenced by its appearance, taste and information concerning country of origin and organic pig production. Meat Science 69, 61-70.

Duncan IJ 1996. Animal welfare defined in terms of feelings. Acta Agriculturae Scandinavica.Section A.Animal Science.Supplementum.

Ekesbo I 1992. Monitoring systems using clinical, subclinical and behavioural records for improving health and welfare. Livestock Health and Welfare, Longman Scientific, Harlow, UK, 20-50.

Ettema JF, and Ostergaard S 2006. Economic decision making on prevention and control of clinical lameness in Danish dairy herds. Livestock Science 102, 92-106.

European Commission 2005. Attitudes of consumers towards the welfare of farmed animals. Special Eurobarometer 229. http://ec.europa.eu/food/animal/welfare/euro_barometer25_en.pdf.

European Commission 2007. Attitudes of EU citizens towards animal welfare. Special Eurobarometer 270. http://www.vuzv.sk/DB-Welfare/vseob/sp_barometer_aw_en.pdf.

European Commission 2013. Overview of Cap Reform 2014-2020. Agricultural Policy Perspectives Brief.

European Commission 2016. Attitudes of Europeans towards Animal Welfare. Special Eurobarometer 442.

European Food Safety Authority 2012. Scientific Opinion on the use of animal-based measures to assess welfare of dairy cows. EFSA Journal 10, 2554-2n/a.

European Food Safety Authority 2009a. Scientific opinion on welfare of dairy cows in relation to behaviour, fear and pain based upon a risk assessment with special reference to the impact of housing, feeding, management and genetic selection. EFSA Journal 7, 1139-1n/a.

European Food Safety Authority 2009b. Scientific opinion on welfare of dairy cows in relation to leg and locomotion problems based upon a risk assessment with special reference to the impact of housing, feeding, management and genetic selection. EFSA Journal 7, 1142-1n/a.

European Food Safety Authority 2009c. Scientific opinion on welfare of dairy cows in relation to metabolic and reproductive problems based upon a risk assessment with special reference to the impact of housing, feeding, management and genetic selection. EFSA Journal 7, 1140-1n/a.

European Food Safety Authority 2009d. Scientific opinion on welfare of dairy cows in relation to udder problems based upon a risk assessment with special reference to the impact of housing, feeding, management and genetic selection. EFSA Journal 7, 1141-1n/a.

European Food Safety Authority 2009e. Scientific report on the effects of farming systems on dairy dairy cattle welfare and disease. EFSA Journal 7, 1143r-11n/a.

Fawcett J, Clark DC, Scheftner WA and Gibbons RD 1983. Assessing anhedonia in psychiatric patients: The Pleasure Scale. Archives of General Psychiatry 40, 79-84.

Färe R, Grosskopf S and Lovell CK 2013. The measurement of efficiency of production,6. Springer Science & Business Media.

Farrell MJ 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society 3, 253-290.

Field A 2009. Discovering statistics using SPSS. Sage publications.

Fishbein M 1976. A behaviour theory approach to the relations between beliefs about an object and the attitude toward the object. In Mathematical Models in Marketing (), pp. 87-88. Springer,

Fishbein M and Ajzen I 2011. Predicting and changing behaviour: The reasoned action approach. Taylor & Francis.

Fisher R and Katz JE 2008. Social desirability bias and the validity of self-reported values. Psychology & Marketing 17, 105-120.

Fisher A and Matthews L 2001. The Social Behaviour of Sheep 8. Social Behaviour in Farm Animals, 211.

Fraser D 1995. Science, values and animal welfare: exploring the inextricable connection. Animal Welfare 4, 103-117.

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

Fraser D 2008. Understanding animal welfare. Acta Veterinaria Scandinavica 50, 1.

Fraser D, Weary DM, Pajor EA and Milligan BN 1997. A scientific conception of animal welfare that reflects ethical concerns. Animal Welfare 6, 187-205.

Fregonesi JA and Leaver JD. 2001. Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems. Livestock Production Science 68, 205-216.

Fuchs C and Diamantopoulos A 2010. Evaluating the effectiveness of brand-positioning strategies from a consumer perspective. European Journal of Marketing 44, 1763-1786.

Gloy BA, Hyde J, LaDue EL 2002. Dairy farm management and long-term financial performance. Agricultural and Resource Economics Review 31, 233-247.

Granello DH and Wheaton JE 2004. Online data collection: Strategies for research. Journal of Counseling & Development 82, 387-393.

Green LE, Hedges VJ, Schukken YH, Blowey RW and Packington AJ 2002. The impact of clinical lameness on the milk yield of dairy cows. Journal of Dairy Science 85, 2250-2256.

Grunert KG 2005. Food quality and safety: consumer perception and demand. European Review of Agricultural Economics 32, 369-391.

Grunert KG 2006. Future trends and consumer lifestyles with regard to meat consumption. Meat Science 74, 149-160.

Grymer J 1980. Displaced abomasum: a disease often associated with concurrent diseases. Compend Cont Educ Pract Vet 11, S290-S295.

Grunert KG, Bredahl L and Bruns K 2004. Consumer perception of meat quality and implications for product development in the meat sectorGÇöa review. Meat Science 66, 259-272.

Hall C and Sandilands V 2007. Public attitudes to the welfare of broiler chickens. Animal Welfare 16, 499-512.

Hansson H, Szczensa-Rundberg M, and Nielsen C 2011. Which preventive measures against mastitis can increase the technical efficiency of dairy farms? Animal 5, 632-640.

Harper G and Henson S 2001. Consumer concerns about animal welfare and the impact on food choice. EU FAIR CT98-3678, Centre for Food Economics Research, The University of Reading.

Harper GC and Makatouni A 2002. Consumer perception of organic food production and farm animal welfare. British Food Journal 104, 287-299.

Harvey D and Hubbard C 2013. Reconsidering the political economy of farm animal welfare: An anatomy of market failure. Food Policy 38, 105-114.

Heath CAE, Browne WJ, Mullan S and Main DCJ 2014. Navigating the iceberg: reducing the number of parameters within the Welfare Quality® assessment protocol for dairy cows. Animal 8, 1978-1986.

Heleski CR, Mertig AG and Zanella AJ 2004. Assessing attitudes toward farm animal welfare: a national survey of animal science faculty members. Journal of Animal Science 82, 2806-2814.

Herbig PA and Kramer H 1994. The effect of information overload on the innovation choice process: Innovation overload. Journal of Consumer Marketing 11, 45-54.

Holm L and Møhl M 2000. The role of meat in everyday food culture: an analysis of an interview study in Copenhagen. Appetite 34, 277-283.

Huxley JN 2013. Impact of lameness and claw lesions in cows on health and production. Livestock Science 156, 64-70.

Ingenbleek PTM, Binnekamp MHA and van Trijp JCM 2006. Betalen voor dierenwelzijn: barri+¿res en oplossingsrichtingen in consumentenmarkten en business-to-business markten,5.06. 02. LEI.

Ingenbleek PT, Immink VM, Spoolder HA, Bokma MH and Keeling LJ 2012. EU animal welfare policy: Developing a comprehensive policy framework. Food Policy 37, 690-699.

Johnsen PF, Johannesson T and Sandøe P 2001. Assessment of herd animal welfare at herd level: many goals, many methods. Acta Agriculturae Scandinavica, Section A-Animal Science 51, 26-33.

Kauppinen T, Vesala KM and Valros A 2012. Farmer attitude toward improvement of animal welfare is correlated with piglet production parameters. Livestock Science 143, 142-150.

Keller KL 1993. Conceptualizing, measuring, and managing customer-based brand equity. the Journal of Marketing, 1-22.

Keller KL and Lehmann DR 2006. Brands and branding: Research findings and future priorities. Marketing Science 25, 740-759.

Kendall HA, Lobao LM and Sharp JS 2006. Public Concern with Animal WellGÇÉBeing: Place, Social Structural Location, and Individual Experience. Rural Sociology 71, 399-428.

Kennedy OB, Stewart-Knox BJ, Mitchell PC and Thurnham DI 2004. Consumer perceptions of poultry meat: a qualitative analysis. Nutrition & Food Science 34, 122-129.

Kimura R 1998. Mutual grooming and preferred associate relationships in a band of free-ranging horses. Applied Animal Behaviour Science 59, 265-276.

Kling-Eveillard F, Knierim U, Irrgang N, Gottardo F, Ricci R and Dockés AC 2015. Attitudes of farmers towards cattle dehorning. Livestock Science 179, 12-21.

Knierim U, Hesse D, von Borell E, Herrmann HJ, Müller C, Rauch HW, Sachser N and Zerbe F 2003. Voluntary animal welfare assessment of mass-produced farm animal housing equipment using a standardised procedure. Animal Welfare 12, 75-84.

Knierim U and Winckler C 2009. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. Animal Welfare 18, 451-458.

Kolodinsky J 2012. Persistence of health labeling information asymmetry in the United States: Historical perspectives and twenty-first century realities. Journal of Macromarketing, 0276146711434829.

Kraus SJ 1995. Attitudes and the prediction of behaviour: A meta-analysis of the empirical literature. Personality and Social Psychology Bulletin 21, 58-75.

Krystallis A, de Barcellos MD, Kügler JO, Verbeke W and Grunert KG 2009. Attitudes of European citizens towards pig production systems. Livestock Science 126, 46-56.

Kuo RJ, Ho LM and Hu CM 2002. Integration of self-organising feature map and K-means algorithm for market segmentation. Computers & Operations Research 29, 1475-1493.

Lagerkvist CJ, Carlsson F and Viske D 2006. Swedish consumer preferences for animal welfare and biotech: a choice experiment.

Laister S, Stockinger B, Regner AM, Zenger K, Knierim U and Winckler C 2011. Social licking in dairy cattle: Effects on heart rate in performers and receivers. Applied Animal Behaviour Science 130, 81-90.

Latruffe L, Balcombe K, Davidova S and Zawalinska K 2004. Determinants of technical efficiency of crop and livestock farms in Poland. Applied Economics 36, 1255-1263.

Lawson LG, Aggerb JF, Lunda M, Coelli T 2004. Lameness, metabolic and digestive disorders, and technical efficiency in Danish dairy herds: a stochastic efficiency production approach. Livestock Production Science 91, 157-172.

Leknes S and Tracey I 2010. Pain and pleasure: masters of mankind. In Pleasures of the brain, p 320-335. Oxford University press, New York, USA.

Lievaart JJ and Noordhuizen JPTM 2011. Ranking experts' preferences regarding measures and methods of assessment of welfare in dairy herds using Adaptive Conjoint Analysis. Journal of Dairy Science 94, 3420-3427.

Liljenstolpe C 2008. Evaluating animal welfare with choice experiments: an application to Swedish pig production. Agribusiness 24, 67-84.

Linstone HA and Turoff M 1975. The Delphi method: Techniques and applications,29. Addison-Wesley Reading, MA.

Lucy MC 2001. Reproductive loss in high-producing dairy cattle: where will it end? Journal of Dairy Science 84, 1277-1293.

Lusk JL, and Norwood FB 2011. Animal welfare economics. Applied economic perspectives and policy, 1-21.

Manning L, Baines RN, and Chadd SA 2006. Quality assurance models in the food supply chain. British Food Journal 108, 91-104.

Malhotra NK 1982. Information load and consumer decision making. Journal of Consumer Research 8, 419-430.

Malotte CK, Jarvis B, Fishbein M, Kamb M, Iatesta M and Hoxworth T 2000. Stage of change versus an integrated psychosocial theory as a basis for developing effective behaviour change interventions. Aids Care 12, 357-364.

Mason GJ and Mendl M 1993. Why is there no simple way of measuring animal welfare? Animal Welfare 2, 301-319.

Mejdell CM 2006. The role of councils on animal ethics in assessing acceptable welfare standards in agriculture. Livestock Science 103, 292-296.

Meuwissen MPM, Van Der Lans IA and Huirne RBM 2007. Consumer preferences for pork supply chain attributes. NJAS-Wageningen Journal of Life Sciences 54, 293-312.

Miele M 2010. Report concerning consumer perceptions and attitudes towards farm animal welfare. European Animal Welfare Platform, Belgium.

Minteer BA, Corley EA and Manning RE 2004. Environmental ethics beyond principle? The case for a pragmatic contextualism. Journal of Agricultural and Environmental Ethics 17, 131-156.

Mishra AK and Morehart MJ 2001. Off-farm investment of farm households: A logit analysis. Agricultural finance review 61, 88-101.

Müller-Lindenlauf M, Deittert C and Köpke U 2010. Assessment of environmental effects, animal welfare and milk quality among organic dairy farms. Livestock Science 128, 140-148.

Napolitano F, Pacelli C, Girolami A and Braghieri A 2008. Effect of information about animal welfare on consumer willingness to pay for yogurt. Journal of Dairy Science 91, 910-917.

Napolitano F, Girolami A and Braghieri A 2010. Consumer liking and willingness to pay for high welfare animal-based products. Trends in Food Science & Technology 21, 537-543.

Nielsen BH, Angelucci A, Scalvenzi A, Forkman B, Fusi F, Tuyttens F, Houe H, Blokhuis H, Sørenson J and Nielsen JR 2014. Use of animal-based measures for the assessment of dairy cow welfare-ANIBAM. EFSA External Scientific Report.

Nocella G, Hubbard L and Scarpa R 2010. Farm animal welfare, consumer willingness to pay, and trust: Results of a cross-national survey. Applied Economic Perspectives and Policy 32, 275-297.

Nunnally JC, Bernstein IH and Berge JMt 1967. Psychometric theory,226. JSTOR.

Olson JC and Jacoby J 1972. Cue utilization in the Quality Perception Process. Proceedings of the Third Annual Conference of the Association for Consumer Research, 167-179 pp.

Olynk NJ, Tonsor GT and Wolf CA 2010. Consumer willingness to pay for livestock credence attribute claim verification. Journal of Agricultural and Resource Economics, 261-280.

Onyiro OM and Brotherstone S 2008. Genetic analysis of locomotion and associated conformation traits of Holstein-Friesian dairy cows managed in different housing systems. Journal of Dairy Science 91, 322-328.

Onwezen MC and van der Weele CN 2016. When indifference is ambivalence: Strategic ignorance about meat consumption. Food Quality and Preference 52, 96-105.

Oosterkamp EB, Bremmer B, Hoste R and De Greef KH 2011. Verkenning van dierlijke tussensegmenten in onze buurlanden; Duurzaam varkensvlees, pluimveevlees en eieren,2011-028. LEI, onderdeel van Wageningen UR.

Özkan Ş, Vitali A, Lacetera N, Amon B, Bannink A, Bartley DJ, Blanco-Penedo I, De Haas Y, Dufrasne I, Elliott J and Eory V 2016. Challenges and priorities for modelling livestock health and pathogens in the context of climate change. Environmental Research 151, 130-144.

Palupi E, Jayanegara A, Ploeger A and Kahl J 2012. Comparison of nutritional quality between conventional and organic dairy products: a meta-analysis. Journal of the Science of Food and Agriculture 92, 2774-2781.

Phillips C, Izmirli S, Aldavood J, Alonso M, Choe BI, Hanlon A, Handziska A, Illmann G, Keeling L and Kennedy M 2010. An international comparison of female and male students' attitudes to the use of animals. Animals 1, 7-26.

Pouta E, Heikkilä J, Forsman-Hugg S, Isoniemi M and Mäkelä J 2010. Consumer choice of broiler meat: The effects of country of origin and production methods. Food Quality and Preference 21, 539-546.

Presi P and Reist M 2011. Review of methodologies applicable to the validation of animal based indicators of welfare. EFSA Supporting Publications 8.

Reece WO 2009. Functional anatomy and physiology of domestic animals. John Wiley & Sons. Iowa, USA.

Robach M 2010. A private sector perspective on private standards: some approaches that could help to reduce current and potential future conflicts between public and private standards. Technical Item Presented at the 78th World Organisation for Animal Health (OIE) General Session, 23-28 May, Paris.OIE, Paris.Dostupno Na: Www.Oie.Int/Fileadmin/Home/Eng/Internationa_Standard_Setting/Docs/Pdf/A_78SG_9.Pdf.

Roche JR, Friggens NC, Kay JK, Fisher MW, Stafford KJ and Berry DP 2009. Invited review: body condition score and its association with dairy cow productivity, health, and welfare. Journal of Dairy Science 92, 5769-5801.

Rousing T and Wemelsfelder F 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Applied Animal Behaviour Science 101, 40-53.

Rodenburg TB, Tuyttens FAM, De Reu K, Herman L, Zoons J and Sonck B 2008. Welfare assessment of laying hens in furnished cages and non-cage systems: assimilating expert opinion. Animal Welfare 17, 355-361.

Samborski V and Van Bellegem L 2016. De biologische landbouw in 2015. Departement Landbouw & Visserij, Brussel, België.

Sato S, Tarumizu K and Hatae K 1993. The influence of social factors on allogrooming in cows. Applied Animal Behaviour Science 38, 235-244.
200

Schön PC, Puppe B and Manteuffel G 2004. Automated recording of stress vocalisations as a tool to document impaired welfare in pigs. Animal Welfare 13, 105-110.

Schütz KE, Cox NR and Tucker CB 2014. A field study of the behavioural and physiological effects of varying amounts of shade for lactating cows at pasture. Journal of Dairy Science 97, 3599-3605.

Schütz KE, ogers AR, oulouin YA, ox NR and ucker CB 2010. The amount of shade influences the behaviour and physiology of dairy cattle. Journal of Dairy Science 93, 125-133.

Schuppli CA, Von Keyserlingk MAG and Weary DM 2014. Access to pasture for dairy cows: Responses from an online engagement. Journal of Animal Science 92, 5185-5192.

Seegers H, Fourichon C and Beaudeau F 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Veterinary Research 34, 475-491.

Serpell JA 2004. Factors influencing human attitudes to animals and their welfare. Animal Welfare 13, 145-151.

Schellekens A and Van de Ven G 2008. Bewust beperkt beweiden: Beter voor koe en boer? Landbouwcentrum voor Voedergewassen, Geel, Belgium. http://www.lcvvzw.be/wp-content/uploads/2008/01/Bewust-beperkt-beweiden.pdf, accessed 23/11/2016.

Shepherd R 1999. Social determinants of food choice. Proceedings of the Nutrition Society 58.04: 807-812.

Simonson I and Tversky A 1992. Choice in context: Trade-off contrast and extremeness aversion. Journal of Marketing Research 29, 281.

Sparks P and Guthrie CA 1998. Self-identity and the Theory of Planned Behaviour: A useful addition or an unhelpful artifice? 1. Journal of Applied Social Psychology 28, 1393-1410.

Spoolder H, De Rosa G, Horning B, Waiblinger S and Wemelsfelder F 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529-534.

Sprenger M, Vangestel C and Tuyttens FAM 2009. Measuring thirst in broiler chickens. Animal Welfare 18, 553-560.

Spruijt BM, van den Bos R and Pijlman FT 2001. A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. Applied Animal Behaviour Science 72, 145-171.

Tauer LW and Mishra MK 2006. Dairy farm cost efficiency. Journal of Dairy Science 89, 4937-4943.

Taylor N and Signal TD 2009. Willingness to pay: Australian consumers and 'on the farm' welfare. Journal of Applied Animal Welfare Science 12, 345-359.

Te Velde H, Aarts N and Van Woerkum C 2002. Dealing with ambivalence: farmers' and consumers' perceptions of animal welfare in livestock breeding. Journal of Agricultural and Environmental Ethics 15, 203-219.

Toma L, McVittie A, Hubbard C and Stott AW 2011. A structural equation model of the factors influencing British consumers' behaviour toward animal welfare. Journal of Food Products Marketing 17, 261-278.

Tuyttens FA, Vanhonacker F, Van Poucke E and Verbeke W 2010. Quantitative verification of the correspondence between the Welfare Quality® operational definition of farm animal welfare and the opinion of Flemish farmers, citizens and vegetarians. Livestock Science 131, 108-114.

Tuyttens FAM, Federici JF, Vanderhasselt RF, Goethals K, Duchateau L, Sans ECO and Molento CFM 2015. Assessment of welfare of Brazilian and Belgian broiler flocks using the Welfare Quality protocol. Poultry science 94, 1758–176.

Vandehaar, MJ, 1998. Efficiency of nutrient use and relationship to profitability on dairy farms. Journal of Dairy Science 81, 272-282.

Van Dierendonck MC, Sigurjónsdóttir H, Colenbrander B and Thorhallsdóttir AG 2004. Differences in social behaviour between late pregnant, post-partum and barren mares in a herd of Icelandic horses. Applied Animal Behaviour Science 89, 283-297.

van Herpen E, Fischer AR and van Trijp HC 2015. How to position 'mildly sustainable' products: The joint impact of assortment display and price setting. Food Quality and Preference 46, 26-32.

Van Laer E, Moons CPH, Ampe B, Sonck B, Vandaele L, De Campeneere S and Tuyttens FAM 2015. Effect of summer conditions and shade on behavioural indicators of thermal discomfort in Holstein dairy and Belgian Blue beef cattle on pasture. Animal 9, 1536-1546.

Van Kleef E, Van Trijp H, Paeps F and Fernandez-Celemin L 2008. Consumer preferences for front-of-pack calories labelling. Public Health Nutrition 11, 203-213.

Van Loo EJ, Caputo V, Nayga RM and Verbeke W 2014. Consumers' valuation of sustainability labels on meat. Food Policy 49, 137-150.

Vanderhasselt RF, Buijs S, Sprenger M, Goethals K, Willemsen H, Duchateau L and Tuyttens FAM 2013. Dehydration indicators for broiler chickens at slaughter. Poultry Science 92, 612-619.

Vanderhasselt RF, Goethals K, Buijs S, Federici JF, Sans ECO, Molento CFM, Duchateau L and Tuyttens FAM 2014. Performance of an animal-based test of thirst in commercial broiler chicken farms. Poultry Science 93, 1327-1336.

van der Voort M, Van Meensel J, Lauwers L, Van Huylenbroeck G and Charlier J 2016. The relation between input-output transformation and gastrointestinal nematode infections on dairy farms. Animal 10, 274-282.

Vanhonacker F, Van Poucke E, Tuyttens F and Verbeke W 2010. Citizens' views on farm animal welfare and related information provision: exploratory insights from Flanders, Belgium. Journal of Agricultural and Environmental Ethics 23, 551-569.

Vanhonacker F and Verbeke W 2009a. Buying higher welfare poultry products? Profiling Flemish consumers who do and do not. Poultry Science 88, 2702-2711.

Vanhonacker F and Verbeke W 2014. Public and consumer policies for higher welfare food products: challenges and opportunities. Journal of Agricultural and Environmental Ethics 27, 153-171.

Vanhonacker F, Verbeke W, Van Poucke E and Tuyttens FA 2007. Segmentation based on consumers' perceived importance and attitude toward farm animal welfare. International Journal of Sociology of Agriculture and Food 15, 91-107.

Van Winden SC and Kuiper R 2003. Left displacement of the abomasum in dairy cattle: recent developments in epidemiological and etiological aspects. Veterinary Research 34, 47-56.

Veissier I, Jensen KK, Botrea R and Sandøe P 2011. Highlighting ethical decisions underlying the scoring of animal welfare in the Welfare Quality® scheme. Animal Welfare 20, 89.

Veissier I, Butterworth A, Bock B and Roe E 2008. European approaches to ensure good animal welfare. Applied Animal Behaviour Science 113, 279-297.

Verbeke W 2009. Stakeholder, citizen and consumer interests in farm animal welfare. Animal Welfare 18, 325-333.

Verbeke W and Vackier I 2004. Profile and effects of consumer involvement in fresh meat. Meat Science 67, 159-168.

Vermeir I and Verbeke W 2006. Sustainable food consumption: Exploring the consumer attitude-behavioural intention gap. Journal of Agricultural and Environmental Ethics 19, 169-194.

Vickrey W 1961. Counterspeculation, auctions, and competitive sealed tenders. The Journal of Finance 16, 8-37.

VLAM 2016. Belgische export van volle en magere melk. http://www.vlam.be/public/uploads/files/feiten_en_cijfers/zuivel/Belgische_export_volle_en_magere_melk_2006-2015.pdf, accessed 23/11/2016/

Von Keyserlingk MAG, Barrientos A, Ito K, Galo E and Weary DM 2012. Benchmarking cow comfort on North American freestall dairies: Lameness, leg injuries, lying time, facility design, and management for high-producing Holstein dairy cows. Journal of Dairy Science 95, 7399-7408.

Vossler CA and Kerkvliet J 2003. A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behaviour for a public referendum. Journal of Environmental Economics and Management 45, 631-649.

Warnick LD, Janssen D, Guard CL and Gröhn YT 2001. The effect of lameness on milk production in dairy cows. Journal of Dairy Science 84, 1988-1997.

Washburn SP, White SL, Green Jr JT and Benson GA 2002. Reproduction, mastitis, and body condition of seasonally calved Holstein and Jersey cows in confinement or pasture systems. Journal of Dairy Science 85, 105-111.

Weinrich R and Spiller A 2016. Developing food labelling strategies: Multi-level labelling. Journal of Cleaner Production 137, 1138-1148.

Welfare Quality® Consortium 2009. Welfare Quality® Assessment Protocol for Cattle. Lelystad: The Netherlands, pp.

Whay HR, Main DCJ, Greent LE and Webster AJF 2003. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. Animal Welfare 12, 205-217.

Willner P 1997. Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. Psychopharmacology 134, 319-329.

Wilson P 2011. Decomposing variation in dairy profitability: the impact of output, inputs, prices, labour and management. Journal of Agricultural Science 149, 507-517.

Wolf CA, Tonsor GT, McKendree MGS, Thomson DU and Swanson JC 2016. Public and farmer perceptions of dairy cattle welfare in the United States. Journal of Dairy Science 99, 5892–5903.

Curriculum Vitae

Personal information

Name:	Sophie de Graaf
Birth date and place:	27th may 1988
E-mail:	sophiedegraaf@gmail.com
ORCID ID:	orcid.org/ 0000-0002-1676-8925

Education

2012 - Current	Agricultural Economics, PhD candidate Ghent University, Belgium
2009 - 2012	Master in Animal Sciences, specialisation: animal health and behaviour, Wageningen University, Wageningen, the Netherlands
2009 - 2012	Master in Animal Sciences, Swedish agricultural university (Sveriges Landbrucksuniversitet), Uppsala, Sweden
2005 – 2009	Bachelor in Animal management, van Hall Larenstein, Leeuwarden, the Netherlands

Publications

Papers in Scientific Journals

de Graaf S, Van Loo EJ, Bijttebier J, Vanhonacker F, Lauwers L, Tuyttens F & Verbeke W 2016, 'Determinants of consumer intention to purchase animal-friendly milk' Journal of Dairy Science 99, 8304-8313.

de Graaf S, Vanhonacker F, Van Loo EJ, Bijttebier J, Lauwers L, Tuyttens FAM & Verbeke W. 2016. Market opportunities for animal-friendly milk in different consumer segments. Sustainability 8, 1302.

de Graaf S, Ampe B, Tuyttens, FAM. In press. Assessing dairy cow welfare at the beginning and end of the indoor period using the Welfare Quality® protocol. Animal Welfare.

Tuyttens F, de Graaf S, Heerkens J, Jacobs L, Nalon E, Ott S, Stadig L, Van laer E & Ampe B 2014, 'Observer bias in animal behaviour research: can we believe what we score, if we score what we believe?' Animal Behaviour, vol 90, blz. 273-280.

de Graaf S, Ampe B, Buijs S, Andreasen S, De Boyer Des Roches A, van Eerdenburg FJCM, Haskell MJ, Kirchner MK, Mounier L, Radeski M, Winckler C, Bijttebier J, Lauwers L, Verbeke W, Tuyttens FAM. 2016. Sensitivity of the integrated scores of the Welfare Quality® dairy cattle on-farm welfare assessment protocol to changes in individual welfare measures. Under review.

de Graaf S, Ampe B, Winckler C, Radeski M, Mounier L, Kirchner MK, Haskell MJ, van Eerdenburg FJCM, De Boyer Des Roches A, Andreasen SN, Bijttebier J, Lauwers L, Verbeke W, Tuyttens FAM 2016. Comparison of expert opinion with Welfare Quality® integrated scores of dairy cattle welfare. Under review.

de Graaf S, Ampe B, Buijs S, Andreasen S, De Boyer Des Roches A, van Eerdenburg FJCM, Haskell MJ, Kirchner MK, Mounier L, Radeski M, Winckler C, Bijttebier J, Lauwers L, Verbeke W, Tuyttens FAM. 2016. Aggregating dairy cattle welfare measures into a multidimensional welfare index based upon expert elicitation. Submitted.

Bijttebier J, de Graaf S, Van Meensel J, Vanhonacker F, Tuyttens F, Verbeke W, Lauwers L. 2016. Cost efficiency of milk production and animal welfare on dairy farms: trade-offs or win-wins?

Abstracts for international conferences

Tuyttens F, de Graaf S, Stadig, L, Jacobs, L, Benaissa, S & Ampe, B 2016, A method for integrating animal welfare measures into an overall welfare index that reflects expert opinion. in I Reimert, J Wijnen, S Buijs & L Bolhuis (eds), Proceedings of the Benelux ISAE conference 2016. Berlicum, blz. 14.

de Graaf S, Lauwers, L, Bijttebier, J, Verbeke, W & Tuyttens, F 2016, Sensitivity of the integrated Welfare Quality® scores of the dairy cattle protocol to changes in individual measures. in Proceedings of the Benelux ISAE conference 2016. Berlicum, blz. 13.
206

de Graaf S, Vanhonacker, F, Bijttebier, J, Lauwers, L, Tuyttens, F & Verbeke, W 2016, Different consumer segments' interest in information about dairy cattle welfare measures. in Proceedings of the 50th Congress of the International Society for Applied Ethology 12-15th July, 2016, Edinburgh, United Kingdom. Edinburgh, blz. 316.

Tuyttens F, Van laer, E, Stadig, L, Jacobs, L, Heerkens, J, de Graaf, S & Ampe, B 2014, Observer bias in applied ethology: can we believe what we score, if we score what we believe? in Proceedings of the 48th Congress of the International Society for Applied Ethology. Wageningen Academic Publishers, Vitoria, blz. 64.

de Graaf S, Tuyttens, F, Ampe, B (ed.), Bijttebier, J (ed.), Lauwers, L (ed.) & Verbeke, W (ed.) 2013, 'Application of the WQ protocol® for assessing dairy cattle welfare on Flemish farms: identification of risk factors and predictors for welfare outcomes', Lille, France, 11/12/13.

de Graaf S, Bijttebier, J, Lauwers, L, Vanhonacker, F, Verbeke, W & Tuyttens, F 2013, Marketing animal welfare as a quality characteristic of milk. in L Goby (ed.), 6th Boehringer Ingelheim Expert Forum on Farm Animal Well-Being. vol. 6, Boehringer Ingelheim, Ingelheim am Rhein, blz. 17 - 19.

de Graaf S, Ampe B, Winckler C, Radeski M, Mounier L, Kirchner MK, Haskell MJ, van Eerdenburg FJCM, De Boyer Des Roches A, Andreasen SN, Bijttebier J, Lauwers L, Verbeke W, Tuyttens FAM 2016. Correspondence of expert opinion with integrated Welfare Quality® scores of the dairy cattle welfare assessment protocol. Welfare Quality Gathering 2016.

Other Reports or publications

Fiems, L, Tuyttens, F, De Sutter, R, de Graaf, S & Sonck, B 2016, Veilig omgaan met runderen. vol. 213, Instituut voor Landbouw- en Visserijonderzoek, Merelbeke.

Bodil Højlund Nielsen, Alessandra Angelucci, Alessandra Scalvenzi, Björn Forkman, Francesca Fusi, Frank Tuyttens, Hans Houe, Harry Blokhuis, Jan Tind Sørensen, Janne Rothmann, Lindsay Matthews, Luc Mounier, Luigi Bertocchi, Marie-Madeleine Richard,

Matteo Donati, Per Peetz Nielsen, Romolo Salini, Sophie de Graaf, Sophie Hild, Stefano Messori, Søren Saxmose Nielsen, Valentina Lorenzi, Xavier Boivin & Peter T. Thomsen, 2014. Use of animal based measures for the assessment of dairy cow welfare ANIBAM EFSA Supporting Publication 2014; 11(9):EN-659, 340 pp. doi:10.2903/sp.efsa.2014.EN-659

de Graaf, S, Tuyttens, F, Bijttebier, J, Lauwers, L, Verbeke, W & Vanhonacker, F 2014, 'Het vermarkten van diervriendelijk geproduceerde melk' Vlaamse Dierenartsvereniging - VDV, vol 4, nr. 196, blz. 30-34.

Professional Memberships

International Society for Applied Ethology (ISAE) & European Federation of Animal Science (EAAP)

Dankwoord

Na viereneenhalf jaar werk ligt hij er dan, mijn doctoraatsboekje. Het is één van de leukste en moeilijkste dingen die ik ooit gedaan heb. Natuurlijk was dit niet gelukt zonder de hulp van een flink aantal mensen.

Ten eerste, Wim Verbeke bedankt voor deze kans, en het delen van je kennis en snelle geest. Bedankt ook Ellen, voor je hulp met achtergronden en analysemethodes. Dan de laatste in het consumentengedeelte: heel veel dank voor Filiep. Je bleef geduldig als ik verdwaald raakte in de theorieën van de consumentenwetenschappen en was altijd bereid te helpen met veel kennis en inzicht.

Ludwig bedankt voor je eeuwige enthousiasme, en je soms licht onduidelijke maar toch altijd duidelijk behulpzame pogingen mij te helpen bij dit onderzoek. Bedankt Jo voor het nalezen van papers, het helpen organiseren en het nadenken over onderzoeksopzet. En tot slot bedankt Jef voor het geduldig uitleggen van de efficiëntie-analyse toen ik het bijna snapte maar net niet helemaal.

Dan mijn derde en laatste promotor: heel veel dank Frank. Dankjewel voor de eindeloze discussies tussen twee koppige mensen en voor het dealen met mijn sporadische chaos. Ik kon er altijd op vertrouwen dat jij me zou helpen en bij zo'n ambitieus project was dat heel fijn. Ik bewonder je wil om overal het meeste uit te willen halen, ook al leidt dit soms tot zuchtende en steunende doctoraatsstudenten.

Dan de mensen van de dierenwelzijnsgroep. Bedankt voor alle pizzalunches, weekendjes weg en andere leuke dingen. Bedankt Dimitry dat je twee winters lang half bevroren op een ladder naar koeien hebt gekeken, zonder te klagen. Leo en Liesje, leden van onze familiegroep van vrienden. Dankjewel voor jullie eerste hulp bij vreselijke zinnen en te lange abstracts, voor het af en toe zeggen dat ik me misschien wat aanstelde, en voor alle cava die we samen naar binnen gegooid hebben. Ik ben heel blij dat we samen in het exotische België zaten, ik zou jullie niet willen missen!

Natuurlijk ben ik Bart Ampe ook veel dank verschuldigd, door jou was statistiek gezellig en mondde een afspraak altijd uit in bijzondere Google-sessies en het verbolgen reageren op de verschillen in woordkeuzes tussen Nederland en België.

I would also like to thank the European dairy cattle welfare experts who filled out two very time-consuming surveys and co-authored three of the papers in this thesis. Marlene, Marie, Miroslav, Alice, Sine, Christoph, Frank and Luc, thank you for all your input, passion and enthusiasm for animal welfare research.

Dan is er nog één iemand die ik specifiek moet bedanken en dat is (thuis)Wim. Bedankt voor je steun en dat je me niet altijd even serieus neemt.