

Recommendations for Field Tests of Milking Machine Performance

Steven D. LeMire, Graduate Research Assistant
Douglas J. Reinemann, Ph. D., Associate Professor
Department of Biological Systems Engineering

Graeme. A. Mein, Ph.D., Visiting Professor
Department of Dairy Science and School of Veterinary Medicine

Morten D. Rasmussen, Ph.D.
Danish Institute of Animal Sciences
Visiting Professor of Biological Systems Engineering, University of Wisconsin - Madison

University of Wisconsin-Madison, Milking Research and Instruction Lab

Written for presentation at the
1999 ASAE annual International Meeting
Toronto, Ontario Canada
July 18-21, 1999

Summary: This paper presents recommendations for field tests on milking machine performance. The subjects of milk production, peak flow, average flow, strip yield, teat conditions, and unit corrections are dealt with. Techniques for maximizing test efficiency such as blocking designs to reduce variability are discussed. Test assumptions are discussed along with practical advantages and limitations of these tests.

Keywords:

Milking Time, Statistical Tests, Statistical Power, Cow Identification

The author(s) is solely responsible for the content of this technical presentation. The technical presentation does not necessarily reflect the official position of ASAE, and its printing and distribution does not constitute an endorsement of views which may be expressed. Technical presentations are not subject to the formal peer review process by ASAE editorial committees; therefore, they are not to be presented as refereed publications.

Quotation from this work should state that it is from a presentation made by (name of author) at the (listed) ASAE meeting.
EXAMPLE C From Authors Last Name, Initials. "Title of Presentation." Presented at the Date and Title of meeting, Paper No X.
ASAE, 2950 Niles Road, St. Joseph, MI 49085-9659 USA.

For information about securing permission to reprint or reproduce a technical presentation, please address inquiries to ASAE.

ASAE, 2950 Niles Rd., St. Joseph, MI 49085-9659 USA
Voice: 616.429.0300 FAX: 616.429.3852

RECOMMENDATIONS FOR FIELD TESTS ON MILKING MACHINE PERFORMANCE

INTRODUCTION

Recommendations on milking equipment and settings are often based on general rules without taking into account the conditions on an individual farm. Proper assessment of a change to the milking machine is often not done. The adoption of data collection systems on dairy farms has greatly improved the ability to monitor the milking process. This paper will focus on simple tools to assess performance of the milking process primarily on farms that have milking parlors and the ability to collect and analyze data. These principles can also be applied, however, to facilities that do not have automated data collection systems.

An optimized milking machine will obtain the most possible milk in the least amount of time with minimal effect on udder health and cow behavior. Previous suggestions for milking-time tests performed in the field include assessment of teat condition, completeness of milking, slipping and falling teatcups, milking time per cow, and peak milk flow rate (Mein and Haman, 1993; Mein and Ried, 1996). Mein and Haman (1993) also suggest that further refinement is required to apply quantitative measures under field conditions and test methods must be carefully specified prior to running the tests to avoid misinterpretation of results. There are several areas of specification of field tests that will be addressed in this paper: accurate and consistent measurement methods, sound experimental design and appropriate statistical analysis.

The performance measures addressed in this paper are:

Average and peak milk flow rate. These are indicators of the speed and efficiency of the milking machine.

Strip yield. The amount of milk that can be removed by hand milking immediately after the milking machine is removed. This is a measure of the completeness of milking.

Teat condition. Both short term and long term changes in the teat are measures of the gentleness of the milking process. It is important to maintain healthy teat tissues to reduce the risk of mastitis infection.

Liner slips and unit falloffs. The number of times the operator must correct the position of the milking machine because of liner slip, or reattach a milking unit after a falloff, influence the efficiency of the work routine as well as the risk of new mastitis infections.

A basic differentiation between correlation and cause is part of any discussion of statistical analysis. Observational studies, which rely on a large sample size and do not involve control of the causal variables, can be used to determine the correlation between the variables studied (e.g. stork population and birth rate). The goal of assessing milking machine performance is to determine cause and effect. In order to conclude a cause and effect relationship the causal variable must be manipulated and all other sources of variability must be controlled.

An example of a common methodology to assess changes in the milking process is to make a change and "see what happens". This approach does not account for chance bias over time that might effect the response. The most likely result of such an effort is an indeterminate conclusion (e.g., the new liners may have affected something but we can't say for sure). This approach does little to advance the knowledge of the dairy manager or investigator.

Field tests on farms are limited by the amount of time, resources and number of cows available. Tests often need to be run in one or two milkings or over one or two days. On a large dairy farm with automated data collection systems one of the practical limitations of performing a milking-time test is the organization of cows to receive the intended treatment. Cows are commonly grouped by production parameters. On dairies without automated data collection systems, a single person will likely collect data. In this scenario, methods to design an experiment with the best resolution that uses the fewest number of cows is highly advantageous. This can be accomplished by using careful experimental design, which controls for the many sources of variability.

Dairy farms, in general, and milking, in particular, are subject to thousands of factors that contribute to variability. Some of the major sources of variability in the milking process are:

1. The operator. It is well known that differences in milking technique and human/animal interaction have a large influence on the milking process (Seabrook, 1994).
2. Environment. Differences in temperature, humidity and other environmental factors both in the housing and milking area can contribute to variability in animal behavior and outcomes of the milking process.
3. Feeding. Changes in feed composition, palatability, and availability will cause changes in milk yield, which will result in changes in the milking process.
4. Cows. There are large differences between cows in their milking characteristics as well as day-to-day variation within cows (Palmer and Jensen, 1994) and changes that occur within cows during the course of a lactation.
5. Milking Machine. Last but not least are changes in the milking machine, which are our main interest in this paper. Changes in the milking machine may occur by intention (adjust regulator vacuum level), by accident (pulsator failure), or may be part of the normal process or wear (aging liners). Our goal is to separate the effects of these changes from other sources of variability.

Statistical analysis can be a powerful tool to distinguish between real effect and random occurrence but all statistical analysis is built upon a set of basic assumptions. If these assumptions are violated the subsequent statistical analysis will have little value. It is important to understand the following concepts to design a successful experiment to determine cause and effect.

Comparison: A statistical test is often a comparison of one set of data to another. It is important to have a clear idea of what the desired comparison is. The causal variables (e.

g., liner A compared to liner B) and the effect (e.g., milking time) must be clearly identified. The size of a meaningful effect must also be determined in order to plan an experiment (e.g., mean change in milking time of 15 seconds). This information is used to properly frame the question (e.g., Does liner B milk cows on average more than 15 seconds faster than liner A?). In this example liner A is being compared to liner B using milking time and as the measure of performance. We would like to conclude that the difference in mean milking times is 15 seconds or more.

Variability: The variability in the effect variable (e.g. milking time) will determine the precision of measuring that outcome. Predictive ability can be improved by either reducing variability or increasing sample size. On farms, the maximum sample size is fixed (number of milking cows). Other practical considerations (the way cows are grouped during milking) may reduce the maximum sample size further. There are a number of ways that variability can be controlled.

Random Assignment: It is important to randomize cows to treatments to avoid unknown or unintended differences between treatment groups. Randomization of cows to treatment can be difficult because of the logistical problems of efficient cow movement. To address this variability, an investigator should block on known sources of variability and randomize on what is not known. When we randomize cows to treatments, we distribute chance bias as chance error.

False Positive and False Negative Error: The type of tests discussed here are means tests, in which we will be making claims about the difference in the average of one set of observations compared with the average of another set of observations. Two types of errors can occur when making these claims. A false positive (Type I) error is the claim that the two population means are different when they are in fact the same. A false negative (Type II) error is the claim that the two population means are the same when in fact, they are different. The probability of a false positive is denoted by α and probability of a false negative is denoted by β . Commonly accepted levels of error in scientific literature are a false positive probability of 0.05 and a false negative probability of 0.20. Statistical power is defined as the probability of finding a given difference if the difference exists and is expressed as $(1-\beta)$.

OBJECTIVES

The objective of this study is to develop recommendations for field tests of milking machine performance. Issues of measurement technique, experimental design and statistical analysis will be addressed.

METHODS

Paired and Independent T-Tests

Two useful statistical tests for field studies are the paired and independent T-tests. In a paired T-test, an individual cow's response to one treatment is compared to that same cow's response to the other treatment. This controls for the considerable cow to cow variation that occurs for any indicator of milking performance. A paired T-test has greater resolution ability, for a fixed sample size, than an independent test (LeMire et al, 1998) because the variability of the response is reduced if there is a positive correlation between cows for the response of interest—this is usually the case for milking responses. This can be a significant advantage if the sample size is limited because of time, logistics or the total pool of cows. A paired T-test test of milking performance will almost always involve testing during two milkings.

A powerful method to account for variability is to perform a paired T-test using a switch back design. An example of a switch back design is as follows: half the cows get treatment A and the other half get treatment B for one milking. At the next milking the treatments are switched (the cow that got A before now get B and the cows that got B before now get A). Cows should be randomized to the treatments for the first milking and get the opposite treatment the next day. Unequal sample sizes are possible, but the tests are most efficient with equal size groups. This approach accounts for the variability both between cows and between days.

If large cow samples are available, an independent test may have advantages because it could be performed in the course of a single milking (e.g. 200 cows are milked with treatment A and another 200 cows are milked at the same milking with treatment B). An independent test assumes all the observations in a sample are independent. This is a reasonable assumption for most milking machine responses if cows are randomly assigned to treatment groups.

Other types of tests such as factorials and Latin squares (E.Engalke, 1996) can be used for simultaneous comparison of multiple variables. The logistical complexity of these tests can easily overwhelm the field tester. Another advantage of T-tests is that the statistical analysis can easily be done with commonly available software programs (Excel, Lotus 123, Quattro Pro, etc.). This paper will, therefore, focus on application of paired and independent T-Tests.

Distributional Assumptions

We use four distributions in this testing: normal, t, exponential, and binomial. We assume that average flow and peak flow, milk weights are normally distributed and do the tests with a t distribution. Strip yields from a properly functioning milking machine appear to be exponentially distributed. We use the binomial distribution for teat condition and claw adjustments. It is important to know how your response is distributed for you to perform the correct test.

Sample Size

An estimate of sample size will help to determine if:

- The experiment is not likely to produce useful results because the required sample size is larger than the available sample,
- Selection of a subset of the available sample will be sufficient and the job of data collection can be made easier.

Equation 1 is used to estimate sample size for a paired T-test.

$$n_{\text{cows}} = \frac{(Z_{\alpha/2} + Z_{\beta})^2 S_{\text{diff}}^2}{\Delta^2} \quad \text{Sample Size for Paired tests [Lyman]}$$

Where: n_{cows} : Number of Cows (Two observations per cow)

$Z_{\alpha/2}$: Z value for Two Tailed false positive probability $\alpha/2$

Z_{β} : Z value for false negative probability β

S_{diff} : Standard deviation of the differences

Δ : Difference in population means

Equation 1. Sample Size Calculation for means test.

In order to estimate the sample size (number of cows) for an independent test, we need to know the mean effect size (Δ , or the difference in means), the standard deviation of the differences and Z values for false positive and false negative probabilities. Z is the standard normal variate. The effect size must be specified by the investigator as the smallest meaningful mean difference. The value for a type I error probability of 0.05 is $Z_{0.05/2} = 1.96$. The value for a type II error probability of 0.20 is $Z_{\beta} = 0.84$. The sample size estimates for both normal and binomial distribution presented here use 0.05 for type I error and 0.20 for type II error. Tables of Z values for other error probabilities are readily available in statistical texts and software packages.

The standard deviation of the data must be determined from a sample of data. This will require 'pre-experiment' data collection without application of any treatments. For example, average milk flow rate would be recorded for two milkings from the available cows. The standard deviation of the difference in average flow rate between these two milkings for each individual cow would be calculated. The standard deviation can be calculated using common spreadsheet software or a calculator with statistical functions.

Equation 1 can also be used to estimate sample size for an independent test by replacing the standard deviation of differences (S_{diff}) with the a pooled standard deviation (S_{pooled}) multiplied by 2. The number of cows also becomes the number of cows in each 'independent' group.

Another way to express these relationships is to calculate the size of the effect that could be found given a fixed number of cows (Equation 2 for a paired T-test test and Equation 3 for an independent test).

$$\Delta_{paired} = \sqrt{\frac{7.84 S_{diff}^2}{n_{cows}}} \qquad \Delta_{independent} = \sqrt{\frac{15.68 S_{pooled}^2}{n_{cows \text{ per group}}}}$$

Equation 2. Effect Size for paired means test.

Equation 3. Effect Size for independent means test.

If the effect size that can be detected is larger than the desired lower limit for a meaningful difference, one should reconsider running the test. This is because the test would be under powered and we would have little ability to conclude that the two sample means are not different.

Blocking and Randomization

One method to control variability is to select a subset of cows (blocking) based on known sources of variability and randomize on unknown sources of variability. This can reduce the required sample size (and work involved collecting data) or may increase the ability to detect a smaller difference because the variability is reduced. A possible negative effect of blocking is that the ability to draw inference to the entire herd is reduced.

Randomization of cows to the treatments can be difficult because of the logistical problems of cow sorting and movement. This is less of a problem in stanchion barns as cows are released or milked individually, allowing the investigator to more easily apply different treatments to different cows. In a free stall barn cows are usually milked in groups and may need to be sorted into pens prior to running the experiment. Without randomization, the investigator will not know if there was some underlying bias in the selection of cows to treatments. Because of this, the investigator is likely to waste time pursuing potentially flawed results. If you are going to take the time to run on farm tests, take the time to randomize the selected cows to the treatments.

Often what is considered random selection is really haphazard selection. For example, in a barn with two pens of cows, one way to select the cows would be to give treatment A to the cows in pen 1 and treatment B to the cows in pen 2. This type of selection is haphazard and could lead to bias and unreliable results.

A simple way to randomize cows to treatments is to take a printout of the available cows, cut the numbers on the paper into even sizes, mix the numbers and blindly pick an equal number for the two groups. The two treatments are then applied to these two groups of cows. This may require that the two groups of cows be sorted prior to the experiment. Without randomization we have less confidence that any differences was due to the treatment and not to an unknown selection bias.

Paired T-tests are best done as switch back experiments. If one treatment is applied at one milking and the other treatment at the next, any unintended change due to daily differences would show up as a difference that could not be separated from the effect of the treatments. In a switch back design, the unintended daily effects is likely to be the same for each group of cows and would not bias the results—assuming random cow assignment to treatment groups.

RESULTS AND DISCUSSION

Examples of the implications of experimental design and blocking are presented for various milking performance indicators for a sample of 48 cows from the UW milking herd. The same procedures apply for larger or smaller sample sizes. It is generally to the advantage of the field tester to use as few cows as possible and spend more time on randomization and obtaining reliable data.

Average Flow Rate

An example of a distribution of average flow for 48 cows is shown in Figure 1. The average flow here is the total milk in pounds over the total milking time in minutes. If cows are randomly assigned to treatment groups, is should be safe to assume that average flow is normally distributed.

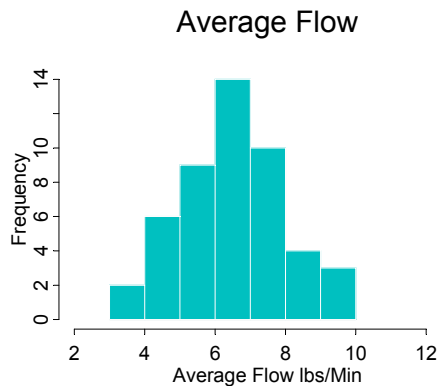


Figure 1. Example of average low in pounds per minute.

The detectable effect size for different blocking scenarios for an independent T-test are given in Table 1. If all 48 cows are used the detectable effect size is a difference of 1.19 lbs./min. If cows are blocked and only those with an average milk flow rate between 5 and 7 lbs./min are used the detectable effect size is reduced to 0.73 lbs./min. Note that blocking reduces both the standard deviation and the required sample size.

Changes in the milking machine may affect the cows differently depending on their average flow rate. Blocking on cows with a flow rate between 8 and 10 lbs./min results in a detectable difference of 0.87 lbs./min. The investigator must make a judgement of the specifics of the treatment and its influence on different cows to choose the best

blocking strategy. If the sample size were increased to 50 cows per group (with a standard deviation of 0.58 lbs./min) the detectable difference would be reduced to 0.33 lbs./min.

Table 1. Estimate of detectable difference in average flow rate with different blocking strategies for a two-tailed independent T-test ($\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (lbs./min)	Number of Cows	Percent of Cows	Cows per group	Detectable effect Size (lbs./min)
3 to 10 lbs./min	1.47	48	100	24	1.19
4 to 8 lbs./min	1.00	39	81	19	0.91
5 to 7 lbs./min	0.58	20	42	10	0.73
8 to 10 lbs./min	0.62	17	35	8	0.87

The detectable effect size for different blocking scenarios for a paired T-test are given in Table 2. By pairing cows the detectable differences are reduced resulting in a more powerful test. The most efficient test for average milk flow rate, would be a paired T-test with cows blocked between 4 and 8 lbs./min. Increasing the sample size to 100 cows group (with a standard deviation of the differences of 0.71 lbs./min) the detectable difference is further reduced to 0.20 lbs./min.

Table 2. Estimate of detectable difference in average flow rate with different blocking strategies for a two-tailed paired T-test ($\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (lbs./min)	Total Cows	Percent of Cows	Detectable Effect Size (lbs./min)
3 to 10 lbs./min	0.90	48	100	0.36
4 to 8 lbs./min	0.71	39	81	0.32
5 to 7 lbs./min	0.70	20	42	0.44
8 to 10 lbs./min	0.74	17	35	0.50

Milk Yield

An example of a milk production distribution is shown in Figure 2. If cows are randomly assigned to treatment groups, it should be safe to assume that milk weight is normally distributed.

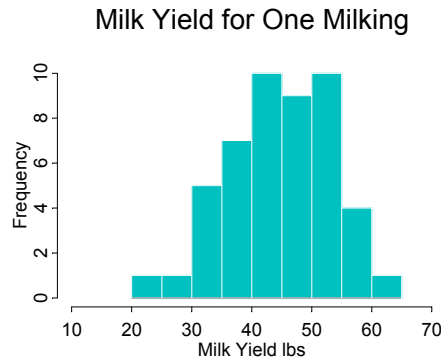


Figure 2. Example of milk yield in pounds.

The milking machine should remove milk from the cow as efficiently as possible. Short-term changes in milk yield caused by the milking machine are likely to be inversely correlated with strip yield. Detectable differences in milk yield for an independent T-test are given in Table 3. Blocking again reduced the sample size and resulted in a smaller detectable difference.

Table 3. Estimate of detectable difference in milk yield with different blocking strategies for a two-tailed independent T-test ($\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (lbs.)	Total Cows	Percent of Cows	Cows per Group	Detectable Effect Size (lbs.)
70 to 20 lbs.	8.7	48	100	24	7.0
60 to 30 lbs.	7.5	45	94	22	6.3
50 to 40 lbs.	3.0	19	40	9	4.0
50 to 70 lbs.	3.0	15	31	7	4.5

The results for a paired T-test test of milk yield are given in Table 4. In this case blocking did not improve the detectable difference. The investigator has more freedom to select cows based on logistical concerns and barn layout for paired switchback tests.

Table 4. Estimate of detectable difference in milk yield with different blocking strategies for a two-tailed paired T-test ($\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (lbs.)	Number of Cows	Percent of Cows	Detectable Effect Size (lbs.)
70 to 20 lbs.	4.5	48	100	1.8
60 to 30 lbs.	4.5	45	94	1.9
50 to 40 lbs.	4.8	19	40	3.1
50 to 70 lbs.	4.4	15	31	3.2

Peak Flow

An example of a peak flow rate distribution is shown in Figure 3 and detectable differences in Table 5 for an independent test and Table 6 for a paired T-test. Peak flow rate here is taken as the maximum 30 second average in milk flow rate. If cows are randomly assigned to treatment groups, it should be safe to assume that peak flow is Normally distributed. For this measure, the most sensitive test would be to use all animals in a paired switch back test.

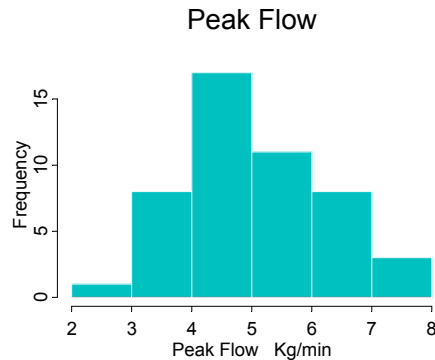


Figure 3. Example of peak flow in Kg/min.

Table 5. Estimate of detectable difference for peak flow in Kg/min and blocking (two-tailed independent test, $\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (Kg/min)	Total Cows	Percent of Cows	Cows per group	Detectable Effect Size (Kg/min)
1 to 9 Kg/min	1.1	48	100	24	0.89
3 to 7 Kg/min	0.94	44	92	22	0.79
4 to 6 Kg/min	0.51	28	58	14	0.54
5.5 to 6.5 Kg/min	0.29	11	23	5	0.51

Table 6. Estimate of detectable difference for peak flow in Kg/min and blocking (two-tailed paired T-test, $\alpha=0.05$, $\beta=0.20$).

Blocking	Standard Deviation (Kg/min)	Total Cows	Percent Cows	Detectable Effect Size (Kg/min)
1 to 9 Kg/min	0.75	48	100	0.30
3 to 7 Kg/min	0.76	44	92	0.32
4 to 6 Kg/min	0.68	28	58	0.36
5.5 to 6.5 Kg/min	0.83	11	23	0.83

Strip Yield

Strip yield can be a measure of over or under milking. Factors known to affect strip yield are detacher settings and liner properties. Field tests of strip yield can be a time consuming process. A test to reduce the number of cows that need to be measured would be an estimate of the percentage number of cows with more than a specified amount of residual milk.

An example of 48 strip yields is shown in Figure 4. The mean of these strip yields is 61 ml of milk. This distribution was obtained from cows milked with a properly functioning milking machine. As we can see this distribution is not normally distributed. Because of this, a T-test would not be appropriate here. It appears that strip yields from a properly function milking machine can be modeled well by an exponential distribution. As over-milking occurs we would expect to see an even sharper slope in the curve. As under-milking occurs, we would expect to see a less steep drop off in the curve. As under milking increased, the curve would eventually move toward a normal distribution.

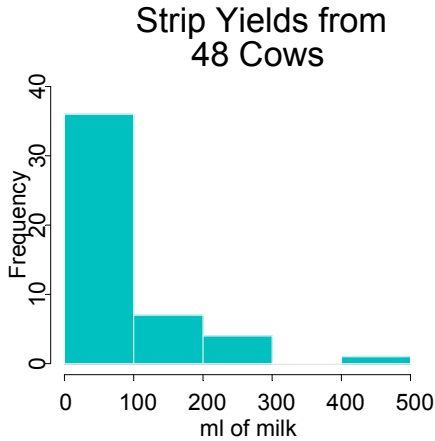


Figure 4 Example strip yield from a properly functioning milking machine.

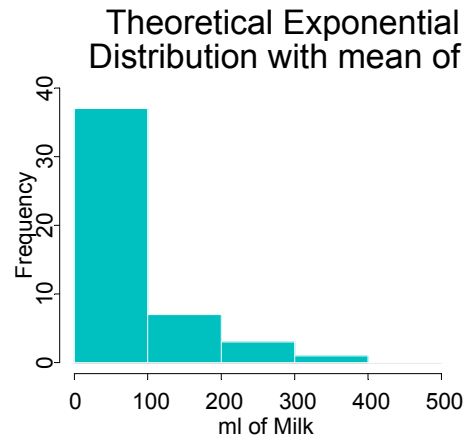


Figure 5 Example theoretical exponential distribution with mean of 61 ml.

If we model the distribution of strip yield as an exponential distribution as shown in

Figure 5: $f(x, \beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ where β is the mean strip, we can estimate the expected

percentage of cows with more than a specified strip yield. For example, if we have an exponential distribution with an average strip yield of 61 ml we would expect about 7 percent of cows to have a strip yield greater than 160 ml strip yield for all four teats. If we find more than 7 percent of cows above this value we can conclude that the mean strip yield of the group is more than 61 ml. An example of how this information could be used is presented in Table 7 using a Binomial test. The probabilities have been rounded up to the nearest cow which errors on the side of a conservative test. As sample sizes get larger, the effect of this rounding will diminish.

Table 7. For a mean strip yield of 61 ml of an effective milking machine (One tailed alpha = 0.05, Beta = 0.20).

Number of Cows Stripped	Expected number of cows over 160 ml	If X number over 160 ml, conclude that mean not 61 ml
10	1	3
20	2	5
30	3	6

Teat Condition

There are several ways to assess the condition of teats. Short-term changes can be assessed by measures such as teat color (normal, red, blue) immediately after milking (Hilerton et al, 1998). Longer-term changes can be assessed using teat end roughness scores. A 4-point scale has been proposed (Britt and Farnsworth, 1996). Teat condition scores are generally not normally distributed and tend to have a skewed distribution. A useful criteria for assessing the teat condition of a group of cows is to measure the number of cows that have a teat condition considered problematic. For example, more than 20 percent of cows with 'rough' teat ends (a score of 3 or 4, using the method proposed by Britt and Farnsworth), or more than 20 percent of cows with red or blue teat color after milking.

A binomial test similar to that presented above for strip yield can be applied to determine if the herd has 20% of the teats in the 3 and 4 range (poor teats). The results of such an analysis are given in Table 8.

Table 8 Cows with 'poor' teat condition (One tailed alpha = 0.05, Beta = 0.20).

Number of Cows scored	Expected Number of with a 'poor' score	If more than this number are found we can conclude that the average of the population is over 20%.
10	2	6
20	4	9
30	6	12
50	10	17
100	20	30

Another way to assess teat condition is to estimate the change in percentage of cows with a 'poor' score when a treatment is changed. The number of cows required to detect a change in the percentage of cows with poor teat condition is shown in Table 9. As we can see, a large number of cows must be sorted to detect small differences in this percentage.

Table 9 . Sample sizes needed to tell a change in teats that score 3 or 4 from 20%(One tailed, alpha = 0.05, beta = 0.20).

From to	Sample Size Per Group
20% to 25%	902
20% to 30%	251
20% to 35%	122
20% to 40%	74
20% to 45%	51
20% to 50%	37

Unit Adjustments

If more than 5-10 slips or falls occur per 100 milkings, a problem may exist that requires correction (Mein and Haman, 1993). Table 10 shows the sample size required to conclude that there was more than 5% slips or falls.

Table 10. Unit Adjustments (One tailed, Alpha = 0.05, Beta = 0.20).

Number of Cows Look At	Expected Number of slips and falls at 5%	Conclude not 5% if over this many slips or falls observed
50	3	7
100	5	11
150	8	15
200	10	18
250	13	21

Another way to assess slips and falls is to estimate the change in percentage of cows when a treatment is changed. The number of cows required to detect a change in the percentage of cows with slips and falls from 5 percent is shown in Table 11.

Table 11. Sample sizes needed to tell a percent change in percent of unit adjustments. (One tailed, Alpha = 0.05, Beta = 0.20)

From to	Sample Size Per Group
5% to 7.5%	1239
5% to 10%	383
5% to 15%	131
5% to 20%	73
5% to 25%	49

CONCLUSIONS

The predictive ability and labor required for field tests of milking machine performance can be improved considerably by proper experimental design and statistical analysis. The recommended procedure for performing a field test is to:

1. Formulate a specific question. This includes identification of the causal variable, the performance measure and the desired detectable mean difference in the performance measure.
2. Estimate the detectable effect size. This includes choosing the type of test (independent, paired T-test or other), identifying the available sample size and estimating the variability (standard deviation) of the performance measure. An assessment of the advantages of blocking to reduce variability can significantly reduce the amount of data that must be collected. Block on known sources of variation and randomize on unknown sources of variation.
3. Collect data in the most accurate manner possible.
4. Perform appropriate statistical analysis to draw a conclusion that the change in the milking machine either had an effect or that the effect was smaller than the pre-defined detectable difference.

REFERENCES

- Britt, J.S. And R. Farnsworth. 1996. A System For Evaluating Teat Anatomy, Skin Condition And Teat Ends. Proc. 35th Annual Meeting, National Mastitis Council, Nashville, TN.
- Engelke, E. , G. Wehowsky, F. Troger, and O. Lindt. 1996. Methodology of Comparing Three Milking Clusters in Large Dairy Herd. *J. Dairy Sci.* 79:384-389.
- Engelke, E. , G. Wehowsky, F. Troger, And O. Lind. 1996. Methodology Of Comparing Three Milking Clusters In Large Dairy Herd. *J. Dairy Sci.* 79:384-389.
- Hillerton, J.E., I. Ohnstad, And J. R. Baines, 1998. Relation Of Cluster Performance To Post Milking Teat Condition. Proc National Mastitis Annual Meeting, 1998, Pp. 75-84.
- Lemire, S.D., D.J. Reinemann, G.A. Mein, M.D. Rasmussen, Statistical Considerations For Milking Time Tests. Paper No. 983128, Written For Presentation At The 1998 ASAE Annual International Meeting, Orlando Florida July 12-15, 1998.
- Mein, G.A., And J. Hamann, 1995. Dynamic Tests For Teacup Action And For The Cluster. Proc. IDF 3rd International Mastitis Seminar, Tel Aviv Israel. Pp. S-7,24-29
- Mein. G.A. And D.A. Ried, 1996. Milking-Time Test And Guidelines For Milking Units Proc. 35th Annual Meeting Of The National Mastitis Council, Nashville Tennessee, 1996. Pp. 235-244.
- Palmer, R. W., And E. L. Jensen, 1994. Removal Of Within-Cow Differences Between Morning And Evening Milk Yields. 1994, *J. Dairy Sci.* 77:2663-2670.
- Seabrook, M. F., 1994. Psychological Interactions Between The Milker And The Dairy Cow. Proc. Third International Dairy Housing Conference: Dairy Systems For The 21st. Century. American Society Of Agricultural Engineers. St. Joseph MI, USA, pp. 49-58.