American Journal of Animal and Veterinary Sciences, 2012, 7 (4), 159-174
ISSN: 1557-4555

doi:10.3844/ajavssp.2012.159.174 Published Online 7 (4) 2012 (http://www.thescipub.com/ajavs.toc)

Management of Data for Health Performance Measurement in the Industrialized Dairy Herd

Mogens Agerbo Krogh and Carsten Enevoldsen

Department of Large Animal Sciences, Faculty of Health and Medical Sciences,
University of Copenhagen, Groennegaardsvej 2, DK-1870 Frederiksberg C, Denmark

Received 2012-05-14, Revised 2012-09-13; Accepted 2012-11-23

ABSTRACT

One part of dairy herd management is to handle disease occurrence by means of health promotion, disease prevention, timely medical treatments, or eradication of disease. Supporting this part of herd management is an essential task for the cattle veterinarian. The study objective was to identify principles and tools for analysis of herd health data in industrialized dairy herds. The analysis takes into account the additional complexity caused by changes in behavior among herd managers and herd personnel due to, for instance, legislative changes to promote animal welfare or food safety. Methods from herd management science were combined with context-specific information about social mechanisms. The results were synthesized into a concrete 7-step plan of action, as follows: (1) As the foundation, use continuously process behavior charts primarily based on animal-level data. (2) Assure strict definition of the measurements considering purpose, collector and meaning in terms of biology and management. (3) Interpret the patterns in the process behavior charts and search for and remove causes of exceptional variation in a dialogue with the herd manager. (4) Search for options to reduce routine variation. Multivariable or multivariate statistical models can give additional information because of their ability to reveal hidden sources of variation. (5) Set targets at the tactical and strategic levels while accounting for costs and benefits with appropriate methods suggested in the study. Issues related to non-financial effects are addressed. (6) Adjust measurement and intervention theory. The previous five steps should initiate an iterative process in which the intervention is evaluated and updated based on the results achieved thus far. (7) Develop a framework in the veterinary practice unit to support the health performance measurement process. The activities in step 7 will almost certainly require expert statistical assistance.

Keywords: Herd Health, Dairy Herd Management, Veterinarian, Health Performance Measurement

1. INTRODUCTION

The size of dairy herds has increased dramatically in many countries and it seems relevant to consider the dairy herd as any other industrialized manufacturing enterprise, service provider, or organization in general. Continuous evaluation of the performance of the production process is an essential part of herd (business) management. One part of herd management is to handle disease occurrence by means of health promotion, disease prevention, timely medical treatments, or eradication of disease. It is an essential task for the cattle veterinarian to support this part of herd management. During the last two decades, computer technology, Automatic Milking Systems (AMS) and other automated data collection tools have dramatically increased the amount of data available for measuring and evaluating performance over time in dairy herds. These data may be especially useful for measuring occurrence of diseases with subtle signs (e.g., ketosis and mastitis), which have become relatively more important because major diseases like tuberculosis and brucellosis have been eradicated. The continuing entry and removal of numerous animals, the interaction between animals and management and feedback mechanisms make the dairy herd a very complicated system or organization, which

Corresponding Author: Mogens Agerbo Krogh, Department of Large Animal Sciences Faculty of Health and Medical Sciences, University of Copenhagen, Groennegaardsvej 2, DK-1870 Frederiksberg C, Denmark Tel: +45 23659010

may make performance measurement and evaluation of performance in the dairy herd more complicated than they may be in most other industries.

Enevoldsen (1993) reviewed technologies and management tools developed for dairy herd health management up to the early 1990s. Principles and tools for measuring and evaluating performance over time were treated in some detail. Inspired by the tools and principles used in manufacturing enterprises and other organizations, including public management, we may find uses of numerous additional tools and principles to be useful. Terms like monitoring, surveillance, control, benchmarking, epidemiological or business intelligence, performance measurement, evaluation, statistical process control and quality control are widely used. However, the definitions and distinctions between them seem to differ among disciplines, the objectives for application are often vague and the interpretation can be complicated.

Krogh and Enevoldsen (2006) describe the so-called VPR platform. It was established in 2003 and gives Danish practicing cattle veterinarians access to a growing number of tools for management of health data. During development of the platform and support of the users, we have identified a number of barriers and needs for efficient support of data management for health performance measurement in the dairy herd. Especially when data were used for very specific decisions, errors in collection and management of data were revealed. Based on this interactive development work with veterinarians in the field together with various research and teaching based on the collected data, we will (objectives):

- Identify principles and tools that are of particular relevance to dairy herd health consultants' continuous evaluation of health performance in the industrialized dairy herd and
- Suggest a coherent set of definitions and tools for management of data for health performance measurement in the industrialized dairy herd

This study is organized into the following main parts: (1) time series analysis and (2) control and a systems approach to herd management.

The work does not present or discuss simple graphical or tabular presentations of data without attempts to address random and systematic variation in the production process, or without support to evaluation of the performance of the process by means of some type of limits or criteria.

1.1. Time series Analysis

In herd management, the most common questions are related to time. We want to know whether there are changes in the production process. Detection of changes requires some kind of comparison of the current (or future) production process with some previous production result. As part of herd management, we use a variety of measurement tools to make observations of some activity (variables) at successive points in time. Such data are called time series data or longitudinal data. The fundamental elements of an analysis of time series data are (Armitage *et al*., 2008) as follows:

- Plot the data before doing any computations
- Look for extreme outliers and search for possible reasons
- Identify obvious long-term trends

The following presents concepts and tools for such a time series analysis of major relevance to dairy herd health management.

1.2. The Process Behavior Chart

Figure 1, the upper panel, shows a typical example of a time series graph meant for measuring performance of a process. In this case, it is a process in a dairy herd, but it could be a process in a factory or a service industry. The data points are the Fat percentage to Protein percentage Ratio (FPR) of individual cows at the first milk test day in the period 5 to 28 days after calving. The diagrams in **Fig. 1**, upper and lower panels, will be described and explained in the following with the terms used by Wheeler (2000), who calls the diagrams in **Fig. 1** Process Behavior Charts (PBC). Above and below the lines connecting the measurements of FPR (the time series graph) are so-called Natural Process Limits (NPL). The purpose of these limits is to separate the routine variation of the process (the natural process) from the exceptional variation. If the process exhibits only routine variation, demonstrated visually as all points inside the limits, the process will also be predictable (within limits). Predictability is an important and favourable characteristic of a process. Consequently, the exceptional points (points outside the limits) must indicate something unpredictable and the cause(s) of the exceptional variation should be continuously explored and, if possible, removed to improve the process (make it predictable). Attempts possibly should be made to reduce routine variation, but doing so will require fundamental changes in the process. This type of change may be necessary if too many results are unacceptable from a product quality point of view. For example, electrical conductivity measurements from an AMS might show only routine variation. Still, an unacceptably large proportion of cows could have mastitis, which would require very time-consuming attention or medication. Therefore, fundamental changes in the AMS or the herd management may be justified. Wheeler (2000) uses the term method of continual improvement to describe the PBC and its intended uses.

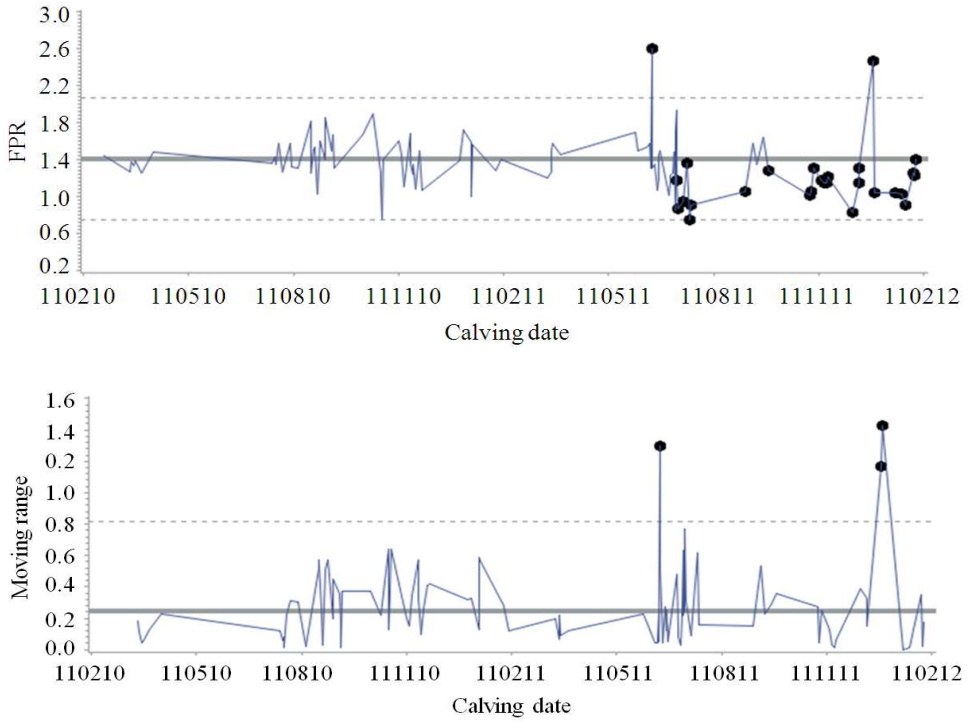

Fig. 1. XmR-chart of Fat-to-Protein Ratio (FPR) in milk recorded between 5 and 28 days after calving. The solid lines are the averages and the dashed lines are the Natural Process Limits (NPL). The upper panel is the X-chart. Observations crossing the NPL and observations that fall for the 'runs-rules' (criteria for trend) are highlighted. The lower panel is the 'moving Range'-chart. Observations crossing the NPL are highlighted

The data points in the lower panel of **Fig. 1** are the numerical differences between successive values in the upper panel. They are called moving Ranges (mR), which directly measure the cow-to-cow variation. The average moving range is the average (arithmetic mean) value of the moving ranges and is shown as the lower horizontal line in the lower panel. The lower and upper NPL in the upper panel are derived from the average moving range in the lower panel by multiplication-constants that depend on the type of data (Wheeler, 2000); in this case, the constant is 2.66. Similarly, the upper range limit for the average mR is obtained by multiplication with the constant 3.27. A more conservative approach is to calculate a median mR, which may be more appropriate if some few values are very high or low. Indications of possible emerging trends are marked in the upper panel. In this case, a series of more than 7 points on one side of the average is regarded as signaling a trend. This pattern represents one of several of the so-called runs rules; of which some are summarized by Kristensen *et al*. (2009).

Based on the first author's personal knowledge about the herd from which **Fig. 1** was derived, the interpretation of the chart can be as follows: In the upper panel, two observations cross the upper limit. These two cows are most likely associated with subclinical ketosis (Krogh *et al*., 2011). Based on the previously described runs rules, there is a trend towards a lower average FPR from June 2011 onwards. In this specific situation, a similar trend was not found in second and older parities (not shown). For this specific herd, this signal of change in the process was most likely related to insufficient training of fresh first-parity cows to the AMS. First-parity cows were left standing outside the AMS for up to 6 h, thus reducing their roughage intake and leading to milk fat depression.

The issue with a chart like that in **Fig. 1** is that we can make two errors: (1) interpret noise (routine variation) as if it were a signal of exceptional variation or (2) fail to detect exceptional variation when it is present. The above-mentioned constants and rules to calculate the limits and define 'signals' are empirical and intended to strike a balance between these two mistakes (Wheeler, 2000). Woodall (2000) stresses that this type of chart is "a tool of exploratory data analysis" (of historical data) and that "no assumptions of normality or independence over time need to be made. In fact, distributional

assumptions cannot even be checked before the chart is initially applied…because one may not have process stability…". Woodall (2000) disputes the effectiveness of the traditionally used runs rules and suggests alternatives, as well as suggesting alternatives to using the mR chart to identify changes in variability in the process. Koutras *et al*. (2007) conclude that the sensitivity improvement achieved by supplementing the classical control chart by runs rules, has a trade off in the false alarm rate. In simple words, runs rules increase sensitivity but also produce more false alarms.

Wheeler (2000) vigorously stresses that no assumptions are required for the PBC. In case there are no signs of exceptional variation or trends, intervention is not warranted. In fact, intervention may distort the process (Wheeler, 2000; Woodall, 2000). Wheeler also vigorously stresses that specific knowledge about the context of the process is needed to discover causes of exceptional variation, which is the primary objective of the method of continual improvement.

In the increasingly automated systems, the users of the information may become detached from the management of data. To completely and fully describe the context, the user needs to know (Wheeler, 2000): Who collected the data? How, when and where were data collected? What do values represent? If computed, how were they computed from raw data? Were there changes in formulas over time? We will add that sometimes it is crucial to know for what purpose the data are collected to understand why data can be misleading. These requirements may be a real challenge to a herd health consultant but also an important learning process.

1.3. Statistical Process Control

1.3.1. Classical Methods

The PBC described above is one simplified version of the Shewhart Control Chart concept, which is among the body of techniques known as Statistical Process Control (SPC), widely used since the 1930s. Kristensen *et al*. (2009) give a detailed description of what they call the classical methods for SPC and their applications to various types of herd management data. The major difference between the PBC and the SPC is that the limits in SPC usually are based on distributional assumptions of the measurements (e.g., normal or binomial) and degree of dependencies between measurements (autocorrelation). For these reasons, these methods are separated from the PBC in this presentation. The validity and importance of these assumptions may be very questionable and hard to judge. Woodall (2000) quotes Hoerl and Palm (1992) as stating that "the underlying model (for SPC) then is only that one has a series of independent random observations from a single statistical distribution. The control chart rules are used to detect deviations from the model, including the model assumptions themselves". In statistical terminology, this concept is called model control. Vries and Reneau (2010) discussed the effectiveness of SPC based on their comprehensive review of applications of the control charts in animal production. Their main conclusion was that an actual search for the true causes of exceptional variation is very difficult and seldom done. Papers on the practical benefits of implemented control chart schemes were not found. Run length distributions (an indicator of SPC effectiveness) were only found in papers describing simulations studies, which may be problematic because simulations usually are based on assumptions about distributions, which we rarely know in a real life setting. Wheeler (2011) claims that autocorrelation (that is, non-independencies of the series of observations) should not influence the limits for NPL. The argument is that autocorrelation will cause a trend (signal) that should be explored and the cause(s) identified and removed. If this exploration and the subsequent intervention are successful, only routine variation remains and routine variation will not contain autocorrelation.

Another major difference between SPC and the PBC is that SPC in many cases shows only data that are filtered or smoothed to better reveal patterns in the data. This process is achieved by calculating one of several types of moving averages. One possible choice is the average of the latest 12 months plotted for each month, which will eliminate erratic fluctuations (smoothing). The moving average may also be weighted so that the latest measurements of the time series are given more weight than the preceding ones. Such weighting is generally recommended to avoid reactions resulting from removal of the oldest historical data. Smoothing may also reveal harmonic variation, which often is caused by seasonal or diurnal factors in the dairy herd. Basically, smoothing serves the same purpose as the runs rules for PBC. Methods for calculating various types of moving averages are available in widely used spreadsheets. However, these simple tools do not always provide limits, probably because calculation of the standard errors becomes more complex. Wheeler (2010) claims that some methods to calculate limits applied in standard software are quite inappropriate.

Woodall (2000) stresses the importance of distinguishing between an initial purely explorative time series analysis like PBC (phase 1) and a subsequent SPC based on the results of the explorative time series analysis (phase 2). In phase 1, we may find justifications for assuming homogeneous processes or certain distributions (e.g., normal or binomial) that permit application of a series of parametric analytical techniques that may be used for prediction and quantification (phase 2, methods addressed below).

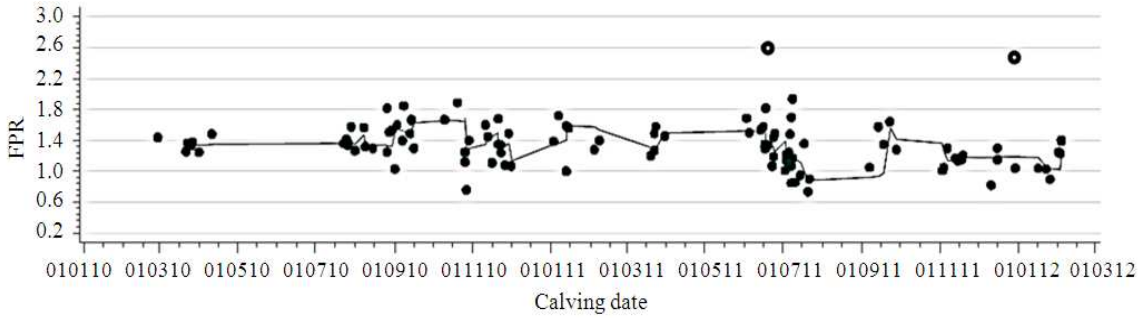

Fig. 2. Fat-to-Protein Ratio (FPR) in milk recorded between 5 and 28 days after calving. Each observation is given by a dot. The solid line is the prediction of the FPR updated at each new observation. Observations with a high probability to be outliers are identified by the model (circles) and do not contribute to the prediction

Woodall (2000) supports the view that the PBC is very robust but also states that "there is a wide difference of opinion on how much robustness is needed in practical applications, so there may always be some disagreement on this issue". Wheeler (2011) probably represents the most extreme view by stating that "We do not need to check for normality or transform the data to make them 'more' normal. We do not have to use sub-grouped data to receive the blessing of the central limit theorem before the chart will work. We do not need to examine our data for autocorrelation".

1.4. Performance Measurement by State Space Models

Figure 2 provides an example of a concept suggested by Thysen (1993). The individual data points are the same as those in **Fig. 1**. The solid line is the filtered prediction of the process at each data point. Outliers (another word for exceptional variation) are indicated by a circle. The solid line (the prediction) can take the following positions: Level shift or 'normal evolution'. An outlier will not affect the prediction; it is filtered out.

Figure 2 is one example of the so-called State Space Models (SSM). Kristensen *et al*. (2009) describe SSM and their potential applications for herd management in detail. The general purpose of a SSM (Kristensen *et al*., 2009) is to estimate the parameters in a mathematical model (e.g., regression coefficients or variances) that combines information from the observed data (e.g., the data points in **Fig. 2**) with some information available before data collection starts (e.g., expected effects of some intervention like changes in milking routines). A major advantage of this type of model is that it is a natural formulation of the Bayesian approach, which means that a priori knowledge can be combined with new information in a systematic fashion. Important assumptions can include types of distributions of error terms (e.g., normal or binomial), type of correlation between measurements, or thresholds for level-shift or outlier. A simple SSM model for dichotomous fertility data is described by Thysen and Enevoldsen (1994). The trend-line is supplemented with a graphical display of the dynamics of the raw data to support a qualitative exploration of potential causes of (exceptional) variation. This concept is implemented in freely available software for herd management support (Thysen and Enevoldsen, 2011), which is applied by a substantial number of Danish cattle veterinarians (we track the use via the download of data from the VPR-platform). The assumption of a binomial distribution behind this concept is not tested. Justification of the binomial distribution would require providing evidence that all cows in the observation period had the same chance of experiencing the events (insemination or pregnancy) (Wheeler, 2000).

In the very simple PBC concept described above, it is the manager's or the consultant's task to react to signals and start a search for causes of exceptional variation. This reaction may require some type of more or less complicated statistical analysis. In the much more complicated SSM, a statistical analysis essentially is embedded in the time series analysis. That approach may give more valid signals but at the cost that the assumptions must be justified, which may be a rather complicated task. In fact, a statistical model control is required and outliers or lack of fit detected by means of model control tools can be considered signals of deviations from the assumed (statistical) theory. In case of signals, the managerial reaction must be directed towards a search for both causes of exceptional variation (a qualitative context-bound search) and an appropriate statistical model. We suggest it will be simpler to start with the virtually assumption-free PBC, especially in the typical dairy health management context where numerous health measurements are available and relevant. Even if a SSM is validated in one context, it is

very likely that distributions and causes of exceptional variation are different in another context. Because statistical model control is a task for experts, this approach may be impractical with many herds and numerous indicator variables in each herd, as is the case for the work context of the herd veterinarian.

1.5. Multivariate Statistical Process Control

With the increasing number of herds with automatic data collection, both the number of health, fertility and production indicators and the measurement frequency increase dramatically. Some of these indicators will be correlated. So-called Multivariate Statistical Process Control is an analytical concept designed to handle the correlations and the large volume of data. By 'multivariate analysis', we mean that several variables are analyzed jointly by creating a new Y-variable (response variable) that is defined by the correlations between the original variables. The new indicator may represent an unobservable (latent) condition that has an interpretation or simply a hidden data structure. The calculations are usually based on so-called principal components. The concept with control limits is the same as in SPC. The variance can also be exposed to time series analysis with the SPC concept. However, the interpretation of out-of-control points becomes more complicated because they cannot be directly linked to one single indicator. The concept was developed several decades ago and is implemented in standard software (e.g., MVPMONITOR procedure, SAS Institute, 2011).

We are not aware of practical applications or interpretations of Multivariate Statistical Process Control for dairy herd management. Enevoldsen *et al*. (1996) applied second-order factor analysis (a similar statistical technique) to condense 22 herd-level indicators of health, fertility and production into 10 and 5 first- and second-order factors, respectively (new variables), but these new variables were not used for time series analysis.

Numerous tests are available for disease diagnosis in the dairy herd (e.g., mastitis pathogens in milk or ketone bodies in urine). In fact, every comparison of performance measurement with the associated target value can be regarded as a diagnostic test. Because diagnostic tests (including performance measurements) will be used for decision support, it is necessary to evaluate the quality in terms of sensitivity and specificity. However, information about these parameters and the associated uncertainty is often insufficient. If information about the validity and precision of a given diagnostic test is insufficient, the herd manager cannot know how an intervention based on the test results will work. Virtually all diagnostic tests are imperfect. However, knowledge about some underlying unobservable state can be obtained by combining tests similar to the multivariate technique described above. Krogh *et al*. (2011) used a Latent Class Analysis (LCA) to handle this problem for diagnosis of ketosis. The LCA might be combined with the SPC tools outlined above.

In some aspects of dairy production, we have a solid theory about the relationships between measurements that allows us to combine a number of measurements into one meaningful combination. This approach is in contrast to the purely data-driven condensation of variables by means of principal components or similar methods. An example is the so-called lactation curve. Krogh and Enevoldsen (2012a) demonstrated an analysis of milk yield recordings in which the shape of the lactation curve is defined by multiple variables in a coherent way that takes into account correlations between variables. In the case of the lactation curve, we have an example of a hierarchy of indicators and applications. We can use some components (e.g., the parameter for acceleration early postpartum) as a direct health indicator, the combination of all parameters into a lactational yield per cow and the summation of yield from all cows into a herd-level indicator of milk delivery.

In recent years, the emergence of social media and other digital stores with vast amounts of text has created a need for automatic detection of emerging trends in, for instance, buying patterns. This search is called text mining. Search engines like Google are based on such tools. The increasing requirements for documentation by means of various reports in the dairy industry may create a need for development of tools for continuous text mining to support health performance measurement. Computerized text analysis has been applied by Allaki (2005) for the veterinary authorities' surveillance of health. Text mining is also implemented in standard software (SAS Institute, 2010).

1.6. Multilevel Statistical Process Control

In a dairy herd, data are produced at multiple organizational levels (e.g., udder-quarter, udder, lactation, cow, group of cows and herd). The data from these levels may be correlated and such dependencies should be accounted for. The correlations could, for instance, be taken into account by pooling the recordings from the four quarters (e.g., electrical conductivity) into one average udder-level measurement. However, important information may be lost by this aggregation. Some of the methods described above may be developed to handle this situation effectively. We are not aware of practical applications for herd management, but industrial applications are reported.

1.7. Control and a Systems Approach To Herd Management

The mainly explorative analytic approaches described in the previous sections will enable us to detect changes within the processes in the production system. However, the historical results from an actual herd will not necessarily tell us whether the resources could have been used better in that herd. That is, was the performance acceptable, really good, or poor? Or was it optimal from a resource use point of view? The following presents relevant approaches to answering this fundamental question. Often this evaluation is called control in the management literature.

1.8. Benchmarking

Benchmarking is one obvious way to select targets. In its simplest form, it could merely be a herd manager asking his neighbor about the performance in his herd as a tool to judge his own results. More systematically, the principle of benchmarking is to identify several other herds with a similar combination of resources as our case herd and compare the performance measurement in our specific case herd with the range of results in these reference herds. This comparison will indicate performance level at best practice. For instance, what is the range of values in the best 25% of a performance indicator (e.g., milk production)? A formal comparison of targets and performance measurements may now allow us to evaluate whether we are on target or not and determine if the system is performing satisfactorily. In addition, dissemination of these targets to the farmers may motivate changes in management (Nir-Markusfeld, 2003). The selected target performance measurements can also be considered a prognosis for the future or a budget.

A fundamental problem in benchmarking is to decide when a potential reference herd really is a comparable herd. It is straightforward to find herds that are comparable with respect to very general characteristics like herd size, breed, type of ration, or milk production level. To further investigate if these herds are truly comparable, the methods described above or the methods described below can be used to delineate the production systems in sufficient detail to judge whether they are comparable.

The principles of benchmarking is used in stochastic frontier analysis in which a 'best performance' frontier is estimated to describe the best performance given a specific set of input factors (Kumbhakar and Lovell, 2003). Also Data Envelopment Analysis (DEA) describes such a frontier but is driven by actual observations (performance measurements), instead of detailed knowledge about production functions. Nielsen and Bramsen (2004) provided an example of DEA in pig production. DEA does not account for uncertainty in the variables. In practical management of Danish dairy production, benchmarking on health indicators so far seems to have used one performance measurement at a time (univariable), which does not account for the correlation between the performance measures.

Correlation between performance measurements in essence means that calculating additional performance measurements will yield only minor additional information. The negative correlations are the most troublesome because targets often are derived from univariable analyses. In the case of lactation curves, Krogh and Enevoldsen (2012a) addressed this issue in detail. An increasing peak yield is strongly associated with a steeper decreasing slope afterwards, but because the correlation varies from herd to herd, the correlation can be a performance measurement per se.

It is obvious that benchmarking is invalidated if the scale of a measurement differs from herd to herd. Milk yield, fat percentage and Somatic Cell Counts (SCC) are examples in which the scales are calibrated in central systems. However, for the cattle veterinarian, animal-level conditions like body condition, lameness and skin lesions are examples in which scoring systems (ratings) are needed. These 'clinical recordings' obviously must be standardized to be useful for benchmarking. Clinical criteria that are constant within herd (e.g., specific for a single manager or veterinarian) may suffice if performance measurement is restricted to historical data within the herd. Kristensen *et al*. (2006) demonstrate typical variation in scores and that agreement in clinical scores quite easily can be improved with training. Consequently, before any target health performance measurement (indicator) can be chosen, the quality of available clinical records must be evaluated. The evaluation essentially includes estimation of sources of variation (random, within-herd, between-herd) and identification of systematic errors in data collection.

Even when score values are described in detail in manuals or protocols, they may be used differently by veterinarians or others doing recordings in the herds (Lastein *et al*., 2009). The veterinarians' perception of the herd health management system could influence the basic clinical recordings. Recordings of disease treatments are also influenced by herd-specific conditions (Vaarst *et al*., 2002), which will make comparability across herds very poor. Krogh and Enevoldsen (2012b) have described a concept to detect this type of measurement error. This approach could be useful in a large veterinary practice that might want to develop a benchmarking system based on recordings from multiple veterinarians in the practice.

Data used for benchmarking are often an aggregation of data for a longer period of time (e.g., a year or a quarter of a year). The same time interval is usually used in routine reports to evaluate the performance of a given concrete herd. In case we have not discovered an important time trend, we may miss a signal or get a misleading signal. Averages, ranges and histograms all obscure time order, which can be misleading (Wheeler, 2000). If, for instance, performance has improved markedly in our case herd, we might be interested only in the value for the latest month. Consequently, an appropriate time series analysis with as few restrictions as possible should always precede traditional statistical analyses like benchmarking or statistical modeling (Armitage *et al*., 2008).

1.9. Planning Tools to Derive Targets for Performance

Even if we have identified 'comparable' herds, specific constraints or personal values may persist that make the concrete herd unique. Therefore and ideally, regular and iterative planning processes should produce herd-specific plans that again should have formulated goals for health, fertility, production, etc., based on the system context and the use of the available input factors like feed, medicine and management. The goals should be specified as targets for the performance measurements that can be derived routinely from the production process (Kristensen *et al*., 2009). A simple approach to setting herd-specific targets is to take historical results and adjust them for expected results of the planned changes in the next planning period. Enevoldsen (1993) demonstrated this simple approach for a series of health and fertility performance measures. The expected results (targets) of changes in plans were based on a mix of general theoretical knowledge and context-specific knowledge about the herd and the management.

Numerous advanced tools are available for planning. Major examples include (Kristensen *et al*., 2009): expert systems (based on norms and logic), linear programming (widely used to formulate feed rations), dynamic programming and Markov decision processes (e.g., used to select the optimal time to replace cows), Bayesian networks and decision graphs (very complicated development of decision trees that represents uncertainties of decision problems) and simulation (computer model of an entire system; e.g., a herd). Ideally, the targets should be estimated from an optimization of the available resources. This optimization can be obtained by means of some of these tools. For dairy herd health management, a very complicated and scientifically well-documented and commercially available herd model is adapted to the needs of practicing cattle veterinarians (Ostergaard *et al*., 2010).

The requirements for performance measurement will depend on the time horizon. In herd management science, decisions about the production system have traditionally been divided into the strategic, tactical and operational levels. Operational decisions typically relate to day-to-day management routines in the production process. The effects of operational decisions can quite quickly be implemented and evaluated and the economic impact of the individual decision is often of minor magnitude for the herd as a whole. The tactical decisions are in the month-to-year time frame. The decision could be to increase the amount of labor and change the feed ration. Strategic decisions are long term. The decision could be to build a new stable, increase the number of dairy cows, or convert to organic farming. The needs for and types of performance measurements are very different at these levels.

Wheeler (2000) provides numerous examples of the errors that can occur if the target setting and comparison with an aggregated single-value performance measurement are used alone in some 'Annual Report' without a detailed preceding time series analysis. In fact, his view seems to be that the aggregated report is unnecessary if an appropriate PBC analysis is conducted. The advantage of this graphical approach is that we avoid definition of arbitrary (non-biological) cut-offs between time periods.

1.10. Causal Analysis Supported by Multivariable Statistical Modeling

The application of the tools for time series analysis usually will create a need for further analysis to identify causes of exceptional variation or emerging trends, or options for reduction of routine variation (that is, to re-engineer the system). A possible need for setting targets may also require additional analysis. Well suited for both purposes are Multivariable Statistical Models (MSM; e.g., logistic and linear regression or analysis of variance), which have been used for research purposes for many years (e.g., Armitage *et al*., 2008). Implementation of MSM at a larger scale for herd management is described by Markusfeld (1993); Enevoldsen (1997a; 1997b) and Nir-Markusfeld (2003). Examples of important information produced by such MSM include: differences in milk production between cows with or without mastitis, differences in chances of pregnancy in cows with or without previous metritis and risk of early culling in cows with or without ketosis. If the analyst has context knowledge about the herd, such information can be valid as estimates of predicted effects of management interventions to reduce disease occurrence.

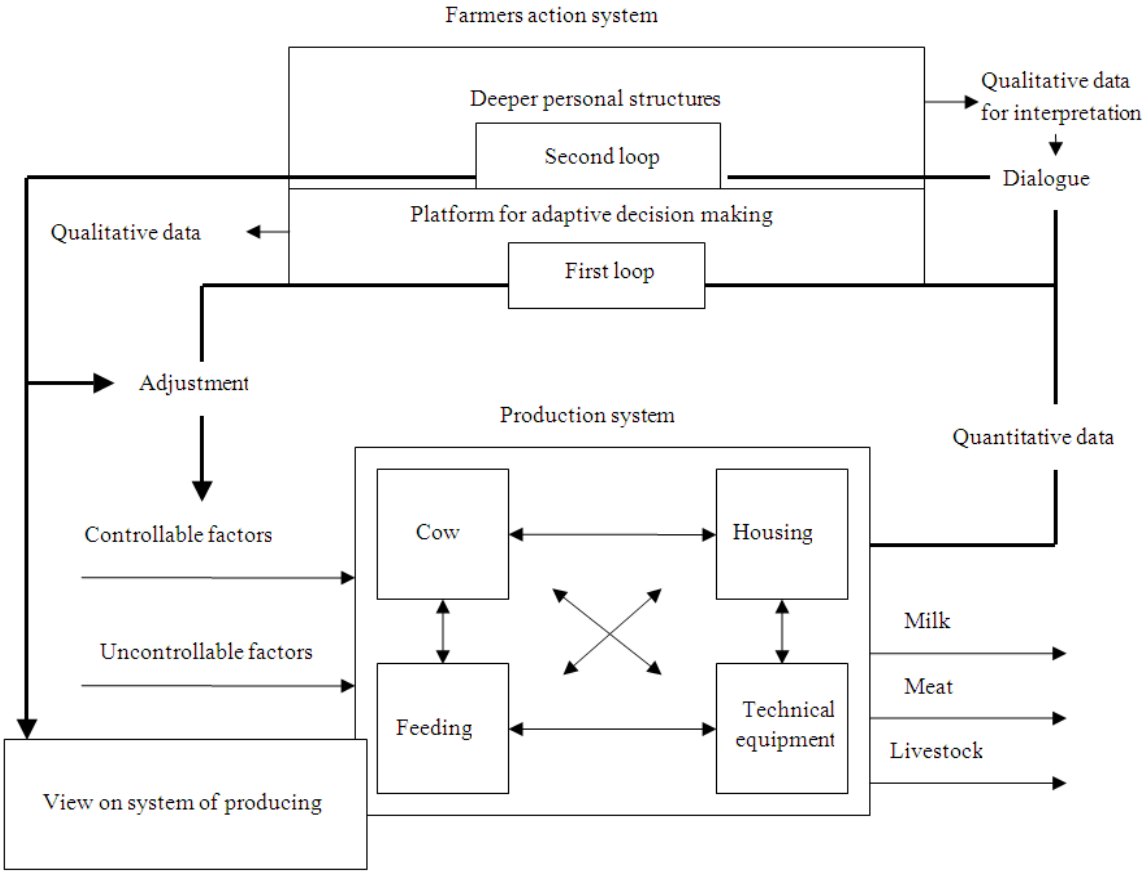

Fig. 3. Factors, relationships, feedback and interactions in a system comprising the production system and the farmer's personal action system (Andersen, 2004, with permission)

A MSM can also be used to estimate a time trend in a performance measurement. Singer and Willett (2003) and Kristensen *et al*. (2009) suggest a range of approaches for modeling change and event occurrence. Multiple levels (e.g., cow, herd and veterinary practice) can also be handled (Krogh and Enevoldsen, 2012c). The advantage compared with the time series analyses described above is that numerous possible confounding factors like parity and stage of lactation can be accounted for in a systematic fashion. Consequently, time trends derived from a MSM may be more valid than time trends derived from the time series analyses. In fact, a MSM may also detect time trends that were not detected by the time series analyses because they were hidden by confounding factors. However, application of MSM relies on several assumptions like distributional properties, independencies of data, or appropriate model specification. Prior application of a PBC may help in identifying situations in which these assumptions are justified. Results of statistical model control may also serve as signals of changes in the process or signals of exceptional variation. Appropriate model control should also detect violation of distributional assumptions.

1.11. Quantitative and Qualitative Methods for a Systems Approach

Andersen (2004) gives an example of the challenges we can face when a herd health consultant works with the herd manager. **Figure 3** represents the synthesis of thorough successive quantitative and qualitative analyses of a single herd conducted at several herd visits and discussions with the herd owner over several months. The production system is composed of cows, housing, feeding and technical equipment. The production process transforms input factors to output (products, milk, meat and livestock). Measurements from the production system (quantitative data) use by the farmer to adjust the flow of input (feedback).

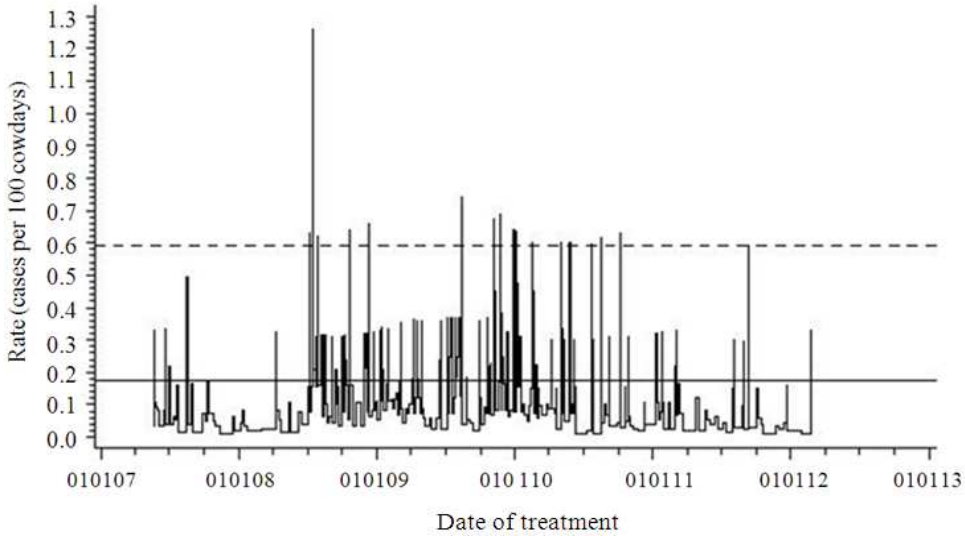

Fig. 4. Rate of Interdigital phlegmon treatment over time in one herd. The average treatment rate (solid line) and natural process limit (dashed line) are calculated on the entire time period. The average moving range is used for calculation of the natural process limit

One view on herd management can be that this adjustment is according to simple decision criteria. However, the case behind **Fig. 3** demonstrated that this particular farmer's action system was very complex and dynamic and involved feedback mechanisms. Personal values and views on the role as farmer in the community played some part. Andersen (2004) described the entire system as a learning system in which double-loop learning took place.

The joint application of some of the tools described above for performance measurement, including tools for setting targets, is demonstrated by Enevoldsen *et al*. (1995), where a systems approach (Kristensen *et al*., 2009) is applied to a concrete case-herd. This approach allows us to express our prior knowledge of the qualitative and quantitative structures of the system we work with. Complicated computer models usually play a major role in a systems approach. However, essential parts of the information needed for input to the computer model must be derived from the herd manager (cf. **Fig. 3**).

The analysis and subsequent synthesis of a theory about such a system as described in **Fig. 3** require much more than routinely collected data. A lengthy dialogue is needed to establish a genuine common understanding between the farmer and the researcher. Several qualitative research techniques are useful for such purposes. However, the information obtained with these qualitative methods can also be very useful for specifying and using MSM to analyze the quantitative data. In the particular case demonstrated in **Fig. 3**, advanced quantitative decision-support tools probably would have been of very limited use if applied without the qualitative knowledge obtained. The qualitative knowledge, in contrast, probably would be quite useful alone.

Kristensen *et al*. (2008) use the term Mixed-Methods Research (MMR) to describe the research approach leading to a model like the one in **Fig. 3**. MMR basically is rooted in the social sciences. Kristensen and Enevoldsen (2008) use a so-called Q-Method to obtain more general knowledge about current subjective views like the manager's views indicated in **Fig. 3**. The latter study also showed that the subjective views on consultancy differed markedly between cattle veterinarians and dairy farmers. This factor illustrates the importance of establishing a genuine common understanding of the entire system. From the quantitative perspective, Wheeler (2000) also stresses the importance of context knowledge by specifying a (somewhat provoking) 'first principle for understanding data': No data have meaning apart from their context.

1.12. Major effects of public management and other organizational constraints on performance data

Figure 4 shows a Process Behavior Chart from a dairy herd during a 4-year period. Limits are empirical and estimated as described for **Fig. 1**. The average treatment rate and the natural process limit, based on average moving range, are calculated on the entire time

period. The performance measurement is the rate of medical treatment for InterDigital Phlegmon (IDP) among the cows in the herd. From **Fig. 4** it is evident that there is a clear change in the treatment rate from July 2008. The issues related to proportions and rates are discussed by Wheeler (2000). The assignable cause of the marked shift(s) was not a change in the biological processes but a change in the criteria for defining the diagnosis. New legislation introduced some disease categories in which farmers legally could get drugs and others in which they could not. For IDP, a farmer could get prescriptions but could not do so for Digital Dermatitis (DD). Not surprisingly, the manager had a strong incentive to use IDP instead of DD in cases of foot problems. For the herd presented in **Fig. 4**, the herd entered the herd health management program and the new legislation in July 2008.

Another example is the use of Somatic Cell Count (SCC) in the milk sold to the milk processor as an indicator of udder health. Because milk payments from the milk processors are reduced in cases of SCC above certain limits, it is quite obvious that farmers have an incentive to discard milk from cows with high SCC values. Consequently, the value of SCC in deliveries as an indicator for the herd's udder health status may be distorted. What happened here is what Wheeler (2000) called the Voice of the Customer. That is, the decision takers in the organization attempt to adjust to the needs of the outside world while the process per se is not changed.

Such distortion of the data is not seen as a problem for the manager or the local consultant because they know what goes on in the process. However, an outside observer without sufficient context knowledge (e.g., a statistician working with large data files for research or a veterinary officer doing follow-up on the legal regulations) may draw naive conclusions about the process, which might lead to unjustified political interventions or causal inference. The upshot could be reduced efficiency of the process or even its misdirection.

A misinterpretation of data like the one outlined above is also recognized in the social sciences and basically viewed in the same way as Wheeler (2000), who gives an example (pp. 70-71) and states that "…pressure to meet any arbitrary numerical goal or target will most often result in the distortion of either the system, or the data, or both". Krogstrup (2011) calls such a local distortive management reaction to outside regulation or requirement a 'perverse side effect' in a thorough discussion of performance measurement, effect evaluation and evidence in (New) Public Management. As an example, targets for the rate of dead cows and calves are now incorporated into Danish legislation. Despite the fact that the targets are extremely high, the first author has experienced that simply setting the targets has made some farmers change behavior. Some farmers became more reluctant to euthanize chronically ill cows, instead keeping them in the herd, hoping for recovery. The consequence is that in some herds, there is a substantial amount of 'accumulated suffering' – cows kept in the herd suffering from various conditions with poor prospects for recovery. This example represents a perverse side effect because the purpose of setting the target was to improve animal welfare. It is clear that inclusion of these sociological aspects will make even more complicated the rather complicated representation of an organization in **Fig. 3**.

Krogstrup (2011) defines the term 'performance measurement' as the combination of measurements of processes (what goes on in terms of, e.g., types of management routines (actions) like heat detection), output of the processes (in terms of what was actually done in the process-routines; e.g., minutes of heat detection every day) and results (outcome; e.g., pregnancy rate). In our herd context, it is implicit that the process is influenced by some intervention and the context (competencies and capacity). That is, by measuring 'output', we measure the intervention that has taken place. The outcome is the result of the output (process). This outcome (results) is what the recipient experiences. Wheeler (2000) basically uses the same demarcation by distinguishing sharply between the voice of the process (performance of the process per se) and the voice of the customer (the quality of the products). A subset of the outcome is the direct or the indirect effect of the intervention; that is, the causal effect(s). Management of an organization can be based on measurements of the outcome; an evaluation of whether the results are on target (in new public management terms, a results contract). In this public management context, the term 'evaluation' may seem similar to the term 'control' described above for herd management. However, Krogstrup (2011) gives a broader definition of evaluation: "A systematic retrospective assessment of output (process), outcome (results), administration and organization of (public) business, which is expected to play a role for practical actions". In this definition, it is essential to note that evaluation includes some judgment that separates important aspects from unimportant aspects. It is also essential that practical use is intended. For an intervention to be practically applicable, we need to know how and when it works. This view is similar to the term 'surveillance' used by Schwabe *et al*. (1977) and Stark and Salman (2001) in epidemiology. They use surveillance as some active goal-oriented process

(Schwabe *et al*., 1977: 'information for action') in contrast to monitoring as some passive data collection (measurement) without evaluation. If no decision or action is possible, then the measurement does not provide information and is thus worthless for management. Kristensen *et al*. (2009) do not make a distinction between monitoring and surveillance and simplify the complexity of views, values, interaction, feedback and learning into a general term like 'utility function' without addressing the problems of identifying this function in practice. To us, the parameterization of a utility function seems to be a big challenge in a veterinary practice context, especially because **Fig. 3** indicates that the utility seems to be dynamic.

With the increasing public focus on regulation of animal production (e.g., animal welfare promotion and reduced usage of antibiotics), it follows that there will be an increasing need for evaluation of the results of the interventions and ideally the effects of the interventions. In large herds with large personnel, some incentive systems based on obtained results may be used. That is, perverse side effects may be an important issue to consider for both local and public management of data collected from the herds. For the purpose of providing documentation of the state of the production system to public authorities, the manager probably does not see 'perverse side-effects' as perverse.

For obvious reasons, we want to know as much as possible about the causal effects of interventions. In a simple-problem context like assessment of the effects of mechanical changes in an AMS on the frequency of cows' visits to the robot, a quantitative estimation of the effect is straightforward with the numerical methods outlined above, if sufficient context knowledge is available. Krogstrup (2011) calls such a problem a tame problem, in contrast to identification or quantification of causal effects (evaluation) in a context like **Fig. 3**. Krogstrup (2011) calls a problem similar to that in **Fig. 3** a wild problem, which mainly is characterized by a vague definition, lack of an optimal solution, unclear causal mechanisms and interaction between context and mechanisms. Krogstrup (2011) gives a thorough discussion of the possibilities for evaluation of such problems. One prerequisite is to specify an intervention theory. Often, the modest ambition will be to explain why some intervention did not work. Basically the formulation of **Fig. 3** will allow us to identify key elements that can be addressed with a mixed-methods approach. Again, context knowledge is essential. Krogstrup (2011) uses the term Context-Mechanism-Outcome, which means that interventions cause mechanisms that then selectively interact with the case-specific circumstances (context) and result in effects that differ in different contexts. A very complicated system like this can be considered self-organizing (Rickles *et al*., 2007). The term complex responsive processes (Stacey, 2011) seems applicable, as well. This concept describes organizational knowledge as being in the relationships between people in an organization.

A clear-cut context-specific intervention theory is also needed to reduce the number of potentially relevant performance measurements that otherwise easily becomes large, causing the overview of the system to be lost. Krogstrup (2011) gives an overview of approaches to evaluate evidence of effects of intervention in the spectrum of contexts, from tame to wild, from the randomized controlled trial, which is regarded as the ideal in medicine but is impossible to apply to wild problems, to the everyday evaluation, or an effect-focused practice. A systematic use of the simple PBC in a herd (which includes more or less qualitative follow-up to remove effects of exceptional variation) could be seen as an example of an effect-focused practice.

1.13. A Definition of (herd) Health in the Context of a Systems Approach

In the preceding text, we have not defined health; we have focused on management of measurements related to disease occurrence. However, our presentation and discussion of these concepts and tools bring us closer to an understanding of health. In standard veterinary textbooks, explicit definitions of health are rare (Gunnarsson, 2006) and Houe *et al*. (2004) also state that health is often defined for a very specific context. Hence, a definition of herd health is at least as problematic. A similar problem exists in humans, for whom the term 'public health' sometimes seems to be defined only as preventive medicine - the science of preventing diseases. However, much broader definitions also have been applied that involve the interaction among society, population and health, intended to improve the health of the population through education and preventive medicine (MacQueen *et al*., 2001).

In a herd health context, the difference between the health of an individual and herd health is that herd health is concerned with the herd as a system, as illustrated in **Fig. 3;** that is, not only the population of animals is of concern but also the 'support' for the population as environment and management. Based on Albrecht *et al*. (1998) and the concepts described above, we propose an analogous definition of herd health, which then can be, "Animal, environment and manager together viewed as a dynamic and complex ecosystem. In this context, an

ecologically informed or process-view of herd health implies the self-regulation through feedback and maintenance of all relevant systems promoting ongoing physical, mental/emotional and social well-being. This latter definition gives us a sharper understanding of what poor herd health is. That is, the loss of the ability to self-regulate and the disintegration of support systems leading to the necessity for intervention. In a process-view, intervention is directed towards restoration of all relevant support systems in order for health again to be self-generated and self-regulated".

In this definition, it is important to acknowledge that being healthy in a herd health context involves the herd managers' conception of the animals' well-being. Thus, the role of the herd manager (context) is pivotal.

2.CONCLUSION

It is our experience from several countries that often the only tools for health performance measurement in dairy herds are simple graphical or tabular presentations of data without attempts to address random and systematic variation in the production process. Also, there is limited or no support for systematic evaluation of the performance of the process by means of some type of limits or criteria for intervention. In the following, we suggest to the herd veterinarian for cattle herds a concrete stepwise approach to using the concepts and tools for management of health performance measurement data presented above to develop a systems approach to herd health management in an industrialized dairy herd.

2.1. Step 1

Develop process behavior charts like that shown in **Fig. 1** for the available routine measurements from standard herd management programs. These charts do not require sophisticated software or hard-to-justify assumptions. Use animal-level data directly whenever possible. Do not wait until ideal data are available; there will always be data available that are useful for health performance measurement.

2.2. Step 2

Make sure you can answer the following questions concerning the definition of the measurements: For what purpose were data collected? Who collected the data? How, when and where were data collected? What do values represent? If computed, how were they computed from raw data? Were there changes in formulas over time? Precise knowledge about these topics in the concrete herd will give a very strong and necessary foundation for interpreting the charts. Knowledge about the specific context and the dynamics in the context will increase. Meeting these requirements may be a real challenge for a herd health consultant but also an important learning process.

2.3. Step 3

Interpret the patterns in each chart, search for assignable causes of exceptional variation (outside limits or trends) and attempt to remove such causes. This systematic process will add further to your knowledge about the herd context, including the manager's more or less subjective views. The charts and your use of them will document your reasons for suggesting interventions to the herd manager and, if needed, to the public veterinary authorities. You will also be able to distinguish clearly between process-related and results-related measurements and experience the difference between them through the dialogue with the manager.

2.4. Step 4

Search for options to reduce the routine variation when the results of the process are unsatisfactory. Some options will be obvious (e.g., repair technical faults in the milking equipment or ensure hoof trimming). However, because of the typically large number of animals and long-time horizon in dairy production, you will profit from some multivariable or multivariate statistical modeling. A range of traditional statistical models and state space models are developed specifically for this purpose (presented and discussed above). Model control of these analyses can also serve as advanced tools to explore causes of exceptional variation. Standard setups are available and the younger generation of veterinarians has been trained in using simple versions. This process will also add substantially to your context knowledge.

2.5. Step 5

Set up targets at the tactical or strategic level. The interventions to reduce the routine variation or simply improve the results by eliminating product out of specifications (e.g., high cell counts) will often require some investments, which are quite easy to estimate. However, the benefits in terms of increased production or decreased disease-associated losses are more complicated to assess. Models for doing such analyses are described above. Some are commercially available and you can get support for interpretation and use. With the knowledge gained during steps 1-4, you will be well equipped to provide relevant and comprehensive input to

these models. The models provide predictions of the important health performance measures and potential profit arising from the interventions you consider. The discussions of the results with the manager will bring you deep into the topics described in **Fig. 3**, which again will provide knowledge about causes of exceptional variation. The entire process in step 5 will also provide estimates of the economic value of each health performance measurement.

2.6. Step 6

Adjust the measurements and the intervention strategy. Steps 1-5 should initiate an iterative process. Some measurements will be dropped, others added, the quality of the measurements assessed, process limits or targets possibly changed, cost-benefit assessed, etc. In essence, you have established a systems approach to dairy herd (health) management like that outlined above.

2.7. Step 7

Develop a framework to support the health performance measurement process at the practice level. This will be particularly useful for establishing a basis for benchmarking because the context knowledge obtained in steps 1-6 will allow identification of the most comparable herds. Above, a tool is presented for identifying rater bias in ratings used for health performance measurements that must be corrected prior to benchmarking, or across-herd analyses to evaluate, for example, the effects of various interventions like those discussed above in the case of metritis diagnosis and treatment. The validity and usefulness of across-herd analyses will be greatly improved compared to data from larger data collections from multiple veterinary practices. A homogeneous set of data will also be useful for evaluation of diagnostic tests applied in practice and development of new health performance measures like those demonstrated in the case of lactation curves. The activities in step 7 will almost certainly require expert statistical assistance.

3. REFERENCES

Albrecht, G., S. Freeman and N. Higginbotham, 1998. Complexity and human health: The case for a transdisiplinary paradigm. Culture Med. Psychiatry, 22: 55-92. PMID: 9657059

Allaki, F.E., 2005. Theorie de la Surveillance de la Sante des Populations. 1st Edn., Universite de Montreal, ISBN-10: 9780494180303, pp: 304.

Andersen, H.J., 2004. Rådgivning-Bevægelse Mellem Data og Dialog. 1st Edn., Mejeriforeningen, ISBN-10: 8789795814, pp: 348.

Armitage, P., G. Berry and J.N.S. Matthews, 2008. Statistical Methods in Medical Research. 4th Edn., John Wiley and Sons, Chichester, ISBN: 9780470775349, pp: 832.

Enevoldsen, C., 1997a. Epidemiological considerations related to within herd multivariable modelling in herd health management. Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics, (ISVEE' 97), Paris, pp: 13-32.

Enevoldsen, C., 1997b. Det israelske rådgivnings-koncept-og en dansk oversættelse.

Enevoldsen, C., 1993. Dairy herd health management (in Danish: Sundhedsstyring i Mælkeproduktionen). Ph.D. Thesis. The Royal Veterinary and Agricultural University.

Enevoldsen, C., J. Hindhede and T. Kristensen, 1996. Dairy herd management types assessed from indicators of health, reproduction, replacement and milk production. J. Dairy Sci., 79: 1221-1236. PMID: 8872716

Enevoldsen, C., J.T. Sorensen, I. Thysen, C. Guard and Y.T. Grohn, 1995. A diagnostic and prognostic tool for epidemiologic and economic analyses of dairy herd health management. J. Dairy Sci., 78: 947-961. PMID: 7790588

Gunnarsson, S., 2006. The conceptualisation of health and disease in veterinary medicine. Acta Vet. Scand., 7: 47-71. PMID: 17090301

Hoerl, R.W. and A.C. Palm, 1992. Discussion: Integrating SPC and APC. Technometrics, 34: 268-272.

Houe, H., A.K. Ersboll and N. Toft, 2004. Introduction to Veterinary Epidemiology. 1st Edn., Biofolia, ISBN-10: 8791319218, pp: 398.

Koutras, M.V., S. Bersimis and P.E. Maravelakis, 2007. Statistical process control using Shewhart control charts with supplementary runs rules. Methodol Comput. Applied Probab., 9: 207-224. DOI: 10.1007/s11009-007-9016-8

Kristensen, A.R., E. Jorgensen and N. Toft, 2009. Herd Management Science. II. Advanced topics. 1st Edn., University Copenhagen, Denmark, ISBN: 9788763460644, pp: 329.

Kristensen, E. and C. Enevoldsen, 2008. A mixed methods inquiry: How dairy farmers perceive the value(s) of their involvement in an intensive dairy herd health management program. Acta Vet. Scand., 50: 50-50. DOI: 10.1186/1751-0147-50-50

Kristensen, E., D.B. Nielsen, L.N. Jensen, M. Vaarst and C. Enevoldsen, 2008. A mixed methods inquiry into the validity of data. Acta. Vet. Scand., 50: 30-30. DOI: 10.1186/1751-0147-50-30

Kristensen, E., L. Dueholm, D. Vink, J.E. Andersen and E.B. Jakobsen *et al*., 2006. Within- and across-person uniformity of body condition scoring in Danish holstein cattle. J. Dairy Sci., 89: 3721-3728. DOI: 10.3168/jds.S0022-0302(06)72413-4

Krogh, M. and C. Enevoldsen, 2006. Organizational and educational support to dairy herd health programs. Proceedings of the 11th Symposium of the International Society for Veterinary Epidemiology and Economics, (ISVEE '06), Cairns, Australia, pp: 986-986.

Krogh, M.A., N. Toft and C. Enevoldsen, 2011. Latent class evaluation of a milk test, a urine test and the fat-to-protein percentage ratio in milk to diagnose ketosis in dairy cows. J. Dairy Sci., 94: 2360-2367. PMID: 21524525

Krogh, M.A. and C. Enevoldsen, 2012a. A framework for integration of benchmarking and within-herd analysis in dairy herd management-analysis of lactation curves as a case. In: Ph.D. Thesis University of Copenhagen, Denmark.

Krogh, M.A. and C. Enevoldsen. 2012b. A tool to detect rater-introduced bias in clinical ratings. In: Ph.D. Thesis, University of Copenhagen, Denmark.

Krogh, M.A. and C. Enevoldsen, 2012c. Evaluation of effects of disease control in a complex dairy herd health management program. In: Ph.D. Thesis, University of Copenhagen, Denmark.

Krogstrup, H.K., 2011. Kampen Om Evidens: Resultatmåling, Effektevaluering Og Evidens. 1st Edn., Hans Reitzel, ISBN: 9788741255163, pp: 170.

Kumbhakar, S. and C.A.K. Lovell, 2003. Stochastic Frontier Analysis. 1st Edn., Cambridge University Press, New York, ISBN-10: 0521481848, pp: 333.

Lastein, D.B., M. Vaarst and C. Enevoldsen, 2009. Veterinary decision making in relation to metritis-a qualitative approach to understand the background for variation and bias in veterinary medical records. Acta Vet. Scand, 51: 36-36. DOI: 10.1186/1751-0147-51-36

MacQueen, K.M., E. McLellan, D.S. Metzger, S. Kegeles and R.P. Strauss *et al*., 2001. What is community? An evidence-based definition for participatory public health. Am. J. Public Health. 91: 1929-1938. DOI: 10.2105/AJPH.91.12.1929

Markusfeld, O., 1993. Epidemiological methods in integrated herd health programs. Acta Vet. Scand, 89: 61-67. PMID: 8237667

Nielsen, K. and J.M. Bramsen, 2004. Interaktiv Benchmarking: Med Eksempler Fra Landbruget. 1st Edn., Fødevareøkonomisk Institut, Denmark.

Nir-Markusfeld, O., 2003. What are production diseases and how do we manage them? Acta Vet. Scand., Suppl., 98: 21-32. PMID: 15259777

Ostergaard, S., J.F. Ettema, A.B. Kudahl and J.T. Sørensen, 2010. Development of a SIMHERD web application for herd health advisors-experiences and perspectives. Proceedings at Farm Animal Health Economics, Nantes, France, Jan. 14-15.

Rickles, D., P. Hawe and A. Shiell, 2007. A simple guide to chaos and complexity. J. Epidemiol. Community Health, 61: 933-937. PMID: 17933949

SAS Institute, 2010. Getting Started with SAS® Text Miner 4.2. 1st Edn., SAS Institute Inc., Cary, NC, USA., pp: 84.

SAS Institute, 2011. SAS/QC 9.3 User's Guide. 1st Edn., SAS Institute, Cary, NC, ISBN-10: 9781607649113, pp: 3344.

Schwabe, C.W., H. Riemann and C.E. Franti. 1977. Epidemiology in Veterinary Practice. 1st Edn., Lea and Febiger, Philadelphia, ISBN: 0812105737, pp: 313.

Singer, J.D. and J.B. Willett, 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. 1st Edn., Oxford University Press, New York, ISBN-10: 0195152964, pp: 672.

Stacey, R.D., 2011. Complex Responsive Processes in Organizations. 1st Edn., Learning and Knowledge Creation in Organizations, Routledge, New York, ISBN-10: 0415249198, pp: 258.

Stark, K.D.C. and M.D. Salman, 2001. Relationships between animal health monitoring and the risk assessment process. Acta Vet. Scand, 42: 71-77. DOI: 10.1186/1751-0147-42-S1-S71

Thysen, I. and C. Enevoldsen, 2011. HerdView-a PC-program for interactive analysis of reproduction and health in dairy herds.

Thysen, I. and C. Enevoldsen, 1994. Visual monitoring of reproduction in dairy herds. Prev Vet. Med., 19: 189-202. DOI: 10.1016/0167-5877(94)90088-4

Thysen, I., 1993. Monitoring bulk tank somatic cell counts by a multi-process Kalman filter. Acta Agric. Scand. Sect. A, Anim. Sci., 43: 58-64. DOI: 10.1080/09064709309410145

Vaarst, M., B. Paarup-Laursen, H. Houe, C. Fossing and H.J. Andersen, 2002. Farmers' choice of medical treatment of mastitis in Danish dairy herds based on qualitative research interviews. J. Dairy Sci., 85: 992-1001. PMID: 12018446

Vries, A. and J.K. Reneau, 2010. Application of statistical process control charts to monitor changes in animal production systems. J. Anim. Sci., 88:11-24. DOI: 10.2527/jas.2009-2622

Wheeler, D.J., 2000. Understanding Variation: The Key to Managing Chaos. 2nd Edn., SPC Press, Knoxville, ISBN-10: 0945320531, pp: 158.

Wheeler, D.J., 2010. The right and wrong ways of computing limits. How does your software measure up? Quality Digest Daily, Jan. 7, 2010. Manuscript No. 205.

Wheeler, D.J., 2011. Myths about process behavior charts. How to avoid some common obstacles to good practice. Quality Digest Daily, Sept. 6, 2011. Manuscript No. 232.

Woodall, W.H., 2000. Controversies and contradictions in statistical process control. J. Quality Technol., 32: 341-378.