© 2020 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2020, 29: 449-461
ISSN 0962-7286
doi: 10.7120/09627286.29.4.449

Can on-farm animal welfare explain relative production differences between dairy herds?

L Frondelius*†, L Jauhiainen†, O Niskanen†, M Mughal‡ and A Sairanen†

† Natural Resources Institute Finland (Luke), Production Systems, Halolantie 31 A, 71750 Maaninka, Finland
‡ University of Eastern Finland, Department of Environmental and Biological Sciences, PO Box 111, 80101 Joensuu, Finland
* Contact for correspondence: lilli.frondelius@luke.fi

Abstract

The aim of this study was to investigate factors responsible for between-herd variation in milk production, when genetic variation is controlled. Quantitative information regarding farms' production environment and animal welfare, as evaluated annually by veterinarians and health and feeding records, were the factors studied. Principal component analysis was used to reduce the production environment variables as well as the welfare data. Two linear regression models were devised. The first model used welfare indicators and veterinary treatments to predict the difference between herds' observed and potential milk yield. The second model explained the residual of the first model by feeding and animal-based indicator data. Together, these two models explained 62% of the variance in milk yield differences between herds. Specifically, feeding of the herd was the most important factor, accounting for 67%, followed by the production environment/animal welfare (30%) and finally animal health, assessed through veterinary treatments, explained the remaining 3% of the variance. A poor welfare rating adversely affected milk production. Similarly, a low score for fatness at slaughter, poor milk quality and high mortality all showed a clear negative association with production. It was found that while feeding remains a major factor, production environment and animal welfare also have significant roles to play when it comes to production. Notably, those farms with major animal welfare problems were shown to display milk yield below the Finnish average.

Keywords: *animal welfare, dairy cattle, feeding, milk production, principal component analysis, routine herd data*

Introduction

There is a paucity of literature looking into the association between on-farm welfare assessments and milk production. Yet, demonstration of a positive association between animal welfare and farm profitability could have the potential to stimulate investment in improvements to cow comfort and welfare by dairy farmers (Villettaz Robichaud *et al* 2018). Coignard *et al* (2014) compared milk yield (test days) with a welfare score attained using the Welfare Quality® (WQ) protocol (Welfare Quality® 2009a). The score for good health (WQ principle) and the average of all four WQ principle scores were associated with milk yield, whereby those farms scoring the lowest had the higher producing cows. On the other hand, lower scores in the WQ criteria expression of social behaviour and positive emotional state were associated with reduced milk yield. Thus, the relationship between animal welfare and milk production appears to be multifaceted.

The development of WQ (Welfare Quality® 2009b) has enabled quantitative measurement of animal welfare to be carried out. However, WQ protocols are considered impractical and time-consuming for routine assessment of animal welfare (eg Knierim & Winckler 2009; Heath *et al* 2014). The estimated duration of a dairy cattle assessment is 5–8 h depending on the size of the herd (Welfare Quality® 2009a). Thus, WQ is not routinely used and cheaper, less time-consuming methods of on-farm welfare assessments are increasingly sought. Several national welfare monitoring systems which are similar in approach to WQ, but more simplified, are used or are under development, for example, in Denmark (Andreasen *et al* 2014), The Netherlands (Metz *et al* 2015) and Italy (Bertocchi & Fusi 2014).

Data related to herd productivity, health and fertility are routinely collected from dairy farms in developed countries, and many studies have shown this routine herd data to be associated with several on-farm welfare indicators (de Vries *et al* 2011). For example, it has been suggested that indicators reflecting issues with fertility and high mortality in different age groups, are potential indicators of poor welfare as well (Sandgren *et al* 2009; Krug *et al* 2015). These indicators may reflect problems in stockmanship and management (Sandgren *et al* 2009). Nyman *et al* (2011) tested routine herd data against nine animal-based, on-farm welfare measures (eg body condition score [BCS], cleanliness, lameness, injuries). In total, 28 herd data indicators

Table 1 Number of herds comprised of different data sources in the three study years.

Data source	Year		
	2014	**2015**	**2016**
Whole population (Finland)	8,567	8,124	7,574
Milk recording (Maitoisa)	6,242	5,865	5,468
Naseva (farms with evaluation of dairy cows included)	4,259	4,300	4,484
Feeding	3,464	3,190	2,868

Veterinarian treatments were available from all the farms included in the data.

showed a significant association with one or more on-farm welfare measures. Furthermore, some of the indicators (cows with late ongoing artificial inseminations, heifers without mating/artificial insemination by 17 months of age, stillbirth rate, cow mortality, mastitis incidence and incidence of feed-related diseases) were able to successfully identify herds with good welfare. De Vries *et al* (2014) studied the relationship between routinely collected herd data and welfare indicators and found that the farms with more lean cows also had a lower milk fat percentage, higher somatic cell count (SCC), worse reproductive performance and higher culling rate. Also, Brouwer *et al* (2015) found that routine data (eg mortality, SCC and infectious diseases) may be used for screening farms for poor health. However, it is worth noting that veterinarians may not always report treatment(s) in the databases thereby running the risk of biasing treatment numbers (Nyman *et al* 2011).

In Finland, producers can register with a veterinary herd health management programme (Naseva) which is administered by ETT Animal Health Association. These farms choosing to participate receive at least one non-emergency visit a year from a healthcare veterinarian to focus on potential targets for improvement in herd health management and animal welfare. The information from routine farm records and on-farm assessments are both utilised in the evaluation. The welfare section is loosely based on the WQ protocol, and over half the measures in this section are animal-based. The management plan needs to be renewed annually by the veterinarian in the course of a farm visit (Hokkanen *et al* 2015; ETT ra 2017). A participation rate of over 60% was recorded for 2016, ensuring that Naseva offers an accurate and wide-ranging insight into the health and welfare status of Finnish dairy herds.

The aim of this paper was to determine the factors which best explain variation in production between herds. To fully understand the effect of factors such as welfare and health on production, it was important to deploy the most comprehensive approach possible. Quantitative information on Finnish dairy production, feeding, health and welfare was assembled and a model was constructed to elucidate the effect of management on milk production.

Materials and methods

The quantitative data required for our study were collected from the dairy herd recording system and its subsidiaries, as well as from the centralised healthcare register for Finnish cattle herds (Naseva [https://www.naseva.fi/PublicContent/IntroductionInEnglish]). The number of dairy farms included in these differing databases (from 2014–2016) are shown in Table 1.

Dairy herd recording system and extensions

Field data for our study were collected from the Finnish Rural Advisory Services database (ProAgria, Vantaa, Finland). Milk production data were recorded in accordance with the International Committee for Animal Recording (ICAR 2017) and included both free- and tie-stall barns. The average milk yield for these farms over the three-year study period was 9,002 kg (8,872, 9,013 and 9,143 kg for 2014, 2015 and 2016, respectively). These milk production data were complemented by feeding records from the Finnish Rural Advisory Service. The base data were augmented by several cow- and farm-based variables, which were used to supplement the content available. These sources of data are described in the following paragraphs.

The Maitoisa application

The models used to help estimate the breeding value need to be able to differentiate between the genetic and management effects on milk production and a solution to this was found in the form of the test-day models (Jensen 2001). Maitoisa ('Milky' in English), is a test-day model which recognises management problems at the farm level and this dairy herd management application produced our outcome variable (Koivula *et al* 2007). The Maitoisa regression model analysis provides an estimate of how much milk a cow should theoretically produce, taking into account: (i) the cow's breeding value; (ii) stage of lactation; (iii) calving age; (iv) season; and (v) parity. The difference between estimated and observed value (Maitoisa variable; kg milk per day) is dependent upon management factors, such as feeding and production environment. At the country level, the average of these differences is zero and so deviation from zero allows the performance of the farm in question to be compared with the national average (Koivula *et al* 2007). For our analysis, instead of using the cow-level variable, we utilised the annual mean of each herd as an outcome variable since this was the highest common denominator.

Feeding extension

Finnish farms make extensive use of the feed rationing programme, CowCompass® (Mtech Digital Solutions Ltd, Finland) which encompasses approximately 70% of the farms participating in the milk recording scheme. The programme includes on-farm feed intake, measured at the herd level, typically 2–4 times per year. Accumulated data (milk production and composition, feed composition) are utilised for ongoing feeding ration adjustment. These annual feed intake and diet ratio measurements do not provide a comprehensive picture of the herd's feeding, but they offer a reflection of the farm's feeding strategy since, for Finnish

https://doi.org/10.7120/09627286.29.4.449 Published online by Cambridge University Press

herds, the seasonality of milk production is only ± 7% (Luke 2018). Separate analysis of the composition of the diet has been addressed via the silage (SDMI) and concentrate intake (CDMI) indexes for which a score of 100 signifies the base level of the calculation (Huhtanen *et al* 2011). Feed rationing is also augmented by the standardised energy corrected milk yield (sECM), which describes the amount of milk produced daily, assuming cows are provided the standard diet of 100 index points (Huhtanen *et al* 2011). The sECM can thus be said to describe the milk production potential of the cows on the particular farm in question. The indices and standardised milk production used in this study are shown in Table 2.

Veterinary treatments

Veterinary treatments consisting of 37 clinical groups are stored within the database and comprise a total of 143 different treatment codes. These codes are entered onto the cows' insemination card by the veterinarian, after which the farmer or artificial inseminator transfers them to the database. For further analyses, 17 general classification groups were created: reproductive dysfunctions and diseases; paralyses; metabolic diseases; diseases of the vascular system; diseases of the kidney and urinary tract; abdominal diseases; gastrointestinal diseases; poisoning; udder diseases, hoof and limb disorders; other injuries; respiratory diseases; diseases of the central nervous system (CNS); skin diseases; endoparasites; infectious diseases; and other diseases. The number of annual treatments in the herd was divided by herd size to provide an indicator of health.

Centralised healthcare register for Finnish cattle herds (Naseva)

During the annual veterinary assessment, the evaluation form and management plan are routinely completed and transferred onto the database along with the herd's health status and production data. The assessment is carried out in two distinct parts: general information on production and management; and on-farm evaluation (comprising animal welfare and incidence/prevention of infectious diseases). The evaluation form consists of 55 quantitatively evaluated measures, in addition to qualitative feedback for each farmer. Animal welfare parameters and incidence of infectious diseases are recorded separately for all the animal groups (eg calves, heifers, dairy cows and bulls) present on the farm and there are also two further subsections where the farm is evaluated as a whole. The quantitative evaluation operates on a three-tier scale, whereby one represents a good situation/practice, two, a satisfying situation/practice and three, that the situation/practice should be improved. Values of 0 (the measure cannot be performed on the farm) or X (the measure does not concern the farm) can also be applied. All the measures concerning dairy cows from the animal welfare and incidence of infectious diseases sections were utilised as expected, and a number of measures from the production and management and prevention of infectious diseases sections were also included, when they related to animal welfare status on the farm. Disbudding, which relates to calves as opposed to dairy cows, was also included. In total, forty-three measures from the Naseva register were included for further analyses. A description of the Naseva measures can be seen in the supplementary material to papers published in *Animal Welfare* section on the UFAW website: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material.

Table 2 Mean (± SD) descriptive statistics of feeding indexes, standardised energy-corrected milk, lactation stage (days in milk) and herd size used in regression Model 2.

Factor	Mean (± SD)
Silage dry matter intake index	105.9 (± 5.92)
Concentrate dry matter intake index	108.3 (± 8.72)
Standardised energy-corrected milk	28.41 (± 2.87)
Days in milk	189.8 (± 32.53)
Herd size (dairy cows)	43.7 (± 30.39)

Principal component analysis

Principal component analysis (PCA) is a statistical procedure that relies upon orthogonal transformation to convert a set of observations of potentially correlated variables into a set of linearly uncorrelated variables called principal components (PC). The 43 Naseva measures had a relatively high rate of correlation, and a PCA was used to compress the information in these data into a smaller number of variables for further use. Additionally, PCA helps to avoid the problem arising from multicollinearity (Jolliffe 2002) of the Naseva variables in further regression analyses.

Prior to selecting PCA variables, the variation caused by herd and year was modelled to reveal which of these indicate short-term change in farm production conditions and which provide information on long-term conditions that may also affect milk production. This was carried out with a variance component model, whereby estimates of variance between farms, and variations within farms were estimated. Relative variance was calculated as a ratio of variance between herds and total variance of the measure. If a relative variance was less than 40%, the measure was categorised as characterising an acute situation, rather than a condition with a long-standing effect on herd milk production and was therefore not included in the PCA. Seven variables were omitted from the PCA as a result of this criterion. These were 'symptoms of clinical salmonella', 'symptoms of paratuberculosis', 'diarrhoea', 'respiratory infections', 'symptoms of *Mycoplasma bovis*', 'on-farm symptoms of ringworm on assessment day' and 'ringworm status of the farm in the previous three years.'

There were two Naseva measures that did not attain a high loading (≤ -0.3 or ≥ 0.3) in any of the PCs but, based on their high values of relative variance, can be expected to explain some of the differences between farms. These measures were 'fatness score in slaughter' (relative

Table 3 Naseva measures included in the principal component analysis, loadings (≥ 0.3) of the measures in two principal components, their estimated variance between herds and relative variance.

Section	Measure	Principal component 1	Principal component 2	Variance estimate	Relative variance (%)[†‡]
Production and management	Diseases, treatments and on-farm deaths have been recorded and evaluated	0.37		0.12	61.0
	Antimicrobial residuals are tested from milk			0.05	59.3
	Bacteriological mastitis tests are used regularly			0.09	68.4
	Calf production			0.07	53.7
Animal welfare	Nutritional state	0.41		0.04	46.2
Feeding	Water availability	0.48		0.20	64.5
	Milk or replacer provision and administration method	0.34		0.15	65.1
Health	General condition and health	0.46	0.38	0.01	46.2
	Animal grouping and group uniformity	0.41		0.05	64.4
	Coat condition	0.43		0.03	47.6
	Hoof health and lameness	0.36	0.41	0.08	53.0
	Skin and joint alterations	0.49		0.16	64.3
Housing conditions	Cleanliness and dryness of animals	0.51		0.13	62.6
	Air quality and temperature	0.49		0.04	58.3
	Space allowance	0.52		0.07	73.2
	Sufficient lying space	0.48		0.07	64.6
	Lying comfort	0.54		0.15	74.5
	Calving management	0.45	–0.55	0.23	83.8
	Tethering and other restrictions of movement	0.46	–0.57	0.23	82.5
	Slipperiness	0.32	0.38	0.05	65.3
Behaviour	Behaviour within the animal group	0.36		0.01	45.0
	Human-animal relationship	0.34		0.00	44.1
	Stereotypic and abnormal behaviour			0.01	39.3
	Handling of the animals and handling facilities	0.33		0.03	67.9
Prevention of infectious diseases	Hygiene in feeding, bedding and manure handling	0.45		0.12	73.8
Incidence of infectious diseases					
Clinical symptoms over the past 12 months	*Streptococcus agalactiae* mastitis			0.03	56.3
	Abortions			0.05	32.6
	Infectious hoof diseases		0.51	0.06	55.8
Ringworm	Ringworm status of the farm during the last three years			0.01	50.7

[†] Relative variance is calculated as a ratio of variance between herds and total variance of the data including yearly measures for all herds.
[‡] Relative variance less than 40% describes mainly momentary conditions, not conditions with a long-standing effect on the herd milk production level. These variables were dropped from the PCA.

variance; 53.5), and 'disbudding' (78.8). In contrast, the variables 'access to outdoor loafing area' (84.0), 'udder health' (62.0), 'milk quality' (61.7), 'mortality' (47.3), and 'access to pasture' (83.5) had loading in PCs, but also a high relative variance. Additionally, these variables come under public scrutiny. And, so, these seven measures were left out of the final PCA and included in the further analysis as 'individual Naseva measures.'

The final PCA comprised 29 measures (Table 3) out of the original 43 and extracted two principal components. Extracting more than two PCs resulted in only one measure with a high loading in each additional PC.

Regression models

In order to check the significance of the potential interactions between welfare indicators (PC and Naseva), veterinary treatments, feeding indicators, sECM, DIM and herd size, all the variables were first added to the same regression model. In this preliminary model, feeding indicators were available for only 3,464 herds (2,212 herds for all years). The inclusion of all variables in the same multiple regression model with veterinarian treatments and Naseva data was inconclusive with respect to animal welfare as the data from the non-matching herds (43%) needed to be removed from the analysis. Instead, to ensure all the possible Naseva information was included, we used two-staged modelling. First, we fitted a model to explain the Maitoisa variable with welfare indicators and veterinarian treatments (Model 1). Second, residuals (ie unexplained variability of the model) from Model 1 were calculated and used as dependent variables, while feeding variables were used as independent variables (Model 2). The mean of residuals was zero, and the mean of the Maitoisa variable was 0.45. This 0.45 was added to all residuals to shift distribution of residuals so that it did not differ in location from the distribution of the Maitoisa variable. Thus, the adjusted residual of Model 1 represents the Maitoisa variable without the variance that was explainable with the Model 1 explanatory variables. To check the validity of the chosen method, the preliminary model was compared with Model 2. The parameter estimates only differed slightly between the two models, but standard errors were higher in the preliminary model compared to Model 2, and thus we chose to use the separate models.

In regression models, statistical significance was set at $P < 0.05$. Variables not reaching significance were omitted from the final models. All statistical analyses were performed using SAS v 9.4 (SAS Institute Inc, Cary, NC, USA), with FACTOR (PCA analysis), GLM (General linear model), MIXED (Mixed linear model) and CORR (Correlations) procedures. Scatterplots were used to check assumptions about normal distribution, the homoscedasticity and a linear relationship between independent and dependent variables.

Model 1: The Maitoisa variable with veterinary treatment and Naseva data

A linear regression model was built to fit the known given values of two PCs based on the Naseva data, the set of individual Naseva measures and veterinary treatment to reveal what proportion of the variance in the Maitoisa variable can be explained by independent variables and what remains inexplicable. Certain veterinary treatments had a similar effect on the Maitoisa variable and were mutually correlated. Thus, these treatments were combined, and the final model had three veterinary treatment variables: treatments for reproductive dysfunctions; treatments for metabolic, udder and CNS diseases; and treatments for other diseases and dysfunctionality. Those Naseva variables with low variation between herds and that did not correlate with other measures (such as 'symptoms of clinical salmonella') were omitted from the regression models. Since the relationship between the Maitoisa and the other variables was not linear (as the model assumes), the majority of the variables were categorised as follows (to achieve linearity):

1) Treatments for reproductive dysfunctions, three levels: A (0.0–0.1 treatments per animal), B (0.1–0.2), C (0.2 or more);

2) Treatments for metabolic, udder and CNS diseases, three levels: A (0.0–0.1 treatments per animal), B (0.1–0.2), C (0.2 or more);

3) Treatments for other diseases and dysfunctions, five levels: A (0.00–0.25 treatments per animal), B (0.25–0.25), C (0.50–0.75), D (0.75–1.00), E (1.00 or more);

4) Principal Component 1 (PC1), five levels: A (score of –1.7 or less), B (–1.7 to –0.6), C (–0.6 to +0.6), D (0.6 to 1.7), E (1.7 or higher);

5) Principal Component 2 (PC2), five levels: A (score of –1.7 or less), B (–1.7 to –0.6), C (–0.6 to +0.6), D (0.6 to 1.7), E (1.7 or higher);

6) Fatness score in slaughter, four levels: A (score 1.00), B (1.01–2.00), C (2.01–2.99), D (3.00);

7) Access to outdoor loafing area, two levels: A (score 1.00–2.00), B (2.01–3.00);

8) Udder health, two levels: A (score 1.00–2.00), B (2.01–3.00);

9) Milk quality, three levels: A (score 1.00), B (1.01–2.99), C (3.00);

10) Mortality, three levels: A (score 1.00–1.99), B (2.00–2.99), C (3.00);

11) Disbudding, two levels: A (score 1.00–2.00), B (2.01–3.00);

12) Access to pasture, three levels: A (score 1.00), B (1.01–2.00), C (2.01–3.00).

Variables represent the mean of all farm-specific observations over the study years. A total of 3,887 farms had all the variables from all three years and were thus included in Model 1. The number of farms classified in each category is shown in Table 4.

Table 4 Linear regression Model I describing the factors influencing the Maitoisa variant of the herd with least square means, standard errors and significant differences. Independent variables were categorised to achieve linearity and are presented with their class intervals.

Variable	Level	Interval	N	Least squares mean	Standard error
Treatments for reproductive dysfunctions (number of treatments per animal)	A	0.0–0.1	1,327	0.06[c]	0.44
	B	0.1–0.2	927	0.41[b]	0.43
	C	> 0.2	1,633	0.69[a]	0.43
Treatments for metabolic, udder and CNS diseases (number of treatments per animal)	A	0.0–0.1	1,555	0.70[a]	0.43
	B	0.1–0.2	1,222	0.65[a]	0.43
	C	> 0.2	1,110	0.24[b]	0.43
Treatments for other diseases and dysfunctions (number of treatments per animal)	A	0.00–0.25	1,004	–0.31[d]	0.44
	B	0.25–0.50	1,187	0.41[c]	0.43
	C	0.50–0.75	863	0.91[b]	0.44
	D	0.75–1.00	475	1.11[b]	0.45
	E	> 1.00	358	1.46[a]	0.45
Principal Component 1 (score)	A	< –1.7	28	0.65[b]	0.59
	B	–1.7–(–0.6)	1,069	0.63[ab]	0.46
	C	–0.6–0.6	2,025	0.51[b]	0.45
	D	0.6–1.7	582	–0.15[c]	0.46
	E	> 1.7	183	–0.48[c]	0.48
Fatness score at slaughter	A	1.00	366	1.10[a]	0.32
	B	1.01–2.00	3,447	0.99[ab]	0.31
	C	2.01–2.99	70	0.48[b]	0.41
	D	3.00	4	–1.72[b]	1.20
Access to outdoor loafing area	A	1.00–2.00	2,134	0.35[b]	0.43
	B	2.01–3.00	1,753	0.55[a]	0.43
Udder health	A	1.00–2.00	2,412	0.52[a]	0.43
	B	2.01–3.00	1,475	0.23[b]	0.43
Milk quality	A	1.00	19	0.08[a]	0.71
	B	1.01–2.99	3,850	0.28[a]	0.38
	C	3.00	18	–2.55[b]	0.67
Mortality	A	1.00–1.99	2,490	0.34[a]	0.43
	B	2.00–2.99	1,226	0.30[a]	0.43
	C	3.00	171	–0.57[b]	0.45
Disbudding	A	1.00–2.00	3,214	0.33[a]	0.43
	B	2.01–3.00	673	0.14[b]	0.43
Access to pasture	A	1.00	15	0.14[ab]	0.70
	B	1.01–2.00	3,108	0.48[b]	0.40
	C	2.01–3.00	764	0.81[a]	0.41

Means within columns showing altered superscripts differ at $P < 0.05$.

https://doi.org/10.7120/09627286.29.4.449 Published online by Cambridge University Press

Model 2: The residual of Model 1 explained by feeding indicator data

In the second linear regression model, a residual of Model 1 was used as a dependent variable. The continuous variables used in linear regression Model 2 were sECM, SDMI index and CDMI index. The sECM describes the production potential of a herd and the other indexes describe both the intensity and quality of the diet. Days in milk (DIM) at two levels (< 200 or ≥ 200 days) and average herd size at two levels (< 60 or ≥ 60 cows) were used as independent (binary) class variables. In this case, the herds' mean DIM captures extended lactation intervals, and herd size indicates any effect related to size. Only those farms with at least one feed intake measurement per year were included, which led to 2,212 farms in Model 2. Variables represent the mean of all farm-specific observations over the three study years. Descriptive statistics of feeding parameters, DIM and herd size are shown in Table 2.

Results

Principal component analysis (PCA)

PCA was used to compress correlated Naseva measures into a potentially lower number of uncorrelated variables. Two principal components were extracted, whereby 29 Naseva measures with an absolute loading value of 0.30 or higher were included (Table 3). PC1 included most of the Naseva measures and these consisted of comfort around resting, nutrition, behaviour, health and animals' overall condition, but also certain specific aspects of management. All these measures had a positive loading on this PC and, thus, a higher PC score equates to poorer welfare since, in Naseva, higher points indicate a worse situation. PC1 was interpreted therefore as a variable of overall negative welfare. PC2 was interpreted as a variable of housing condition and management, and comprised measures, such as tethering and management of calving, but also various major health issues, such as hoof health. Most of the health- and welfare-related variables had a positive loading in this PC, but calving management and tethering had a negative loading, which complicated the interpretation of this PC. In the case of PC2, a higher PC score equates to a poorer situation regarding animal-based welfare measures, but a better one for management-based measures.

Regression analysis

The remaining variables in the final model in the first regression analysis (Model 1) explaining the Maitoisa variable are summarised in Table 4. PC2 did not have any association with the Maitoisa and was omitted from the final regression Model 1.

Those farms with a greater number of veterinary treatments for reproductive dysfunctions and other combined diseases showed a higher positive Maitoisa variable (P < 0.001) compared to farms with less treatments in these groups. Conversely, a higher number of treatments for metabolic, udder and central nervous system diseases was associated with a lower value for Maitoisa (P < 0.001).

A higher PC1 score was associated with a lower Maitoisa value (P < 0.001). Two classes with the highest factor scores

Table 5 **Linear regression Model 2 describing the factors influencing the Maitoisa variant of the herd with coefficient, standard error and significant *P*-values.**

Factor	Coefficient	Standard error	*P*-value
Silage dry matter intake index	0.061	0.0053	< 0.001
Concentrate dry matter intake index	0.094	0.0040	< 0.001
Standardised energy-corrected milk	0.423	0.0117	< 0.001
Days in milk (less than 200 days)	–0.414	0.0749	< 0.001
Herd size (more than 60 cows)	–0.392	0.0705	< 0.001

(≥ 0.6) had a negative Maitoisa, while farms with low or intermediate factor scores showed a positive Maitoisa. In individual Naseva measures — 'fatness score in slaughter', 'milk quality' and 'mortality' — the class with the highest values differed from other classes (P < 0.05, P < 0.001 and P < 0.001, respectively), and had a negative Maitoisa. Higher scores for 'disbudding' were also associated with a lower Maitoisa (P < 0.05). Farms with higher scores for 'access to outdoor loafing area' and 'access to pasture' had higher positive Maitoisa (P < 0.01) while higher scores in 'udder health' were associated with a lower Maitoisa (P < 0.01).

In Model 2, higher sECM, SDMI and CDMI indexes showed a positive correlation with Maitoisa (P < 0.001) (Table 5). As for DIM and herd size, the association was negative (P < 0.001).

To summarise, the first model, with Naseva evaluations and health records as explanatory variables, accounted for 18% (R^2) of the variance in the Maitoisa. The second model explained 54% (R^2) of the residual of Model 1, meaning 44% of original variation (0.54 × [100–18]% = 44%). CMDI, SDMI and sECM explained 13, 4 and 34%, respectively, of the remaining variation in Model 2 and 11, 3 and 28% of the original variation. DIM and herd size both accounted for less than 1%. The two models together explained 62% of the original Maitoisa variance. From this, feeding indicators comprised 67%, Naseva evaluations 30% and veterinary treatment the remaining 3%.

Discussion

The Maitosa random regression test-day model (Koivula *et al* 2007) allows measures of milk production between herds to be compared by quantifying individual animal factors. A lower result represents a lower milk yield compared to the average production which a genetically similar herd should be expected to achieve. This difference in production can occur as a result of less-intensive feeding and/or poorer cow management. On the other hand, a higher result is an indication of above average herd production. The measure was originally intended to be used on a daily basis to show the short-term situation at a given farm. For our study, annual herd averages were utilised to provide comparable values between herds.

Quantitative information on the production environment and farms' welfare status were analysed along with health and feeding records to estimate those factors most responsible for variance in the Maitoisa test-day model. We were able to explain the majority of the variance (62%), but a large proportion still remained unexplained due to the heterogeneity between herds. From the explained part, the most important factor was found to be feeding parameters with sECM, lactation length and herd size (67%), followed by the production environment and animal welfare (30%). Health problems, indicated by the number of treatments per animal, explained the remaining 3% of the variance.

The feeding results will be discussed with an emphasis on the production environment and animal welfare and their relationship with veterinary treatments and the health management programme (Naseva). A number of the Naseva evaluation welfare measures have not previously been specifically reported in the literature and so our discussion will consider the indirect connections between comparable welfare measures and milk production.

Feeding and standardised energy-corrected milk yield

Taking the total Maitoisa variation, 44% was attributable to feeding parameters with sECM, DIM and herd size. The most important single parameter was sECM (28%) which describes a cow's observed milk production potential. In principle, the Maitoisa should account for the animal effect and the sECM should have minimal influence. However, the Maitoisa explains the expected genetic value of the herd (genotypic variation) while the sECM is based on measured intake and milk production (phenotypic variation). Thus, the sECM calculation is more accurate compared to the Maitoisa, ie the observed sECM includes feed availability or microbiological quality, both of which are impossible to include in the feeding records.

Feeding parameters (CDMI and SDMI) describe feeding intensity. Feed indexes are assumed to be herd independent because within the most common breeds the interaction between diet and herd genotype is weak (Hammami *et al* 2009). The feed indexes captured 11 and 3% of the explained Maitoisa variation for the CDMI and SDMI index, respectively. The former is directly related to the amount and composition of concentrates in the diet (Huhtanen *et al* 2008) while the latter is related to digestible value and quality of roughage preservation (Huhtanen *et al* 2007). An increase in both of these indices correlates directly with milk yield. The variation in concentrate intake in the data was ± 1.7 kg DM per day which equates to ± 1 kg milk per day. Correspondingly, the variation in SDMI index was ± 5.9 index points, which is equal to 0.6 kg silage dry matter intake and approximately ± 0.5 kg milk (Huhtanen *et al* 2007). From a management perspective, adjusting concentrate supplementation is easier than adjusting silage quality, a concept which explains the superior effect of the CDMI index in the Maitoisa. Dummy class variables 'DIM' and 'size of the herd' captured only a minimal amount (both 1%) of the variation but were kept in the model as they were statistically significant.

Naseva and veterinary treatments

Although most of the variation in milk production can be attributed to feeding and sECM, production environment and animal management also have an effect. A high standard of welfare is a prerequisite for cow longevity and elevated milk yield. In most instances, reduced welfare is associated with decreased production (eg Rajala-Schultz *et al* 1999; Breuer *et al* 2000; Huxley 2013). However, evidence is growing that high yielding animals have greater susceptibility to problems related to health and welfare (Oltenacu & Broom 2010). A review article by Ingvartsen *et al* (2003) reported that the significant increase in lactational performance over the last few decades has occurred as a direct result of intensive genetic selection. Moreover, there has also been further intensification in feeding and management practices. These trends have been accompanied by an alarming increase in health and fertility problems, although there is a lack of clear evidence demonstrating a causal relationship with increased yield. In herd-level data there are always a variety of confounding factors that are impossible to take into consideration, and treatment rate may not always necessarily mirror disease incidence rate (Lind *et al* 2012). In our study, farms with greater incidence of reproductive dysfunctions and combined other diseases, per animal, had a higher positive Maitoisa variable, indicating that cows in high yielding farms have a greater tendency for health problems, particularly those related to fertility. This is in accordance with Coignard *et al* (2014) who found milk yield to be negatively associated with the score for the WQ principle,'good health', leading them to speculate that a high milk yield increases cows' susceptibility to health disorders, disorders that subsequently affect milk yield. However, Rearte *et al* (2018) showed that although a negative association exists between reproductive performance and milk yield, the relationship is weak and dependent on herd-level milk production.

On the other hand, the Maitoisa variable was decreased by metabolic diseases together with udder diseases and diseases of the CNS, suggesting that these health issues impinge directly on milk yield. That said, it is worth noting that Maitoisa did not drop below zero, ie the relative difference in production remained above average. Similarly, Rajala-Schultz *et al* (1999) found milk fever and ketosis to be associated with decreased milk yield but that this occurred more often in high production cows. The individual Naseva measure, 'udder health', also negatively affected the Maitoisa value since farms in lower and, thus, better classes in this measure showed a higher positive difference. In contrast, Ingvartsen *et al* (2003) found mastitis was the only disease to demonstrate a clear link between increased milk yield and risk of infection. Yet, here, even in the high 'udder health' class, the Maitoisa remained positive. Ingvartsen *et al* concluded that, overall, the relationship between production-related disorders and lactational performance was highly complicated.

https://doi.org/10.7120/09627286.29.4.449 Published online by Cambridge University Press

Generally speaking, care must be taken with the use of veterinary treatment databases in studies such as these. Data loss appears to be an occupational hazard and the extent to which records are complete varies depending on the disease (Espetvedt *et al* 2012; Lind *et al* 2012). Additionally, these databases are devised to only report instances of animals treated medically as opposed to recording all animals that are sick (Lind *et al* 2012).

The Maitoisa variable was associated with a general status of animal welfare, on-farm. PC1 was interpreted as being a variable of a negative overall welfare status. As all the Naseva measures' loadings in the PC1 were positive and, in Naseva, higher points equate to a worse situation, a higher PC score is also indicative of poorer welfare on the farm in question. Farms with low and intermediate scores (< 0.6) in this PC had a positive Maitoisa value, which did not differ significantly between classes. As a contrast, farms in the two classes with the highest PC scores showed a negative Maitoisa, and there was a clearly defined increase in negative Maitoisa as the PC increased. In the following paragraphs we will discuss the variables within the PC1 separately, although their association with Maitoisa remain highly interconnected.

In our data, lameness and hoof health were represented in both veterinary treatments and the Naseva data. From the Naseva data, the measure, 'hoof health and lameness' was included in PC1, and was assessed on-farm through an evaluation of the amount of lame animals and distinct changes in hoof conformation (ETT ra 2012). Lameness negatively impacts upon milk yield, with the decline often beginning prior to lameness being diagnosed clinically (Rajala-Schultz *et al* 1999). PC1 negatively affected the Maitoisa variable, leading us to assume that in our study lameness and poor hoof health had a negative association with milk production. However, in veterinary treatments, limb and hoof disorders were combined within 'other treatments', which had a positive association with Maitoisa. Although Rajala-Schultz *et al* (1999) found lameness to lower milk yield, older (parity ≥ 4) cows diagnosed as lame were actually producing more milk than healthy conspecifics. Also, Huxley (2013) reported high yielding animals to be more likely to become lame, with lameness reducing milk yield between 270 to 574 kg during one, single lactation. Thus, it is likely that as with fertility and other health problems, a high milk yield may also cause a predisposition for lameness. It is worth noting, however, that our data included only veterinary treatments of limb and hoof disorders and information from hoof trimmers was lacking — a factor which may have led to an underestimation of the prevalence of hoof disorders.

Both 'nutritional state' and 'water availability' from Naseva's section on feeding were positively loaded in the PC1 and, thus, negatively associated with the Maitoisa variable. For 'nutritional state', a rough estimation of the BCS is made and the amount of thin animals (ETT ra 2012). Results in the literature are somewhat contradictory regarding the association between BCS and milk yield (for reviews, see Roche *et al* 2009; de Vries *et al* 2011), perhaps due to the non-linear nature of this interaction (Roche *et al* 2009): up to the optimum BCS milk yield is positively associated with BCS, but a lower BCS leads to larger differences in milk yield. In our study, the individual Naseva measure 'fatness score in slaughter' also showed a negative association with the Maitoisa variant. As the score of this measure increased, so the positive Maitoisa decreased, and farms scoring 3 (ie more than 50% of the slaughtered cows showed the lowest fatness score [ETT ra 2012]), had a negative Maitoisa. In conclusion, our data showed, on average, that farms with a high proportion of thin cows tend to have a lower milk yield.

Virtually all the Naseva measures from the 'housing conditions' section were included in the PC1. Many of the measures in this section, such as 'cleanliness and dryness of the animals' and 'sufficient lying space', describe comfort around resting. Daily lying time has been suggested to have a direct link with milk yield (Cook 2008). Insufficient lying space and poor stall maintenance negatively affect milk production, at least in situations of overstocking (Bach *et al* 2008). Irrespective of milk yield, Cook (2008) suggested all cows require a minimum of 12 h of rest daily. Housing conditions, eg stall type and bedding material, also affect the prevalence and severity of integument alterations (eg Kester *et al* 2014; Zaffino Heyerhoff *et al* 2014). Skin and joint alterations are measured in the health section of the Naseva assessment and were included in PC1 with a positive loading, thereby affecting the Maitoisa negatively. On the other hand, Potterton *et al* (2011) reported the odds ratios of hair loss and ulceration to be higher for those cows yielding more than 34 kg per day compared with very low producing animals generating less than 21 kg a day.

Behavioural measures from Naseva, that were included in PC1, were 'behaviour within the animal group', 'human-animal relationship' and 'handling of the animals and handling facilities.' Coignard *et al* (2014) found that lower scores in the WQ criteria 'expression of social behaviour' and 'positive emotional state' were connected with lower milk yield. The Naseva measure 'behaviour within the animal group' comprises aspects of both these WQ criteria since it evaluates the agonistic social behaviour, but also the play behaviour, calmness and apathy of the animals (ETT ra 2012). Social competition within a herd can affect cow behaviour. For example, in a regrouping situation, standing time increases and the feeding behaviour of both dominant and subordinate cows is altered, which leads to a reduction in milk production (Hasegawa *et al* 1997; Phillips & Rind 2001). Indeed, Grant and Albright (2001) concluded one of the most potentially important factors affecting dry matter intake and productivity in cows was enabling access to feed for every cow in the group whenever she so desires. Dairy cows undergo frequent handling by humans during their life, and thus the relationship between animals and humans is a crucial component of dairy cattle welfare (Raussi 2003).

High cow reactivity and fear of people is linked to reduced milk production (Rushen *et al* 1999; Breuer *et al* 2000; Hemsworth *et al* 2000). In the study of Rushen *et al* (1999), the presence of a person with previous aversive handling of the cows during milking was found to increase the residual milk by 70%. Similarly, Breuer *et al* (2000) noted several behavioural variables indicative of fear of people were correlated negatively with milk yield, and 19% of the variation in milk production between farms was explained by fear of humans. However, their study was conducted in a region of Australia where herds have a year-round, pasture-based production system, meaning that the results may not be directly comparable to indoor production systems.

Access to pasture and to an outdoor loafing area were included in the regression analysis as individual measures from Naseva. Both had a negative effect on the Maitoisa variable, maintaining it, however, at a positive level. Several studies have illustrated a negative association between access to pasture and milk production (for a review, see de Vries *et al* 2011). Besides the direct effects of access to pasture or an outdoor loafing area on milk yield, there can also be more indirect associations based on the type of barn. It is mandatory in Finland for tie-stall barns to enable outdoor access for dairy cows — on pasture or to an outdoor loafing area — for at least 60 days between May and September (Council Regulation on the Protection of Cattle 592/2010). In free-stall barns it is permissible for dairy cows to be kept inside all year round. Thus, it could be that the farms attaining lower (and thus better) scores in these measures are tie-stall barns. In Naseva, farms score a 3 if they do not fulfill the legislation and 2 if there is no outdoor access for free-stall animals or if tie-stall animals have only shortened access to pasture (for more than 60 but less than 90 days) (ETT ra 2012).

'Disbudding' is a management-based measure in Naseva. It evaluates anaesthetic and analgesic use during disbudding (ETT ra 2012). Higher scores lowered the Maitoisa, which translates as farms that disbud calves without analgesics and/or anaesthetics having lower milk yields than those where all the calves are disbudded by a veterinarian using sedatives, local anaesthesia and analgesics. The association between disbudding management and milk production is almost certainly indirect and better 'disbudding' scores relate instead to the farm's welfare as a whole. In Finland, approximately 72% of farms disbud their calves and it is more common among those that are part of Naseva (Hokkanen *et al* 2015). It is also more common on farms with a healthcare agreement to call a veterinarian to perform the procedure.

For the individual Naseva measure 'mortality', farms scoring 3 (ie their mortality rate is more than 10% [ETT ra 2012]), had a negative Maitoisa. In contrast, farms with lower scores showed a positive Maitoisa. Mortality has been suggested as one of the key indicators of poor on-farm welfare (Sandgren *et al* 2009; Brouwer *et al* 2015) and, according to our results, it could also indicate lowered milk production, although the association is almost certainly indirect. Sarjokari *et al* (2018) reported Finnish farms to have, on average, a mortality rate of 6%, varying between 0 and 15.7%. The most commonly reported reasons for death on-farm (including both euthanasia and unassisted deaths) were accidents, calving difficulties, milk fever, mastitis, and foot and claw disease. Sarjokari *et al* (2018) also reported unnecessary regroupings during the dry period and unnecessary variations in barn facilities, such as floor types, as creating a greater likelihood of on-farm cow death.

Along with mortality, 'milk quality' was also associated with a lowered Maitoisa (farms scoring 3 had the most negative Maitoisa, lowering it to –2.42 kg, on average). In the Naseva evaluation, a score of 3 signifies that the SSC of the bulk milk on the farm has repeatedly exceeded 250,000 cells ml^{-1} (ETT ra 2012). Milk quality, as such, cannot be considered a measure of animal welfare, but more a measure of management and hygiene (Barkema *et al* 1998, 1999). In concordance with our results, it has been shown that farms with a low SCC (< 150,000 cells ml^{-1}) in bulk milk have higher milk production and, per decrease of 100,000 cells ml^{-1} in bulk milk SCC, the mean cumulative 305-day fat-corrected milk production increases by 272 kg (Barkema *et al* 1998).

As with veterinary treatment records, debate also occurs as to the reliability of the Naseva assessment (Mughal *et al* 2017). Flaws in the Naseva assessment, relative to other welfare assessments, consist of very cursory instructions and a lack of fixed sample sizes in measures. Another disadvantage is that the scoring system only differentiates farms that are satisfactory or need to improve from the rest of the farm population (good situation/practice). Thus, our data are unable to demonstrate the effect of positive welfare differences on milk production. However, based on our results, it seems at least that those farms with extreme Naseva results and assessed as needing to improve, show a clearly negative Maitoisa, thereby demonstrating a lower than average milk production compared to Finland as a whole. The advantage of Naseva is that more than 60% of Finnish dairy farms participate in the programme, meaning that these results can be considered relatively comprehensive.

Model structure

Two-stage modelling (ie using residuals as dependent variables) has been relatively common practice in a number of disciplines, such as accounting. Here, it allowed us to keep 43% more observations in the most critical part (animal welfare) of the study. However, recent work has revealed the potential for this two-stage approach to cause coefficient bias (Chen *et al* 2018). The magnitude of said bias (in terms of coefficients and standard errors) is a function of the correlations between model regressors. In our case, the regressors used in the first and second model were not correlated, and biased results should not be expected. However, it is important to be aware of this risk and the two-step process must be carefully considered before using.

https://doi.org/10.7120/09627286.29.4.449 Published online by Cambridge University Press

Animal welfare implications

Despite the major role that feeding plays in milk production, animal welfare and the production environment also play a key role in explaining production differences between dairy farms with a relatively high milk yield. Of course, incorporating this many variables as well as a PCA into the analysis makes it extremely hard to draw many specific conclusions regarding the associations between milk production and individual measures of welfare. However, what seems irrefutable is that poor animal welfare is associated with farms not reaching their full production potential, leading to loss of income. However, a growth in milk yield can conceivably be linked with an increase in certain health issues as well as veterinary treatments. Demonstrating this connection between animal welfare and farm profitability could potentially motivate farmers to invest in improvements in animal welfare. Further research is required into how routinely collected welfare data could help us achieve this goal.

Conclusion

The information provided by routinely updated welfare evaluations was found to explain some differences in production between dairy herds. The most important factors were feeding related, accounting for 67% of the total explained variance, followed by the production environment/animal welfare (30%) and animal health which explained the remaining 3%. Dairy farms with major issues regarding animal welfare and mortality were found to be especially prone to producing less milk than average in Finland.

Acknowledgements

The authors would like to thank ETT ra and the Naseva management group for supplying data from the Centralised healthcare register, and the Association of ProAgria Centres for providing data from the Dairy herd recording system. Also, thanks to Dr Jaakko Mononen for commenting on the manuscript during the revision process. This study was funded by the Maatilatalouden kehittämisrahasto under the Finnish Ministry of Agriculture and Forestry (grant number 149/03.01.02/2016).

References

Andreasen SN, Sandøe P and Forkman B 2014 Can animal-based welfare assessment be simplified? A comparison of the Welfare Quality® protocol for dairy cattle and the simpler and less time-consuming protocol developed by the Danish Cattle Federation. *Animal Welfare 23*: 81-94. https://doi.org/10.7120/09627286.23.1.081

Bach A, Valls N, Solans A and Torrent T 2008 Associations between nondietary factors and dairy herd performance. *Journal of Dairy Science 91*: 3259-3267. https://doi.org/10.3168/jds.2008-1030

Barkema HW, Schukken H, Lam TJGM, Beiboer ML, Benedictus G and Brand A 1998 Management practices associated with low, medium, and high somatic cell counts in bulk milk. *Journal of Dairy Science 81*: 1917-1927. https://doi.org/10.3168/jds.S0022-0302(98)75764-9

Barkema HW, Van Der Ploeg JD, Schukken YH, Lam TJGM, Benedictus G and Brand A 1999 Management style and its association with bulk milk somatic cell count and incidence rate of clinical mastitis. *Journal of Dairy Science 82:* 1655-1663. https://doi.org/10.3168/jds.S0022-0302(99)75394-4

Bertocchi L and Fusi F 2014 *Guidelines for the assessment of welfare and biosecurity in dairy cattle in loose housing systems. Italian National Animal Welfare Reference Centre (CReNBA).* Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna: Brescia, Italy

Breuer K, Hemsworth PH, Barnett JL, Matthews LR and Coleman GJ 2000 Behavioural response to humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science 66*: 273-288. https://doi.org/10.1016/S0168-1591(99)00097-0

Brouwer H, Stegeman JA, Straatsma JW, Hooijer GA and 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*: 76-82. https://doi.org/10.1016/j.prevetmed.2015.09.009

Chen W, Hribar P and Melessa S 2018 Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research 56*: 751-791. https://doi.org/10.1111/1475-679X.12195

Coignard M, Guatteo R, Veissier I, Lehébel A, Hoogveld C, Mounier L and Bareille N 2014 Does milk yield reflect the level of welfare in dairy herds? *Veterinary Journal 199*: 184-187. https://doi.org/10.1016/j.tvjl.2013.10.011

Cook NB 2008 Time budgets for dairy cows: how does cow comfort influence health, reproduction and productivity? *Proceedings of Penn State Dairy Cattle Nutrition Workshop* pp 53-60. 12-13 November 2008, Grantville, PA, USA

Council Regulation on the Protection of Cattle 2010 *Council Regulation on the Protection of Cattle (592/2010).* https://www.finlex.fi/fi/laki/ajantasa/2010/20100592

de Vries M, Bokkers EAM, Dijkstra T, van Schaik G and de Boer IMJ 2011 Invited review: Associations between variables of routine herd data and dairy cattle welfare indicators. *Journal of Dairy Science 94*: 3213-3228. https://doi.org/10.3168/jds.2011-4169

de Vries M, Bokkers EAM, van Schaik G, Engel B, Dijkstra T and de Boer IJM 2014 Exploring the value of routinely collected herd data for estimating dairy cattle welfare. *Journal of Dairy Science 97*: 715-730. https://doi.org/10.3168/jds.2013-6585

Espetvedt MN, Wolff C, Rintakoski S, Lind A and Østerås O 2012 Completeness of metabolic disease recordings in Nordic national databases for dairy cows. *Preventive Veterinary Medicine 105*: 25-37. https://doi.org/10.1016/j.prevetmed.2012.02.011

ETT ra 2012 *Naseva terveydenhuoltokäynnin ohjeet eläinlääkäreille versio 2/2012.* https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwils4ni9PDcAhXKw6YKHawDJMQFjAAegQICBAC&url=https%3A%2F%2Fwww.naseva.fi%2FPublicContent%2FGetAttachment%3FStaticContentFileId%3D2&usg=AOvVaw3486WP1oMmMekltwsJROzd

ETT ra 2017 *Centralised Health Care Register for Finnish Cattle Herds.* https://www.naseva.fi/PublicContent/IntroductionInEnglish

Grant RJ and Albright JL 2001 Effect of animal grouping on feeding behavior and intake of dairy cows. *Journal of Dairy Science 84(S)*: E156-E163. https://doi.org/10.3168/jds.S0022-0302(01)70210-X

Hammami H, Rekik B and Gengler N 2009 Genotype by environment interaction in dairy cattle. *Biotechnology, Agronomy, Society and Environment 13*: 155-164

Hasegawa N, Nishiwaki A, Sugawara K and Ito I 1997 The effects of social exchange between two groups of lactating primiparous heifers on milk production, dominance order, behavior and adrenocortical response. *Applied Animal Behaviour Science 51*: 15-27. https://doi.org/10.1016/S0168-1591(96)01082-9

Heath CAE, Lin Y, Mullan S, Brown WJ and Main DCJ 2014 Implementing Welfare Quality® in UK assurance schemes: evaluating the challenges. *Animal Welfare 23*: 95-107. https://doi.org/10.7120/09627286.23.1.095

Hemsworth PH, Coleman GJ, Barnett JL and Borg S 2000 Relationships between human-animal interaction and productivity of commercial dairy cows. *Journal of Animal Science 78*: 2821-2831. https://doi.org/10.2527/2000.78112821x

Hokkanen A-H, Wikman I, Korhonen T, Pastell M, Valros A, Vainio O and Hänninen L 2015 Perceptions and practices of Finnish dairy producers on disbudding pain in calves. *Journal of Dairy Science 98*: 823-831. https://doi.org/10.3168/jds.2013-7668

Huhtanen P, Rinne M, Mäntysaari P and Nousiainen J 2011 Integration of the effects of animal and dietary factors on total dry matter intake of dairy cows. *Animal 5*: 691-702. https://doi.org/10.1017/S1751731110002363

Huhtanen P, Rinne M and Nousiainen J 2007 Evaluation of the factors affecting silage intake of dairy cows: a revision of the relative silage dry-matter intake index. *Animal 1*: 758-770. https://doi.org/10.1017/S175173110773673X

Huhtanen P, Rinne M and Nousiainen J 2008 Evaluation of concentrate factors affecting silage intake of dairy cows: A development of the relative total diet intake index. *Animal 2*: 942-953. https://doi.org/10.1017/S1751731108001924

Huxley JN 2013 Impact of lameness and claw lesions in cows on health and production. *Livestock Science 156*: 64-70. https://doi.org/10.1016/j.livsci.2013.06.012

ICAR 2017 *Introduction to the ICAR Guidelines*. https://www.icar.org/index.php/icar-recording-guidelines/

Ingvartsen KL, Dewhurst RJ and Friggens NC 2003 On the relationship between lactational performance and health: Is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. *Livestock Production Science 83*: 277-308. https://doi.org/10.1016/S0301-6226(03)00110-6

Jensen J 2001 Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science 84*: 2803-2812. https://doi.org/10.3168/jds.S0022-0302(01)74736-4

Jolliffe IT 2002 *Springer Series in Statistics, Second Edition*. Springer: NY, USA

Kester E, Holzhauer M and Frankena K 2014 A descriptive review of the prevalence and risk factors of hock lesions in dairy cows. *The Veterinary Journal 202*: 222-228. https://doi.org/10.1016/j.tvjl.2014.07.004

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

Koivula M, Nousiainen JI, Nousiainen J and Mäntysaari EA 2007 Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *Journal of Dairy Science 90*: 2563-2568. https://doi.org/10.3168/jds.2006-517

Krug C, Haskell MJ, Nunes T and Stilwell G 2015 Creating a model to detect dairy cattle farms with poor welfare using a national database. *Preventive Veterinary Medicine 122*: 280-286. https://doi.org/10.1016/j.prevetmed.2015.10.014

Lind A, Thomsen PT, Ersbøll AK, Espetvedt MN, Wolff C, Rintakoski S and Houe H 2012 Validation of Nordic dairy cattle disease recording databases – Completeness for locomotor disorders. *Preventive Veterinary Medicine 107*: 204-213. https://doi.org/10.1016/j.prevetmed.2012.06.009

Luke 2018 *Milk and milk products statistics*. http://stat.luke.fi/en/milk-and-milk-product-statistics

Metz JHM, Dijkstra T, Franken P and Frankena K 2015 Development and application of a protocol to evaluate herd welfare in Dutch dairy farms. *Livestock Science 180*: 183-193. https://doi.org/10.1016/j.livsci.2015.07.002

Mughal M, Frondelius L, Pastell M, Lindeberg H, Tuunainen E and Mononen J 2017 Comparing welfare assessment results from Welfare Quality® and a Finnish healthcare scheme for cattle. In: de Jong IC and Koene P (eds) *Proceedings of the Seventh International Conference on the Assessment of Animal Welfare at Farm and Group Level* pp 138. 5-8 September 2017, Ede, The Netherlands. Wageningen Academic Publishers: Wageningen, The Netherlands

Nyman A-K, Lindberg A and Sandgren C 2011 Can pre-collected register data be used to identify dairy herds with good cattle welfare? *Acta Veterinaria Scandinavica, 53(S1):* S8. https://doi.org/10.1186/1751-0147-53-S1-S8

Oltenacu PA and Broom DM 2010 The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal Welfare 19(S)*: 39-49

Phillips CJC and Rind MI 2001 The effects on production and behavior of mixing uniparous and multiparous cows. *Journal of Dairy Science 84*: 2424-2429. https://doi.org/10.3168/jds.S0022-0302(01)74692-9

Potterton SL, Green MJ, Harris J, Millar KM, Whay HR and Huxley JN 2011 Risk factors associated with hair loss, ulceration, and swelling at the hock in freestall-housed UK dairy herds. *Journal of Dairy Science 94*: 2952-2963. https://doi.org/10.3168/jds.2010-4084

Rajala-Schultz PJ, Gröhn YT and McCulloch CE 1999 Effects of milk fever, ketosis, and lameness on milk yield in dairy cows. *Journal of Dairy Science 82:* 288-294. https://doi.org/10.3168/jds.S0022-0302(99)75235-5

Raussi S 2003 Human-cattle interactions in group housing. *Applied Animal Behaviour Science 80*: 245-262. https://doi.org/10.1016/S0168-1591(02)00213-7

Rearte R, LeBlanc SJ, Corva SG, de la Sota RL, Lacau-Mendigo IM and Giuliodorill MJ 2018 Effect of milk production on reproductive performance in dairy herds. *Journal of Dairy Science 101*: 7575-7584. https://doi.org/10.3168/jds.2017-13796

https://doi.org/10.7120/09627286.29.4.449 Published online by Cambridge University Press

Roche JR, Friggens NC, Kay JK, Fisher MW, Stafford KJ and Berry DP 2009 Body condition score and its association with dairy cow productivity, health and welfare. *Journal of Dairy Science 92*: 5769-5801. https://doi.org/10.3168/jds.2009-2431

Rushen J, de Passillé AMB and Munksgaard L 1999 Fear of people by cows and effects on milk yield, behavior, and heart rate at milking. *Journal of Dairy Science 82*: 720-727. https://doi.org/10.3168/jds.S0022-0302(99)75289-6

Sandgren CH, Lindberg A and Keeling LJ 2009 Using a national dairy database to identify herds with poor welfare. *Animal Welfare 18*: 523-532

Sarjokari K, Hovinen M, Seppä-Lassila L, Norring M, Hurme T, Peltoniemi OAT, Soveri T and Rajala-Schultz PJ 2018 On-farm deaths of dairy cows are associated with features of free-stall barns. *Journal of Dairy Science 101*: 6253-6261. https://doi.org/10.3168/jds.2017-13420

Villettaz Robichaud M, Rushen A and de Passillé AM 2018 Is the profitability of Canadian free-stall farms associated with their performance on an animal welfare assessment? *Journal of Dairy Science 101*: 2350-2358. https://doi.org/10.3168/jds.2017-13315

Welfare Quality® Consortium 2009a *Welfare Quality® Assessment Protocol for Cattle*. ASG Veehouderij BV: Lelystad, The Netherlands. http://www.welfarequalitynetwork.net/network

Welfare Quality® Consortium 2009b An overview of the development of the Welfare Quality® project assessment systems. In: Keeling L (ed) *Welfare Quality Reports No 12*. Cardiff University: UK

Zaffino Heyerhoff JC, LeBlanc SJ, de Vries TJ, Nash CGR, Gibbons J, Orsel K, Barkema HW, Solano L, Rushen J, de Passillé AM and Haley DB 2014 Prevalence of and factors associated with hock, knee, and neck injuries on dairy cows in freestall housing in Canada. *Journal of Dairy Science 97*: 173-184. https://doi.org/10.3168/jds.2012-6367