**University of Natural Resources and Life Sciences, Vienna**

**Department of Sustainable Agricultural Systems**
**Division of Livestock Sciences**

# Effects of Sample Size and Sampling Strategies on Reliability of Selected Welfare Measures in Dairy Cattle

## Marshallplan scholarship final report

Vienna
March 2014

Author: Julia Trieb, H0740666
Supervisor: Univ.Prof. Dr.med.vet. Christoph Winckler
Co-Supervisor: Assoc. Prof. Ph.D. Cassandra Tucker
Exchange University: University of California, Davis

# Acknowledgements

I am especially grateful to the Marshall Plan Scholarship Foundation, who enabled me with their financial support to realize this project.

I also would like to express my gratitude to my master thesis supervisor Prof. Christoph Winckler, who supported and encouraged me in every single step of my way and also to my co-supervisor Assoc. Prof. Cassandra Tucker, who helped me to realize this project at UC Davis.

I really appreciate the help from John Champagne, DVM, MPVM and Thomas Graham, DVM, MPVM,PhD – without both of you, I still would be trying to find some appropriate farms for my data collection. It also would not have been possible to realize this project without the help from all the farmers in California willing to participate in this study.

Special thanks go to my family and friends in Austria- you were my rock in turbulent waters. Without all of you, I would never have been able to reach the point, where I am now. Thank you for the endless conversations, which gave me courage and self-confidence again and again.

I am also really grateful that I got to know Genene, Scott and also Roux. You were the best substitute family, I can imagine. Without you, my experiences in California would not have been the same.

“Imperare sibi maximum imperium est.” –
Seneca. Epistulae morales ad Lucilium, XIX, CXIII

# Contents

# 1 Introduction

Product quality in a broader sense also includes the way, how products of animal origin are made. In this context, animal health and welfare have achieved increased public and scientific interest.

It is now generally accepted, that for a valid assessment of the welfare state of animals, both for scientific and for on-farm assessment purposes, animal based parameters are of high importance. The EU-funded project Welfare Quality therefore developed the Welfare Quality ® assessment protocol for cattle (Welfare Quality, 2009), which relies to a major part on animal-based measures of welfare. It also provides the basis for several research projects in cattle in many different countries.

Reliability, validity and feasibility are crucial features of measures used in welfare assessment. This study focused on aspects of reliability and feasibility with regard to a number of animal-based measures of the Welfare Quality assessment protocol for dairy cattle.

Reliability relates to the extent of possible measurement error due to the measuring system – often the observer – and measurement procedure. When the assessment has to be performed on a large scale and under commercial conditions, the issue of feasibility of measures is another essential selection criterion. An important issue is the time necessary to carry out a measure (e.g. long-term observations to detect changes in time budget are less feasible), the need for specific devices to perform the measure, or the requirement for specific skills to perform the measure (e.g. expertise in taking blood samples for metabolic disorders).

Taking these issues into account, the Welfare Quality ® assessment protocol for dairy cattle provides guidelines in terms of sample sizes, but remains rather vague with regard to sampling strategies. The sample sizes have been calculated considering theoretical statistical assumptions such as prevalence (i.e. 50%), tolerated deviance from the true prevalence and confidence intervals. To our knowledge, these sample sizes have not been tested for appropriateness in a commercial farm setting.

However, studies investigating sample sizes for finishing pig farms Mullan et al. (2009), showed that the prevalence can reliably only be estimated using very large samples thus reducing feasibility. Especially for low prevalence measures in pig

farms, such as lameness, even a large sample size could not approach the true prevalence of those parameters.

Apart from the sample size issue, often the question remains open, how a certain sample may be obtained. For example, in dairy cattle Main et al. (2010) used the milking order divided in thirds and found, that there are differences of the prevalence in all thirds. In the last third, the lameness prevalence was 11.9% higher than in the first third. Scoring cows from the second third provided the best estimates of prevalence as compared with the true prevalence. Also Sauter-Louis et al. (2004) found, that there is a relationship between the milking order and the occurrence of lameness with the significantly higher prevalence of lame cows in the last quarter of the milking order.

Therefore the objective of this study was to evaluate the impact and effects of different sample sizes and sampling strategies on the reliability of selected welfare measures in dairy cattle.

The following research questions have been addressed:

1) With which precision does the sample size specified by the Animal Welfare Quality Assessment Protocol for Cattle represent the true prevalence of clinical health parameters?

2) Are there preferred sampling strategies in order to obtain reliable estimates of the prevalence of parameters o clinical health?

# 2 Animals, Materials and Methods

## *2.1 Materials*

In total ten Californian dairy farms were included. On each farm one representative pen was selected. It was always the high production group as defined by the herd manager; hospital pens were not taken into account.

Table 1: Location, breed, number of scored animals and housing system of the farms

| Farm | Location | Breed | Number of scored animals | Housing system |
|---|---|---|---|---|
| 1 | Tulare County | J | 204 | dry lot |
| 2 | Tulare County | HF | 227 | dry lot |
| 3 | Tulare County | HF | 144 | free stall |
| 4 | Tulare County | HF | 255 | dry lot |
| 5 | Tulare County | HF | 145 | freestall |
| 6 | Tulare County | HF | 216 | freestall |
| 7 | Tulare County | HF | 81 | dry lot |
| 8 | Tulare County | HF | 191 | freestall |
| 9 | Sacramento County | HF | 123 | freestall |
| 10 | Sacramento County | HF | 241 | freestall |

As shown in Table 1, eight farms were located in the Tulare County, CA and two farms were located in the Sacramento County, CA. All farm visits took place between October 2013 and December 2013.

Six farms had free stalls with deep sand bedded cubicles as housing system and four farms used dry lots. Nine farms had Holstein (HF) as the predominant breed, but one farm had Jersey (J) as the primary breed. In total 1807 animals were scored.

Data collection per farm required one day.

On the farms, every animal within the chosen pen was scored to get the true prevalence of the welfare measures. For this purpose, all animals were locked in headlocks at the feed bunk. The ear tag numbers was used to identify the animals.

Also, the position of each animal at the feed bunk was recorded and served as 'feed bunk order' for some of the computer based sampling strategies.

While the animal was standing at the feed bunk, the integument alterations at the body regions hindquarter, tarsus and carpus as well as the body condition score, and signs of diarrhea, discharge from the eyes, nose and vulva as well as cleanliness of the animal was recorded (for details see next chapter). The cow was then released one by one to score lameness and integument alterations at the regions neck/shoulder/back and flank/side/udder.

When all cows had been individually assessed and released from the headlocks, a random sampling procedure according to the guidelines from the Animal Welfare Quality ® Assessment Protocol for Cattle took place. In this step, the ear tag numbers has been noted. Therefore 96 animals (except farm 5 and 9, where 55 have been selected) from all areas within the pen has been selected including standing, feeding and lying animals.

In a third step, during the milking following the scoring in the pen, the milking order was taken. This was again done using the ear tag numbers. No clinical recordings took place in the milking parlor, since all animals had already been examined in the home pen.

## *2.2 Measures*

All measures were scored as described in the Welfare Quality ® Assessment Protocol for cattle (Welfare Quality, 2009).

Body condition: For body condition a 0 to 2 scale was used, where 0 was defined as a regular body condition, 1 as very lean and 2 as very fat taking the cavity around the tall head, the loin, the vertebrae as well as tall head, hip bones, spine and ribs as decisive body regions into account.

Cleanliness: Cleanliness was recorded on three body parts – udder, lower hind legs and hind quarters. 0 was defined as no dirt or minor splashing and 2 was defined as plaques of dirt.

Lameness: Gait scoring was used as a measure of lameness using a 0 to 2 scale, where 0 was identified as not lame, 1 as lame and 2 as severely lame. Animals with a score '1' had an imperfect temporal rhythm, which created a limp in their gait. The score '2' is identified as a strong reluctance to bear weight on one limb or if more than one limb is affected.

Integument alterations: Four different body regions were distinguished (hindquarter, neck/shoulder/back, carpus, flank/side/udder and lower hindleg.
The number of two different categories of alterations (hairless patches, lesions/swellings) per body region was counted.
Diarrhea: Signs of diarrhea were scored as a 0/2 measure, where 0 meant no evidence and 2 meant evidence of diarrhea.
Vulvar discharge: As for diarrhea 0 was defined as no evidence of discharge and 2 as evidence of discharge.
Nasal discharge/ocular discharge: As for diarrhea 0 was defined as no evidence of discharge and 2 as evidence of discharge.
Hampered respiration: Hampered respiration was defined as the absence or presence of visible difficulties in breathing excluding panting due to heat stress.

## *2.3 Statistical analysis*

Four different approaches were chosen for the comparisons.

- Every second to every tenth animal:

  Out of the scoring order at the feed bunk, every second to every tenth animal was chosen to create a new data set and thus estimated prevalence.

- On-farm random sampling:

  On the farms, random sampling of different sample sizes (see Table 2) was carried out. For this purpose, animals from all over the pen were picked and their ear tag number noted down. The final data set from on-farm random sampling finally comprised three samples per farm (WQa, WQb and Wqmax). In total up to 96 animals were selected. In Farm 7 and Farm 9 it was not possible to achieve this number for technical reasons; in this case 55 animals were chosen.

  As regards the different sample sizes given in Table 2, WQa refers to the suggested sample size for the respective number of animals. This sample size is based on an assumed prevalence of 50%,a deviance of 10% and a confidence interval of 95%. If this number is not feasible, then Welfare Quality (2009) suggests to score the sample size given as WQb based on a confidence interval of 90% and a deviance of 10%. WQmax is the maximum

sample size as provided by Welfare Quality, which applies to herd sizes above 300 animals.

- Computer-based random sampling:

  Using the SAS procedure PROC survey select, the same numbers of animals as used in the on farm random sampling was selected (Table 2).

- Milking order divided in thirds:

  The milking order was, based on results from other studies, divided into thirds. Those thirds also have been used for the comparisons between the true prevalence of the measures and the estimated prevalence given from the subsamples.

Table 2: Sample sizes for the ten farms according to Welfare Quality (2009) used for on-farm and computer-based random sampling

| Farm | Number of animals | WQa | WQb | WQmax |
|---|---|---|---|---|
| 1 | 204 | 65 | 51 | 96 |
| 2 | 227 | 68 | 52 | 96 |
| 3 | 144 | 57 | 46 | 96 |
| 4 | 235 | 69 | 53 | 96 |
| 5 | 145 | 59 | 47 | 96 |
| 6 | 216 | 67 | 52 | 96 |
| 7 | 81 | 44 | 37 | 55 |
| 8 | 191 | 64 | 50 | 96 |
| 9 | 123 | 54 | 43 | 55 |
| 10 | 241 | 69 | 53 | 96 |

For data analysis the statistical package SAS 9.2 (SAS Institute) was used. The association between the true prevalence and the prevalence out of the sampling strategies was determined using a linear regression model (PROC REG). Only for those sample sizes/strategies, where the prevalence of the sampling strategy met the three criteria $R^2 > 0.9$, slope not significantly different from 1 and intercept not significantly different from 0, the estimated prevalence was assumed to reliably represent the true prevalence.

The measure ‘vulvar discharge’ was removed from analysis, because there were no cows showing signs of vulvar discharge. Furthermore there were only 4 cows with a body condition score of 2 (very fat); again, the prevalence of very fat animals was not further considered for analysis. Only one cow showed ‘hampered respiration’ and this measure was also excluded from analysis.

# 3 Results

Descriptive results of the prevalences found in the farms studied are given in Table 3. For every measure the values for true prevalence, mean, standard deviation, as well as minimum and maximum values are provided. The results for those measures, where the presence has been counted, has not been calculated as the number of hairless patches or lesions and swellings, but as the percentage of affected animals.

Table 3: True prevalence, mean, standard deviation as well as minimum (min) and maximum values (max) for all measures.

| Measure | True Prevalence | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|---|
| BCS = 1 (very lean) | 13.4 % | 14.8 % | 0.4 | 1.2 % | 35.3 % |
| Lameness (gait score > 0) | 11.8 % | 11.5 % | 0.4 | 1.4 % | 21.5 % |
| Diarrhea | 8.1 % | 8.9 % | 0.3 | 0.0 % | 18.6 % |
| Carpus hairless | 14.3 % | 15.7 % | 0.4 | 0.0 % | 74.5 % |
| Carpus lesion | 36.8 % | 39.6 % | 0.6 | 10.6 % | 53.7 % |
| Carpus swelling | 2.8 % | 2.9 % | 0.2 | 0.0 % | 7.9 % |
| Flank/side/udder hairless | 4.6 % | 5.1 % | 0.5 | 0.0 % | 15.2 % |
| Flank/side/udder lesion | 1.7 % | 1.9 % | 0.2 | 0.0 % | 4.9 % |
| Flank/side/udder swelling | 0.8 % | 0.7 % | 0.1 | 0.0 % | 2.5 % |
| Hindquarter hairless | 4.8 % | 5.4 % | 0.4 | 0.7 % | 22.8 % |
| Hindquarter lesion | 2.1 % | 2.3 % | 0.2 | 0.0 % | 5.9 % |
| Hindquarter swelling | 0.8 % | 0.7 % | 0.1 | 0.0 % | 2.6 % |
| Neck/shoulder/back hairless | 5.9 % | 6.6 % | 0.5 | 0.0 % | 35.7 % |
| Neck/shoulder/back lesion | 1.2 % | 1.3 % | 0.2 | 0.0 % | 6.4 % |
| Neck/shoulder/back swelling | 1.7 % | 2.0 % | 0.2 | 0.0 % | 9.5 % |

| | | | | | |
|---|---|---|---|---|---|
| Tarsus hairless | 15.0 % | 14.3 % | 0.4 | 3.9 % | 28.5 % |
| Tarsus lesion | 29.5 % | 29.8 % | 0.6 | 6.2 % | 87.7 % |
| Tarsus swelling | 2.9 % | 2.8 % | 0.6 | 0.0 % | 5.5 % |
| Udder Cleanliness (dirty) | 16.2 % | 16.8 % | 0.4 | 0.0 % | 37.3 % |
| Hindquarter Cleanliness (dirty) | 33.3 % | 32.7 % | 0.5 | 16.7 % | 48.9 % |
| Tarsus cleanliness (dirty) | 18.4 % | 19.3 % | 0.4 | 2.6 % | 46.9 % |
| Nasal discharge | 19.4 % | 20.5 % | 0.4 | 0.41 % | 48.0 % |
| Ocular discharge | 22.1 % | 23.6 % | 0.4 | 0.0 % | 74.5 % |

In the following sections, the results from the different sampling strategies, as shown in Table 3, are presented.

## *3.1 Every 2nd to 10th animal*

There are 14 out of 23 measures, where at least one subsample identified in this approach met the three criteria $R^2 > 0.9$, slope = 1 and intercept = 0. The estimated prevalence of the measures carpus swelling, flank/side/udder hairless, lesion and swelling, hindquarter lesion and swelling as well as neck/shoulder/back lesion and tarsus swelling did not match true prevalence. For those measures, where the true prevalence is less than 5%, this sampling strategy did not fulfill the defined criteria. Generally, for those measures, where the true prevalence is above 5%, at least the sampling strategy with scoring every 2nd and 3rd animal approached the three criteria except the measure hindquarter hairless, where the estimated prevalence matched the true prevalence only with the subsample 'every 3rd animal'

The results point out, that the sample size has a major impact on the estimated prevalence of a herd. There is a relationship between the level of the prevalence and the reliability of the results obtained from this sampling strategy. It can be said, that the smaller the true prevalence of a measure is, the higher must be the sample size.

Graph 1 shows the regression line of the measure carpus hairless with the sampling strategy every 2nd animal. The estimated prevalence met all three criteria, which means that the estimated prevalence equaled the true prevalence. Whereas the $R^2$ for the measure hindquarter cleanliness in Graph 2 is 0.67 and therefore the sampling strategy for this measure did not fulfill the requirements.

## *3.2 On-farm random sampling*

For 8 measures, the estimated prevalence matched the true prevalence in all three different sample sizes within this sampling strategy. For the measures carpus swelling and hindquarter hairless, the sample size WQmax with 96 animals met the three criteria $R^2 > 0.9$, slope = 1 and intercept = 0. Additionally, the estimated prevalence of lameness and diarrhea in the two sample sizes WQa and WQmax matched their true prevalence. For the other 11 of 23 measures no sample size fulfilled the criteria.

## *3.3 Computer based random sampling*

The estimated prevalence within the subsamples of this sampling strategy did not match the true prevalence of 11 measures, although the true prevalence of some of this measures, e.g. hindquarter cleanliness (33.3 %) rather high.
On the other hand, for 5 measures such as hindquarter hairless and lesion the true prevalence was 5.8 % and 2.1 %, the subsampling strategy WQmax, where 96 animals have been chosen, did match the requirements. Ocular discharge and tarsus lesions matched the requirements in the two sample sizes WQa and WQb.

## *3.4 Milking order*

The results for this sampling strategy show, that the estimated prevalence as obtained from the last third of animal sin the milking parlor for the measures lameness, diarrhea, carpus lesion, flank/side/udder hairless, hindquarter hairless and swelling, neck/shoulder/back lesion, as well as tarsus lesion, udder and tarsus cleanliness and nasal discharge matched the true prevalence. Ocular discharge

fulfilled the three criteria in the second third. Carpus hairless was the only measure, which matched the true prevalence in all thirds.

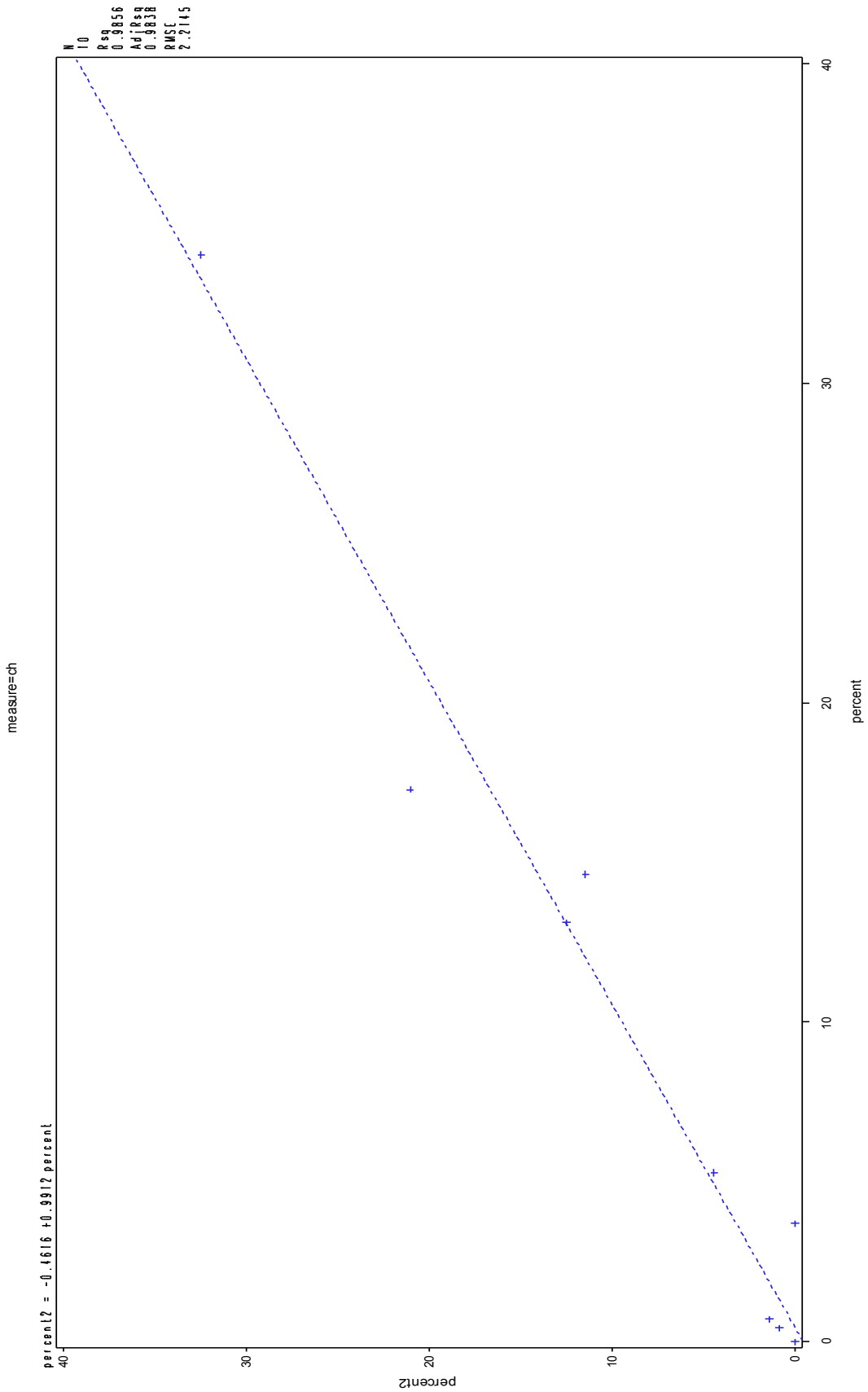
In this sampling strategy, the proportion of the true prevalence regarding the three criteria is mostly evenly distributed all over the results. However it is interesting, that except for one measure, all other measures met the three criteria in the 3rd third of this sampling strategy.

Table 4 gives an overview of the measures, the true prevalence all over the farms and the four different sampling strategies. Only where the measure meets the three criteria $R^2 > 0.9$, slope = 1 and intercept = 0, the estimated prevalence equals the true prevalence.

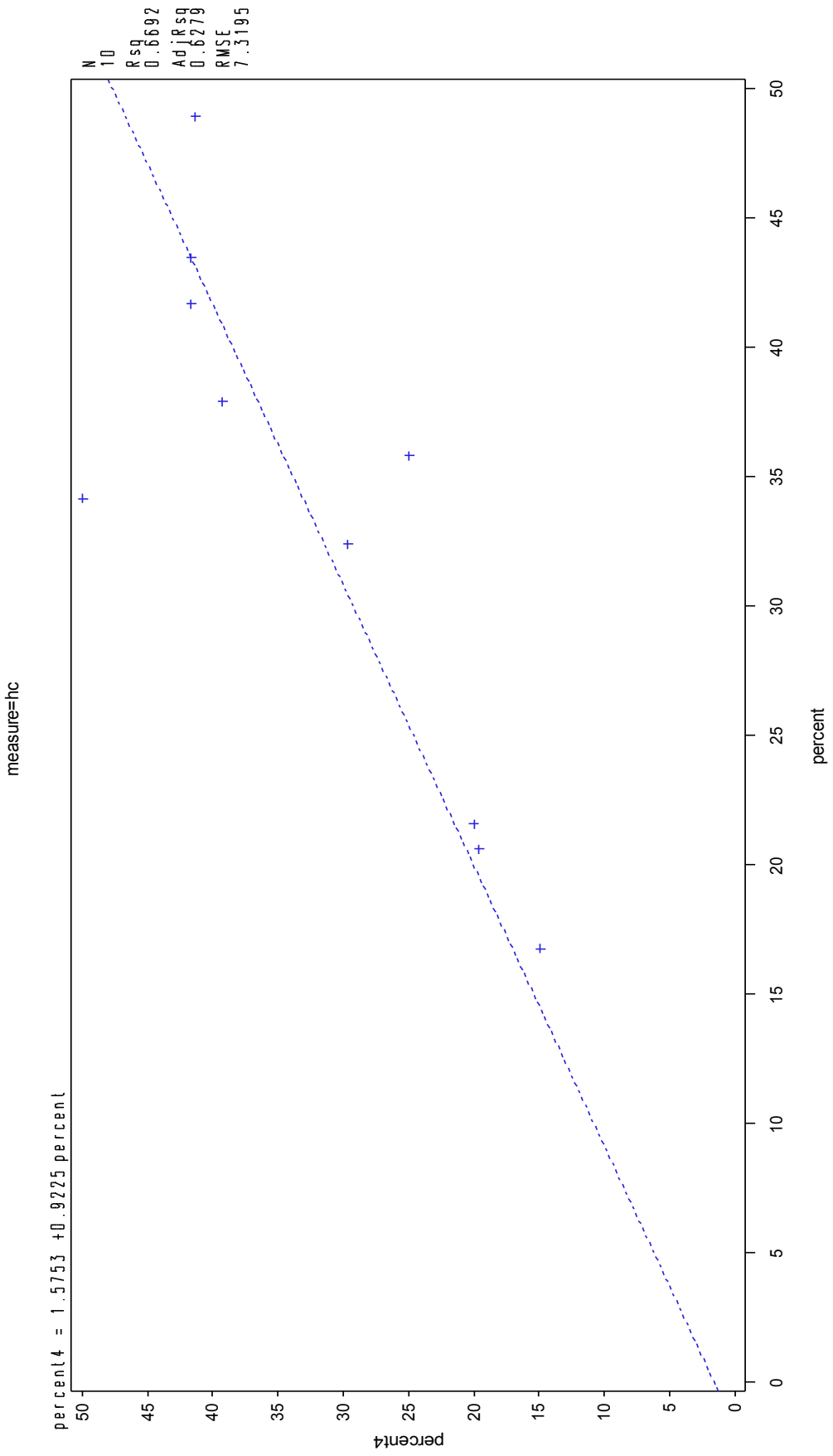
Table 4: Results of all comparisons between the measures and the different sampling strategies

| Measure | Every xth cow (2nd to 10th) | On farm random sampling | Computer based random sampling | Milking order (thirds) |
|---|---|---|---|---|
| BCS = 1 (very lean) | 2nd, 5th | all | all | none |
| Lameness (gait score > 0) | 2nd, 3rd | WQa, WQmax | WQmax | 3rd |
| Diarrhea | 3rd | WQa, WQmax | none | 3rd |
| Carpus hairless | 2nd, 3rd, 4th, 7th, 8th | all | all | all |
| Carpus lesion | 2nd, 3rd | all | WQmax | 3rd |
| Carpus swelling | none | WQmax | none | none |
| Flank/side/udder hairless | none | none | none | 3rd |
| Flank/side/udder lesion | none | none | none | none |
| Flank/side/udder swelling | none | none | none | none |
| Hindquarter hairless | 3rd | WQmax | WQmax | 3rd |

| | | | | |
|---|---|---|---|---|
| Hindquarter lesion | none | none | WQmax | 3rd |
| Hindquarter swelling | none | none | none | none |
| Neck/shoulder/back hairless | 2nd | none | WQmax | none |
| Neck/shoulder/back lesion | none | none | none | 3rd |
| Neck/shoulder/back swelling | 2nd | none | none | none |
| Tarsus hairless | 2nd | WQa, WQmax | none | none |
| Tarsus lesion | 2nd, 3rd, 4th, 5th, 6th | all | WQa, WQb, | 3rd |
| Tarsus swelling | none | none | none | none |
| Udder cleanliness (dirty) | 2nd, 3rd, 5th | all | all | all |
| Hindquarter cleanliness (dirty) | 3rd | none | none | none |
| Tarsus cleanliness (dirty) | 2nd, 4th | all | all | 3rd |
| Ocular discharge | all | none | WQa, WQb, | 2nd |
| Nasal discharge | 2nd, 3rd, 4th, 7th | all | all | 3rd |




Graph 1: Regression line of the measure carpus hairless with the sampling strategy every 2nd animal




Graph 2: Regression line of the measure hindquarter cleanliness with the sampling strategy every 4th animal

# 4 Discussion

To our knowledge, this is the first project investigating the reliability of prevalence estimates for mostly health-related animal based welfare measures as derived from different sampling strategies. This also means that there are almost no results available in the literature to compare with.

The range of true prevalences varied at the farm level. This might result out of the use of different housing systems and their impact on animal welfare measures. The mean true lameness prevalence was rather low with 11.8% compared with e.g. Main et al. (2010), who found an overall lameness prevalence of 39.1 %.
Also comparable is the measure 'tarsus lesion', where the prevalence in the study from Gratzer (2011) ranged between 2.1% and 9.4%, whereas we found a true prevalence of 29.5%. This might again result out of the use of different housing systems.
On the contrary, Gratzer (2011) found in different countries a median proportion very lean cows, which was lower than 10%, expect in the United Kingdom, where the proportion was 13.0%. These results are similar to the results found in this study, where the prevalence of cows with a poor body condition was 13.4%. At this point it needs to be mentioned, that the sample size was much higher.

The results show, that there are differences between the different sampling strategies. However, in general rather large samples such as every 2$^{nd}$ or 3$^{rd}$ cow or the WQa and WQmax sample sizes seem to be the most promising approaches. Particularly interesting is a comparison between the two random sampling strategies. In both strategies, the sample size is constant, but the results demonstrate, that for some measures, the on-farm random sampling may be less likely to generate representative samples. There are more measures whose prevalences were reliably estimated using the computer-based random sampling than using the on-farm random sampling. Although the assessor tried to identify animals in the home pen as unbiasedly as possible, more conspicuous animals may have been picked up or some animals simply avoided to get or stay near the assessor. On the other hand, under practical conditions, a computer-based random sampling would markedly increase the efforts needed to actually identify the animals which have been selected

by the computer out of the (large) group of animals. Differences between the two random sampling strategies were obviously not due to the magnitude of the prevalence, since both ‘tarsus hairless’ and ‘body condition = 1’ showed prevalences of 13.4 % and 15.0 %, respectively., For ‘tarsus hairless’, all three sample sizes of the on-farm random sampling fulfilled the criteria, but for the computer-based random sampling none of the three sample sizes represented the true prevalence. For ‘body condition = 1’, both random sampling strategies were compatible with the true prevalence.

The results of the sampling strategy ‘milking order’ showed a rather high degree of accordance between the estimated and the true prevalence of the thirds, even though the number of farms is rather small. While this effect may be explained for lameness (male cows being less able to move and less competitive), it is to some extent surprising for measures such as cleanliness at the lower hind leg or skin alterations of the neck.
However, it needs to be taken into account, that in this study the scoring at the milking parlor was only a theoretical one since only the ear tag numbers were noted and the information on the animal-based measures had been recorded in the home pen. For some parameters such as skin alterations in the front part of the body and especially in large herds it is unrealistic to score all measures in the milking parlour.

The results show, that except the two measures hindquarter cleanliness and ocular discharge, where the true prevalence was high, the percentage needs to be above 6.5 % to equal the estimated prevalence at the on-farm random sampling strategy. Also the effect of the level of prevalence needs to be mentioned at this point. The estimated prevalences of measures with a low prevalence are less reliable.
The different sample sizes given from Welfare Quality (2009) are calculated for a ‘worst-case situation’ with a prevalence of 50%. Therefore the accordance between the true and the estimated prevalence for measures with a low prevalence should be easier for the both random sampling strategies, where the sample sizes from Welfare Quality (2009) were used.

Depending on the measure and the prevalence of the measure, different sampling strategies need to be used. For measures with a high prevalence, the choice

between several different sampling strategies is possible, whereas for measures with a low prevalence there are fewer strategies, which can be used.
Some low prevalence parameters, such as joint swellings or lesions, on the other hand must not be disregarded or oversimplified. Therefore an appropriate sampling strategy for those measures needs to be found. Due to the fact, that in welfare assessment it is less likely that assessors will be able to apply different sampling strategies, a compromise needs to be found.
Because of this, one suggestion for a sampling strategy could be to focus on measures with low prevalences and to adjust also the sample size to get a good accordance between the estimated and the true prevalence of those measures. Nevertheless also the time and cost factor must not be disregarded and therefore it is questionable, how feasible this would be.

All in all it can be concluded, that the size of the true prevalence *per se* cannot be the decisive factor for the choice of the sampling strategy. Furthermore it can be said, that with a larger sample size, there is a higher degree of consistency between the estimated and the true prevalence of a measure. Last but not least, also the feasibility of a sampling strategy – such as the time and cost factors – should not be disregarded.

References

GRATZER E.T. (2011). Animal health and welfare planning in austrian organic dairy farming.Dissertation Universität für Bodenkultur, Wien

MULLAN S., BROWNE W. J., EDWARDS S. A., BUTTERWORTH A., WHAY H. R., MAIN D. C. J. (2009). The effect of sampling strategy on the estimated prevalence of welfare outcome measures on finishing pig farms. Applied Animal Behaviour Science 119, 39–48

MAIN D.C.J., BARKER Z.E., LEACH K.A., BELL N.J., WHAY H.R., BROWNE W.J. (2010). Sampling strategies for monitoring lameness in dairy cattle. J. Dairy Sci. 93: 1970-1978

SAUTER-LOUIS C.M., CHESTERTON R.N., PFEIFFER D.U. (2004). Behavioural characteristics of dairy cows with lameness in Taranaki, New Zealand. New Zealand Veterinary Journal 52:3, 103-108

WELFARE QUALITY CONSORTIUM (2009). Welfare Quality ® assessment protocol for cattle. ISBN: 978-90-78240-04-4