Using Residue Tests in Support of a Quality Milk Program

William M. Sischo, *DVM, MPVM, PhD*
Department of Veterinary Science
Pennsylvania State University
University Park, PA 16802

Abstract

The certification process for antibiotic residue tests for raw, comingled bovine milk has been a long and arduous process. In terms of protecting the consumer from antibiotic tainted milk perhaps the process developed by FDA and administered by AOAC International can be deemed a success. Unfortunately for the producer and veterinarian the process has left some problems. First, the tests have not been evaluated in a population context. The number of samples required for certification was small and unlikely to represent the range of bulk or individual milk quality that will be experienced in the field. The quality of the estimates of the population test parameters resulting from the certification is questionable meaning that we will be conducting an uncontrolled field trial when these tests come on line in January 1995. Second, the certification process which focussed on developing convenient tests to be used in the rapid screening of milk by processors allowed tests to be certified that will detect some antibiotics below regulatory tolerance levels and in some cases above tolerance levels. Although there will be some accounting for these discrepancies in the labels for these tests, the presence and use of these tests suggests that producers and veterinarians will be facing the very real probability that legal milk will be dumped and producers penalized for the test's mistake. Third, some AOAC certified tests (approved for bulk milk) will be marketed as farm and cow tests. Although the labels for these tests will explicitly describe their approval for bulk milk only, the implicit message is that the test can be used appropriately for individual animal milk. There will be no data to support use of these test on individual animals and it will be necessary to subject these tests to small sample size protocols to develop some expertise in using these tests on the farm. Additionally we will need to develop an epidemiologic approach in order to effectively interpret test results for the dairy producer. Finally, because the breadth of the testing program will be increasing (the present official test, the *Bacillus stearothermophilus* disk assay detected only a portion of the 6 beta lactams targeted in the new program) there will be an increase in the number of violations detected beginning January 1995 with no reason other than increased ability to detect the antibiotics that had gone undetected.

We will have to live with these tests and we should make efforts to understand how these tests can be best used. The goal of the dairy industry needs to be towards the continuing production of nutritious, good tasting, and safe milk. Although the tests as they stand today are fraught with problems they should be utilized as necessary as part of a farm Total Quality Management program and within the context of the MDBQAP. They are not intended nor should they ever be used to define a quality product. Milk quality in all its dimensions begins with an on-farm program to promote animal and particularly udder health. When disease does occur alternatives to antibiotics should be employed when possible and if antibiotics become necessary then they should be used in a rational manner. This approach should include on-farm training in handling and administering the drugs, developing treatment protocols to provide guidelines for antibiotic use, and importantly having record keeping and identification systems that track antibiotic use and can be used by every member of the dairy management team to know which animals have been treated. The final link in the system will be the screening kits which will be appropriately used to verify that treated cows for which milk has been held from sale according to the label guidelines is negative for antibiotics and the milk is saleable from that aspect.

Introduction

One goal of a dairy production Total Quality Management program is to prevent the occurrence of antibiotics in raw milk leaving the farm. One effective approach to meet this goal is to implement Hazard Analysis Critical Control Point (HACCP) procedures.[1] The essence of a HACCP program is to identify the "critical control points" to prevent the occurrence of a hazard, develop management procedures in accordance with the critical control points, monitor the HACCP program and make any necessary changes to management procedures in response to the outcome of the monitoring. A HACCP program for antibiotic avoidance (Milk and Dairy Beef Quality Assurance Program--MDBQAP) has been developed jointly by the American Veterinary Medical Association and the National Milk Producers Federation.[2] The program (outlined in a producer manual) defines 10 critical control points for preventing antibiotic residues. Although 9 of the 10 points focus primarily on disease prevention and on-farm management of antibiotics and treated cows, one point is: "Use drug residue screening tests."

Although it is clear in the producer manual that milk from individual animals as well as bulk milk should be tested to ensure antibiotic-free milk is leaving the farm, it is less clear whether we have any tests that can be reliably used on individual animals let alone on bulk tank milk. This problem is mentioned in the 1994 producer manual[a]: "Much discussion has occurred in the

[a]*Milk and Dairy Beef Residue Prevention Protocol, 1994 Producer Manual, published by Agri-Education Inc. Stratford IA.*

last two years over the validity of tests designed for milk. A central element in this discussion is whether most tests, which are designed to detect residues in bulk tank milk samples, can be used to detect residues in individual cow samples." The objective of this paper is to review the tests currently on the market and provide some epidemiologic guidelines for their use to support an on-farm HACCP program for residue avoidance.

Historical Perspectives

The need for rapid tests to detect antibiotic residues in milk first surfaced in November 1990. The Government Accounting Office prepared a report to the House Committee on Government Operations implying that the Food and Drug Administration (FDA) did not have access to appropriate technology to ensure the nation's milk supply free of antibiotics.[3] One of the suggestions in the report was that FDA should, "prioritize and expedite its current efforts to develop and evaluate new screening and confirmatory test methods for animal drug residues in milk....". This report not only stimulated FDA to consider mechanisms for certifying tests, the report also helped define the need and market for rapid screening tests.

In 1991 the MDBQAP was established as the core of an industry-sponsored national dairy residue avoidance program. As mentioned previously the program recommends on-farm screening of milk from cows and bulk tanks as a critical control point. In 1992 the Pasteurized Milk Ordinance was changed to require all tankers to be tested for beta lactam drugs as they entered milk plants. Although screening tests had been available prior to these events, the GAO report coupled with the changes in PMO and the interest in on-farm testing stimulated the development of rapid screening tests by private sector. By the end of 1992, rapid screening tests were widely available to both milk receivers, veterinarians, and dairy producers, but only one of these tests actually had been evaluated by FDA and approved as an official test for detecting beta lactam antibiotics bulk milk. Few of the tests had any public data on their performance either *in-vitro* or in field trials. In 1992 and 1993 several field evaluations of some of the screening tests were conducted. The results suggested that the tests were prone to false positives not only in cows with clinical mastitis but also in clinically normal cows.[4-8]

The Need for Effective Evaluation of the Residue Screening Tests

The last three years have been tumultuous for the manufacturers of tests. The major criticism directed at the screening tests was the lack of rigor applied to their evaluation and the unsupported recommendations by the manufacturers for their broad use. Specifically criticisms were that the tests (whether they were intended for use in quality control laboratories or on the farm) were evaluated on spiked milk samples and not publicly evaluated on field samples. The manufacturer's own evaluations focussed on a test's ability to detect very low antibiotic concentrations even below accepted safe levels of antibiotics, but there was little or no emphasis in the evaluations to assess cross-reactivity of the tests to natural inhibitory substances in the milk or even milk constituents. In addition, there was little or no regard for population or epidemiologic assessment of the tests.

Despite the limited scope of the evaluations, several manufacturers still suggested that the tests could be used to screen for antibiotics in co-mingled milk from silos, hauling tankers, and bulk tanks as well as milk from cows and individual quarters. Undoubtedly the need for testing was dictated by public and regulatory pressure and the dairy industry cannot be blamed for utilizing the available tools, but the inadequate evaluation of the tests coupled with poor epidemiologic interpretation of the test results resulted in poor decisions concerning milk disposition from the farm up to the corporate level. As a result, good milk was discarded and some producers unnecessarily penalized.

The real concern over the use of these tests was raised because of the need to evaluate the status of milk of cows on the farm. Herd level evaluations of some of the early screening tests had published in the 1980's indicated that the tests used on individual cow's milk were prone to reporting false positives,[9,10] but it was work by Dr. Jim Cullor at the University of California that brought the industry's attention to serious problems associated with the beta lactam tests.[11] Cullor's research raised two concerns, first that the dairy industry had mistakenly relied on industry claim's for the reliability of the tests and second were asking producers through the MDBQAP to use tests in situations they were not designed for, to test bulk tanks and individual cow's milk. The high rates of pretreatment false positives in Cullor's study in cows with clinical mastitis questioned whether the tests had any role for on-farm screening. Subsequent studies verified and extended Cullor's findings for beta lactam screening tests to non-clinical cows and highlighted a problem of false positive for screening tests for tetracyclines and sulfa drugs.[5-8]

Approximately the same time as Cullor's initial report, the GAO submitted another report to Congress suggesting that the potential for drug residues in milk far exceeded FDA's detection abilities.[12] The report also recommended that FDA be more active in moving tests onto the market. Following these events, FDA developed specifications for certification of tests to be used for screening raw, comingled bovine milk for 6 beta-lactam antibiotics.[13] FDA also established a working

relationship with AOAC International to supervise, conduct independent evaluations and eventually certify the tests for use in screening milk.

The Certification Process

The FDA specifications for screening tests focussed on four areas: First, the specifications named the beta lactams antibiotics and the tolerance levels (ppb) that were to be detected by the screening tests (Penicillin G, 5ppb; Ceftiofur, 50ppb; Cloxacillin, 10ppb; Cephapirin, 20ppb; Ampicillin, 10ppb; and Amoxicillin, 10ppb). Second, the estimated sensitivity of the tests (the proportion of positive samples detected by the screening test) were to be such that the lower limit of the 95% confidence limit not be below 90%, and the required sensitivity would be attained for at least one of the 6 designated beta lactams. If a test had the ability to detect antibiotic concentrations below the "tolerance" levels with a 90% sensitivity the manufacturer is required to state this on the label. All evaluations were conducted on a minimum of 30 "fortified" (spiked) milk samples. The third area of evaluation for the tests was "selectivity."[b] The assessment and goals for selectivity mirrored that used for sensitivity, and was determined using 30 zero control samples. The fourth evaluation area determined dose-response curves for the tests. Essentially this was an evaluation of the screening test's limits of detection, i.e. how likely was the test to call a sample positive when the antibiotic concentration in the sample was below tolerance levels. A perfect test would be negative for all samples below tolerance concentrations. This evaluation was conducted on 6 samples of zero control milk fortified with antibiotics over a range of zero to tolerance levels at four evenly spaced doses across that range.

Additional experiments were also performed to assess "ruggedness of the tests", i.e. performance of the tests on frozen samples, cross reactivity of the tests to other antibiotics, interference from bacteria and somatic cells, and stability and lot-to-lot consistency. All of these experiments were conducted by the manufacturer and data submitted to AOAC for evaluation.

Another key component of the AOAC test protocol was evaluation of the screening tests by a contracted, independent laboratory. The results of the evaluations submitted by the kit manufacturers were verified by the contract laboratory using the FDA-AOAC protocols. If a screening test met all aspects of the AOAC evaluation it would be certified as "Performance Tested", the stamp of approval from AOAC and FDA-Center for Veterinary Medicine. Screening tests that did not meet all the evaluation criteria could still be recognized as valid and certified as "Performance Tested" but must acknowledge their limitations on the package label.

The Evaluations Process and the Results

The evaluation process began in January of 1992 with the formation of the AOAC Research Institute. The institute's primary duty was to administer the test kit confirmation program. The initial period for accepting applications was February 2 through April 15, 1992. In March, the start date for the acceptance of applications to the certification process was changed to begin on April 1. In May of 1992 the application process was temporarily suspended following Cullor's report[11] and after concerns were voiced by the manufacturers and others about the certification program. In May, a meeting was held with the manufacturers and others to discuss refinements to the data submission requirements. In October 1992, a "Memorandum of Understanding" was signed between FDA and AOAC. This memorandum officially recognized the AOAC testing program as an appropriate evaluation of beta lactam tests to be used in state milk monitoring programs. At the time of the signing AOAC re-opened the application period for test kits. Manufacturers were notified that they would have until January 1993 to submit data for AOAC consideration. In January, AOAC reported that applications for 12 test kits were received, 3 were subsequently withdrawn after preliminary review which left nine kits to advance to the independent testing and review phase. In March, the AOAC announced plans to institute a program to evaluate "cowside" tests. In May of 1993, initial results from the testing were distributed to manufacturers and after comments and questions the tests were scheduled for an additional round of testing. Additional delays in the certification pushed reporting evaluation results into October of 1993. The first results of evaluation were reported for eight tests with two additional tests reported as certified. By August 1994, 15 tests had been certified as "Performance Tested" (Table 1) and states had begun programs to train personnel in the use of these official tests. Also in August, AOAC announced a certification program for screening tests to evaluate raw, comingled, bovine milk for tetracyclines and sulfa antibiotics. FDA-CVM also began a limited evaluation of beta-lactam screening tests for use on-farm.

From the long history of the evaluations it is clear that the process has been difficult with the by-product that the intent of the program has been compromised. The intent of the program was to certify rapid tests for regulatory use. The realization of the certification process is that we now have tests that detect some antibi-

[b]*Selectivity was defined as the proportion of "truly negative samples that are found by the assay to be negative". The term selectivity seems to have been coined for the test kit evaluations and has the same meaning as specificity. Both sensitivity and specificity are population-based assessments of tests and not laboratory assessments. It would seem that the criticism directed at the first tests has resulted in new jargon and not better evaluations to credibly deal with one of the original criticisms--lack of population-based assessments of the tests.*

otics below tolerance levels and other antibiotics above tolerance levels. With the anticipated regulatory use of these tests, some tankers with below action levels of beta-lactam antibiotics will be regarded as shipping milk with illegal levels of antibiotics--surely an awkward position for the milk receivers and those people working in the states charged with enforcing the PMO. Another intent of the certification process was to foster commercial development of screening tests that could be used interchangeably in a variety of testing situations. The reality of certification is that we now have tests that detect antibiotics at different levels from each other. Results of screening will vary from site-to-site setting up the potential that negative milk shipped from one state could be found positive in another state and rejected as contaminated. This concept is inimical to the intent of the NCIMS and Pasteurized Milk Ordinance which is directed at standardizing the regulations for milk shipments between states.

Table 1. Summary of evaluations of beta lactam screening assays for use under Appendix N of the Pasteurized Milk Ordinance. Evaluations were supervised by AOAC International in accordance to guidelines developed by FDA, AOAC, NCIMS, NMPF kit manufacturers, and AABP. Data presented in the table were first published as FDA memorandum M-a-85 (July 22, 1994).

Test Name	Company	DRUG					
		Pen G (5ppb)	Ampicillin (10ppb)	Amoxicillin (10ppb)	Cloxacillin (10ppb)	Cephapirin (20ppb)	Ceftiofur (50pb)
Charm II Tablet Competitive Assay	Charm	4.8	9	10	**70**	**4.5**	**25**
Charm Farm	Charm	5	10	10	**40**	20	**25**
Charm II Tablet Sequential Assay	Charm	4.8	8	10	**50**	**4.5**	**23**
Charm II Tablet Transit Test	Charm	4.8	9	10	**80**	**4.5**	**13**
Charm II Rapid Inhibition Test	Charm	3	**4.5**	**4.5**	**25**	16	50
Charm I- Cowside II Tablet	Charm	4.8	10	10	**50**	**8**	40
Charm II Tablet Quantitative Assay	Charm	4.8	8	10	10	**4.5**	**23**
Charm *B. stearotherm.* Tablet Disk Assay	Charm	5	6.5	10	**48**	11	**75**
Delvo Test P	Gist-Brocade	3	10	8	**30**	**8**	50
Delvo-X-Press	Gist-Brocade	5	10	10	**50**	**10**	**10**
Lactek B-L	IdeTek	5	8	10	8	16	**ND**
Lactek CEF	IdeTek	**ND**	**ND**	**ND**	**ND**	**ND**	50
Penzyme III Test	SKF	5	10	8	**80**	**8**	**80**
Penzyme Milk Test	SKF	5	10	8	**80**	**8**	**80**
SNAP Test	IDEXX	5	10	10	**50**	**8**	50

The intent of the certification also was to certify tests for use on raw, comingled bovine milk, not for milk from a cow or even from a farm's bulk tank. The reality is that tests carrying the AOAC "Performance Tested" standard will be packaged and marketed as farm tests (Table 1). Finally the intent of the certification was to develop a battery of tests that could be used across the country to administer a program to ensure an antibiotic-free milk supply. The reality of the certification is that within the limits of the AOAC's and manufacturer's the tests have been fairly evaluated and validated. The results of these evaluations suggest that within the matrix of "raw, comingled bovine milk" the tests will error on the "safe" side, that is detect levels of some beta-lactam antibiotics below tolerance levels. But since the tests have not been evaluated on a population basis, there is a great deal of uncertainty as to how well the tests will work in the field on bulk milk from a variety of sources and produced under a myriad of management schemes. There is no epidemiologic context for the application of these tests, and come January 1995 we will be subjecting these tests to a large, uncontrolled field trial.

Using Antibiotic Residue Tests on the farm to Produce Quality Milk--"Testing the Tests"

It is clear there is no official mandate for the certified tests to be used on cow or quarter samples. The AOAC validation is clearly inadequate to validate the tests for individual animal use and FDA is yet in the process of drafting guidelines to evaluate the tests for cow samples. This leaves an obvious void in the dairy industry's efforts to implement the MDBQAP since a critical on-farm monitoring point requires the use of a screening test. Until the void is filled, there are some guidelines that you can use to help an individual herd owner independently assess tests and interpret the results for use on their farm.

First, the National Mastitis Council Research Committee has advocated that antibiotic screening tests should be negative when evaluating untreated cows. This is a minimum requirement to assess test performance. A simple protocol for evaluating this aspect of the tests has been developed by Dr. Jim Cullor.[14] The system he has proposed (Testing the Tests) can be implemented on a farm and follows a sampling scheme similar to the AOAC program. The protocol evaluates premilking, composite milk samples from 30 untreated cows with clinical mastitis but without systemic symptoms.[c] This evaluation establishes a pretreatment, farm-specific baseline for test specificity.

[c]*The Testing-the Test protocol involves sampling 30 cows with clinical mastitis in one quarter. The test sample is a composite sample of 5 mls of premilking gland secretion from each of the four quarters. The screening test to be evaluated should be run against triplicate samples of the quarter composite sample, negative control (undiluted bulk tank milk), and a positive control (5 mls of bulk tank milk combined with 1 ml of beta-lactam antibiotic).*

The next critical areas is assess for the suitability of an on-farm test is its ability to distinguish a treated cow with detectable levels (at established tolerance levels) from a cow with below tolerance levels of antibiotics in the milk. Cullor has applied the Testing the Tests pretreatment protocol to the situation immediately following the final treatment and after the labelled withholding time. There are a few obstacles to implement this protocol which involve interpreting test results since there will be no "gold standard" to identify the true status of a cow's milk. The first obstacle is not a problem inherent in the validation approach but associated with the tests themselves and that is the limits of detection for the tests may not match the legal requirements for shipping milk. This is particularly true for Cephapirin and Ceftiofur which many of the tests detect at levels below tolerance levels and will result in false positives or "non-actionable positives" as one investigator has described them.[15] But many of the tests also detect cloxacillin above the tolerance level and will result in false negatives. While the simple solution to this would be to match the test to the types of antibiotics being used on the farm this ignores the important consideration of matching on-farm testing with testing at the processing plant. The other obstacles to interpreting post treatment results of Testing-the-Tests protocols following treatment include the inherent pharmacology of the antibiotic and associated cow-to-cow variation in drug metabolism and excretion and importantly random error in the outcome due to the small sample size. All things being the same, from evaluation to evaluation results will vary because of random error. Ultimately, without more tools and information the decision of which screening test to use will be difficult at best. At this point an epidemiologic point of view would provide tools to make a more informed decision.

Using Antibiotic Residue Tests on the farm to Produce Quality Milk--The Role of Chance

The design of the Test the Tests evaluation assumes that we are going to work with a fairly uniform population of cows, i.e. the antibiotic type and dose and the animal's rates of metabolism and excretion of the antibiotic are constant. If these assumptions are true then in order to proceed in our evaluation of the tests we will have to know what the estimated population prevalence of antibiotic residues in milk will be for the two sampling periods following treatment. This information should be available from the manufacturer but if it isn't readily available we can make some assumptions that will help our interpretation. It is safe to assume that the prevalence of antibiotic residues at the first milking after the last treatment would be high but variable depending on the antibiotic used. One estimate would be that 50% of the cows still have antibiotics in their milk at the first milkout following their last treatment. In this scenario what role does chance play in our trial for the expected number of cows with antibiotics in their milk?

In our population of cows that have been treated, one-half of the cows will have antibiotics in their milk immediately after their last treatment, but because we are only sampling 30 cows from this population we will have random variation affecting the results of our trial. On average we expect that 15 of the cows will have antibiotics in their milk, but how different from 15 can we expect in our study and still be sampling from our population with 50% prevalence? The outcome of antibiotics present or absent has a binomial distribution, like the outcome of a coin toss or a pregnancy examination. The shape of a binomial distribution depends on the expected prevalence in our population (50% for our first trial) and the size of the trial (30 cows). Table 2 shows the range of values and the probability they would occur for our trial. The most likely values will be 14,15 and 16 cows detected with antibiotics (each observed in 14% of the trials), but the range of possible values extends from 9 to 21 cows. In other words, with a perfect test in our 30 cow trial we might identify as few as 9 and as many as 21 cows and still be sampling from our cow population with a 50% prevalence of antibiotic residues.

Table 2. The number of cows that might be observed with antibiotic residues (using a perfect test) from a series of trials involving 30 cows and a true population prevalence of antibiotic residues of 50%. The binomial sampling distribution is used.

Number of cows positive for antibiotics	Proportion of trials with this number of cows positive for antibiotics	Proportion of trials with more than this number of cows positive for antibiotics
9	0.01	0.98
10	0.03	0.95
11	0.05	0.9
12	0.08	0.82
13	0.11	0.71
14	**0.14**	**0.57**
15	**0.14**	**0.43**
16	**0.14**	**0.29**
17	0.11	0.18
18	0.08	0.1
19	0.05	0.05
20	0.03	0.02
21	0.01	0.01

Another problem that we can address with the binomial distribution is the fact that none of the tests have perfect sensitivity or specificity. From another point of

view, in our cow population there will be a prevalence of cows with false test results, and for our 30 cow trial we will have to expect some variation in what we might see in our results. This fact (the false test result) is not an important problem when we have a prevalence of antibiotic residues of approximately 50%, but it is an important consideration when the prevalence is either high or low. As an example, if our test has a 98% specificity or a 2% probability of producing false positives, then what is the range of test positive animals that would be compatible with an evaluation involving 30 animals with a zero prevalence of antibiotics in their milk? The binomial distribution for a trial involving 30 animals and using 2% as the estimate of the prevalence of false positives is shown in Table 3. In our trial of 30 cows the most probable and expected result will be zero cows detected with antibiotics (55% of the trials), but there is also a strong statistical possibility that at least one cow (33% of the trials) and as many as three cows (2% of the trials) will be detected.

Table 3. The number of cows that might be observed with antibiotic residues from a series of trials involving 30 untreated cows using a screening test with a 98% specificity. The binomial sampling distribution is used.

Number of cows positive for antibiotics	Proportion of trials with this number of cows positive for antibiotics	Proportion of trials with more than this number of cows positive for antibiotics
0	**0.55**	**0.45**
1	**0.33**	**0.12**
2	0.1	0.02
3	0.02	0

Using the Antibiotic Residue Test Kits On-Farm--An Evaluation

Table 4 shows hypothetical results from three different screening tests (all tests have similar limits of detection for the antibiotic we are using on the farm) evaluated according to the Test the Tests protocol.[14] The expected prevalence of cows with antibiotic residues at the first sampling time (pretreatment) is zero. From Table 4 it is apparent that Test A cannot distinguish a cow with clinical mastitis from one that has been treated with antibiotics in this herd and should not be considered a reliable test. Both tests B and C are able to distinguish untreated cows from treated cows.[d]

The prevalence of cows with antibiotic residues in theirmilk at the second sampling (first milking following the final treatment) is expected to be relatively high. In our scenario we will assume a "true" prevalence of approximately 50%. Test B detects fewer cows with antibiotics in their milk (4/30) than we expected given the range of possible values for the binomial distribution (9-21 positive out of 30 cows tested) from Table 2. This is an unacceptable result in that the test is unable to identify a treated cow from an untreated cow. Test C appears to be able to distinguish between treated and untreated cows.

Table 4. Hypothetical results from evaluation of antibiotic screenings tests. Samples are obtained from 30 cows with clinical mastitis prior to treatment, at the milking following the final treatment, and at the first milking following label withholding times. The hypothetical tests evaluated have the same limit of detection for the antibiotic in question.

Test Evaluated	Pretreatment*	Immediate Post Treatment	Following Label Withholding
A	11/30	28/30	20/30
B	0/30	4/30	0/30
C	1/30	13/30	1/30

*Number of samples found positive using the antibiotic screening test out of 30 samples evaluated

In the final sampling time (following the appropriate withholding time) we would expect a zero prevalence of cows with antibiotics in their milk. Test C appears to be able to perform under this scenario. On this farm, the best performing test is C and would be the test of choice.

For these evaluations we are relying heavily on prior knowledge of the behavior of the antibiotic at the population level and the statistical distribution of outcome (probability of identifying cows with and without antibiotics in their milk) in our 30 cow trial. Because of our small sample size and farm-to-farm variation there is a fairly high probability that there will be deviation from the population values. The consequence to this fact is that small scale evaluations such as Testing-the-Tests protocol will be subject to a level of interpretation; but, given the state of the certification process today, the Testing-the-Tests protocol is the most reasonable way to approach the problem of deciding which test to use on the farm. Ultimately, it is in the best interest of the industry that we as a profession demand that any test marketed as a farm or individual animal test be properly evaluated in a population context and not only in single farm trials with relatively few cows.

[d]*Although Test C indicates one positive cow in the untreated group (assuming there have been no mistakes in defining cows as untreated in this trial) this result is still compatible with our hypothesis of no treated cows since we are using a test with imperfect specificity. The specificity of the test is still high (97%, 95% confidence interval ranging from 81-100%). The wide range to the confidence interval is mainly due to the small mumber of cows involved in the evaluation trial and not likely due to any real difference between the specificity of the two tests.*

Using Antibiotic Residue Tests on the Farm to Produce Quality Milk--Predictive Value

Once you have decided which test you will use on a given farm it will be important to carefully apply the tests to situations where you can maximize the value of the test results. Traditionally we have assessed the performance of a test by determining the test's population sensitivity (the probability of a test correctly identifying a cow with antibiotic above tolerance levels in her milk) and specificity (the probability of correctly identifying a cow without antibiotics--or at levels below tolerance level--in her milk). But sensitivity and specificity only partially characterize the usefulness of a test. The most useful measure to the practitioner of a test's utility is predictive value.

Predictive value positive for the residue tests is the probability that a positive test result is associated with an animal producing milk with tolerance levels of antibiotics. Predictive value negative is the probability that a negative test result is associated with an animal below tolerance levels of antibiotics. Predictive value is a function of population sensitivity and specificity and the prevalence (or probability) of the condition being tested. For a given sensitivity and specificity, predictive values will change as prevalence changes. As the probability of a cow being positive for antibiotics increases the predictive value positive of the test increases while predictive value negative decreases. As the probability of cow being positive for antibiotics decreases the predictive value positive for the test decreases and the predictive value negative increases.

These characteristics of predictive value are shown in Figures 1-3. The screening test we will use has for our example has a population sensitivity of 99% and specificity of 95%. Figure 1a depicts a population of cows recently treated with antibiotics. The prevalence of cows with action levels of antibiotics in their milk is 50%, i.e. half of the cows are truly positive for antibiotics in their milk. Figure 1b depicts the screening test results using our example test. In this case the predictive values positive and negative are quite good, that is both negative and positive test results are mostly associated with negative and positive cows, respectively. The key word is "mostly" since there is a small probability that a positive test results is wrong. Figure 2a depicts a population of cows treated with antibiotics following the recommended withholding time. The prevalence of cows with tolerance levels of antibiotics is 5%. Figure 2b depicts the screening test results using our same example residue test. Using the same test but under conditions of low prevalence, the predictive value positive drops to 50% while our predictive value negative approaches 100%. In this situation a negative result is nearly an absolute answer, but a positive test result is difficult to interpret. Figure 3a depicts a population of cows very recently treated. The prevalence of cows with tolerance levels of antibiotics in their milk is 99%. Figure 3b shows how cows would be classified using our residue test, under these conditions of high prevalence the predictive value positive of the test approaches 100% while the predictive value negative becomes 33%. In this scenario, a negative test result is more likely to be associated with a treated animal than with an untreated animal while the positive result is nearly an absolute answer.

The importance of understanding predictive value cannot be understated since it will be a guide to help apply the tests. One of the cardinal guidelines for deciding to use a screening test is, the test should provide the practitioner with more useful information af-

Figure 1: A hypothetical population of cows recently treated with antibiotics (treated cows are black, untreated cows are white) is shown in 1a. The prevalence of cows with greater than tolerance levels of antibiotics in their milk is 50%, i.e. half of the cows are truly positive for antibiotics in their milk. The screening test used to screen the population has sensitivity of 99% and specificity of 95%. Figure 1b depicts the screening test results cross-classified by true status of the cows.

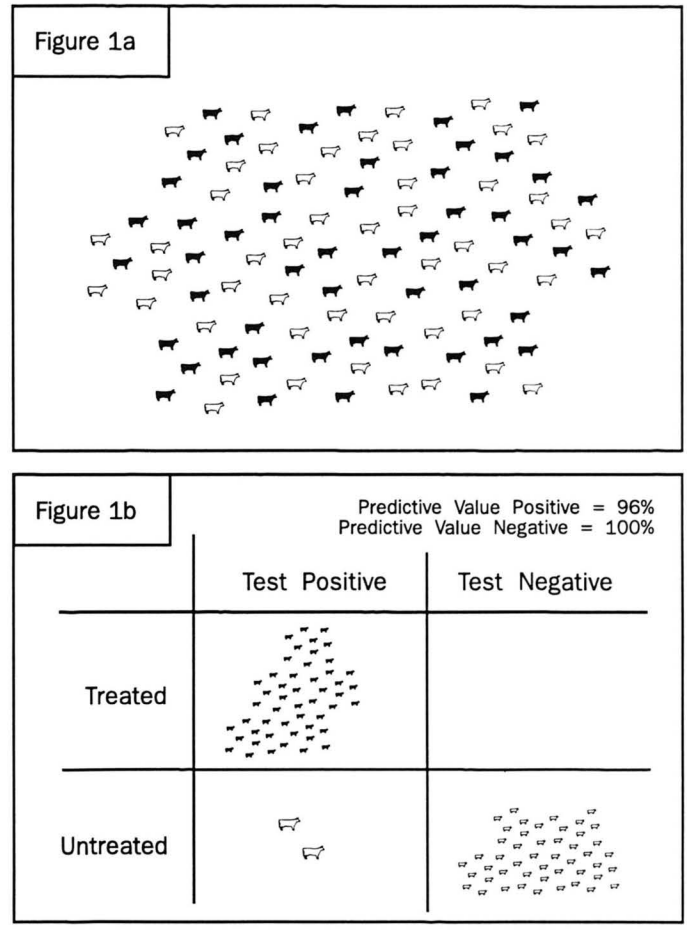
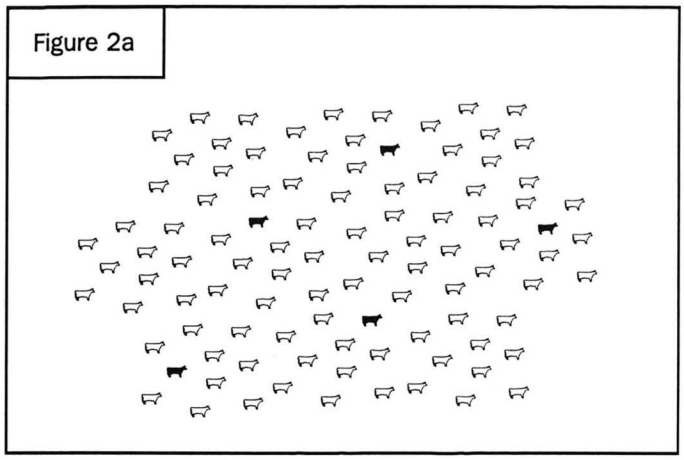
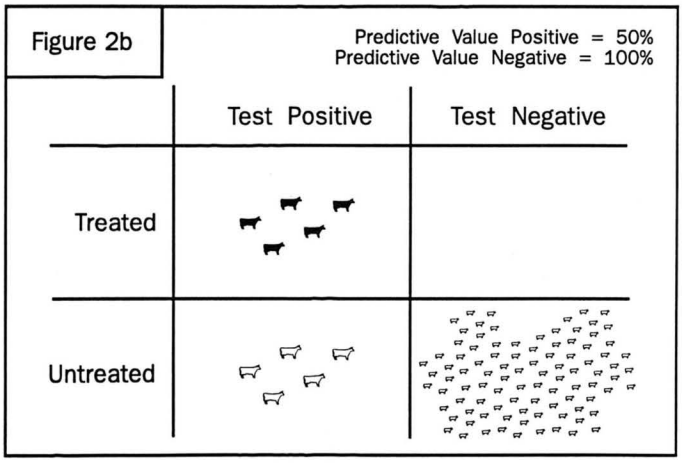


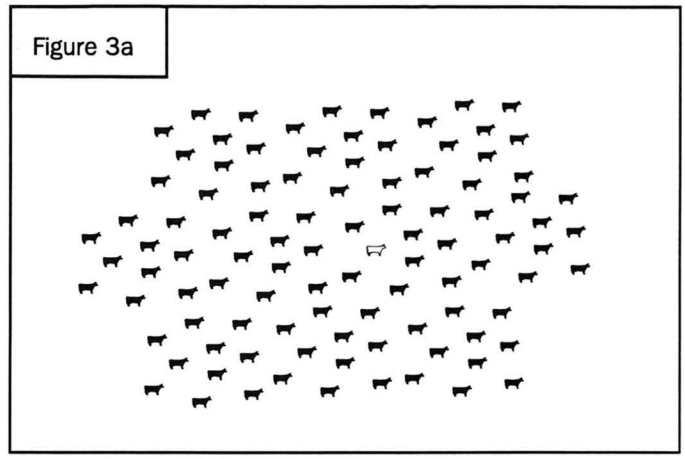
Figure 2: A hypothetical population of cows treated with antibiotics (treated cows are black, untreated cows are white) evaluated after the labelled withholding time has been observed is shown in 2a. The prevalence of cows with greater than tolerance levels of antibiotics in their milk is 5%, i.e. 5 of 100 cows are truly positive for antibiotics in their milk. The screening test used to screen the population has sensitivity of 99% and specificity of 95%. Figure 2b depicts the screening test results cross-classified by true status of the cows.

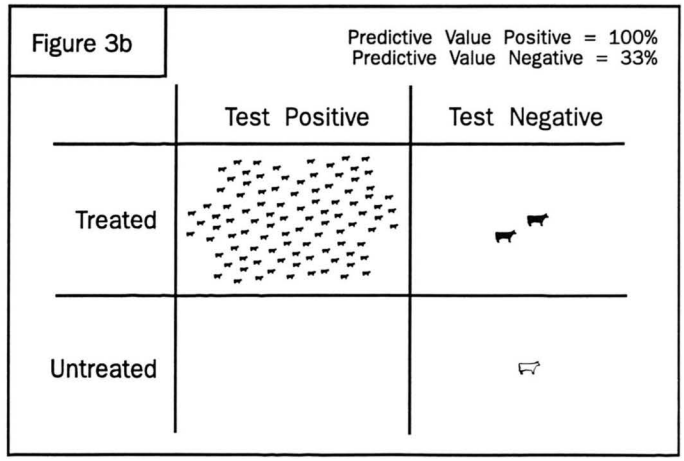




ter running than the test then was available before running the test. To satisfy this guideline it will be important to apply the antibiotic screening tests only in situations where the predictive values of the tests are nearly absolute answers. In which case the only appropriate situation to use the tests will be to evaluate the status of cows following label withholding time. It would be inappropriate to use these tests to test cows free of antibiotics prior to the label withholding time, since using the tests in this manner will result in some false negatives which will jeopardize the dairy producer's milk market. Even worse would be to use these tests to randomly screen untreated cows. The predictive value of a positive test in this circumstance is near zero and will only result in the false conclusion that every herd has cows producing milk with antibiotics which is being sold to the consumer.

Figure 3: A hypothetical population of cows treated with antibiotics (treated cows are black, untreated cows are white) evaluated immediately after the final treatment is shown in 3a. The prevalence of cows with greater than tolerance levels of antibiotics in their milk is 95%, i.e. 95 of 100 cows are truly positive for antibiotics in their milk. The screening test used to screen the population has sensitivity of 99% and specificity of 95%. Figure 3b depicts the screening test results cross-classified by true status of the cows.





Conclusion

The certification process for antibiotic residue tests for raw, comingled bovine milk has been a long and arduous process. In terms of protecting the consumer from antibiotic tainted milk perhaps the process developed by FDA and administered by AOAC International can be deemed a success. Unfortunately for the producer and veterinarian the process has left some problems. First, the tests have not been evaluated in a population context. The number of samples required for certification was small and unlikely to represent the range of

bulk or individual milk quality that will be experienced in the field. The quality of the estimates of the population test parameters resulting from the certification is questionable meaning that we will be conducting an uncontrolled field trial when these tests come on line in January 1995. Second, the certification process which focussed on developing convenient tests to be used in the rapid screening of milk by processors allowed tests to be certified that will detect some antibiotics below regulatory tolerance levels and in some cases above tolerance levels. Although there will be some accounting for these discrepancies in the labels for these tests, the presence and use of these test suggests that producers and veterinarians will be facing the very real probability that legal milk will be dumped and producers penalized for the test's mistake. Third, some AOAC certified tests (approved for bulk milk) will be marketed as farm and cow tests. Although the labels for these tests will explicitly describe their approval for bulk milk only, the implicit message is that the test can be used appropriately for individual animal milk. There will be no data to support use of these test on individual animals and it will be necessary to subject these tests to protocols such as Test the Tests.[14] Finally, because the breadth of the testing program will be increasing (the present official test, the *Bacillus stearothermophilus* disk assay detected only a portion of the 6 beta lactams targeted in the new program) there will be an increase in the number of violations detected beginning January 1995 with no reason other than increased ability to detect the antibiotics that had gone undetected.

We will have to live with these tests and we should make efforts to understand how these tests can be best used. The goal of the dairy industry needs to be towards the continuing production of nutritious, good tasting, and safe milk. Although the tests as they stand today are fraught with problems they should be utilized as necessary as part of a farm Total Quality Management program and within the context of the MDBQAP. They are not intended nor should they ever be used to define a quality product. Milk quality in all its dimensions begins with an on-farm program to promote animal and particularly udder health. When disease does occur alternatives to antibiotics should be employed when possible and if antibiotics become necessary then they should be used in a rational manner. This approach should include on-farm training in handling and administering the drugs, developing treatment protocols to provide guidelines for antibiotic use, and importantly having record keeping and identification systems that track antibiotic use and can be used by every member of the dairy management team to know which animals have been treated. The final link in the system will be the screening kits which will be appropriately used to verify that treated cows for which milk has been held from sale according to the label guidelines is negative for antibiotics and the milk is saleable from that aspect.

References

1. Bauman HE. The HACCP concept and microbiological hazard categories. *Food Tech* 1974; 28(9):30-34. 2. Herrick JB. Milk quality assurance and dairy practitioners. *J Am Vet Med Assoc* 1991; 199:1268-1269. 3. Zadjura EM, Rahl MJ, Aguilar ML, Layden WM, Weaver SW, Karpman MB. Drug residues in milk. 1990; Washington D.C. United States General Accounting Office. RCED-91-26: 1p. GAO report on safety of milk. 4. Tyler JW, Cullor JS, Erskine RJ, Smith WL, Dellinger J, McClure K. Milk antimicrobial drug residue assay results in cattle with experimental, endotoxin-induced mastitis. *J Am Vet Med Assoc* 1992; 201:1378-1384. 5. Sischo WM, Burns CM. Field trial of four cowside antibiotic-residue screening tests. *J Am Vet Med Assoc* 1993; 202:1249-1254. 6. Van Eenennaam AL, Cullor JS, Perani L, Gardner IA, Smith WL, Dellinger J, *et al.* Evaluation of milk antibiotic residue screening test in cattle with naturally occurring clinical mastitis. *J Dairy Sci* 1993; 76:3041-3053. 7. Hoffmeister A, Suhren G, Braun H, Heeschen W. Sulfonamidnachweis in milch mit mikrobiologischen, immunologischen und chemischphysikalischen methoden. Deutsche Milchwirtstchaft 1992; 43:724-728. 8. Carlsson A, Bjorck L. Liquid chromatography verification of tetracycline residues in milk and influence of milk fat lipolysis on the detection of antibiotic residues by microbial assays and the Charm II test. *J Food Prot* 1992; 55:374-378. 9. Jones GM, Seymour EH. Cowside antibiotic residue testing. *J Dairy Sci* 1988; 71:1691-1699. 10. Carlsson A, Bjorck L. The effect of some indigenous antibacterial factors in milk on the growth of *Bacillus stearothermophilus var calidolactis*. Milchwissenschaft 1987; 42:282-285. 11. Cullor JS, Van Eenennaam A, Delinger J, Perani L, Smith W, Jensen L. Antibiotic residue assays: Can they be used to test milk from individual cows? *Vet Med* 1992; 477-494. 12. Zadjura EM, Layden WM, Karpman MB, Turner P, Mitchell JF, Rose J, *et al.* Animal drug residues in milk. 1992; Washington D.C.: Government Accounting Office. RCED-92-209: 1p. 13. AOAC Research Institute. Specific data submission requirements for test kits to test for beta-lactam residues in milk. Anonymous 1992. 14. Cullor JS. Testing the tests intended to detect antibiotic residues in milk. *Vet Med* 1994; 462-472. 15. Charm SE. Determination of non actionable positives associated with antibiotic tests. *Dairy Food Env Sanit* 1994; 14:151-154.