A validation study of the pre-recorded data-based herd status index for dairy herd welfare identification

BY

NAZIYA MAUYENOVA TUYAKOVNA

Department of Animal Science

McGill University, Montreal

AUGUST 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Science

In

Animal Science

ABSTRACT

Animal welfare concerns have led to the development of welfare standards, i.e., assessment protocols to ensure that animals' needs are provided. Such protocols currently in use mainly focus on outcome measures directly assessed on an individual, reflecting the measure of welfare as the state of individual animals (and not as a group). However, the farm visits necessary for implementing such assessment protocols require a considerable amount of time and associated high costs. Lately, on the other hand, groups of researchers have proposed pre-screening assessment tools using pre-recorded data with the objective of remote detection of herds with the highest welfare issues, and, therefore, reducing the number of farm visits to those herds in need of intervention. Nevertheless, applying a new assessment method requires the evaluation of its validity relative to an existing assessment method. Such an evaluation is needed to ensure that the proposed method is reliable and corresponds to its objective for which it was developed. Hence, this thesis aimed to determine the validity of the herd status index (HSI) to identify an overall state of dairy cattle welfare at the herd level by identifying its performance level and the correspondence of its indicators relative to the proAction® on-farm outcome welfare assessment method.

Farm-level data for five outcome measures of welfare – lameness, body condition, hock, neck, and knee injuries scores, collected as a part of the proAction® Quality Assurance Program were integrated with pre-recorded test-day dairy herd improvement (DHI) data for the three years before the on-farm assessment, extracted from the Lactanet Inc. (Sainte-Anne-de-Bellevue, QC, Canada) database. Two-stage cluster analysis was performed to partition study herds into subgroups based on five-dimensions – outcome measures of welfare, which resulted in four distinct groups of herds classified as groups with the least (C1), the highest (C4), and average (C1, C3) welfare issues. The clusters significantly differed ($P < 0.05$) from each other regarding all five-dimensions, except for the prevalence of neck injuries of herds in C1 and C2. Followed by

the cluster analysis, the HSI was calculated for each of the study herds based on the method developed by Warner *et al.* (2020). The findings showed that the HSI based on twelve pre-recorded DHI indicators could identify herds with the highest welfare issues relative to herds' classification based on proAction® on-farm welfare assessment data.

With regards to individual pre-recorded indicators of the HSI, five out of twelve indicators – involuntary replacement and mortality rates, herd management and transition cow indexes, and prevalence of cows with high SCC > 400,000 cells/ml in milk significantly differed between herds in C2 and C4, that was in complete correspondence with the classification of herds based on outcome measures of welfare. Thus, these five indicators' contribution to the HSI's overall performance level is substantial and corresponds to the objective, making the HSI a comparatively valid method to identify herds with the highest welfare issues. However, in terms of the remaining indicators of the HSI, this study revealed no significant results between study clusters; therefore, it may indicate that the contribution of these indicators into the overall performance level of the HSI is not essential and they should be reconsidered based on scientific evidence and can be replaced by those which were shown to hold high potential.

RESUMÉ

Les préoccupations en matière de bien-être animal ont conduit à l'élaboration de normes de bien-être animal, c'est-à-dire de protocoles d'évaluation pour garantir que les besoins des animaux sont satisfaits. Ces protocoles actuellement utilisés se concentrent principalement sur des mesures de résultats directement évaluées sur les individus reflétant que la mesure du bien-être est l'état de chaque animal (et non l'état du groupe). Cependant, les visites à la ferme nécessaires à la mise en œuvre de tels protocoles d'évaluation nécessitent un temps considérable et des coûts élevés. Récemment, des groupes de chercheurs ont proposé des outils d'évaluation utilisant des données préenregistrées dans le but de détecter à distance les troupeaux présentant les problèmes de bien-être les plus élevés et, par conséquent, de réduire le nombre de visites à la ferme nécessaires, notamment en ciblant ces troupeaux nécessitant une intervention. Néanmoins, l'application d'une nouvelle méthode d'évaluation nécessite l'évaluation de sa validité par rapport à une méthode existante. Une telle évaluation est nécessaire pour s'assurer que la méthode proposée est fiable et correspond à l'objectif pour lequel elle a été développée. Ainsi, cette thèse vise à déterminer la validité de l'indice de statut du troupeau (HSI) pour l'identification d'un état global de bien-être des bovins laitiers au niveau du troupeau en identifiant son niveau de performance et la correspondance de ses indicateurs par rapport à l'évaluation du bien-être réalisée dans le cadre du programme proAction®.

Les données au niveau du troupeau pour cinq mesures de résultat du bien-être - boiterie, état corporel, blessures au jarret, au cou et au genou, recueillies dans le cadre du programme proAction® ont été intégrées aux données préenregistrées DHI (Dairy Herd Improvement) pour les trois années précédant l'évaluation à la ferme, extraites de la base de données Lactanet Inc. (Sainte-Anne-de-Bellevue, QC, Canada). Une analyse par grappes en deux étapes a été réalisée pour répartir les troupeaux de l'étude en sous-groupes présentant les caractéristiques de bien-être

les plus similaires sur la base de cinq dimensions - mesures des résultats du bien-être, ce qui a abouti à quatre groupes distincts de troupeaux classés comme des groupes avec le moins (C1), le plus élevé (C4) et les problèmes de bien-être moyen (C1, C3). Les grappes données différaient significativement ($P < 0,05$) les unes des autres en ce qui concerne les cinq dimensions, à l'exception de la prévalence des blessures au cou des troupeaux en C1 et C2. Suite à l'analyse par grappes, le HSI a été calculé pour chacun des troupeaux de l'étude selon la méthode développée par Warner *et al.* (2020). Les résultats de cette thèse ont montré que le HSI basé sur douze indicateurs préenregistrés DHI peut identifier les troupeaux avec les problèmes de bien-être les plus élevés par rapport à la classification des troupeaux basée sur les données d'évaluation du bien-être proAction® à la ferme.

En ce qui concerne les indicateurs individuels préenregistrés du HSI, cinq des douze indicateurs - taux de remplacement involontaire et de mortalité, gestion du troupeau et indices des vaches en transition, et prévalence des vaches avec un CCS élevé > 400000 cellules/ml dans le lait différaient significativement entre les troupeaux des groupes C2 et C4, qui correspondaient parfaitement à la classification des troupeaux basée sur les mesures de résultat du bien-être (proAction®). Ainsi, la contribution de ces cinq indicateurs au niveau de performance global du HSI est substantielle et correspond à l'objectif du HSI, ce qui fait du HSI une méthode relativement valable pour identifier les troupeaux présentant les problèmes de bien-être les plus élevés. Cependant, en ce qui concerne les indicateurs restants du HSI, cette étude n'a révélé aucun résultat significatif entre les grappes d'études. Par conséquent, cela peut indiquer que la contribution de ces indicateurs au niveau de performance global du HSI n'est pas essentielle et qu'ils devraient être reconsidérés sur la base de preuves scientifiques et pourraient être remplacés par ceux dont il a été démontré qu'ils ont un potentiel élevé.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Elsa Vasseur, for her support and guidance throughout my Master's degree. Thank you for giving me the excellent opportunity to be a part of your lab and conduct research among the most aspiring and supportive people. I am also incredibly grateful for your support in aspects of my personal life. I would also like to thank my co-supervisor, Dr. Roger Cue, for your invaluable guidance and help throughout my academic years. Thank you for helping me with the study methodology, statistical analyses, and written thesis. I cannot express enough thanks to my committee member, Daniel Warner, for your valuable contribution, particularly for helping to extract the study data, develop the methodology, and perform data analysis.

I also would like to express many thanks to the CowLife McGill lab members for always being supportive. I have learned valuable lessons from each of you! Special thanks to Gabriel Machado Dallago for your help in developing the methodology, guidance on statistical analyses, and always being encouraging.

My sincere gratitude goes to Dairy Farmers of Canada and Lactanet Inc. for providing the data to implement my research. I would also like to express my appreciation to the government of Kazakhstan for funding to pursue my academic goals.

I would like to thank my parents, Tuyak and Panu, my brother Bakhtiyar, my sisters, Kyzgaldak, Nazym, and Zhuldyz, for your love, support, and prayers. When there are people who believe in you, the impossible things become possible! Special thanks to my dear friend DP for your continued support and motivation, for always being there, particularly these challenging times when the entire world faced difficulties due to the COVID-19 pandemic.

Thank you all again! It has been an honor to work with you!

TABLE OF CONTENTS

TABLE OF CONTENTS: FIGURES

TABLE OF CONTENTS: TABLES

TABLE OF CONTENTS: SUPPLEMENTAL MATERIAL

1. INTRODUCTION

The importance of maintaining "good" welfare for economic and ethical reasons could be facilitated by the implementation and its verification of welfare standards on dairy herd care and management (Rushen *et al.*, 2011). To this end, in Canada, Dairy Farmers of Canada have developed the proAction® Quality Assurance Program that addresses the quality milk production, food safety, dairy cattle welfare, and environmental aspects of the dairy industry (DFC, 2013). Regarding the welfare of a dairy cow, under the "Animal care" assessment protocol of the proAction® initiative, Canadian dairy farms are examined for the adherence to the determined national welfare standards in the "Code of Practice for the Care and Handling of Dairy Cattle" (NFACC-DFC, 2009) regarding animal care, housing, and management. The assessment protocol mainly consists of outcome measures of welfare (also named animal-based measures), particularly lameness, injuries, and body condition, and its verification requires the examination of an individual cow from a pre-selected sample for those measures.

Acknowledging that outcome measures are considered to reflect the direct response of a dairy cow to its environment, particularly the conditions and care she is given, the preference for outcome measures to assess dairy cattle welfare may be justified. However, despite the numerous studies assessing such measures' validity, particular concerns remain on the practical implementation of those outcome measures regarding large-scale herds' visual examination (Knierim & Winckler, 2009). Every individual animal's visual examination for each outcome measure is labor-intensive, hence, costly (Sørensen *et al.*, 2007; Vasseur, 2017). Another concern regarding outcome-based welfare assessment is the consistency of assessment results among different assessors as even with repeated training and continuous retraining, intra- and inter-observer repeatability are not perfect (Vasseur, 2017). Finally, since only those previously included in the sample are evaluated, this approach may pose a risk to cows that need immediate close observation by excluding them from the sample (de Vries *et al.,* 2014). Also, time constraints may prevent timely identification of health problems and, consequently, delay the application of necessary measures (De Vries *et al.*, 2016).

Considering the mentioned challenges in performing on-farm welfare assessment using visual observation of outcome measures, several authors have proposed an alternative approach using pre-recorded data from national databases that do not require farm visits (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014; Otten *et al.*, 2016). Such an alternative, in its turn, allows increasing the effectiveness of welfare assessment programs by providing timely assistance for herds in higher need and a considerable reduction in farm visits (De Vries *et al.*, 2016). Moreover, the uniformity of the indicators registered to such databases enables one to compare herds based on corresponding aspects of farming practices, and, thus, adopt the practical skills from the high-performance herds (Brouwer *et al.*, 2015). Also, the uniform collection of such data across different countries makes it possible to inter-exchange the management practices between them. Finally, it will allow dairy herds' welfare to be continuously monitored and systematic issues related to housing and management to be detected (Otten *et al.*, 2016). When it comes to multi-factorial data, the composite index is the most appealing method to synthesize a multitude of information into a single index (Santeramo, 2017). In 2020, such an approach was applied to the development of the herd status index (HSI) in Canada, which has been proposed as a pre-screening tool to identify the overall state of welfare at the herd level (Warner *et al.*, 2020) in the context of advisory services, in order to target herds in need of a follow-up on-farm assessment to identify potential risk factors to cow welfare in individual farms.

Simultaneously, due to the multidimensionality of pre-recorded data with a wide range of variables at both the animal and herd levels, the determination of the exact variables that correspond to specific outcome measures of welfare is challenging. A number of pre-recorded dairy herd improvement (DHI) indicators were associated with some outcome measures of welfare (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014). Several studies, for instance, have established positive relationships between prevalence of lameness and some DHI indicators such as age at first calving (Rutherford *et al.*, 2009), culling and mortality rates (Booth *et al.*, 2004; Bicalho *et al.*, 2007; Thomsen & Sørensen, 2009). Nevertheless, some other studies' results were in contradiction with previous results, indicating the existence of no associations. Milk quality-related indicators are one of the earliest measures recorded and used as a part of health monitoring. A recent study (Ginestreti *et al.*, 2020) has estimated the potential of milk quality data to screen herds with poor welfare by conducting correlation analysis; however, the results showed weak or no correlations with outcome measures of welfare.

The results of those previous studies are not consistent and contradict each other. At the same time, some show the existence of relationships among outcome measures of welfare and pre-recorded DHI indicators and indicate their potential to use for the identification of welfare state at the herd level; others found no or a few associations, thus questioning the value of such data and the assessment method based on that. Moreover, even though some literature has shown the relationships amongst existing welfare measures and pre-recorded DHI indicators, it is essential to note that the relationship at the individual level is not necessarily the same at the herd level (De Vries *et al.*, 2011; Ginestreti *et al.*, 2020). Another concern is the method used to develop an assessment method based on such data. Previous studies (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014) have applied different approaches to develop a scoring system based on selected DHI indicators that demonstrated moderate levels of sensitivity and specificity that was dependent on the choice of indicators. Hence, it is appropriate to question whether the assessment approach based on pre-recorded data is a valid method for the identification of dairy cattle welfare at the herd level relative to its ability to encompass all three essential aspects of welfare as defined by Fraser (2003) (i.e., biological functioning, affective state, and natural behavior). It is also crucial to determine if the indicators are consistent with currently recognized well-being indicators (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014). Therefore, in this sense, this thesis aims at answering these questions by evaluating the validity of the HSI relative to proAction® outcome measures of welfare in Quebec dairy herds.

2. REVIEW OF LITERATURE

2.1 The definition of animal welfare

Animal welfare is interpreted as a state of an individual animal that is dependent on a range of internal and external factors. Since animal welfare has initially arisen as an issue due to the impaired welfare of farm animals and increased public concern, it has undergone certain changes pertaining to its interpretation and what is considered as "good" welfare. With regards to the definition of animal welfare, one of the earliest publications defines it as the level of an animal's coping with its living environment (Broom, 1986), which is the definition used by the World Organization for Animal Health (World Organization for Animal Health, 2008), as well. Today, most scientists focus on three main welfare areas: biological functioning, affective state, and natural behavior (Von Keyserlingk *et al.*, 2009). While biological functioning and affective state focus on animal's health and freedom from pain, hunger, or thirst, natural behavior accentuates the importance of animal's ability to live according to its natural environment (Fraser *et al.*, 1997; Fraser, 2003; Von Keyserlingk *et al.*, 2009). Based on this three-area approach, an animal is considered to have "good" welfare when it is physically and mentally healthy, in other words, is not experiencing pain or stress, and capable of fulfilling its behavioral needs. Nevertheless, depending on the area of interest, the focus areas are prioritized, and one area may be given more importance while reducing the importance of others (Fraser, 2008). Hence, considering that all these three areas are essential for ensuring "good" welfare, in principle, an animal's physiological, behavioral, and emotional needs must be met in animals under society's care (Rushen *et al.*, 2011).

2.2 The current state of Canadian dairy cattle welfare

Intensification of the dairy industry has been an area of concern and critique related to what consequences it has led to and what aspects it has impacted over the last decades, and animal welfare is one of the fields that have been targeted (Fraser, 2005). Driven mainly by economic reasons, the intensification of the dairy industry has focused on increasing milk production relative to a reduction of input factors (Clay *et al.*, 2020), and consequently, milk production per cow has seen a phenomenal increase. As an example, in 2019, the Canadian dairy sector produced 92,26 million hectoliters of milk with an average milk yield of 10,519 kg/cow per year, which is 42% higher compared to the milk yield in 1990 (7,412 kg/cow per year) (Canadian Dairy Information Center, 2019). At the same period, however, dairy farms have experienced a reduction in farm numbers and a higher concentration of cows with an increased average herd size of up to 85 cows per herd in 2019 (Jelinski *et al.*, 2015; Barkema *et al.*, 2015; Robbins *et al.*, 2016). Furthermore, dairy farms have shifted from traditional farms with seasonal pasture access to full-confinement in stall-based housing systems, which is comparatively labor, resource, and cost-effective for large-scale farms (Barkema *et al.,* 2015). As a result, nowadays, most Canadian dairy farms are mainly equipped with tie- and free-stall housing systems, with 73% and 27%, respectively (Canadian Dairy Information Center, 2020).

However, all the changes that have occurred due to the intensification have not gone without negative consequences for the health and welfare of dairy cows (Barkema *et al.,* 2015). The shift to total confinement in stall-based housing systems is recognized as one of the main factors, significantly reducing animal welfare due to movement constraints and inappropriate stall design that may not meet a cow's needs (von Keyserlingk *et al.*, 2012; Bouffard *et al.*, 2017). Reproductive problems, mastitis, and lameness are highly prevalent welfare issues in both tie- and free-stall housing systems and the leading causes of cow removal from the herd (Rozzi *et al.*, 2007), accounting for a total of about 40% of culling in Canadian dairies (Canadian Dairy Information Center, 2020). For instance, lameness is one of the most concerning welfare issues among current dairy farms with the herd-level prevalence of 25% and 21% in the tie- and free-stall housing systems, respectively (Bruijnis *et al.*, 2013; Bouffard *et al.*, 2017; Jewell *et al.*, 2019a) and which results in experiencing pain (Croyle, 2020) leading to further negative consequences such as reduced production and reproduction (Chapinal *et al.*, 2013).

Furthermore, the prevalent welfare issues, in particular, mastitis and lameness, are costly (Espejo *et al.*, 2006) and have significant negative impacts on farm profitability due to reduced longevity, low performance, and disease treatment costs (von Keyserlingk *et al.*, 2012; Delgado *et al.*, 2017). Longevity is an important indicator of welfare and dairy farm profitability, which is defined as a period from first calving until removing a cow from the herd (Delgado *et al.,* 2017). It is also referred to as a productive lifetime since it is a period when a cow starts to produce milk and brings economic value to a farm. The average productive lifetime has seen a substantial decline from 3.5-4.0 lactations in 1970 to currently 2.5-3.0 lactations (Delgado *et al.,* 2017), which is often inferred to happen due to impaired welfare as a result of intensive production. A cow's longevity directly depends on culling, which is either involuntary - caused by health conditions or the cow's ability to reproduce, or voluntary – a conscious decision made by a farmer based on economic interest (Smith *et al.*, 2000; Delgado *et al.*, 2017). The decision to voluntarily remove a cow from the herd is mainly based on economic considerations when a cow does not meet production expectations. When early culling is inevitable, it poses an economic risk for a farm since it is estimated that a breakpoint is only reached when a cow is at three or higher lactations (Delgado *et al.,* 2017). Hence, longevity is an essential indicator of welfare in the case when a cow is involuntarily culled and an important indicator of farm profitability, improvement of which is highly beneficial for both a cow and a farmer (Bruijnis *et al.*, 2013; Larsson *et al.*, 2016).

Based on the preceding, the current state of dairy cattle welfare can be considered a successive circle of complex issues that can be solved by enhancing animal welfare. Hence, the development of a welfare assessment program that enables detection of welfare-associated risk factors and the obtention of reliable estimators of the true welfare status is crucial for animal health, farm profitability, and consumer satisfaction (Whay *et al.*, 2012; Tremetsberger *et al.*, 2015).

2.3 Measures of welfare and on-farm welfare assessment

The importance of maintaining "good" welfare for economic and ethical reasons on farms may require the implementation of welfare standards (Rushen *et al.*, 2011) for dairy herd care and management. However, the complex nature of animal welfare makes it challenging to provide an assessment that could give the closest image of welfare and can encompass a range of its dimensions (i.e., biological functioning, affective state, and natural behavior: (Fraser, 2003)).

The individualistic welfare assessment is a widely accepted approach. Taking into account that animal welfare is considered a state of an individual animal, outcome measures are increasingly preferred for assessing animal welfare due to their direct link to the animal itself. Therefore, outcome measures are believed to reflect the animal's response to the input factors (Whay *et al.,* 2003), while the latter hold low potential due to their indirect nature and complex interaction with other resource and management factors (Knierim & Winckler, 2009). In general, it is recognized that both input and outcome measures are essential indicators of animal welfare and that the most valid assessment is reached when those measures are combined (Johnsen *et al.*, 2001). In 2013, based on such an individualistic approach, Dairy Farmers of Canada developed the "Animal Care" protocol as a part of the proAction® Quality Assurance Program. The "Code of practice for the Care and Handling of Dairy Cattle" (Code of Practice) developed by the National Farm Animal Care Council and Dairy Farmers of Canada (NFACC-DFC, 2009) is the main

Canadian reference document that provides both requirements and recommendations regarding animal health, housing, feeding and management of dairy cattle. Dairy farms are assessed against compliance with the requirements and recommendations given in the Code of Practice. In Europe, the Welfare Quality® (2009) assessment protocol for dairy cattle is the most common reference related to welfare evaluation on farms. Both proAction® Quality Assurance Program and the Welfare Quality® protocol mainly consist of individual outcome measures of welfare, among which lameness, injuries, body condition, and cleanliness are the most commonly used.

2.3.1 Input measures of welfare

Input measures of welfare are referred to housing and management practices, which are measured in an animal's environment and which effect on animal welfare outcomes have been recognized. The input factors, in particular, stall dimensions, such as stall width, height, the quality of the materials used, are essential for cow welfare and comfort (Code of Practice, 2009). Epidemiological studies have shown the risk factors associated with the prevalence of lameness or body injuries with housing or management aspects such as stall design aspects or stocking density (Chapinal *et al.*, 2013; Solano *et al.*, 2015; Jewell *et al.*, 2019a). Input measures can be relatively easily recorded (Johnsen *et al.,* 2001; Whay *et al.,* 2003); therefore, they have been used frequently for regulatory purposes, e.g., in European countries (Otten, 2014). Many countries have developed recommendations related to the environment of the animal. The Canadian Code of Practice (2009) contains recommendations regarding housing and management practices, such as, stall dimensions, quality of bedding material, and stocking density (NFACC-DFC, 2009) that have to be met to ensure that cows' comfort needs are met. However, due to advisory character and financial constraints related to infrastructural changes, the recommendations in the Code of Practice (2009), only voluntary upon the recent implementation of the proAction® initiative, were not necessarily followed by producers (Bouffard *et al.*, 2017). Research on dairy herds housed in tie-stall housing systems showed that not following the recommendations defined in the Code of Practice (2009) led to a higher risk of lameness, neck and knee injuries, and reduced the lying time (Bouffard *et al.*, 2017). A follow-up study (Boyer *et al.*, 2020) evaluated the effect of increased stall width and chain length from recommendations, which revealed positive results such as improved resting, lessened injuries prevalence, and improved ease of movement, indicating that some changes regarding stall dimensions may lead to improved welfare and comfort.

2.3.2 Outcome measures of welfare

Outcome measures, also referred to as animal-based measures, evaluate the animal's responses to its environment (i.e., housing and management) it is provided with, therefore considered as "true" measures of dairy cattle welfare. Key welfare outcome measures mainly include health-related indicators such as an incidence of lameness and different types of body injuries, such as hock, neck, and knee injuries, since they cause moderate to severe levels of pain and reduce cow comfort (Code of Practice, 2009). For instance, lameness has shown to affect health negatively, hence, the welfare of a dairy cow, which results in experiencing pain (Chapinal *et al.*, 2009; Croyle, 2020), leading to further negative consequences such as reduced milk production and reproduction performances (Solano *et al.*, 2015; Jewell *et al.*, 2019b). Body condition is another critical welfare parameter that is indicative of a farm's nutrition and management practices (Roche *et al.*, 2009). The effect of body condition, either excessively thin or excessively fat, on dairy cattle's health and welfare has been reported (Garnsworthy, 2006; Roche *et al.*, 2009; Drackley, 2016). Research has shown that under-conditioned cows at calving have low production and reproduction. In contrast, cows with high BCS at calving have shown to be at a high risk of metabolic disorders, and, consequently, reduced reproductive performance and increased disease susceptibility as a cause of negative energy balance (Garnsworthy, 2006). As an illustration of it, Rasmussen *et al.* (1999) demonstrated that cows with a BCS ≥ 3.5 at calving had a higher risk of ketosis than those with BCS ≤ 2. Moreover, it may suggest that under- and over-conditioned cows are likely to have reduced productive lifetime, further leading to substantial financial losses for a farm (Garnsworthy, 2006).

Furthermore, some research has shown that lameness and BCS could be inter-related. Espejo *et al.* (2006), for instance, reported that cows with low body conditions had a higher prevalence of clinical lameness as compared to those with normal or higher body conditions, which according to authors might have taken place due to the reduced nutrient intake as a cause of impaired mobility. The study by Randall *et al.* (2015) was in line with those findings, showing that cows with BCS < 2 (on a categorical scale with 0.25 increments) had the highest risk for having lameness, while BCS ≥ 2.5 was suggested as an optimal score in order to avoid the risks of lameness in dairy cows. A more recent study (Jewel *et al.*, 2019) has demonstrated that low BCS (BCS ≤ 2.5) is one of the risk factors that may lead to lameness, thus suggesting the improvement of the condition to reduce the lameness prevalence.

Overall, the outcome measures (i.e., lameness, BCS) included in the assessment protocols may represent different aspects of dairy cattle welfare; however, specific criteria need to be followed to ensure outcome-based welfare assessment method validity.

2.4 The correspondence of outcome measures of on-farm welfare assessment to the main assessment criteria: validity, reliability, and feasibility

EFSA (2012) states that the assessment based on outcome measures is an acceptable welfare assessment method. When it comes to any assessment approach, validity, reliability, and feasibility are the main criteria that have to be met (Flower & Weary, 2006; Knierim & Winckler, 2009), which refer to the correspondence to its original purpose; specifically, when referring to its implementation, and the consistency of the assessment results. Acknowledging that welfare outcome measures are considered to reflect the animal's response to its environment (i.e., housing and management), the preference for outcome measures to assess dairy cattle welfare may be justified. However, despite the commonly-used outcome-based assessment methods (Knierim & Winckler, 2009), specific concerns remain on the consistency of outcome measures in the current use and practicality of their scoring, specifically, with regards to the visual examination of large-scale herds performed by different assessors (Vasseur, 2017; Sandgren *et al.*, 2009). Finally, visual assessments of only animals from a pre-selected sample may carry a risk of leaving animals in a higher need on the side (Lundmark *et al.*, 2015). Given the mentioned facts, the existing on-farm welfare assessment consisted predominantly of individual cow-based welfare indicators, does not necessarily meet all the above-mentioned criteria.

2.4.1 Validity: meeting the objectives

In order for a selected indicator to be used as a measure of welfare, it needs to be validated (Rushen *et al.*, 2011) before its inclusion into an assessment protocol. Validity is an important criterion, which implies that the developed method meets its original purpose (Knierim & Winckler, 2009). Regarding key outcome measures of welfare, numerous experimental studies have been conducted to estimate their validity.

With regards to lameness, there are different methods available to detect lameness in dairy cows. However, gait scoring on a 5-point scale (Flower & Weary, 2006) and stall lameness scoring (SLS) focusing on specific behaviors of a cow (Gibbons *et al.*, 2014) are commonly used methods in Canada on farms with free- and tie-stall housing systems, respectively. Gait scoring is a visual examination of cows for the absence or presence of behavioral criteria such as gait types, back position, and others, that rates cows from 1 (sound) to 5 (severely lame) (Flower & Weary, 2006). The validity of the gait scoring to detect lameness was demonstrated by Flower and Weary (2006), who reported that the gait scoring was able with a 92% accuracy to distinguish cows with sole ulcers from healthy ones. A study conducted by Rushen *et al.* (2007) validated lameness detection by gait scoring and weight-bearing with and without a local anesthetic, results of which showed that after the injection of an anesthetic the lame cows had better gait scores and placed more weight on a leg with an injury as opposed to before injection. However, Thomsen *et al.* (2008) argued that the commonly used 5-point scale system does not necessarily correspond strictly to its corresponding definition of each point-scale, i.e., absence and the level of severity of lameness.

Since the cows kept on farms with tie-stalls are restricted in their movements compared to those housed on farms with free-stall housing systems, gait scoring is not applicable unless cows are untied and walked to be assessed for locomotion. Therefore, SLS has been developed to measure lameness in tied cows, which includes observing specific behaviors of a cow such as weight shifting, standing on the edge of the stall, and uneven weight bearing (Gibbons *et al.*, 2014). Gibbons *et al.* (2014) have demonstrated that SLS is a valid method to detect lameness on-farms with tie-stalls relative to gait scoring, where SLS was able to identify lame cows with the specificity of 0.77 and a mean accuracy of 71.7%. In contrast, a study by Palacio *et al.* (2017) showed a substantial difference (9.6%) in the prevalence of lameness identified by SLS and gait scoring, therefore illustrating the potential risk of SLS to underrate the prevalence of lameness of cows in tie-stalls.

2.4.2 Reliability: the consistency of outcome

Another criterion of the assessment is the reliability of its outcome. Intra- and inter-assessor reliability is an essential constituent of the assessment method referring to the consistency of results obtained when the assessment of the object/subject is performed by the same and different assessors, respectively (EFSA, 2012; Bokkers *et al.,* 2012). It is the main challenge to achieve an agreement between different assessors in performing the assessment of outcome measures, which requires specific skills, therefore, a certain level of qualification of assessors. EFSA (2012) emphasizes the importance of training assessors in order to achieve reliable results. Given this issue, numerous attempts have been made, i.e., visual assessment charts and training programs have been developed to ensure consistency and reliability of assessment results. Nevertheless, despite the efforts made to reach high intra- and inter- assessor repeatability, the issue remains open. Among the key outcome measures used in the dairy cattle welfare assessment, body condition scoring used to evaluate, either visually or by palpating, the proportion of body fat, where a low score indicates emaciation and a high score an excessive proportion of fat (Roche *et al.*, 2004; Vasseur *et al.*, 2013). In 2013, a group of researchers estimated the effect of training on the consistency of assessment results between different assessors. The assessors obtained 1-wk classroom and live training sessions, where assessors were provided with a previously developed body condition scoring chart facilitated based on Elanco Animal Health BCS chart (Elanco Animal Health, 1996), and which included photography and detailed description of each body part to be evaluated. The training level was set to be achieved at a coefficient of 0.80 or higher. The results revealed that all trained assessors achieved a high agreement of 0.80 with trainers by the end of the training session, therefore showing the effectiveness of training before the actual assessment. In a recent study, Croyle *et al.* (2018) estimated the effectiveness of training by conducting 3d training to score six outcome measures of welfare, particularly hock injuries, lameness, BCS, and cleanliness. The study results demonstrated a high coefficient of inter-assessor agreement mean of 1.00, 0.90, and 0.85 for BCS, cleanliness, and hock injuries, respectively. However, the coefficient of inter-assessor agreement for lameness scoring was the lowest (0.66) among all indicators included in the study, which slightly increased up to 0.74 during the repeated assessment due to a follow-up video training. The authors concluded that training contributes to the improvement of inter-assessor repeatability of key animal-based welfare measures. However, with regards to the lameness scoring, it remains challenging to achieve a high inter-assessor agreement compared to the rest of the key welfare indicators, which is in accordance with the previous findings (Thomsen *et al.*, 2008). As opposed to commonly used lameness, injury, or BCS, the emotional state assessment (i.e., active, positively occupied, happy, distressed) has shown lower reliability levels, which do not necessarily increase with assessors' training (Bokkers *et al.*, 2012). This contributes to explain that current welfare assessment protocols mainly consist of outcome indicators related to visual observation of animal's physical health and partially of input factors, for which validity has been demonstrated, and reliability could be achieved with training.

2.4.3 Feasibility: time constraints and associated costs

Feasibility implies a practical implementation of the assessment in reasonable time and expenses. Practical implementation of on-farm outcome-based welfare assessment is a considerable limitation as the assessment is time-demanding and requires high costs (Sørensen *et al.*, 2007; Rushen *et al.*, 2011; Vasseur, 2017). Assessment includes several implementational steps: a) selection of a sample size before the actual assessment, the scale of which depends on an average number of cattle in the milking herd; b) followed by the actual assessment of each cow from the pre-selected sample for targeted measures of welfare. However, the visual examination of an individual focal animal for each of the measures is labor-intensive, hence, costly (Sørensen *et al.*, 2007). For instance, a study estimated the time needed for the implementation of the Welfare Quality® assessment protocol that is at 60% consists of outcome measures of welfare (De Vries *et al.*, 2016), showed that a dairy herd with an average herd size of 60-100 cows requires up to 5.6-6.6 h (Knierim & Winckler, 2009). Another challenging aspect of the on-farm welfare assessment implementation is its costs. Sorensen *et al.* (2007) estimated that a herd with a size ranging between 60 to 120 cows requires 2,205 euros to implement the on-farm assessment.

Considering the limitations posed by on-farm welfare assessment, it is clear that there is room for alternative measures, by, i.e., using secondary data or replacement of animal-based measures by those which are comparatively easy to collect (Knierim & Winckler, 2009). In view of the presented limitations of on-farm outcome-based welfare assessment, several groups of researchers have proposed an alternative approach using, when available, secondary data such as data collected on farms by different agencies (milk recording agency, veterinarian associations) and collated into national databases (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014; Otten *et al.*, 2016). The next section of this literature review will look at available literature on alternative welfare assessment approach based on secondary data, its potential, and posed challenges.

2.5 Pre-recorded data-based welfare assessment at the herd level as a potential welfare assessment approach

2.5.1 National dairy recording databases

The dairy industry has faced management challenges due to an increase of the industry in scale due to a high concentration of cows per herd that ultimately has led to a shift from traditional data gathering methods (i.e., paper farm records) into automatized recording databases. Such databases are valuable resources allowing to monitor and evaluate the genetic, production, and reproduction performances, health and facilitate herd management (Tomaszewski, 1993; Lescourret *et al.*, 1993; Olsson *et al.*, 2001). Since its instauration, national milk recording databases allow recording and collation of an increased amount of data. Besides, the type and usage of the indicators routinely recorded in such databases have been diversified. Today, national recording dairy databases typically consist of routinely collected records relating to the herd demographics, housing and nutrition practices, production and reproduction performances, health, longevity, and profitability for lactating cows and youngstock (Warner *et al.*, 2020). In Canada, the central dairy herd improvement database is led and supported by Lactanet Inc. (Sainte-Anne-de-Bellevue, QC, Canada). In 2019, according to Agriculture and Agri-food Canada, a total of 6787 Canadian dairy herds with a total number of 622,0 thousand cows were enrolled in the DHI program, including both supervised and unsupervised services, which is about 65% of all Canadian dairy herds. Almost half of the herds registered are in Quebec (n=3360 herds/243,2 thousand cows), followed by the province of Ontario (n=2349 herds/209,6 thousand cows) (Agriculture and Agri-food Canada, 2019).

2.5.2 Pre-recorded data-based welfare assessment

The availability of a wide range of variables registered in the national dairy recording databases has allowed diversifying its usage, such as developing alternative welfare assessment methods. A few authors (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014; Otten *et al.*, 2016) have attempted to develop a welfare assessment approach proposed as a pre-screening tool to identify the overall state of dairy cattle welfare at the herd level instead of, or in combination with an on-farm assessment. This approach could be useful, especially in the context of advisory services and herd improvement initiatives. Such an alternative may enable focusing on herds with a higher need for close observation, thus reducing the number of farm visits, consequently decreasing the time needed to assess all herds. A study estimated the reduction in the number of farms visited by 43% to 67% due to pre-screening herds for their overall welfare state (De Vries *et al.*, 2016). Furthermore, the uniformity of the indicators recorded in such databases enables one compare herds based on corresponding aspects of farming characteristics and, thus, encourages the adoption of best practices from the high-performance herds (Brouwer *et al.*, 2015). Also, the standardization of data collection and possibilities of comparisons across countries may allow for exchange management practices between countries and encourage farmers to enroll in national recording databases. Finally, a pre-recorded data-based assessment will allow monitoring the welfare of dairy herds constantly and detect systematic management practices or housing-related issues that may not always be identifiable through a once per year outcome-based assessment conducted through a certification program.

Overall, the feasibility and accessibility of pre-recorded data make it appealing to use. Nevertheless, secondary data-based welfare assessment at the herd level raises questions regarding the quality of such data (Mörk *et al.*, 2009), its validity (i.e., captures its objective), and if the selected indicators correspond to currently recognized outcome measures of welfare (Sandgren *et al.*, 2009; Nyman *et al.*, 2011; de Vries *et al.*, 2014). In addition, there are concerns about the reliability of the methods applied (i.e., weighting and aggregation) to the development of the composite index (Otten, 2014). The following section will look at the alternative herd welfare assessment approach based on pre-recorded data and its potential and posed challenges.

2.6 Representativeness of pre-recorded data-based indicators relative to outcome measures of on-farm welfare assessment

A new assessment method's validity needs to be evaluated relative to an existing one to ensure that the proposed approach is reliable and meets its original purpose. In the case of dairy cattle welfare, taking into account its multidimensionality, it is critical first to establish the link between indicators of a new assessment method and valid measures of welfare used as a part of the welfare assessment protocols. The majority of studies estimating the link between indicators routinely collected as a part of the DHI agency and outcome welfare measures have reported moderate to high levels of the potential of DHI indicators to assess the welfare of a dairy cow (De Vries *et al.*, 2011; Otten *et al.*, 2019; Ginestreti *et al.*, 2020). The relationships between welfare measures and DHI indicators were reviewed by De Vries *et al.* (2011), who reported that 23 out of 27 selected indicators from DHI database were associated with 16 measures of welfare, among which culling, milk production, and reproduction had the relationship with the largest number of welfare measures.

The indicators of herd demographics routinely collected as a part of DHI data, in particular, culling and on-farm mortality rates have shown to be important indicators of health, hence, the welfare of dairy cattle (Thomsen *et al.*, 2006), as well as of economic interest (Compton *et al.*, 2017). Over the last decades, both culling and mortality rates have increased in dairy herds worldwide (Roche *et al.*, 2020). Alongside the increased cow mortality, calves' mortality rate, particularly calf mortality during the perinatal period – 42 hours after born, is high and poses challenges for dairy farms (Mee, 2008). Regarding the relationship of culling and mortality rates with outcome measures of welfare, for instance, Sandgren *et al.* (2009) reported that on-farm cow and calf mortality rates were associated with the prevalence of under-conditioned cows (BCS ≤2). The findings by de Vries *et al*. (2014) were in line, showing that the prevalence of cows with low body conditions had a relationship with higher on-farm cow mortality and replacement rates. Some research showed that cows diagnosed with lameness are at a higher risk of premature culling. Bicalho *et al.* (2007), for instance, found that lame cows hold 45% of the likelihood to be culled or to die as compared to non-lame cows. Lameness, in its turn, is negatively associated with the reproductive performance of a dairy cow. Alawneh *et al.* (2011) reported the increased reproduction problems among clinically lame cows, while Rutherford *et al.* (2009) showed the positive association of lameness with early calving age.

Bulk tank milk analysis is a widely used method for assessing milk quality and monitoring dairy cattle health and reproduction status (Brandt *et al.*, 2010; Ginestreti *et al.*, 2020). SCC, fat and protein content, protein-fat ratio, lactose, urea, and BHB concentration are standard parameters of such analyses used to identify dairy cattle's health status. The level of SCC in milk, collected as a part of a milk quality test, is used to assess udder health and detect mastitis in dairy cows (Frössling *et al.*, 2017). Studies estimating the relationship between SCC and outcome welfare measures reported inconsistent results. For instance, some research results found that a high level of SCC in milk was associated with the percentage of cows with low body conditions (Berry *et al.*, 2007; de Vries *et al.*, 2014).

Several studies have estimated if those measures can be used to assess the welfare of dairy cattle. For instance, a study in Denmark (Otten *et al.*, 2019) showed that the on-farm cow mortality rate has a high potential to identify herds with high lameness prevalence. Otten *et al.* (2019) estimated the potential of four pre-recorded indicators, particularly mortality, SCC, the proportion of lean cows at slaughter, and age at first calving to detect herds with lameness. Only cow mortality and SCC were found to have a significant association with a high prevalence of lameness. A more recent study (Ginestreti *et al.*, 2020) showed that the milk quality data parameters could not be used to identify dairy herd welfare at the herd level. The study estimated the associations between milk quality data indicators and welfare measures included in the Italian on-farm welfare assessment protocol, where animals were given a score for three different areas (a) management, b) housing and c) animal-based measures. The study results revealed weak ($r < 0.4$) associations for SCC and no associations for total bacteria count, urea and protein, fat content with an area C (i.e., animal-based measures).

The study results estimating the relationship between pre-recorded indicators with outcome welfare measures are not consistent and contradict each other; while some showed the existence of a relationship, others found no association. The relationship of an individual pre-recorded indicators with outcome welfare measures is essential; however, considering the availability of a range of pre-recorded variables, it is of critical importance to select and combine indicators of DHI in order them to represent essential aspects of dairy cattle welfare and jointly identify herds with welfare deficiencies. Therefore, we will further look at the available literature on the development of welfare assessment tools based on pre-recorded indicators.

2.7 Development of welfare assessment methods based on pre-recorded data-based indicators

In the view of a range of variables available and their association with outcome measures of welfare, several attempts have been made to develop assessment tools based on pre-recorded indicators. In particular, few research groups, mainly in European countries, have generated different approaches to identify herd welfare status, either good or poor or both, by combining several indicators from national recording databases.

A study in Sweden (Sandgren *et al.,* 2009) estimated the potential of selected indicators from the Swedish national dairy recording database for identifying herds with poor welfare. The authors used a total of nine outcome measures of welfare (cleanliness and body condition in calves, cows, and young stock, lameness, injuries/inflammations, rising behavior) expressed as herd prevalence. A total of thirteen herds out of 55 Swedish dairy herds were classified as having poor welfare, where the latter was defined as herds being amongst the bottom 10% on ≥2 outcome measures. Three indicators amongst selected pre-recorded indicators: herd prevalence of cows with late ongoing artificial inseminations, the prevalence of heifers without mating/artificial insemination by the age of 17-mo, and calf mortality (2-6-mo) were able to correctly identify eight out of thirteen herds with "poor" welfare with a sensitivity of 0.62. In continuation of the previous study, Nyman *et al.* (2011) conducted a study to evaluate if pre-recorded dairy register data could be used to identify herds with "good" welfare. In the context of this study, "good" welfare was defined when herds had zero scores amongst the bottom 10% based on the same outcome measures used in Sandgren *et al.* (2009), and a total of twenty-eight study herds were classified as herds with good welfare. The study reported that six selected pre-recorded indicators (percentage of cows with late ongoing artificial insemination (>120 days), percentage of heifers without mating/artificial insemination by 17-mo, stillbirth rate, cow mortality, the incidence of mastitis, and feed-related diseases) were able to detect twenty-seven out of twenty-eight herds with good welfare. The results revealed a high sensitivity level of 96% for identifying herds with "good" welfare, which is considerably higher than in the previous study (Sandgren *et al.*, 2009), however only 56% of specificity (the probability of correctly identifying herds that are not cases). Further, the authors estimated if combining two models assigned to identify herds with "poor" and "good" herd welfare would enable discrimination between herds with poor and good welfare status. However, the study findings showed that the combined model had led to more misclassification of study herds than when the two models were used separately (Nyman *et al.,* 2011).

Further, research in the Netherlands (Brouwer *et al.*, 2015) estimated the validity of a Continuous Cattle Health Monitor (CCHM) developed to detect herds with welfare deficiencies and monitor dairy cattle welfare. There were a total of eleven routinely collected indicators included in the CCHM (i.e., cow and young stock mortality, SCC (10^3 cells/ml per quarter), the incidence of subclinical mastitis, and others) (Brouwer *et al.*, 2015). Authors concluded that CCHM was able to discriminate between herds with good vs. poor welfare, with a 50% sensitivity and specificity of 76% when the cut-off value for “poor” welfare was set at <60 points. However, when the cut-off value was set at <70 points, the sensitivity (100%) of the CCHM substantially increased (specificity of 51%), with 52% of herds correctly classified as herds with poor welfare. A more recent study in Denmark (Otten *et al.*, 2016) investigated the performance of the welfare index developed based on twenty-four indicators from the Danish Cattle database relative to welfare index consisted of twelve welfare outcome measures (leg, hindquarter, and udder cleanliness, carpus, tarsus, and body integument alterations, claw conformation, BCS, lameness, avoidance distance, rising behavior, and hair coat). The study reported a significant association of pre-recorded data-based index with outcome-based index ($P < 0.05$). Furthermore, the study reported data-based indicators that showed significant relationships with outcome measures of welfare. For instance, calf mortality was significantly associated with lameness, hock injuries, and low BCS ($P < 0.1$), and bulk tank milk SCC had a significant link with herd prevalence of lameness ($P < 0.05$).

A study in the Netherlands (de Vries *et al.*, 2014) estimated the predictive potential of pre-recorded data to detect herds with welfare issues relative to welfare measures from the Welfare Quality® protocol for cattle (2009) using different sensitivity and specificity levels. As opposed to the results of two previous studies (Sandgren *et al.,* 2009; Nyman *et al.,* 2011), the Dutch model showed considerably low sensitivity of 71% and specificity of 72%, with an accuracy of 71%. Overall, in the above-presented studies, the sensitivity and specificity levels varied depending on the indicators included, as well as the criteria of classification of herds. When the method is developed to detect either positive or negative cases, a high specificity is preferred since it reduces the number of false positives, meaning that, i.e., if the method is developed for identifying herds with poor welfare, then high specificity enables to avoid incorrectly identifying herds with no cases (i.e., no welfare deficiencies). However, a study by de Vries *et al.* (2014) reported that when the specificity level was set at the maximum level of 97.5% (i.e., incorrectly detect herds with no cases), the sensitivity was under 40%, implying that the majority of herds with severe welfare issues were overlooked. Likewise, when setting the maximum sensitivity of 97.5% (i.e., correctly detect herds with cases (welfare deficiencies)), it led to the specificity of below 40%, meaning that more than a half of the herds with cases were incorrectly classified as having high welfare deficiencies. However, the authors suggested that by setting the high sensitivity maximum of 97.5% or 70%, the compulsory farm visits can be reduced by 16% and 45%, respectively, contributing to the time efficiency. So far, the research has demonstrated that dairy herds with severe welfare deficiencies can be detected using secondary-data indicators, which would allow increasing the efficiency of routine on-farm welfare assessment by reducing the number of compulsory farm visits.

2.8 The herd status index and associated challenges

The indexing method is popular among multidimensional studies, which allows the synthesizing of a multitude of information in a compact way (Santeramo, 2017), and has been used in a range of different fields such as environmental studies, economics, population welfare assessment. Since national dairy recording databases consist of multidimensional data with a wide range of measures, and also considering the complexity of animal welfare, the indexing method is probably the most appealing method to unify such a multitude of information into a single index (Knierim & Winckler, 2009).

In 2020, such an approach was utilized to develop the herd status index (HSI) to identify welfare status at the herd level in Canadian herds (Warner *et al.*, 2020). The HSI represents several dimensions reflecting essential aspects of welfare and is grouped under a) longevity, b) nutrition, management, profitability, and c) young stock and reproduction. There are thirteen indicators selected out of 72 potential variables available from the Quebec DHI data, as illustrated in Table **3.3** (Warner *et al.*, 2020). Due to varying measurement units, the indicators are normalized between 0 to 1 before weighting and unifying into a single index. The HSI is quite a straightforward approach that averages each of its indicators, and the herds are ranked between 0 to 100th percentiles, the low and high percentile referring to "poor" and "good" state of overall welfare, respectively. Warner *et al.* (2020) demonstrated that the HSI was comparatively stable for herds that had low- (>p10) and high-ranks (<p90) with standard deviations of 0.066 and 0.062, respectively, while the herds that had the HSI of between 25 and 75 had a considerably higher standard deviation of 0.162. Thus, the authors concluded that the HSI could be used to target those herds with high and least welfare issues (Warner *et al.*, 2020), therefore allowing to reduce farm visits for visual observation for outcome-based measures of welfare. However, the HSI has never been validated against true herd welfare status based on outcome measures of welfare measured on farm, such as lameness or injury prevalence.

Nevertheless, despite the recognized effectiveness of the indexing approach, however, the construction of the composite index and its level of performance can be challenging and requires adherence to specific constructional rules, i.e., selection of particular methods and thorough validity analysis, since the selection of different methods can lead to different results. The selection of variables that represent its dimension is crucial in constructing the composite index (Oţoiu & Grădinaru, 2018). Aggregation is another essential step following the weighting, which condenses the information conveyed by indicators into a single index. The process of aggregating heterogeneous information is challenging and exposed to numerous threats (Oţoiu & Grădinaru, 2018). For instance, the adoption of different aggregation procedures may alter the rankings based on composite indicators (Santeramo, 2017). Therefore, measures included in the scheme are one challenge, while the aggregation of measures is another to process the vast information on single measures into an overall welfare interpretation.

2.9 Hypothesis and implications

The application of any new assessment method requires evaluating its validity relative to an existing assessment method. Such evaluation is needed to ensure that the proposed alternative is reliable and corresponds to the objective for which it was developed. Therefore, taking into account the complex nature of animal welfare that is dependent on a range of internal as well as external factors, is crucial to ensure that the HSI can encompass multidimensionality and capture the complex nature of animal welfare at the herd level relative to the on-farm outcome-based assessment method that represents a comparatively valid assessment method (such as herd prevalence of lamness or injury measured on animals). Besides, despite the high applicability of an indexing method to multidimensional issues, the construction of a composite index and its level of performance can be challenging and requires adherence to strict rules. In particular, the selection of dimensions and corresponding indicators and a method applied to weighting and aggregation of selected indicators are the most crucial steps that may lead to different results.

It was hypothesized that the HSI constructed based on pre-recorded DHI indicators can reflect the true status of dairy cattle welfare at the herd level relative to an existing proAction® on-farm outcome-based welfare assessment method in Quebec dairy herds. The Proaction® database includes five farm-level outcome measures which are recorded to assess on-farm animal welfare. The validity of those five measures – lameness, body condition, hock, neck, and knee injuries, to assess dairy welfare status had been demonstrated by research and are commonly used in on-farm dairy welfare assessments wolrdwide (detailed in Section 2.4).

2.10 General and specific objectives

2.10.1 General objectives

The main objective of this study is to evaluate the validity of the HSI, constructed based on pre-recorded DHI indicators, for the identification of dairy cattle welfare at the herd level relative to an existing proAction® on-farm outcome-based welfare assessment method in Quebec dairy herds. To approach this objective, one could potentially use a multiple regression model which could have multiple collinearity problem, or alternatively, one could look to create groups or clusters of herds which have similar characteristics and see their group (clusters) association with welfare measures and the HSI indicators. For the objective of this thesis a clustering approach was applied.

2.10.2 Specific objectives

1) Investigate the differences amongst clusters relative to outcome measures of welfare and pre-recorded indicators of herd status index;

2) Investigate the differences amongst clusters relative to herd status index and perform a comparative analysis in relation to outcome measures of welfare.

3. MATERIALS AND METHODS

A cross-sectional study was conducted to evaluate the validity of the HSI for the identification of dairy cattle welfare at the herd level relative to proAction® on-farm outcome-based welfare assessment (DFC, Canada). Data collected from Quebec dairy herds in the framework of the proAction® Quality Assurance Program (DFC, Canada) were integrated with DHI data from the Lactanet Inc. (Sainte-Anne-de-Bellevue, QC, Canada) database.

3.1 Data sources

3.1.1 Outcome measures of welfare

To ensure that the entire dairy production cycle starting at the farm all the way to consumers is in line with quality and safety standards Dairy Farmers of Canada established proAction® Quality Assurance Program (DFC, 2013). This compulsory certification program covers six industry-related areas – milk quality, food safety, animal care, traceability, biosecurity, and environment. Each of these parts of the program sets specific standards in terms of its respective areas. Regarding the animal care module, Dairy Farmers of Canada in collaboration with National Farm Animal Care Council have developed the “Animal Care” dairy cattle welfare assessment protocol based on the recommendations and requirements from The Code of Practice for the Care and Handling of Dairy Cattle (Code of Practice) (NFACC-DFC, 2009). There are overall five outcome welfare measures – lameness, body condition, hock, neck, and knee injuries, which validity to assess dairy welfare status had been demonstrated by research and are commonly used in on-farm dairy welfare assessments wolrdwide (detailed in Section 2.4). More comprehensive information on Canadian dairy cattle welfare assessment can be found on the official website of DFC (https://dairyfarmersofcanada.ca), however, the brief description of scoring methods is given below.

The welfare assessment implementation among Canadian dairy farms in accordance with the “Animal Care” protocol consists of the following steps: a) a selection of a sample size before the actual on-farm assessment, the scale of which depends on an average herd number of lactating cows; b) the actual assessment of all cows from the pre-selected sample for mentioned outcome measures. For the implementation of the protocol, dairy farms are visited by an independent technician, and it is required that dairy farms are assessed every two years (DFC, 2019).

Table 3.1 Description of classification of proAction® on-farm assessment outcome welfare measures (BCS ≤ 2, hock, neck, and knee injuries, and lameness) on a point scale

Measures	**Classification**	**Reference**
Lameness	Acceptable ≤ 2 Monitor = 3 Requires corrective action ≥ 4	(Flower & Weary, 2006) (Gibbons et al., 2014)
Body condition score	Requires corrective action ≤ 2 Acceptable > 2	(Vasseur *et al.*, 2013)
Hock injuries	Requires corrective action ≤ 1 Acceptable ≥ 2	(Gibbons *et al.*, 2012)
Knee injuries	Requires corrective action ≤ 1 Acceptable ≥ 2	
Neck injuries	Requires corrective action ≤ 1 Acceptable ≥ 2	

Regarding the scoring methods, depending on the severity level of given welfare measures, a cow’s welfare status is classified as “Acceptable” or “Corrective action”. Table **3.1** illustrates the description of the scoring methods. Hock and knee injuries are assessed based on a 4-point and neck injuries on a 3-point scoring system. For instance, during a visual assessment of individual cows for the presence of hock lesions and swellings, a cow is given scores of 2 or 3 and classified as “Acceptable” when a cow does not have swellings, or scores of 0 or 1 and classified as “Corrective action” when a cow has medium or major swelling areas (Gibbons *et al.*, 2012).

Assessment of a cow's lameness is a visual examination for the presence and the level of severity of limp, rating cows from 1 to 5 (sound to severely lame) in free-stalls (Flower & Weary, 2006) and the evaluation of certain types of behaviors of a cow such as weight shifting, standing on the edge of the stall, and uneven weight bearing in tie-stalls (Gibbons *et al.*, 2014).

For body condition scoring, cows are examined for the proportion of body fat rating cows on a 5-point scoring system (emaciated to fat) (Vasseur *et al.*, 2013), and labeling cows with a body condition score ≤ 2 as "Corrective action" and "Acceptable" when cows are given a score of >2. Finally, based on taken measures, a producer is provided with a herd prevalence report for each measure. Apart from an individual herd prevalence report, all herds are classified under "green", "yellow", and "red" zones that correspond to the top 25%, medium 50%, and bottom 25%, respectively, and the latter indicating a need for welfare improvement by taking recommended necessary measures.

Before the actual welfare assessment based on a new developed method, DFC had conducted a pilot assessment project during 2014-2015 that included 120 dairy farms across Canada. In the present study, animal measures of the first herd assessment were considered. Given herds were assessed by 30 independent technicians from September 2016 to February 2019, as a part of the "Animal Care" module of the proAction® Quality Assurance Program (DFC, Canada, 2013).

Data for five outcome measures of welfare – lameness, body condition, hock, neck, and knee injuries scores for a total of 4770 Quebec dairy herds were extracted from the proAction® database (DFC, Canada). As mentioned in the Section 2.4 of the Literature review, these particular five measures of welfare are included in European and American dairy cattle welfare assessment protocols, and research has shown their comparatively sufficient level of validity to assess the welfare of farm animals. Nevertheless, it is important to note that the threshold used for determining the level of “good” vs. “poor” welfare might differ from country to country, as well as the interpretation of the results.

3.1.2 Pre-recorded data-based indicators

The DHI records contain a range of variables on herd demographics, production and reproduction performances, health and management, at both the animal and the herd level. In 2019, according to Agriculture and Agri-food Canada, a total of 6787 Canadian dairy herds with a total number of 622,000 cows were enrolled in the DHI program, including both supervised and unsupervised services, which is about 65% of all Canadian dairy herds. Almost half of the herds registered are in Quebec (n=3360 herds/243,200 cows), followed by the province of Ontario (n=2349 herds/209,600 cows) (Agriculture and Agri-food Canada, 2019).

Table 3.2 Year-season classification of the 3682 Quebec dairy herds, enrolled in Canadian dairy herd improvement program, from the Lactanet Inc. database (Sainte-Anne-de-Bellevue, QC, Canada). This classification pertains to the farm visit when the proAction® data were collected.

Year-season	Period	n of herds
Year-season I	December 2016 to February 2017	77
Year-season II	March to May 2017	284
Year-season III	June to July 2017	221
Year-season IV	September to November 2017	478
Year-season V	December to February 2018	393
Year-season VI	March to May 2018	602
Year-season VII	June to July 2018	629
Year-season VIII	September to November 2018	543
Year-season IX	December 2018 to February 2019	455
Total		3682

The DHI test-day data for the herds included in the proAction® data were extracted from the Lactanet Inc. (Sainte-Anne-de-Bellevue, QC, Canada) database, the indicators of which were used for the calculation of the HSI. Based on Warner et al. (2020) HSI calculation method, to account for possible fluctuations that might occur due to unpredictable circumstances, data for the period of three years before the proAction® on-farm assessment date were extracted and classified by year-season combinations pertaining to the farm visit when the proAction® data were collected, starting in December 2016 to February 2019. Overall, this resulted in nine separate year-season extractions for a total of 3682 herds (78%) that matched with the proAction® data set, as shown in (Table **3.2**).

Table 3.3. Description of the thirteen indicators of the herd status index (Warner *et al.*, 2020) registered in Canadian dairy herd improvement program (Lactanet Inc. database, Sainte-Anne-de-Bellevue, QC, Canada)

Category	**Indicators**	**Description**
Longevity	Cow longevity, %	Cows at 3 and more lactations
	Involuntary replacement rate, %	Total replacement rate, % – Sold for production, %
	Cow mortality, %	Cows dead
Nutrition, production, and profitability	Cows with low MUN[1], %	Indicator of lack of dietary protein availability in the rumen, low milk urea nitrogen: < 5 mg/dL of milk
	Cow with high milk P:F[2], %	P:F ratio: protein-to-fat ratio; indicator of subacute ruminal acidosis; high milk P:F ratio: > ratio of 1.1
	HMI[3]	Indicator of the optimal use of the genetic potential based on standardized milk (0.2594 × kg milk + 12.1975 × kg milk fat + 7.707 × kg milk protein)
	TCI[4]	Indicator of fresh cow performance (the difference between predicted and actual first test 305-d milk, (patented by the Wisconsin Alumni Research Foundation, AgSource Cooperative Services)
	Cow lifetime profit rank	Indicator of estimated profitability of animals in the herd, (milk revenues minus rearing, maintenance and production related expenses)
Young stock and reproduction	Calf mortality, %	Calves died at 0–24 hours
	Age at first calving, months	
	Abortion rate, %	
	Cows with high BHB[5], %	BHB: ß-hydroxy butyrate; measure of risk for hyperketonemia; high milk BHB: > 0.2 mmol/L of milk
	Cows with high SCC[6], %	Indicator of mastitis; high milk SCC: > 400,000 SCC/mL of milk

[1]MUN – milk urea nitrogen; [2]P:F – protein-fat ratio; [3]HMI – herd management index ; [4]TCI – transition cow index; [5]BHB - ß-hydroxy butyrate; [6]SCC – somatic cell count

Regarding the indicators from pre-collected data registered in the DHI, only those indicators were extracted, which were necessary for the calculation of the HSI and the indicators related to herd management and demographics. The HSI represents several dimensions reflecting essential aspects of welfare and is grouped under a) longevity, b) nutrition, management, and profitability, and c) young stock and reproduction, and is composed of a total of thirteen corresponding pre-recorded indicators, as illustrated in Table **3.3**. More detailed information on the HSI development can be found in Warner *et al.* (2020).

3.2 Data cleaning and integration of data sets

Prior to performing the statistical analysis, data sets were edited in accordance with the pre-defined requirements of the study. Figure **3.1** illustrates the four-step data cleaning process: 1) cleaning of proAction® data set before its integration; 2) application of criteria to DHI data set to meet the requirements of the HSI methodology; 3) merging nine year-season extractions; 4) integration of proAction® and DHI data sets. As was mentioned above, the initial proAction® data set consisted of 4770 dairy herds, 40 of which were found to be duplicates due to previous involvement in a pilot assessment project; hence, they were removed from the data set. Besides, since data were extracted by year-season combinations with the first year-season starting as of December 2016, herds assessed before this date (n=10) were eliminated from the initial data. Removal of herds due to the mentioned reasons resulted in there being 4720 herds before integrating with the DHI data.

The next step was to ensure DHI data met specific criteria before merging with proAction® data. As was mentioned before, there were nine separate year-season extractions of DHI data for a total of 3682 herds. Before merging all nine extractions into one data set, herds with less than 10 cows for SCC, BHB, MUN, P:F per test date were labeled as missing values for those indicators,

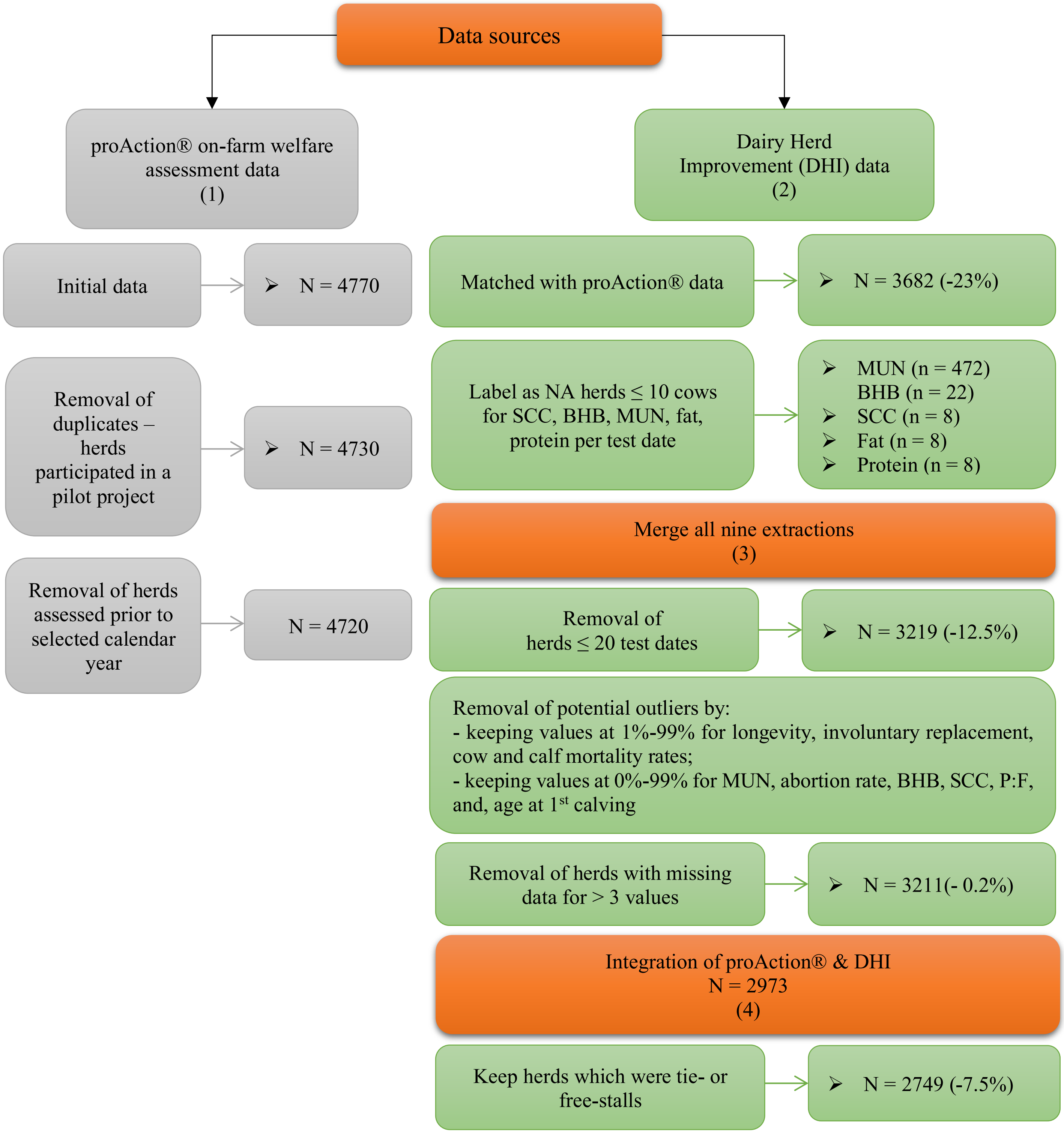

Figure 3.1 Flow chart for the four-step data cleaning process and the integration of proAction® on-farm assessment data (DFC, Canada) with test-day dairy herd improvement data from the Lactanet Inc. ddatabase (Sainte-Anne de-Bellevue, QC, Canada)

the overall number of which was equal to 518 observations, with the highest for low MUN (n = 472), followed by high BHB (n = 22), and less than 10 observations for the remaining indicators. Then, all nine extractions were merged into one data set by computing one value for each variable and each herd as the mean of the variable over the number of test dates available. The herds with less than 20 test dates (n = 463) were removed to ensure that observations have sufficient test-day data over the three years.

The values outside the normal range were identified to avoid erroneous information. In the case of average cow longevity, involuntary replacement, cow and calf mortality rates, observations at the 0 and 100th percentiles were labeled as missing values to remove potential outliers from the data set. In contrast, for the indicators of low MUN, high BHB, high SCC, high P:F, abortion rate, and age at first calving observations between the 0 and 99th percentiles were kept since most of the values for these indicators included value of 0, or between 0 and 1. Finally, to avoid hazards related to gaps in the data set, herds with missing data for more than 3 values (n = 8) were eliminated. All these data cleaning steps resulted in n = 3211 herds before integration of DHI data with the corresponding proAction® data.

In the final step, previously cleaned proAction® and DHI data were integrated based on the unique herd identification number, which resulted in a merged total of 2973 herds. In addition, only herds with tie- or free-stall housing systems were kept for further analyses, excluding those with different housing types (n = 224, i.e. pack bedding, pasture). Ultimately, after completing the data cleaning process, the final study data used for statistical analysis was 2749 dairy herds, with an average herd size of 62 cows, ranging from a minimum size of 11 to a maximum of 550 cows. The majority of herds were tie-stall systems, which accounted for 81% (n = 2219), while the remaining were free-stall (19%; n = 530), with average milk production of 9,194 ± 1,321 kg/year (mean ± SD) and 9,355 ± 1,299 kg/year (mean ± SD) for tie- and free-stall herds, respectively.

3.3 Descriptive statistics

Descriptive statistics describing the herd prevalence of five proAction® on-farm assessment outcome welfare measures and twelve out of the thirteen initial pre-recorded DHI indicators of the HSI (the indicator of a lifetime profit rank was removed due to multicollinearity issues, see in Section 4.2), presented in Table **3.4** and Table **3.5**, respectively.

Among outcome measures, the highest herd prevalence was for hock injuries (17.7 ± 14.9%), followed by the herd prevalence of lameness (8.9 ± 9.9%) and knee injuries (7.2 ± 8.8%), while the herd prevalence of neck injuries and low BCS remained under 5% (4.2 ± 7.3% and 1.8 ± 4.1%, respectively).

Table 3.4 Descriptive statistics of the herd prevalence of five proAction® on-farm assessment outcome measures of welfare (Mean ± SD) for the study population of 2749 Quebec dairy farms enrolled in Canadian dairy herd improvement program

Outcome measures	**Mean**	**SD**	**Min**	**Max**
BCS ≤ 2, %	1.8	4.15	0	39.1
Hock injuries, %	17.8	14.93	0	91.3
Neck injuries, %	4.2	7.28	0	52.4
Knee injuries, %	7.2	8.85	0	71.4
Lameness, %	8.9	9.95	0	66.7

With regards to the herd prevalence of the HSI indicators (see table Table **3.3** for definition of variables and thresholds) for 2749 study herds, the mean prevalence of cow longevity was 39.5 ± 7.95%, and involuntary replacement, cow and calf mortality rates of 31.4 ± 9.18%, 3.6 ± 3.58%, and 7.9 ± 5.03%, respectively. The mean herd prevalence of cows with high BHB (1.9 ± 1.31%), high P:F (3.3 ± 2.46%), and low MUN (3.1 ± 3.89%) was under 3.5%, while for the indicator of cows with high SCC mean herd prevalence was considerably higher (12.2 ± 4.32%). Regarding Herd management and Transition Cow Indexes, the mean herd prevalence for both indicators was under 150, however, with a negative value for the Herd management index. The mean age at first calving was 26.1 ± 1.72mo, while the mean abortion rate was the lowest amongst all indicators (0.7 ± 2.87%).

Table 3.5. Descriptive statistics of the herd prevalence of twelve pre-recorded DHI indicators of the HSI (Mean ± SD) for the study population of 2749 Quebec dairy farms enrolled in the Canadian dairy herd improvement program

Category	Variable	Mean	SD	Min	Max
Longevity	Cow longevity, %	39.5	7.95	18.1	62.1
	Involuntary Replacement Rate, %	31.4	9.18	9.1	60.2
	Cow mortality, %	3.6	3.58	0	16.9
Nutrition, production and profitability	Cow with low MUN[1], %	3.1	3.89	0	25.6
	Cow with high milk P:F[2], %	3.3	2.46	0	13.5
	HMI[3]	-145.1	1156.1	-4112.3	6229.7
	TCI[4]	149.7	391.8	-1594.0	2151.0
Young stock and reproduction	Calf mortality, %	7.9	5.03	0	25.0
	Age at first calving, months	26.1	1.72	22.1	33.8
	Abortion rate, %	0.7	2.87	0	20.0
	Cows with high BHB[5], %	1.9	1.31	0	7.9
	Cows with high SCC[6], %	12.2	4.32	1.3	24.3

[1]MUN – milk urea nitrogen; [2]P:F – protein-fat ratio; [3]HMI – herd management index ; [4]TCI – transition cow index; [5]BHB - ß-hydroxy butyrate; [6]SCC – somatic cell count

3.4 Statistical Analysis

The R statistical software (version 3.5.0; R Foundation for Statistical Computing, Vienna, Austria; https://cran.r-project.org) was used to perform data cleaning (R packages “tidyverse”, “dplyr” (Wickham *et al.*, 2019), “ggplot2” (Wickham, 2016), and “ggpubr” (Kassambara, 2018)) and the statistical analyses, with the level of statistical significance set at $P < 0.05$.

3.4.1 Objective 1 - investigate the differences amongst clusters relative to outcome measures of welfare and pre-recorded indicators of herd status index

Cluster analysis was performed on outcome welfare measures to obtain subgroups of herds. One of the first steps required when implementing cluster analysis is the standardization of input variables due to heterogeneity of measurement units to avoid instances where a variable's influence on the cluster solution is greater than it should be (Hair *et al.*, 2013). However, since the input variables for the clustering, outcome welfare measures (i.e., proAction® variables), in this case, were assessed at the same scale and same measurement units, they were not standardized. Correlation coefficients between outcome measures were calculated by Pearson correlations using the function *rcorr* in the R package "Hmisc" (Harrell Jr & Dupont, 2017) to verify for potential multicollinearity.

The challenge when performing cluster analysis is that it will partition a given data set based on identified input variables, even if no natural subgroups exist. Therefore, the validation of clusters is an essential step that ensures the meaningfulness and relevance of clustering results (Balijepally et al., 2011; Kassambara, 2017). For this reason, before performing the clustering algorithms, the Hopkins' statistic was computed to assess the clustering tendency (Kassambara, 2017) using the function *hopkins* from the R package "clustertend" (Lawson & Jurs, 1990) by randomly splitting the study data, and comparing the obtained Hopkin's coefficient between random and study data.

There are various clustering algorithms available; however, several authors recommend a combination of the two methods: perform a hierarchical clustering to determine cluster seed points for a nonhierarchical clustering method (Balijepally *et al.*, 2011; Hair *et al.*, 2013; Kassambara, 2017). The advantage of the hierarchical clustering algorithm with Ward's minimum variance linkage is that it reduces the within-group variance, which leads to a better grouping, meaning that herds that are closest and most similar on given variables are grouped. In comparison with the hierarchical clustering algorithm, the k-means algorithm is less sensitive to outliers and irrelevant variables; however, it requires the specification of the number of clusters (k-numbers) beforehand by a researcher (Balijepally *et al.*, 2011; Hair *et al.*, 2013). Hence, it is recommended to perform a hierarchical clustering algorithm to determine the k-numbers for further use as seeds for the k-means algorithm (Balijepally *et al.*, 2011). Therefore, to compensate for each clustering algorithm's limitations, a two-stage clustering algorithm was used in this study.

As the first stage of the clustering algorithm, the distance matrix was calculated using the function *dist* from the R package "stats" (R Core Team, 2019), followed by forming clusters themselves and determining the optimal number of clusters for its further usage as k-seeds. The selection of a distance measure is an important step that affects the final cluster results (Kassambara, 2017). Euclidean distance was used to determine the similarity between each of the 2749 study herds based on five-dimensions. Once the similarity measure was calculated, the next step was to perform the hierarchical agglomerative clustering algorithm using the function *hclust* from the R package "stats" (R Core Team, 2019) to generate the optimal structure of clusters based on the distance matrix. The hierarchical agglomerative clustering algorithm generates several cluster solutions, starting with each observation considered a cluster of its own, herds in this case, and merging the most similar clusters into a single cluster (Hair *et al.*, 2013). As a linkage method, Ward's minimum variance method was used, which has shown to be the most effective linkage method by several studies (Saraçli *et al.*, 2013). Hierarchical cluster results were validated by cophenetic distance performed by computing the correlation coefficient between the Euclidean and cophenetic distances using the function *cor* in the R package "stats" (R Core Team, 2019). The selection of the optimal number of clusters consisted of two-steps: the visual observation of the dendrogram using the function *fviz_dend* in the R package "factoextra", and by the "majority rule": applying three different methods - the Elbow method, Silhouette analysis, and Gap statistics using the function *fviz_nbclust* in the R package "factoextra".

Once the first stage algorithm is completed and seed points were selected, the k-means partitioning clustering algorithm was performed using the function *kmeans* in the R package "stats" (R Core Team, 2019) to partition study herds into subgroups based on outcome measures by specifying k-seeds identified in the previous clustering stage. The k-means classifies observations in a given data set so that observations in the same group are highly similar, while observations in different groups are highly dissimilar (Kassambara, 2017).

Descriptive statistics for the input variables were performed to describe each cluster of herds. A multi-way ANOVA was performed using the functions *lm, Anova, emmeans, and contrast* in the R packages "car" (Fox & Weisberg, 2018), "rstatix" (Kassambara, 2020), and "emmeans" (Lenth *et al.*, 2019)) to investigate the potential differences between subgroups regarding outcome-based welfare measures and pre-recorded indicators of the HSI using a statistical model given below. The clusters were formed solely based on outcome measures of welfare (i.e., proAction® variables) hence the effect of housing and year-season were included as additional factors in our model:

$$Y_{ijkp} = \mu + \text{cluster}_i + \text{housing}_j + \text{year-season}_k + \varepsilon_{ijkp}$$

Where: Y_{ijkp} is the dependent variable: a) outcome-based measures of welfare and b) pre-recorded indicators of the HSI of the i^{th} cluster in the j^{th} housing type and k^{th} year-season; μ = overall mean; cluster_i is the fixed effect of the i^{th} cluster, a categorical variable with four levels: *1, 2, 3, 4;* housing_j is the fixed effect of the j^{th} housing, a categorical variable with two levels: *tie- and free-stall housing systems;* year-season_k is the fixed effect of the k^{th} year-season of collection of proAction®

data (for year-season classification see Table **3.2**), a categorical variable with nine levels:*year-season I – winter:* Dec 2016 to Feb 2017; *year-season II – spring:* Mar 2017 to May 2017; *year-season III – summer:* Jun 2017 to Aug 2017; *year-season IV – fall:* Sep 2017 to Nov 2017; *year-season V – winter:* Dec 2017 to Feb 2018; *year-season VI – spring:* Mar 2018 to May 2018; *year-season VII – summer:* Jun 2018 to Aug 2018; *year-season VIII – fall:* Sep 2018 to Nov 2018; *year-season IX – winter:* Dec 2018 to Feb 2019; ε_{ijkp} = the random residual error associated with the experimental unit (herds); $\varepsilon_{ijkp} \sim N(0, \boldsymbol{\sigma}^2{}_e)$.

Multiple comparisons among clusters, housing types, and year-seasons were performed using Bonferroni's adjustment test for multiple comparisons of least-square means with an error level of 0.05. Housing types and year-seasons were included in the statistical model, but they were not the focus of the study. Since we were not specifically interested in the main effects of year and season, they were combined together with the interaction effect as a simple factor year-season.

3.4.2 Objective 2 - investigate the differences amongst clusters relative to herd status index and perform a comparative analysis relative to outcome welfare measures

The HSI was calculated for each study herd based on the method previously developed by Warner *et al.* (2020). Prior to the HSI calculation, its composite indicators were standardized to percentile ranks due to the heterogeneity of measurement units. Correlation coefficients using the function *rcorr* in the R package "Hmisc" (Harrell Jr & Dupont, 2017) and variance inflation factor (VIF) using the function *omcdiag* and *imcdiag* in the R package "mctest" (Muhammad *et al.*, 2020) were calculated to assess potential multicollinearity between composite indicators. A VIF value of a maximum 10 and a tolerance value of a minimum 0.10 were set as the levels for a variable to remain as an HSI indicator. Indicators were then aggregated into a composite index (HSI) per each herd based on a linear additive aggregation method by simply summing equally weighted indicators and dividing by the total number of indicators.

Pearson correlation analysis was performed for the associations between the HSI and the five outcome welfare measures using the function *rcorr* in the R package "Hmisc" (Harrell Jr & Dupont, 2017). The study looked for potentially significant differences among clusters regarding the HSI to evaluate the validity of the HSI relative to an existing proAction® outcome-based on-farm welfare assessment method, using Bonferroni's adjustment test for multiple comparisons of least-square means with an error level of 0.05, for which the following statistical model was used.

$$Y_{ijkp} = \mu + \text{cluster}_i + \text{housing}_j + \text{year-season}_k + \varepsilon_{ijkp}$$

Where: Y_{ijkp} = the dependent variable, the HSI of the i^{th} cluster in the j^{th} housing type and k^{th} season; μ = overall mean; cluster_i is the fixed effect of the i^{th} cluster, a categorical variable with four levels: *1, 2, 3, 4*; housing_j is the fixed effect of the j^{th} housing, a categorical variable with two levels: *tie- and free-stall housing systems; year-season*$_k$ is the fixed effect of the k^{th} year-season of collection of proAction® data (for year-season classification see Table **3.2**), a categorical variable with nine levels: *year-season I – winter:* Dec 2016 to Feb 2017; *year-season II – spring:* Mar 2017 to May 2017; *year-season III – summer:* Jun 2017 to Aug 2017; *year-season IV – fall:* Sep 2017 to Nov 2017; *year-season V – winter:* Dec 2017 to Feb 2018; *year-season VI – spring:* Mar 2018 to May 2018; *year-season VII –* summer: Jun 2018 to Aug 2018; *year-season VIII – fall:* Sep 2018 to Nov 2018; *year-season IX – winter:* Dec 2018 to Feb 2019; ε_{ijkp} = the random residual error associated with the experimental unit (herds); $\varepsilon_{ijkp} \sim N(0, \boldsymbol{\sigma}^2_e)$.

Multiple comparisons between clusters, housing types, and year-seasons were performed using Bonferroni's adjustment test for multiple comparisons of least-square means with an error level of 0.05. Housing types and year-seasons were included in the statistical model, but they were not the focus of the study. Since we were not specifically interested in the main effects of year and season, they were combined together with the interaction effect as a simple factor year-season.

4. RESULTS

4.1 Objective 1 - investigate the differences amongst clusters relative to outcome measures of welfare and pre-recorded indicators of herd status index

4.1.1 Cluster analysis

Pearson correlation coefficients, presented in Table **4.1**, revealed significant *($P<0.05$)* but weak positive correlations ($r < 0.3$) between the prevalence of neck injuries and the prevalence of low BCS and prevalence of hock injuries. Also, there were some weak to moderate positive relationships ($r < 0.3$) between lameness and the remaining outcome measures of welfare, that were not found to be significant.

Before performing the two-stage cluster algorithms, Hopkins' statistic test was calculated to assess the spatial randomness of the data set, which resulted in Hopkin's coefficient of 0.18, hence, assuring that the data set contains meaningful clusters (Kassambara, 2017).

Table 4.1. Pearson correlation coefficients between five proAction® on-farm assessment outcome measures of welfare for the study population of 2749 Quebec dairy farms enrolled in the Canadian dairy herd improvement program

Variable	BCS ≤ 2	Hock injuries	Neck injuries	Knee injuries	Lameness
BCS ≤ 2	1				
Hock injuries	0.169	1			
Neck injuries	0.084	0.125	1		
Knee injuries	0.182	0.241	0.216	1	
Lameness	0.323	0.296	0.247	0.308	1

Once ensuring the existence of essential clusters in the data set, the hierarchical clustering algorithm using Ward's minimum variance linkage method was performed to determine the optimal number of clusters. A distance matrix was calculated using the Euclidean distance method to determine the similarity between the 2749 study population based on five-dimensions. The correlation coefficient between the Euclidean and cophenetic distances was calculated to verify the hierarchical clustering algorithm's output, which was equal to 0.63, implying that the dendrogram reflects the study data comparatively well.

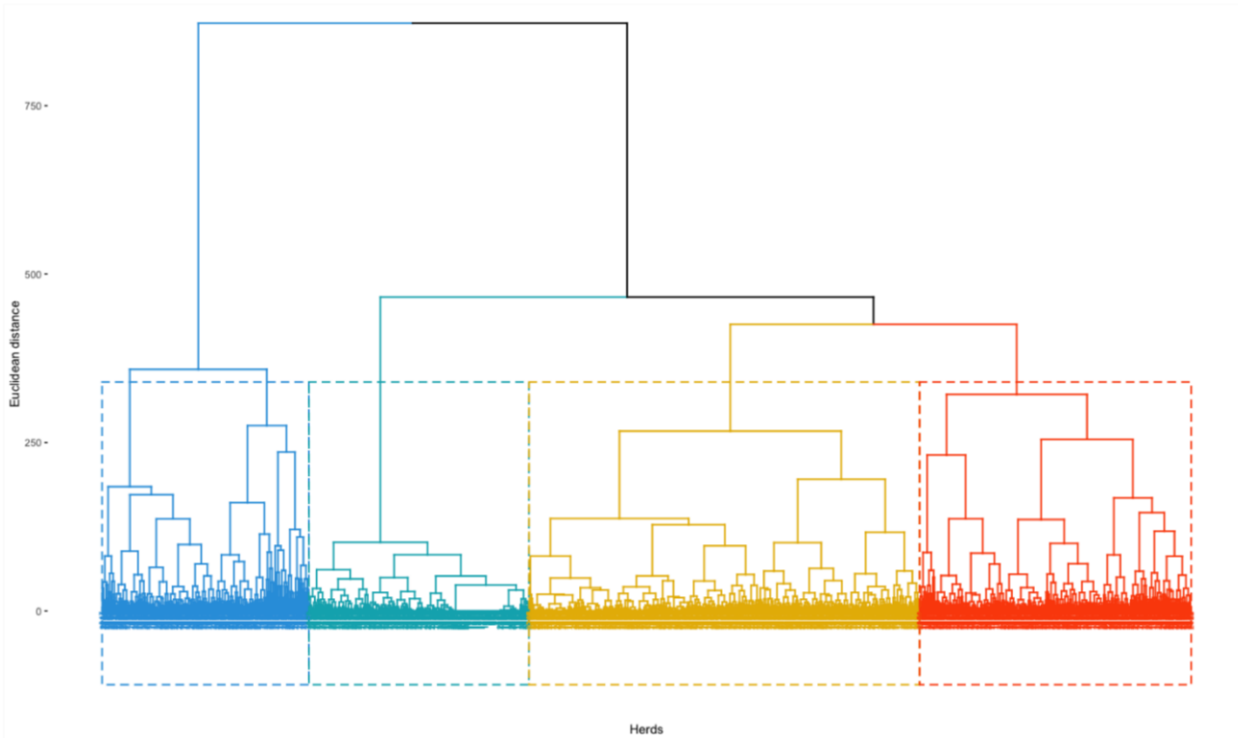

Figure 4.1 The dendrogram of the hierarchical clustering algorithm using Ward's minimum variance linkage method for the study population of 2749 Quebec dairy herds enrolled in Canadian dairy herd improvement program.

The selection of the optimal number of clusters consisted of two-steps: the visual observation of the dendrogram and the calculation of three different indexes. Figure **4.1** llustrates the dendrogram – visual output of the hierarchical clustering algorithm, from which we can intuitively draw a threshold line and divide observations into k = 4 clusters. The selection of the optimal number of clusters is the challenging step that affects the following results, thus requires a more thorough analysis and strict rules for its determination. Visual inspection of a dendrogram is a good way to have an idea about the clustering; however, this method alone is not efficient and not reliable (Eremenko, 2018). Therefore, along with the dendrogram's visual inspection, we computed three different methods – Elbow, Silhouette analysis, and Gap statistics (bootstrap = 500) to determine the optimal number of clusters based on the "majority rule". The silhouette analysis method suggested k = 2 as an optimal k-number, while the Elbow method and Gap statistics suggested k = 4 as an optimal number of clusters for given study data. Therefore, based on visual inspection of a dendrogram, in combination with the "majority rule", and also considering that Gap statistics are considered as a more sophisticated method (Kassambara, 2017), a cluster number of k = 4 was selected as an optimal number of clusters for this study.

Once the first stage algorithm is completed and seed points selected, the k-means partitioning clustering algorithm was performed to partition study herds based on outcome welfare measures by specifying k-seeds of 4, identified based on the hierarchical clustering algorithm. The main advantage of the k-means clustering is that it minimizes the total intra-cluster variation, meaning that it groups herds that are the most similar to each other and the most dissimilar to herds from other clusters.

4.1.2 Herd characteristics by cluster groups

Herd profile characteristics in each of the clusters are described in Table **4.2**. The size of clusters differed considerably**,** with Cluster 2 (C2) having the largest number of herds with about 40% of study herds (n = 1070), followed by Cluster 1 (C1) (n = 746) and Cluster 4 (C4) (n = 570), and with the least number of herds concentrated in Cluster 3 (C3) (n = 363). Herds in C4 demonstrated the lowest mean herd size (54.6 ± 29), milk production (8980 ± 1313.5 kg/year), milk fat (360.5 ± 51.7kg) and protein (294.5 ± 42.8kg). In contrast, herds in C1, with concentration of 27% out of total 2749 study herds, had the highest mean herd size (65.2 ± 36.4), milk production (9482 ± 1224.5 kg/year), milk fat (379.7 ± 49.3kg) and milk protein (311.3 ± 39.7kg). While there were some differences among herd profile parameters in C1 and C4, herds in C2 and C3 had quite similar values.

Table 4.2 Herd profile characteristics (Mean ± SD, median, range) for the study population of 2749 Quebec dairy herds partitioned into four study clusters using two-stage clustering algorithm

Measure	Cluster 1 (n=746)	Cluster 2 (n=1070)	Cluster 3 (n=363)	Cluster 4 (n=570)
N of herds				
- tie-stalls	583 (78%)	795 (74.2%)	305 (84%)	536 (94%)
- free-stalls	163 (22%)	275 (25.7%)	58 (16%)	34 (6%)
Herd size	65.2 ± 36.4 57 20 – 460	64.8 ± 40.5 56 11 – 550	59.7 ± 26.6 53 16 – 175	54.6 ± 29.0 48 13 – 375
Milk production, kg	9482 ± 1224.5 9506 5581 – 12975	9133 ± 1399.7 9272 4495 – 14065	9354 ± 1159.2 9371 5518 – 12365	8980 ± 1313.5 9054 3742 – 12136
Milk protein, kg	311.3 ± 39.7 312.0 199.0 – 442.0	301.4 ± 43.6 305.0 144.0 – 442.0	307.3 ± 38.3 309.0 182.0 – 402.0	294.5 ± 42.8 298.0 130.0 – 396.0
Milk fat, kg	379.7 ± 49.3 378.0 233.0 – 595.0	367.6 ± 52.9 370.0 148.0 – 578.0	373.3 ± 47.9 374.0 217.0 – 534.0	360.5 ± 51.7 365.0 169.0 – 499.0

The objective of the study was to group herds based on welfare outcomes regardless of herd characteristics as they are not relevant per se for our objectives; therefore, those herd characteristics will not be part of the discussion of this study and are presented here for informative purposes only.

4.1.3 Differences amongst clusters in herd prevalence of outcome measures of welfare

The multiple comparisons results showed statistically significant differences in the least-squares means (LSmeans; also referred to as estimated marginal means) amongst all five outcome welfare measures except only for the difference between C1 and C2 in terms of the herd prevalence of neck injuries. Overall, it can be noted that C2 and C4 had the highest differences in means relative to all outcome measures of welfare, while C1 and C2 had the lowest differences in means for four outcome measures and no statistically significant difference for the mean herd prevalence of neck injuries (Table **4.3**). Therefore, based on the results of multiple comparisons amongst clusters with regards to outcome welfare measures, the dairy herds in C4 could be considered as herds with the most prevalent welfare issues. Herds in C1 and C2, in contrast, could be considered as subgroups of herds with the least welfare issues relative to standardized welfare measures of a dairy cow.

Table 4.3. Mean herd prevalence of proAction® on-farm assessment outcome measures of welfare (BCS ≤ 2, hock, neck, and knee injuries, lameness) for the study population of 2749 Quebec dairy herds partitioned into four study clusters using two-stage clustering algorithm

Outcome measure	Cluster 1 (n = 746)			Cluster 2 (n = 1070)			Cluster 3 (n = 363)			Cluster 4 (n = 570)		
	LSmean	SE	DF	LSmean	SE	DF	LSmean	SE	DF	LSmean	SE	DF
BCS ≤ 2, %	1.3[b]	0.17	2736	0.7[a]	0.14	2736	3.1[c]	0.23	2736	3.9[d]	0.2	2736
Hock injuries, %	22.7[c]	0.29	2736	5.6[a]	0.25	2736	46.7[d]	0.39	2736	18.2[b]	0.33	2736
Neck injuries, %	2.3[a]	0.28	2736	2.3[a]	0.25	2736	4.6[b]	0.39	2736	8.2[c]	0.33	2736
Knee injuries, %	2.3[b]	0.29	2736	2.2[a]	0.25	2736	4.8[c]	0.40	2736	8.2[d]	0.34	2736
Lameness, %	4.7[b]	0.32	2736	3.3[a]	0.28	2736	9.5[c]	0.44	2736	14.7[d]	0.37	2736

[a-d] Different letters indicate significant differences between clusters ($P < 0.05$)

The difference in mean prevalence lameness amongst all four clusters was found to be statistically significant (Table **4.3**), with the largest difference between C2 and C4 (-15%; $P < 0.05$), followed by C1 and C4 (-14%; $P < 0.05$), and C2 and C3 (-11%; $P < 0.05$), while the dairy herds in C1 and C2 differed the least (+1.33%; $P = 0.05$) as opposed to herds in C1 and C4. The results of all six pairwise comparisons of low body condition herd prevalence were shown to be statistically significant (Table **4.3**). The same pattern found as in the case with lameness; in particular, the difference in mean prevalence between C2 and C4 (-3%; $P < 0.05$) was the highest, while the C1 and C2 (0.5%; $P < 0.05$) had the least difference in the mean prevalence of herds with low body condition scores.

Regarding the prevalence of hock injuries, all clusters varied considerably with significant differences ($P < 0.05$; Table **4.3**). In particular, herds in C3 had a significantly higher mean herd prevalence of hock injuries (-41.2%; $P < 0.05$) compared to C2. In contrast, the difference in the mean prevalence of hock injuries between C1 and C4 was shown to be the least. With regards to the remaining measures of dairy cattle welfare, all four clusters had statistically significant differences in mean herd prevalence of neck and knee injuries, except for the mean herd prevalence of neck injuries between C1 and C2 (Table **4.3**). The same pattern is observed, as in the case of the lameness and low body condition scores, relative to the difference in the mean herd prevalence of neck and knee injuries between C2 and C4 being the highest (-6%; $P < 0.05$) and (-11%; $P < 0.05$), respectively. Also, herds had the lowest difference in the mean prevalence of knee injuries between C1 and C2 (+1.5%; $P < 0.05$).

4.1.4 Differences amongst clusters in herd prevalence of pre-recorded indicators of the HSI

The multiple comparisons showed statistically significant differences of mean prevalences relative to some of the pre-recorded HSI indicators (Table 4.4).

The multiple comparisons results revealed statistically significant differences of mean herd prevalence of cow longevity between C2 and C3 (+1.7%; $P = 0.004$) and the mean cow mortality prevalence between C2 and C4 (-0.7%; $P = 0.001$). Regarding the involuntary replacement rate, there were no statistically significant differences observed amongst clusters.

The mean herd prevalences of cows with low milk urea nitrogen (MUN), Herd Management Index (HMI), and Transition Cow Index (TCI) are the composite indicators of the HSI that correspond to its dimension of nutrition, production, and management. The results (Table 4.4) showed statistically significant differences in mean prevalences amongst clusters relative to HMI and TCI; however, there were no statistically significant differences observed in the case of the mean herd prevalence of low MUN. The difference in HMI between C1 and C4 was the highest, with C1 having a considerably high HMI (+463; $P < 0.0001$), followed by the difference between C2, C3 and C4 (+289; $P < 0.0001$), (+2609; $P < 0.003$). However, there were no significant differences between the remaining clusters. The significant differences were observed in mean TCI between C4 and C1 (+130.5; $P < 0.0001$), C2 (+115; $P < 0.0001$) and C3 (+135; $P < 0.0001$), while the last three clusters did not significantly differ relative to this indicator.

The mean herd prevalence of cows with high BHB in milk was significantly different between C1 and C2 (-0.2%; $P = 0.008$), C2 and C3 (+0.3%; $P = 0.0007$), as well as between C3 and C4 (-0.3%; $P = 0.0029$), while there were no significant differences observed amongst the remaining clusters. In the case of the mean herd prevalence of cows with high SCC in milk, the herds in C1 (-0.9%; $P = 0.0007$) and C2 (-0.3%; P < 0.05) had lower mean herd prevalence of cows with high SCC compared to C4. Regarding the mean herd prevalence of P:F, the clusters did not significantly differ in herd prevalence of high protein-fat ratio in milk (Table 4.4)

Multiple comparisons results on housing types and year-seasons can be found in Supplemental Table **8.1** – Supplemental Table **8.6**.

Table 4.4. Mean herd prevalence of twelve pre-recorded indicators of the herd status index for the study population of 2749 Quebec dairy herds partitioned into four study clusters using two-stage clustering algorithm

Indicators	Cluster 1 (n = 746)			Cluster 2 (n = 1070)			Cluster 3 (n = 363)			Cluster 4 (n = 570)		
	LSmean	**SE**	**DF**	**LSmean**	**SE**	**DF**	**LSmean**	**SE**	**DF**	**LSmean**	SE	DF
Longevity												
Cow longevity, %	38.9[ab]	0.34	2684	39.7[b]	0.29	2684	38.0[a]	0.47	2684	38.8[ab]	0.39	2684
Involuntary Replacement Rate, %	31.2[ab]	0.39	2663	31.1[a]	0.34	2663	31.5[ab]	0.54	2663	32.3[b]	0.45	2663
Cow mortality, %	3.81[ab]	0.15	2690	3.54[a]	0.13	2690	4.0[ab]	0.21	2690	4.26[b]	0.18	2690
Nutrition, production and profitability												
Cows with low MUN[1], %	2.8	0.18	2417	3.1	0.15	2417	2.9	0.24	2417	3	0.21	2417
Cow with high milk P:F[2], %	3.4	0.10	2708	3.5	0.09	2708	3.4	0.14	2708	3.3	0.12	2708
HMI[3]	48.1[c]	48.80	2680	-125.9[b]	42.50	2680	-145.7[b]	67.40	2680	-415.1[a]	57.30	2680
TCI[4]	179.4[b]	16.40	2734	163.8[b]	14.30	2734	183.5[b]	22.70	2734	48.9[a]	19.30	2734
Young stock and reproduction												
Calf mortality, %	8.1	0.21	2711	7.6	0.19	2711	8.2	0.30	2711	8.3	0.25	2711
Age at first calving, months	25.9[a]	0.07	2711	26.1[b]	0.06	2711	25.9[ab]	0.10	2711	26.2[ab]	0.09	2711
Abortion rate, %	0.9	0.12	2703	0.7	0.11	2703	0.7	0.17	2703	1.0	0.14	2703
Cows with high BHB[5], %	1.9[a]	0.06	2703	2.1[b]	0.05	2703	1.8[a]	0.08	2703	2.1[b]	0.07	2703
Cows with high SCC[6], %	11.5[a]	0.18	2711	11.8[a]	0.16	2711	12.1[ab]	0.25	2711	12.4[b]	0.21	2711

[a-d] Different letters indicate significant differences between clusters ($P < 0.05$)

[1]MUN – milk urea nitrogen; [2]P:F – protein-fat ratio; [3]HMI – herd management index ; [4]TCI – transition cow index; [5]BHB - ß-hydroxy butyrate; [6]SCC – somatic cell count

4.2 Objective 2 - investigate the differences amongst clusters relative to herd status index and perform a comparative analysis in relation to outcome measures of welfare

Multicollinearity analysis revealed high correlation coefficients between HMI, lifetime profit rank and TCI, and VIF factors of 3.5, 3.7, and 1.6, respectively. Since lifetime profit rank had many missing values and consisted of different variables that are not readily available, the lifetime profit rank was removed, resulting in twelve composite indicators. Removal of the lifetime profit rank indicator resulted in a considerably low VIF factor of 1.5 for the HMI (Supplemental Table **8.7**) and no strong correlation between remained indicators (Supplemental Table 8.8), thus eliminating an issue of multicollinearity. The remaining twelve indicators were aggregated into a single index per each herd based on a linear additive aggregation method by simply summing equally weighted composite indicators and dividing by the total number of composite indicators. Overall, the HSI calculation for each study herd (n = 2749) resulted in the general HSI with a range of between a minimum value of 0.12 and a maximum of 0.86.

The study results showed statistically significant differences in mean HSI between C4 and the remaining clusters, C1 (0.036; $P < 0.05$), C2 (0.028; $P < 0.05$), and C3 (0.027; $P < 0.05$), whereas there were no significant differences between the first three clusters (Table **4.5**).

Table 4.5. Herd status index for the study population of 2749 Quebec dairy herds enrolled in Canadian dairy herd improvement program partitioned into four clusters using two-stage clustering algorithm

Clusters	LSmean	SE	DF
Cluster 1	0.51^{b}	0.005	2736
Cluster 2	0.50^{b}	0.004	2736
Cluster 3	0.49^{b}	0.006	2736
Cluster 4	0.47^{a}	0.005	2736

$^{a\text{-}d}$ Different letters indicate significant differences between clusters ($P < 0.05$)

Furthermore, Pearson correlation analyses revealed no strong linear relationships between the HSI and outcome welfare measures (Table **4.6**). Multiple comparison results on housing types and year-seasons can be found in Supplemental Table 8.9.

Table 4.6. Pearson correlation analysis between the HSI and five proAction® on-farm assessment outcome measures of welfare (BCS≤2, hock, neck, and knee injuries, lameness) for the study population of 2749 Quebec dairy farms enrolled in Canadian dairy herd improvement program

Variable	R	P-value
BCS ≤ 2	-0.044	0.022
Hock injuries	-0.001	0.968
Neck injuries	-0.064	0.001
Knee injuries	-0.175	0.0001
Lameness	-0.099	0.0001

5. DISCUSSION

The objective of this study was to evaluate the validity of the HSI, constructed based on pre-recorded indicators, for the identification of dairy cattle welfare at the herd level relative to an existing proAction® on-farm outcome-based welfare assessment method. It was hypothesized that the HSI could reflect the overall state of dairy cattle welfare at the herd level relative to a recognized on-farm outcome-based welfare assessment method. Five outcome measures of welfare from proAction® on-farm welfare assessment data and thirteen pre-recorded indicators, extracted from the Quebec DHI data, were used to meet the objective of the study.

5.1 Segregation of study herds based on five-dimensions – outcome measures of welfare

Cluster analysis identified four distinct groups of herds with considerably differing sizes. About 21% of study herds were concentrated in the C4, in which herds had the highest prevalence of welfare measures among all four clusters, except for the prevalence of hock injuries. Hence, C4 was considered as a group of herds with the highest prevalence of welfare issues. On the other hand, C2 with a total of about 40% of study herds was described as having the least welfare issues based on the lowest mean herd prevalence of all five measures; however, the difference between C1 and C2 in herd prevalence of neck injuries was found to be not significant. The clusters C1 and C3 had a high prevalence of cows with hock injuries, while the values for remaining measures were in between those in C2 and C4.

5.2 The performance of the HSI on identifying dairy herd welfare state relative to the proAction® on-farm outcome measures of welfare

Among all four clusters, herds in C4 had the lowest mean HSI of 0.47 ± 0.005, and that was found to be significantly different from the means HSI of the remaining groups, whereas the means HSI for the remaining clusters were slightly higher (C1: 0.51 ± 0.005; C2: 0.50 ±0.004; C3:

0.49 ± 0.006), however with no significant differences amongst these clusters. With regards to individual pre-recorded indicators used for the calculation of the HSI, herds in C2 significantly differed from at least one of the remaining clusters in a total of seven pre-recorded indicators – longevity, involuntary replacement and mortality rates, HMI and TCI, cows with high SCC > 400,000 cells/ml in milk, and high concentration of BHB > 0.2 mmol/L of milk, five out of which had significant differences from those in C4 and remaining two with herds in C3. Warner et al. (2020) demonstrated that the HSI was comparatively stable for herds that had low- (>p10) and high-ranks (<p90) with standard deviations of 0.066 and 0.062, respectively, while the herds that had the HSI of between 25th and 75th percentiles had a considerably higher standard deviation of 0.162. Thus, the authors concluded that the HSI could be used to target those herds with high and least welfare issues (Warner et al., 2020). This study confirmed that the HSI based on twelve pre-recorded indicators from the national recording database could identify dairy cattle welfare with a high prevalence of welfare issues and partially herds with the lowest prevalence of welfare issues.

5.2.1 Pre-recorded indicators of the HSI under the category of longevity

The HSI is calculated as a simple average of the selected indicators and which allows unifying equally weighted components and include or exclude indicators of different measurement units once they are normalized, therefore allowing to rank given subjects. Thus, considering the constant development of welfare measures, digitalization, and undergoing research, a change of indicators might be required.

Regarding the individual pre-recorded indicators used for the construction of the HSI, the herds in C4 had the highest mean prevalence cow mortality of 4.26 ± 0.18% and the highest mean prevalence involuntary replacement rate of 32.3 ± 0.45%, both pre-recorded indicators showing significant differences with herds in C2 only. De Vries *et al.* (2011) reported that the culling was among pre-recorded indicators related to the highest number of outcome welfare measures. In Canada, lameness is one of the top reasons for involuntary culling, with a total of 7.0% (Canadian Dairy Information Center, 2019). Otten *et al.* (2016) demonstrated the potential of pre-recorded indicators to discriminate between lame and non-lame cows. In our study, Pearson correlation coefficients showed a significant but weak association ($r < \pm 0.4$) between the overall HSI and the mean herd prevalence of lameness. The mean herd prevalence of lameness, in our study, was substantially low (8.9 ± 9.9%) as compared to those found in recent epidemiological studies where the herd prevalence of lameness was 21% (Solano et al., 2013) and 25% (Bouffard et al., 2017) on farms with free- and tie-stall systems, respectively. The low lameness prevalence in our data might be due to potential underrate of the actual prevalence, either due to the scoring method (i.e., studies have shown the potential risk of underrate when using stall lameness scoring (Palacio et al., 2017), cow sample (i.e., only cows in the lactating herds were assessed excluding sick cows) or observer bias (Vasseur *et al.*, 2013).

On the other hand, C2 – herds with the least welfare issues showed the highest mean prevalence of cow longevity (39.7 ± 0.29%) and overall high HSI (0.50 ± 0.004%), which only was different from the HSI of herds in C4. Higher longevity has been previously associated with good welfare and farm profitability (Bruijnis *et al.*, 2013; Alvåsen *et al.*, 2018; Robichaud *et al.*, 2019); therefore, longer longevity is desirable, implying healthy cows and profitable farms. Therefore, the high mean herd prevalence of cow longevity may indicate that herds in C2 are healthier compared to the remaining groups, which is in accordance with the classification of herds in this cluster by outcome measures as a group with the least welfare issues and the highest prevalence of cow longevity indicator.

5.2.2 Pre-recorded indicators of the HSI under the category of nutrition, production, and profitability

The herd prevalence of cows with low MUN < 5mg/dL of milk, HMI, TCI, and high P:F>1.1 are the HSI indicators that are used to evaluate nutritional, production, and management parameters. HMI is an indicator of optimal use of the genetic potential calculated based on standardized milk. TCI is used to monitor a transition cow's health status and performance level relative to the industrial standards and is patented by the Wisconsin Alumni Research Foundation (AgSource Cooperative Services). To our knowledge, no studies included HMI and TCI as potential indicators to identify herd welfare status. Both HMI and TCI for herds in C4 were significantly different from those in the remaining cluster herds with the lowest values of -415.1 ± 57.30 for HMI and 48.9 ± 19.30 for the TCI, which indicate the ineffectiveness of optimal use of the genetic potential and a farm's transition cow program on farms grouped under C4. The MUN is a parameter used for the evaluation of the nutritional aspect. Sandgren *et al.* (2009) showed that increased herd prevalence of cows with low body condition score (≤ 2) and injuries were associated with high or low MUN level in milk. However, our study revealed no significant results for the herd prevalence of cows with low MUN < 5 mg/dL of milk between study clusters. A possible explanation for the absence of the significant differences amongst four cluster groups in our study might be a comparatively high number of farms (n=319 missing values (12%)) having no information for the given indicator. Perhaps including an indicator of the herd prevalence of cows with high MUN might improve its potential in combination with an indicator of herd prevalence of cows with low MUN.

5.2.3 Pre-recorded indicators of the HSI under the category of young stock and reproduction

Milk quality analysis is a widely used method in monitoring dairy cattle health and reproduction (Brandt *et al.*, 2010; Ginestreti *et al.*, 2020). SCC has been used in several studies to assess herd-level welfare based on pre-recorded data (de Vries *et al.*, 2014; Ginestreti *et al.*, 2020) that have shown inconsistent results. In this study, among milk analysis parameters, mean herd prevalence of cows with high SCC > 400,000 cells/ml in milk was the highest for the C4 (12.4 ± 0.21%), which was significantly different from the herds in C1 (11.5 ± 0.18%) and C2 (11.8 ± 0.16%), while the latter two had no difference. International standards for the acceptable count of SCC vary between 350,000 cells/ml up to 750,000 cells/ml, with a level of 400,000 cells/ml in Canada (Kelly *et al.*, 2018). However, some studies recommend a threshold at 200,000 cells/ml for the level of SCC to be counted, which enables identifying about 85% of infected cows, thus avoiding false positive and false negative test results (Kelly *et al.*, 2018). In the HSI context, the level of SCC is 400,000 cells/ml; therefore, reducing the threshold to SCC > 200,000 cells/ml in milk might help improve its potential to identify the welfare state at the herd-level. Furthermore, research has shown an effect of high SCC on milk production, that is, cows with high levels of SCC produce less milk (Cinar *et al.*, 2015), or reduced SCC was shown to be associated with increased herd milk production (Kelly *et al.*, 2018). Thus, it may be more reasonable to add an indicator of milk production that would give a broader picture in combination with a parameter of high SCC in milk. Another important parameter of milk quality data is a concentration of BHB in milk, for which excessive levels indicate metabolic disorders (Benedet *et al.*, 2019). The mean herd prevalence of elevated BHB > 0.2 mmol/L of milk differed for herds in C4 and C2, whereas herds in C3 had the lowest mean herd prevalence that significantly differed from those in C2. Therefore, in this study, the indicator of elevated BHB level in milk could not identify clusters of herds with the highest and lowest welfare issues relative to on-farm outcome welfare measures.

In the context of the HSI – age at first calving, calf mortality and abortion rates were classified under the category of young stock and reproduction, which in our study did not differ amongst clusters. However, previous studies on the development of pre-screening tools for identifying herds with “poor” welfare have shown a high potential and high sensitivity to identify herds with poor welfare when using a number of reproduction-related pre-recorded indicators. A study conducted in Sweden (Sandgren *et al.,* 2009) reported that fertility-related indicators, in particular, a combination of herd prevalence of cows with late ongoing artificial inseminations, the prevalence of heifers without mating/artificial insemination by the age of 17-mo, and calf mortality (2-6-mo) had the highest sensitivity, i.e., potential to identify herds with poor welfare status correctly. A high number of values at 0 (herd prevalence of abortion rate (2542, 92.4%), calf mortality rate (243, 8.8%)) may explain the absence of significant results for the indicators of herd prevalence of abortion rate and calf mortality. Therefore, it might be reasonable to consider replacing indicators under the category of young stock and reproduction of the HSI with those reproduction-related indicators that were shown to have the potential to identify herds with welfare and collected regularly.

Overall, the findings of this study demonstrated the potential of the HSI to target herds with the highest prevalence of welfare issues; however, there was a low potential of the HSI to distinguish herds with good welfare. The on-farm welfare assessment focuses on identifying the prevalence of welfare issues of dairy herds using those welfare indicators that measure welfare defficiencies (aim on identifying welfare issues). We may say that, for instance, the low prevalence of given welfare measures can indicate that those herds have good welfare, but it is not correct since other welfare issues not assessed as part of the target assessment may co-exist with measured ones (i.e., we measured lameness, but not resting time or human-animal relationship). If we want to identify herds with good welfare, we need to use data that assess welfare using a more comprehensive range of indicators and indicators of good welfare. It directly depends on how we define "good" welfare; what is the threshold for labeling a herd as having a good state of welfare? For instance, previous studies have applied a threshold, i.e., if the herd was not scored amongst the bottom 10% based on nine outcome welfare measures, then it was considered among the herds with "good" welfare (Sandgren *et al.*, 2009). Brouwer *et al.* (2015) used a threshold CCHM < 60 points = poor cattle health and ≥60 = sufficient cattle health. Again, the multidimensional concept of welfare (i.e., biological functioning, affective state, and natural behavior) makes it challenging to measure a comprehensive welfare level. It is also challenging to determine the level at which a "good" welfare threshold can be set. In our study, the group (C2) with the lowest prevalence of welfare issues amongst the remaining three clusters had a prevalence of welfare issues ranging from 0.7 ± 0.14% to 5.6 ± 0.25%. One could question if those herd prevalences are representative of a good level of welfare (i.e., shouldn't we target less than 1, 5, 10 %?).

Furthermore, five out of twelve HSI indicators significantly differed, while the remaining indicators showed no difference or difference only with one group. The absence of significant differences between clusters may be explained by the study data being at the herd level since welfare issues observed at an individual level can not necessarily be observed at a group level (De Vries *et al.*, 2011; Ginestreti *et al.*, 2020). That is, for instance, individual-level SCC can be diluted at the herd level (Ginestreti *et al.*, 2020). Also, for the calculation of the HSI in the current study, DHI data was used three years before the proAction® on-farm assessment with an assumption that welfare at the herd level is more stable than at an individual level and the difference in a few years does not result in substantial changes. However, the state of herd welfare might have changed as a result of, i.e., housing system-related improvements. For instance, Keil *et al.* (2006) reported that providing cows with a minimum of 50h access to outdoors during a 4-wk period reduced the prevalence of hock lesions, thus showing that positive welfare changes may be observed in a short period. Another important aspect to consider is the selection of the indicators that have shown a high potential to identify outcome measures of welfare. This may indicate that the HSI's performance to identify overall welfare status at the group level can be improved by including additional pre-recorded indicators for which potential scientific evidence exists. For example, similar studies have shown the potential of reproduction-related indicators to have high sensitivity in identifying herds with poor welfare (Sandgren *et al.,* 2009). However, the selected reproduction indicators used for the development of the HSI did not show significant results in our study. Therefore, it may be reasonable to replace the indicators of the HSI with those that have shown comparatively high potential in order to enhance the performance level of the HSI to detect herds with welfare deficiencies. Also, defining a different threshold for health-related pre-recorded indicators, particularly cows with high SCC, high BHB, and low MUN may improve the HSI's performance level. One might argue that more than twelve indicators could have been selected among those over 70 variables available. However, our study does not support this view since, from the statistical perspective, a greater number of indicators does not improve its performance level (Otten *et al.,* 2019). Thus, when it comes to choosing among the vast range of indicators available, we believe that it is more constructive to select few but essential indicators with high potential to assess welfare and that encompass essential aspects of dairy cattle welfare (i.e., biological functioning, affective state, and natural behavior), rather than increasing the numbers of indicators.

Finally, other approaches might be used to classify herds based on outcome measures rather than the cluster analysis used in this study, which might yield different results. One of the challenges when performing cluster analysis is determining the optimal number of clusters, based on which given data is partitioned into subgroups of units. Many authors claim that a two-stage clustering algorithm is a more reliable approach to apply (Balijepally *et al.*, 2011; Hair *et al.*, 2013), which was used in this study. We used a visual approach based on dendrograms obtained through hierarchical clustering algorithms with a combination of three well-known methods for the determination of optimal cluster numbers. Based on our results, we also assume that the cluster numbers of 2 may have been optimal since the C4 was outstanding relative to the overall HSI and most of its indicators, whereas the overall HSI for C1, C2, and C3 were not significantly different, and illustrated no or few differences in terms of pre-recorded indicators of the HSI. Finally, bias related to an on-farm welfare assessment method itself may have led to a not true herd prevalence of the welfare issues. For example, the herd sample size is based on herd size, but herd size is not considered in the analysis since assessment results are based on herd prevalence, i.e., the proportion of cows with welfare problems; there is a possible risk that herds with small size proportionally have higher welfare issues and the opposite for large herds (Nyman *et al.,* 2011). Moreover, although the assessors of welfare implementation under the proAction® Quality assurance program received training to assess cows, the reliability of its outcome was not verified during the data collection.

6. CONCLUSION

In conclusion, the study supports that the HSI can identify herds with the highest prevalence of welfare issues, and therefore, could be useful for advisory services and, thus, for early intervention to take the necessary measures. Moreover, such an approach can also be used in certification programs and contribute to avoiding the risks related to data collection of welfare outcome measures such as observer bias. However, the study also demonstrated the considerably low potential of the HSI to detect herds with good welfare. Furthermore, the indicators' contribution to the HSI's performance level differed considerably, with the considerable significant differences between clusters with the highest and the lowest welfare issues for five indicators, while the remaining indicators showed no significant results.

Further research is needed on the selected indicators of the HSI and those indicators for which potential scientific evidence exists to enhance the performance of the HSI for the identification of dairy cattle welfare at the herd level. Alongside developing valid methods (i.e., technologies to assess other dimensions of animal welfare such as behavioral aspect or affective state), such unifying index should also take into account positive measure of welfare to broaden the use of such method to assess overall dairy herd welfare status.

7. REFERENCES

Alawneh, J., Laven, R., & Stevenson, M. (2011). The effect of lameness on the fertility of dairy cattle in a seasonally breeding pasture-based system. *Journal of Dairy Science*, *94*(11), 5487-5493.

Alvåsen, K., Dohoo, I., Roth, A., & Emanuelson, U. (2018). Farm characteristics and management routines related to cow longevity: a survey among Swedish dairy farmers. *Acta Veterinaria Scandinavica*, *60*(1), 38.

Balijepally, V., Mangalaraj, G., & Iyengar, K. (2011). Are we wielding this hammer correctly? A reflective review of the application of cluster analysis in information systems research. *Journal of the Association for Information Systems*, *12*(5), 1.

Barkema, H. W., von Keyserlingk, M. A., Kastelic, J. P., Lam, T., Luby, C., Roy, J.-P., LeBlanc, S. J., Keefe, G. P., & Kelton, D. F. (2015). Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science*, *98*(11), 7426-7445. https://www.journalofdairyscience.org/article/S0022-0302(15)00617-7/pdf

Benedet, A., Manuelian, C., Zidi, A., Penasa, M., & De Marchi, M. (2019). Invited review: β-hydroxybutyrate concentration in blood and milk and its associations with cow performance. *animal*, *13*(8), 1676-1689.

Berry, D., Lee, J., Macdonald, K., Stafford, K., Matthews, L., & Roche, J. (2007). Associations among body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle. *Journal of Dairy Science*, *90*(2), 637-648.

Bicalho, R., Vokey, F., Erb, H., & Guard, C. (2007). Visual locomotion scoring in the first seventy days in milk: Impact on pregnancy and survival. *Journal of Dairy Science*, *90*(10), 4586-4591.

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

Booth, C., Warnick, L., Gröhn, Y., Maizon, D., Guard, C., & Janssen, D. (2004). Effect of lameness on culling in dairy cows. *Journal of Dairy Science*, *87*(12), 4115-4122. https://www.journalofdairyscience.org/article/S0022-0302(04)73554-7/pdf

Bouffard, V., De Passille, A. M., Rushen, J., Vasseur, E., Nash, C. G. R., Haley, D. B., & Pellerin, D. (2017). Effect of following recommendations for tiestall configuration on neck and leg lesions, lameness, cleanliness, and lying time in dairy cows. *Journal of Dairy Science*, *100*(4), 2935-2943. https://www.journalofdairyscience.org/article/S0022-0302(17)30145-5/pdf

Brandt, M., Haeussermann, A., & Hartung, E. (2010). Invited review: Technical solutions for analysis of milk constituents and abnormal milk. *Journal of Dairy Science, 93*(2), 427-436.

Broom, D. M. (1986). Indicators of poor welfare. *British veterinary journal*, *142*(6), 524-526. https://www.sciencedirect.com/science/article/abs/pii/0007193586901090?via%3Dihub

Brouwer, H., Stegeman, J., Straatsma, J., Hooijer, G., & van Schaik, G. (2015). The validity of a monitoring system based on routinely collected dairy cattle health data relative to a standardized herd check. *Preventive veterinary medicine*, *122*(1-2), 76-82. https://www.sciencedirect.com/science/article/pii/S0167587715300143?via%3Dihub

Bruijnis, M., Meijboom, F., & Stassen, E. (2013). Longevity as an animal welfare issue applied to the case of foot disorders in dairy cattle. *Journal of Agricultural and Environmental Ethics*, *26*(1), 191-205.

Chapinal, N., Barrientos, A., Von Keyserlingk, M., Galo, E., & Weary, D. (2013). Herd-level risk factors for lameness in freestall farms in the northeastern United States and California. *Journal of Dairy Science*, *96*(1), 318-328. https://www.journalofdairyscience.org/article/S0022-0302(12)00826-0/pdf

Chapinal, N., De Passille, A., Weary, D., Von Keyserlingk, M., & Rushen, J. (2009). Using gait score, walking speed, and lying behavior to detect hoof lesions in dairy cows. *Journal of Dairy Science*, *92*(9), 4365-4374.

Cinar, M., Serbester, U., Ceyhan, A., & Gorgulu, M. (2015). Effect of somatic cell count on milk yield and composition of first and second lactation dairy cows. *Italian Journal of Animal Science*, *14*(1), 3646.

Clay, N., Garnett, T., & Lorimer, J. (2020). Dairy intensification: Drivers, impacts and alternatives. *Ambio*, 1-14.

Compton, C., Heuer, C., Thomsen, P. T., Carpenter, T., Phyn, C., & McDougall, S. (2017). Invited review: A systematic literature review and meta-analysis of mortality and culling in dairy cattle. *Journal of Dairy Science*, *100*(1), 1-16.

Croyle, S. (2020). *Steps towards understanding the consistently high prevalence of lameness and hock injuries on Canadian dairy farms*

De Vries, M., Bokkers, E., Dijkstra, T., Van Schaik, G., & De Boer, I. (2011). Invited review: Associations between variables of routine herd data and dairy cattle welfare indicators. *Journal of Dairy Science*, *94*(7), 3213-3228. https://www.journalofdairyscience.org/article/S0022-0302(11)00314-6/pdf

de Vries, M., Bokkers, E., van Schaik, G., Engel, B., Dijkstra, T., & De Boer, I. (2014). Exploring the value of routinely collected herd data for estimating dairy cattle welfare.

Journal of Dairy Science, *97*(2), 715-730. https://www.journalofdairyscience.org/article/S0022-0302(13)00802-3/pdf

De Vries, M., Bokkers, E., Van Schaik, G., Engel, B., Dijkstra, T., & De Boer, I. (2016). Improving the time efficiency of identifying dairy herds with poorer welfare in a population. *Journal of Dairy Science*, *99*(10), 8282-8296. https://www.journalofdairyscience.org/article/S0022-0302(16)30453-2/pdf

Delgado, H., Cue, R., Haine, D., Sewalem, A., Lacroix, R., Lefebvre, D., Dubuc, J., Bouchard, E., & Wade, K. (2017). Profitability measures as decision-making tools for Québec dairy herds. *Canadian Journal of Animal Science*, *98*(1), 18-31.

Drackley, J. (2016). The importance of BCS management to cow welfare, performance and fertility. *WCDS Advances in Dairy Technology*, *28*, 195-206.

Eremenko, K. (2018). *Confident Data Skills: Master the Fundamentals of Working with Data and Supercharge Your Career*. Kogan Page Publishers.

Espejo, L., Endres, M. I., & Salfer, J. A. (2006). Prevalence of lameness in high-producing Holstein cows housed in freestall barns in Minnesota. *Journal of Dairy Science*, *89*(8), 3052-3058. https://www.journalofdairyscience.org/article/S0022-0302(06)72579-6/pdf

Flower, F., & Weary, D. (2006). Effect of hoof pathologies on subjective assessments of dairy cow gait. *Journal of Dairy Science*, *89*(1), 139-146.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

Fraser, D. (2003). Assessing animal welfare at the farm and group level: the interplay of science and values.

Fraser, D. (2008). Understanding animal welfare. *Acta Veterinaria Scandinavica*, *50*(1), 1-7.

Fraser, D., Weary, D. M., Pajor, E. A., & Milligan, B. N. (1997). A scientific conception of animal welfare that reflects ethical concerns.

Fraser, D. G. (2005). *Animal welfare and the intensification of animal production: An alternative interpretation* (Vol. 2). Food & Agriculture Org.

Frössling, J., Ohlson, A., & Hallén-Sandgren, C. (2017). Incidence and duration of increased somatic cell count in Swedish dairy cows and associations with milking system type. *Journal of Dairy Science*, *100*(9), 7368-7378.

Garnsworthy, P. C. (2006). Body condition score in dairy cows: targets for production and fertility. *Recent advances in animal nutrition*, *40*, 61.

Gibbons, J., Haley, D. B., Cutler, J. H., Nash, C., Heyerhoff, J. Z., Pellerin, D., Adam, S., Fournier, A., de Passillé, A. M., & Rushen, J. (2014). A comparison of 2 methods of assessing lameness prevalence in tiestall herds. *Journal of Dairy Science*, *97*(1), 350-353.

Gibbons, J., Vasseur, E., Rushen, J., & De Passillé, A. (2012). A training programme to ensure high repeatability of injury scoring of dairy cows. *Animal Welfare-The UFAW Journal*, *21*(3), 379.

Ginestreti, J., Strano, R. M., Lorenzi, V., Fusi, F., Angelucci, A., Ferrara, G., Galletti, G., Bergagna, S., Bolzoni, G., & Zanardi, G. (2020). Bulk tank milk quality data is unlikely to give useful information about dairy cow welfare at herd level. *Journal of Dairy Research*, 1-4.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). *Multivariate Data Analysis*. Pearson Education Limited. https://books.google.ca/books?id=VvXZnQEACAAJ

Harrell Jr, F. E., & Dupont, C. (2017). Hmisc: Harrell miscellaneous. R package version 4.0-3. *Online publication*.

Jelinski, M. D., LeBlanc, S., & Kennedy, R. (2015). Demographics of the Canadian dairy industry from 1991 to 2011. *The Canadian Veterinary Journal*, *56*(7), 701. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466820/pdf/cvj_07_701.pdf

Jewell, M., Cameron, M., Spears, J., McKenna, S., Cockram, M., Sanchez, J., & Keefe, G. (2019a). Prevalence of hock, knee, and neck skin lesions and associated risk factors in dairy herds in the Maritime Provinces of Canada. *Journal of Dairy Science*, *102*(4), 3376-3391. https://www.journalofdairyscience.org/article/S0022-0302(19)30114-6/pdf

Jewell, M., Cameron, M., Spears, J., McKenna, S., Cockram, M., Sanchez, J., & Keefe, G. (2019b). Prevalence of lameness and associated risk factors on dairy farms in the Maritime Provinces of Canada. *Journal of Dairy Science*, *102*(4), 3392-3405.

Johnsen, P., Johannesson, T., & Sandøe, P. (2001). Assessment of farm animal welfare at herd level: many goals, many methods. *Acta Agriculturae Scandinavica, Section A-Animal Science*, *51*(S30), 26-33.

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA. https://books.google.ca/books?id=-q3snAAACAAJ

Kassambara, A. (2018). ggpubr:"ggplot2" based publication ready plots. *R package version 0.1*, *7*. [Record #119 is using a reference type undefined in this output style.]

Keil, N., Wiederkehr, T., Friedli, K., & Wechsler, B. (2006). Effects of frequency and duration of outdoor exercise on the prevalence of hock lesions in tied Swiss dairy cows. *Preventive veterinary medicine*, *74*(2-3), 142-153.

Kelly, A. L., Leitner, G., & Merin, U. (2018). Milk quality and udder health: Test methods and standards.

Knierim, U., & Winckler, C. (2009). On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare*, *18*(4), 451-458.

Larsson, J., Fall, N., Lindberg, M., & Jacobson, M. (2016). Farm characteristics and management routines related to neonatal porcine diarrhoea: a survey among Swedish piglet producers. *Acta Veterinaria Scandinavica*, *58*(1), 77.

Lawson, R. G., & Jurs, P. C. (1990). New index for clustering tendency and its application to chemical problems. *Journal of chemical information and computer sciences*, *30*(1), 36-41.

[Record #120 is using a reference type undefined in this output style.]

Lescourret, F., Genest, M., Barnouin, J., Chassagne, M., & Faye, B. (1993). Data modeling for database design in production and health monitoring systems for dairy herds. *Journal of Dairy Science*, *76*(4), 1053-1062.

Lundmark, F., Berg, C., Wahlberg, B., & Röcklinsberg, H. (2015). 'One animal is no animal'–consequences of measuring animal welfare at herd level. *Know your food*, 31-35.

Mee, J. F. (2008). Newborn dairy calf management. *Veterinary Clinics of North America: Food Animal Practice*, *24*(1), 1-17.

Mörk, M., Lindberg, A., Alenius, S., Vågsholm, I., & Egenvall, A. (2009). Comparison between dairy cow disease incidence in data registered by farmers and in data from a disease-recording system based on veterinary reporting. *Preventive veterinary medicine*, *88*(4), 298-307.

Muhammad, I. U., Muhammad, A., & Muhammad, M. I. U. (2020). Package 'mctest'.

Nyman, A.-K., Lindberg, A., & Sandgren, C. H. (2011). Can pre-collected register data be used to identify dairy herds with good cattle welfare? Acta Veterinaria Scandinavica,

Olsson, S., Baekbo, P., Hansson, S., Rautala, H., & Østerås, O. (2001). Disease recording systems and herd health schemes for production diseases. *Acta Veterinaria Scandinavica*, *42*(1), 51-60.

Oţoiu, A., & Grădinaru, G. (2018). Proposing a composite environmental index to account for the actual state and changes in environmental dimensions, as a critique to EPI. *Ecological indicators*, *93*, 1209-1221.

Otten, N. D. (2014). *Identification of dairy herds with animal welfare problems*. SL Grafik.

Otten, N. D., Rousing, T., Houe, H., Thomsen, P. T., & Sørensen, J. T. (2016). Comparison of animal welfare indices in dairy herds based on different sources of data. *Animal Welfare*, *25*(2), 207-215.

Otten, N. D., Toft, N., Thomsen, P. T., & Houe, H. (2019). Evaluation of the performance of register data as indicators for dairy herds with high lameness prevalence. *Acta Veterinaria Scandinavica*, *61*(1), 49.

Palacio, S., Peignier, L., Pachoud, C., Nash, C., Adam, S., Bergeron, R., Pellerin, D., De Passille, A., Rushen, J., & Haley, D. (2017). Assessing lameness in tie-stalls using live stall lameness scoring. *Journal of Dairy Science*, *100*(8), 6577-6582.

Randall, L., Green, M., Chagunda, M., Mason, C., Archer, S., Green, L., & Huxley, J. (2015). Low body condition predisposes cattle to lameness: An 8-year study of one dairy herd. *Journal of Dairy Science*, *98*(6), 3766-3777.

Rasmussen, L. K., Nielsen, B. L., Pryce, J., Mottram, T., & Veerkamp, R. (1999). Risk factors associated with the incidence of ketosis in dairy cows. *ANIMAL SCIENCE-GLASGOW-*, *68*, 379-386.

Robbins, J., Von Keyserlingk, M., Fraser, D., & Weary, D. (2016). Invited review: Farm size and animal welfare. *Journal of Animal Science, 94*(12), 5439-5455. https://academic.oup.com/jas/article-abstract/94/12/5439/4703581?redirectedFrom=fulltext

Robichaud, M. V., Rushen, J., de Passillé, A., Vasseur, E., Orsel, K., & Pellerin, D. (2019). Associations between on-farm animal welfare indicators and productivity and profitability on Canadian dairies: I. On freestall farms. *Journal of Dairy Science*, *102*(5), 4341-4351.

Roche, J., Dillon, P., Stockdale, C., Baumgard, L., & VanBaale, M. (2004). Relationships among international body condition scoring systems. *Journal of Dairy Science, 87*(9), 3076-3079.

Roche, J. R., Friggens, N. C., Kay, J. K., Fisher, M. W., Stafford, K. J., & Berry, D. P. (2009). Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *Journal of Dairy Science*, *92*(12), 5769-5801. https://www.journalofdairyscience.org/article/S0022-0302(09)71299-8/pdf

Roche, S., Genore, R., Renaud, D., Shock, D., Bauman, C., Croyle, S., Barkema, H., Dubuc, J., Keefe, G., & Kelton, D. (2020). Describing mortality and euthanasia practices on Canadian dairy farms. *Journal of Dairy Science*, *103*(4), 3599-3605.

Rozzi, P., Miglior, F., & Hand, K. (2007). A total merit selection index for Ontario organic dairy farmers. *Journal of Dairy Science*, *90*(3), 1584-1593. https://www.journalofdairyscience.org/article/S0022-0302(07)71644-2/pdf

Rushen, J., Butterworth, A., & Swanson, J. (2011). Animal behavior and well-being symposium: Farm animal welfare assurance: Science and application. *Journal of Animal Science*, *89*(4), 1219-1228.

Rushen, J., Pombourcq, E., & de Passillé, A. M. (2007). Validation of two measures of lameness in dairy cows. *Applied Animal Behaviour Science*, *106*(1-3), 173-177.

Rutherford, K. M., Langford, F. M., Jack, M. C., Sherwood, L., Lawrence, A. B., & Haskell, M. J. (2009). Lameness prevalence and risk factors in organic and non-organic dairy herds in the United Kingdom. *The Veterinary Journal*, *180*(1), 95-105. https://www.sciencedirect.com/science/article/abs/pii/S1090023308001263?via%3Dihub

Sandgren, C. H., Lindberg, A., & Keeling, L. (2009). Using a national dairy database to identify herds with poor welfare. *Animal Welfare*, *18*(4), 523-532.

Santeramo, F. G. (2017). Methodological challenges in building composite indexes: Linking theory to practice. In *Emerging Trends in the Development and Application of Composite Indicators* (pp. 127-139). IGI Global.

Saraçli, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, *2013*(1), 203.

Smith, J., Ely, L., & Chapa, A. (2000). Effect of region, herd size, and milk production on reasons cows leave the herd. *Journal of Dairy Science*, *83*(12), 2980-2987.

Solano, L., Barkema, H., Pajor, E., Mason, S., LeBlanc, S., Heyerhoff, J. Z., Nash, C., Haley, D., Vasseur, E., & Pellerin, D. (2015). Prevalence of lameness and associated risk factors in

Canadian Holstein-Friesian cows housed in freestall barns. *Journal of Dairy Science*, *98*(10), 6978-6991.

Sørensen, J. T., Rousing, T., Møller, S. H., Bonde, M., & Hegelund, L. (2007). On-farm welfare assessment systems: what are the recording costs? *Animal Welfare*, *16*(2), 237-239.

Thomsen, P., Munksgaard, L., & Tøgersen, F. (2008). Evaluation of a lameness scoring system for dairy cows. *Journal of Dairy Science*, *91*(1), 119-126.

Thomsen, P., & Sørensen, J. T. (2009). Factors affecting the risk of euthanasia for cows in Danish dairy herds. *Veterinary Record*, *165*(2), 43-45. https://veterinaryrecord.bmj.com/content/165/2/43.long

Thomsen, P. T., Kjeldsen, A., Sørensen, J. T., Houe, H., & Ersbøll, A. K. (2006). Herd-level risk factors for the mortality of cows in Danish dairy herds. *Veterinary Record*, *158*(18), 622-626.

Tomaszewski, M. A. (1993). Record-keeping systems and control of data flow and information retrieval to manage large high producing herds. *Journal of Dairy Science*, *76*(10), 3188-3194.

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

Vasseur, E. (2017). Animal behavior and well-being symposium: optimizing outcome measures of welfare in dairy cattle assessment. *Journal of Animal Science*, *95*(3), 1365-1371. https://academic.oup.com/jas/article-abstract/95/3/1365/4703832?redirectedFrom=fulltext

Vasseur, E., Gibbons, J., Rushen, J., & De Passillé, A. (2013). Development and implementation of a training program to ensure high repeatability of body condition scoring of dairy cows. *Journal of Dairy Science*, *96*(7), 4725-4737.

Von Keyserlingk, M., Rushen, J., de Passillé, A. M., & Weary, D. M. (2009). Invited review: The welfare of dairy cattle—Key concepts and the role of science. *Journal of Dairy Science*, *92*(9), 4101-4111. https://www.journalofdairyscience.org/article/S0022-0302(09)70735-0/pdf

von Keyserlingk, M. A., Barrientos, A., Ito, K., Galo, E., & Weary, D. M. (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*(12), 7399-7408. https://www.journalofdairyscience.org/article/S0022-0302(12)00760-6/pdf

Warner, D., Vasseur, E., Villettaz Robichaud, M., Adam, S., Pellerin, D., Lefebvre, D. M., & Lacroix, R. (2020). Development of a Benchmarking Tool for Dairy Herd Management Using Routinely Collected Herd Records. *Animals*, *10*(9), 1689.

Whay, H., Barker, Z., Leach, K., & Main, D. (2012). Promoting farmer engagement and activity in the control of dairy cattle lameness. *The Veterinary Journal*, *193*(3), 617-621.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. springer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

8. SUPPLEMENTAL MATERIAL

Supplemental Table 8.1 Mean herd prevalence of proAction® on-farm assessment outcome welfare measures (cows with low BCS≤ 2, hock, neck, and knee injuries, lameness) for the study population of 2749 Quebec dairy herds partitioned into four study clusters using two-stage clustering algorithm on farms with free- and tie-stall housing systems

Outcome measure	**Free-stall housing system (n=530 herds)**			**Tie-stall housing system (n=2219 herds)**		
	LSmean	**SE**	**DF**	**LSmean**	**SE**	**DF**
BCS ≤ 2, %	2.4	0.19	2736	2.1	0.11	2736
Hock injuries, %	23.8[a]	0.33	2736	22.8[b]	0.18	2736
Neck injuries, %	3.4[a]	0.33	2736	5.3[b]	0.18	2736
Knee injuries, %	7.5[a]	0.37	2736	8.6[b]	0.20	2736
Lameness, %	7.1[a]	0.35	2736	11.6[b]	0.20	2736

[a-b] Different letters indicate significant differences between housing systems ($P < 0.05$)

Supplemental Table 8.2 Pre-recorded indicators of the herd status index for the study population of 2749 Quebec dairy herds partitioned into four clusters using two-stage clustering algorithm on farms with free- and tie-stall housing systems

Outcome measure	Free-stall housing system (n=530 herds)			Tie-stall housing system (n=2219 herds)		
	LSmean	**SE**	**DF**	**LSmean**	**SE**	**DF**
Cows with low MUN[1], %	2.7[a]	0.20	2417	3.2[b]	0.11	2417
Cows with high P:F[2], %	3.6[b]	0.12	2708	3.2[a]	0.07	2708
Age at 1st calving, months	25.8[a]	0.08	2711	26.2[b]	0.05	2711
Cows with high BHB[3], %	2.1[b]	0.06	2703	1.8[a]	0.04	2703
Cows with high SCC[4], %	11.4[a]	0.21	2711	12.5[b]	0.12	2711

[a-b] Different letters indicate significant differences between housing systems ($P < 0.05$)

[1]MUN – milk urea nitrogen; [2]P:F – protein-fat ratio; [3]BHB - ß-hydroxy butyrate; [4]SCC – somatic cell count

Supplemental Table 8.3 Mean herd prevalence of the low body condition score (BCS≤2) in different seasons using Bonferroni's adjustment test (an error level of 0.05)

Pairwise comparisons	Estimate	SE	df	t-ratio	P-value
S1-Winter - S2-Spring	-1.4	0.57	2736	-2.43	0.552
S1-Winter - S3-Summer	0.2	0.59	2736	0.26	1.000
S1-Winter - S4-Fall	-0.7	0.55	2736	-1.28	1.000
S1-Winter - S5-Winter	-0.9	0.56	2736	-1.58	1.000
S1-Winter - S6-Spring	-1.3	0.54	2736	-2.33	0.724
S1-Winter - S7-Summer	-0.3	0.54	2736	-0.64	1.000
S1-Winter - S8-Fall	-0.3	0.55	2736	-0.60	1.000
S1-Winter - S9-Winter	-0.5	0.55	2736	-0.86	1.000
S2-Spring - S3-Summer	1.5	0.40	2736	3.82	0.005
S2-Spring - S4-Fall	0.7	0.34	2736	2.03	1.000
S2-Spring - S5-Winter	0.5	0.35	2736	1.41	1.000
S2-Spring - S6-Spring	0.1	0.32	2736	0.41	1.000
S2-Spring - S7-Summer	1.0	0.32	2736	3.23	0.046
S2-Spring - S8-Fall	1.1	0.33	2736	3.19	0.052
S2-Spring - S9-Winter	0.9	0.34	2736	2.68	0.267
S3-Summer - S4-Fall	-0.9	0.37	2736	-2.31	0.755
S3-Summer - S5-Winter	-1.0	0.39	2736	-2.68	0.268
S3-Summer - S6-Spring	-1.4	0.36	2736	-3.93	0.003
S3-Summer - S7-Summer	-0.5	0.36	2736	-1.39	1.000
S3-Summer - S8-Fall	-0.5	0.37	2736	-1.31	1.000
S3-Summer - S9-Winter	-0.6	0.38	2736	-1.67	1.000
S4-Fall - S5-Winter	-0.2	0.32	2736	-0.57	1.000
S4-Fall - S6-Spring	-0.5	0.28	2736	-1.98	1.000
S4-Fall - S7-Summer	0.4	0.28	2736	1.29	1.000
S4-Fall - S8-Fall	0.4	0.29	2736	1.31	1.000
S4-Fall - S9-Winter	0.2	0.30	2736	0.77	1.000
S5-Winter - S6-Spring	-0.4	0.30	2736	-1.24	1.000
S5-Winter - S7-Summer	0.5	0.30	2736	1.80	1.000
S5-Winter - S8-Fall	0.6	0.31	2736	1.82	1.000
S5-Winter - S9-Winter	0.4	0.32	2736	1.30	1.000
S6-Spring - S7-Summer	0.9	0.26	2736	3.52	0.016
S6-Spring - S8-Fall	0.9	0.27	2736	3.48	0.019
S6-Spring - S9-Winter	0.8	0.28	2736	2.81	0.182
S7-Summer - S8-Fall	0.0	0.27	2736	0.08	1.000
S7-Summer - S9-Winter	-0.1	0.28	2736	-0.46	1.000
S8-Fall - S9-Winter	-0.1	0.29	2736	-0.52	1.000

Supplemental Table 8.4 Mean herd prevalence of the hock injuries in different seasons using Bonferroni's adjustment test (an error level of 0.05)

Pairwise comparisons	Estimate	SE	df	t-ratio	P-value
S1-Winter - S2-Spring	-0.5	0.98	2736	-0.53	1.000
S1-Winter - S3-Summer	-0.5	1.02	2736	-0.51	1.000
S1-Winter - S4-Fall	0.9	0.94	2736	1.00	1.000
S1-Winter - S5-Winter	1.1	0.96	2736	1.15	1.000
S1-Winter - S6-Spring	1.2	0.93	2736	1.29	1.000
S1-Winter - S7-Summer	1.7	0.93	2736	1.87	1.000
S1-Winter - S8-Fall	1.6	0.94	2736	1.72	1.000
S1-Winter - S9-Winter	1.1	0.94	2736	1.18	1.000
S2-Spring - S3-Summer	0.0	0.69	2736	-0.01	1.000
S2-Spring - S4-Fall	1.5	0.58	2736	2.54	0.407
S2-Spring - S5-Winter	1.6	0.61	2736	2.68	0.269
S2-Spring - S6-Spring	1.7	0.55	2736	3.10	0.070
S2-Spring - S7-Summer	2.2	0.55	2736	4.06	0.002
S2-Spring - S8-Fall	2.1	0.57	2736	3.73	0.007
S2-Spring - S9-Winter	1.6	0.58	2736	2.79	0.189
S3-Summer - S4-Fall	1.5	0.64	2736	2.30	0.785
S3-Summer - S5-Winter	1.6	0.67	2736	2.45	0.520
S3-Summer - S6-Spring	1.7	0.61	2736	2.79	0.194
S3-Summer - S7-Summer	2.3	0.62	2736	3.65	0.010
S3-Summer - S8-Fall	2.1	0.63	2736	3.38	0.027
S3-Summer - S9-Winter	1.6	0.64	2736	2.54	0.399
S4-Fall - S5-Winter	0.2	0.54	2736	0.31	1.000
S4-Fall - S6-Spring	0.2	0.48	2736	0.52	1.000
S4-Fall - S7-Summer	0.8	0.48	2736	1.64	1.000
S4-Fall - S8-Fall	0.7	0.50	2736	1.34	1.000
S4-Fall - S9-Winter	0.2	0.51	2736	0.33	1.000
S5-Winter - S6-Spring	0.1	0.51	2736	0.16	1.000
S5-Winter - S7-Summer	0.6	0.51	2736	1.21	1.000
S5-Winter - S8-Fall	0.5	0.53	2736	0.95	1.000
S5-Winter - S9-Winter	0.0	0.54	2736	0.01	1.000
S6-Spring - S7-Summer	0.5	0.44	2736	1.22	1.000
S6-Spring - S8-Fall	0.4	0.46	2736	0.92	1.000
S6-Spring - S9-Winter	-0.1	0.48	2736	-0.16	1.000
S7-Summer - S8-Fall	-0.1	0.46	2736	-0.26	1.000
S7-Summer - S9-Winter	-0.6	0.48	2736	-1.28	1.000
S8-Fall - S9-Winter	-0.5	0.49	2736	-1.00	1.000

Supplemental Table 8.5 Mean herd prevalence of the knee injuries in different seasons using Bonferroni's adjustment test (an error level of 0.05)

Pairwise comparisons	Estimate	SE	df	t-ratio	P-value
S1-Winter - S2-Spring	-0.5	1.10	2736	-0.43	1.000
S1-Winter - S3-Summer	-1.8	1.14	2736	-1.57	1.000
S1-Winter - S4-Fall	0.4	1.06	2736	0.38	1.000
S1-Winter - S5-Winter	0.8	1.08	2736	0.76	1.000
S1-Winter - S6-Spring	0.6	1.03	2736	0.55	1.000
S1-Winter - S7-Summer	0.7	1.04	2736	0.68	1.000
S1-Winter - S8-Fall	0.7	1.05	2736	0.65	1.000
S1-Winter - S9-Winter	1.2	1.06	2736	1.17	1.000
S2-Spring - S3-Summer	-1.3	0.77	2736	-1.71	1.000
S2-Spring - S4-Fall	0.9	0.65	2736	1.34	1.000
S2-Spring - S5-Winter	1.3	0.68	2736	1.90	1.000
S2-Spring - S6-Spring	1.0	0.62	2736	1.69	1.000
S2-Spring - S7-Summer	1.2	0.62	2736	1.91	1.000
S2-Spring - S8-Fall	1.1	0.64	2736	1.80	1.000
S2-Spring - S9-Winter	1.7	0.65	2736	2.61	0.327
S3-Summer - S4-Fall	2.2	0.71	2736	3.07	0.079
S3-Summer - S5-Winter	2.6	0.75	2736	3.51	0.017
S3-Summer - S6-Spring	2.4	0.69	2736	3.44	0.022
S3-Summer - S7-Summer	2.5	0.69	2736	3.63	0.011
S3-Summer - S8-Fall	2.5	0.71	2736	3.50	0.017
S3-Summer - S9-Winter	3.0	0.72	2736	4.21	0.001
S4-Fall - S5-Winter	0.4	0.61	2736	0.70	1.000
S4-Fall - S6-Spring	0.2	0.53	2736	0.32	1.000
S4-Fall - S7-Summer	0.3	0.54	2736	0.58	1.000
S4-Fall - S8-Fall	0.3	0.55	2736	0.50	1.000
S4-Fall - S9-Winter	0.8	0.57	2736	1.47	1.000
S5-Winter - S6-Spring	-0.3	0.57	2736	-0.44	1.000
S5-Winter - S7-Summer	-0.1	0.57	2736	-0.19	1.000
S5-Winter - S8-Fall	-0.1	0.59	2736	-0.24	1.000
S5-Winter - S9-Winter	0.4	0.61	2736	0.69	1.000
S6-Spring - S7-Summer	0.1	0.50	2736	0.29	1.000
S6-Spring - S8-Fall	0.1	0.51	2736	0.21	1.000
S6-Spring - S9-Winter	0.7	0.53	2736	1.25	1.000
S7-Summer - S8-Fall	0.0	0.52	2736	-0.07	1.000
S7-Summer - S9-Winter	0.5	0.54	2736	0.98	1.000
S8-Fall - S9-Winter	0.6	0.55	2736	1.02	1.000

Supplemental Table 8.6 Mean herd prevalence of the involuntary replacement rate in different seasons using Bonferroni's adjustment test (an error level of 0.05)

Pairwise comparisons	Estimate	SE	df	t-ratio	P-value
S1-Winter - S2-Spring	0.04	1.36	2663	0.03	1.0000
S1-Winter - S3-Summer	0.54	1.42	2663	0.38	1.0000
S1-Winter - S4-Fall	0.77	1.32	2663	0.58	1.0000
S1-Winter - S5-Winter	0.88	1.34	2663	0.66	1.0000
S1-Winter - S6-Spring	3.26	1.29	2663	2.52	0.4198
S1-Winter - S7-Summer	4.03	1.30	2663	3.11	0.0680
S1-Winter - S8-Fall	3.92	1.31	2663	3.00	0.0986
S1-Winter - S9-Winter	4.47	1.32	2663	3.39	0.0251
S2-Spring - S3-Summer	0.50	0.94	2663	0.54	1.0000
S2-Spring - S4-Fall	0.73	0.78	2663	0.93	1.0000
S2-Spring - S5-Winter	0.85	0.82	2663	1.03	1.0000
S2-Spring - S6-Spring	3.23	0.75	2663	4.33	0.0006
S2-Spring - S7-Summer	3.99	0.75	2663	5.32	<.0001
S2-Spring - S8-Fall	3.89	0.77	2663	5.02	<.0001
S2-Spring - S9-Winter	4.44	0.79	2663	5.61	<.0001
S3-Summer - S4-Fall	0.23	0.87	2663	0.26	1.0000
S3-Summer - S5-Winter	0.34	0.91	2663	0.38	1.0000
S3-Summer - S6-Spring	2.72	0.84	2663	3.26	0.0414
S3-Summer - S7-Summer	3.49	0.84	2663	4.15	0.0012
S3-Summer - S8-Fall	3.38	0.86	2663	3.93	0.0032
S3-Summer - S9-Winter	3.93	0.88	2663	4.49	0.0003
S4-Fall - S5-Winter	0.12	0.73	2663	0.16	1.0000
S4-Fall - S6-Spring	2.50	0.65	2663	3.86	0.0041
S4-Fall - S7-Summer	3.26	0.65	2663	5.00	<.0001
S4-Fall - S8-Fall	3.16	0.68	2663	4.68	0.0001
S4-Fall - S9-Winter	3.71	0.70	2663	5.34	<.0001
S5-Winter - S6-Spring	2.38	0.69	2663	3.44	0.0214
S5-Winter - S7-Summer	3.15	0.70	2663	4.51	0.0002
S5-Winter - S8-Fall	3.04	0.72	2663	4.24	0.0008
S5-Winter - S9-Winter	3.59	0.74	2663	4.88	<.0001
S6-Spring - S7-Summer	0.77	0.60	2663	1.27	1.0000
S6-Spring - S8-Fall	0.66	0.62	2663	1.06	1.0000
S6-Spring - S9-Winter	1.21	0.65	2663	1.88	1.0000
S7-Summer - S8-Fall	-0.11	0.63	2663	-0.17	1.0000
S7-Summer - S9-Winter	0.44	0.65	2663	0.68	1.0000
S8-Fall - S9-Winter	0.55	0.67	2663	0.83	1.0000

Supplemental Table 8.7 Individual multicollinearity diagnostics of the pre-recorded indicators of the herd status index registered in dairy herd improvement program (Lactanet Inc. database, Sainte-Anne-de-Bellevue, QC, Canada) across 2749 Quebec dairy herds

Variable	**VIF[1]**	**TOL[2]**
Cow longevity, %	1.129	0.886
Involuntary Replacement Rate, %	1.144	0.874
Cow mortality, %	1.083	0.924
Cow with low MUN[3], %	1.032	0.969
Cow with high milk P:F[4], %	1.013	0.986
HMI[5]	1.499	0.667
TCI[6]	1.536	0.650
Calf mortality, %	1.499	0.667
Age at first calving, months	1.536	0.650
Abortion rate, %	1.499	0.667
Cows with high BHB[7], %	1.536	0.650
Cows with high SCC[8], %	1.499	0.667

[1]VIF – variance inflation factor (a threshold for a variable to remain as an indicator of the herd status index); [2]TOL – tolerance (a threshold for a variable to remain as an indicator of the herd status index); [3]MUN – milk urea nitrogen; [4]P:F – protein-fat ratio; 5HMI – herd management index ; [6]TCI – transition cow index; [7]BHB - ß-hydroxy butyrate; [8]SCC – somatic cell count

Supplemental Table 8.8 Pearson correlation coefficients of the twelve test-day pre-recorded indicators of the herd status index registered in dairy herd improvement program (Lactanet Inc. database, Sainte-Anne-de-Bellevue, QC, Canada)

Variable	Cow longevity, %	Involuntary Replacement Rate, %	Cow mortality, %	Cow with low MUN, %	Cow with high milk P:F, %	HMI	TCI	Calf mortality, %	Age at 1st calving, months	Abortion rate, %	Cows with high BHB, %	Cows with high SCC, %
Cow longevity, %	1											
Involuntary Replacement Rate, %	-0.324	1										
Cow mortality, %	-0.078	0.159	1									
Cow with low MUN[1], %	0.019	-0.028	-0.017	1								
Cow with high milk P:F[2], %	-0.031	0.033	0.012	0.049	1							
HMI[3]	-0.04	-0.022	-0.076	-0.131	-0.059	1						
TCI[4]	0.068	-0.071	-0.147	-0.145	-0.061	0.545	1					
Calf mortality, %	-0.014	0.051	0.047	0.012	0.006	-0.021	-0.057	1				
Age at 1st calving, months	-0.011	0.038	0.115	0.083	0.068	-0.291	-0.306		1			
Abortion rate, %	-0.022	0.008	0.01	-0.013	0.018	0.011	0.008	0.03	-0.03	1		
Cows with high BHB[5], %	0.05	-0.02	0.044	0.112	-0.007	-0.219	-0.205	-0.024	0.165	-0.019	1	
Cows with high SCC[6], %	-0.025	0.016	0.154	0.111	0.054	-0.233	-0.24	0.061	0.251	-0.011	0.169	1

[1]MUN – milk urea nitrogen; [2]P:F – protein-fat ratio; [3]HMI – herd management index ; [4]TCI – transition cow index; [5]BHB - ß-hydroxy butyrate; [6]SCC – somatic cell count

Supplemental Table 8.9 Mean herd prevalence of the herd status index in different seasons using Bonferroni's adjustment test (an error level of 0.05)

Pairwise comparisons	Estimate	SE	df	t-ratio	P-value
S1-Winter - S2-Spring	-3.6	173.7	2680	-0.02	1.000
S1-Winter - S3-Summer	-117.1	180.1	2680	-0.65	1.000
S1-Winter - S4-Fall	127.1	167.6	2680	0.76	1.000
S1-Winter - S5-Winter	62.2	170.7	2680	0.37	1.000
S1-Winter - S6-Spring	191.9	164.7	2680	1.17	1.000
S1-Winter - S7-Summer	241.7	164.9	2680	1.47	1.000
S1-Winter - S8-Fall	229.7	166.3	2680	1.38	1.000
S1-Winter - S9-Winter	343.6	168	2680	2.05	1.000
S2-Spring - S3-Summer	-113.5	118.4	2680	-0.96	1.000
S2-Spring - S4-Fall	130.7	98.8	2680	1.32	1.000
S2-Spring - S5-Winter	65.8	104	2680	0.63	1.000
S2-Spring - S6-Spring	195.5	94.4	2680	2.07	1.000
S2-Spring - S7-Summer	245.3	94.7	2680	2.59	0.346
S2-Spring - S8-Fall	233.3	97.5	2680	2.39	0.604
S2-Spring - S9-Winter	347.2	100.2	2680	3.46	0.020
S3-Summer - S4-Fall	244.1	109.1	2680	2.24	0.913
S3-Summer - S5-Winter	179.3	113.9	2680	1.57	1.000
S3-Summer - S6-Spring	308.9	105.1	2680	2.94	0.119
S3-Summer - S7-Summer	358.8	105.4	2680	3.40	0.024
S3-Summer - S8-Fall	346.8	107.7	2680	3.22	0.047
S3-Summer - S9-Winter	460.7	110.2	2680	4.18	0.001
S4-Fall - S5-Winter	-64.8	92.7	2680	-0.70	1.000
S4-Fall - S6-Spring	64.8	81.7	2680	0.79	1.000
S4-Fall - S7-Summer	114.6	82.1	2680	1.40	1.000
S4-Fall - S8-Fall	102.6	84.7	2680	1.21	1.000
S4-Fall - S9-Winter	216.6	88	2680	2.46	0.500
S5-Winter - S6-Spring	129.6	87.5	2680	1.48	1.000
S5-Winter - S7-Summer	179.5	87.8	2680	2.04	1.000
S5-Winter - S8-Fall	167.5	90.1	2680	1.86	1.000
S5-Winter - S9-Winter	281.4	93.3	2680	3.02	0.093
S6-Spring - S7-Summer	49.8	75.9	2680	0.66	1.000
S6-Spring - S8-Fall	37.8	78.5	2680	0.48	1.000
S6-Spring - S9-Winter	151.8	82.1	2680	1.85	1.000
S7-Summer - S8-Fall	-12	78.9	2680	-0.15	1.000
S7-Summer - S9-Winter	101.9	82.5	2680	1.24	1.000
S8-Fall - S9-Winter	113.9	84.5	2680	1.35	1.000