Testing to Fulfill HACCP (Hazard Analysis Critical Control Points) Requirements: Principles and Examples

I. A. GARDNER
Department of Medicine and Epidemiology,
School of Veterinary Medicine,
University of California, Davis 95616

ABSTRACT

On-farm HACCP (hazard analysis critical control points) monitoring requires cost-effective, yet accurate and reproducible tests that can determine the status of cows, milk, and the dairy environment. Tests need to be field-validated, and their limitations need to be established so that appropriate screening strategies can be initiated and test results can be rationally interpreted. For infections and residues of low prevalence, tests or testing strategies that are highly specific help to minimize false-positive results and excessive costs to the dairy industry. The determination of the numbers of samples to be tested in HACCP monitoring programs depends on the specific purpose of the test and the likely prevalence of the agent or residue at the critical control point. The absence of positive samples from a herd test should not be interpreted as freedom from a particular agent or residue unless the entire herd has been tested with a test that is 100% sensitive. The current lack of field-validated tests for most of the chemical and infectious agents of concern makes it difficult to ensure that the stated goals of HACCP programs are consistently achieved.

(**Key words**: HACCP, hazard analysis critical control points, sensitivity, specificity)

Abbreviation key: **HACCP** = hazard analysis critical control points.

INTRODUCTION

Modern **HACCP** (hazard analysis critical control points) principles require the monitoring of the production process. The goal is to use the results of the monitoring to identify and quantify the magnitude of the risk and to correct risk factors wherever possible. As part of on-farm HACCP programs, dairy producers and veterinarians need to implement testing regimens that allow them to determine whether the critical tolerances or limits for potential foodborne drug and chemical residues and infectious agents, as identified in food safety objectives stated in the *Healthy People 2000* (3), have been exceeded. An integral part of the implementation of on-farm HACCP programs is the development and application of diagnostic tests that monitor bulk tank milk, milk samples from individual cows, and the dairy environment for chemical and drug residues and microbes. However, many of the existing tests have not been adequately evaluated, and testing plans for specific agents have not been defined.

Application of HACCP principles to dairy production has been described, and an example of a HACCP program for mastitis control has been proposed (7, 8). The suggested HACCP program and associated testing are based on many decades of research and address a condition that is perceived as costly by the dairy producer. In contrast, many of the foodborne infections that potentially are transmitted via meat, milk, and other dairy products are not associated with clinical disease, production losses, or tangible direct costs to the dairy producer. Even in the case of salmonellosis, which can cause high rates of morbidity and mortality when the agent is introduced into a naive herd, most infections are subclinical in endemically infected herds (18). The benefits of HACCP programs often are unclear to livestock producers, and substantial education is necessary to change this skepticism (6). Consequently, whether dairy producers will voluntarily implement HACCP and its associated testing programs in the absence of regulations is unclear; voluntary adoption of the 10-point Milk and Dairy Beef Quality Assurance Program to prevent antibiotic residues has occurred in only about 10% of the nation's dairy herds (4).

This paper describes performance characteristics of tests, measures of test validity for individual and herd tests, and sample sizes needed to meet certain testing objectives. Tests for the detection of antibiotic residues and ELISA for *Salmonella dublin* are

Received September 18, 1996.
Accepted April 14, 1997.

presented as examples of tests currently in use for on-farm monitoring.

PERFORMANCE CHARACTERISTICS OF ON-FARM TESTS

Low cost, ease of performance, and rapidity of results are key criteria when tests are used in food-producing industries in which products are rapidly processed and shipped. Performance characteristics (sensitivity, specificity, and reproducibility) of tests usually are of lesser concern, yet are important, especially if the costs of misdiagnosis to dairy producers, processors, and consumers are substantial. Ideally, a test should correctly reflect the true status of an animal or sample, and the assay should also yield identical or similar results when run in different laboratories and by different technicians. Reproducibility is very important for test kits, which will potentially be used in a variety of locations including dairies, processing plants, veterinary practices, and diagnostic laboratories.

Sensitivity (the probability that the test gives a positive result in a truly positive individual) and specificity (the probability that the test gives a negative result in a truly negative individual) are recognized by researchers and veterinarians as standard measures of test performance, as indices to compare two tests for the same agent or residue, and as a way to select combinations of tests that optimize their respective strengths. Estimates of sensitivity and specificity usually are obtained from experimental and observational (field) studies and reflect how well the new test performs compared against a standard or definitive test. The definitive test often is invasive, expensive to run, or otherwise impractical to use but provides results that are as close to perfect as possible (13). Sometimes studies to estimate these performance characteristics necessarily are complicated because the tests are used in many ways in the field. A combination of laboratory and field methods has been recently proposed to evaluate antibiotic residue tests in milk from individual cows (11).

Although sensitivity and specificity are the most commonly used measures of test validity, receiver operating characteristic curves (22) are being increasingly used for tests measured on ordinal or continuous scales (e.g., titers or ELISA optical densities) to show how well a test performs over a range of possible cutoff values. The validity of two or more tests can be compared by estimating areas under receiver operating characteristic curves rather than by comparing results at a single cutoff value.

After working with colleagues on the development and validation of diagnostic tests in many species, I have been more strongly convinced of the following. First, all tests are imperfect. If claims are made that a test is 100% sensitive and specific, the study was probably done with limited sample sizes or unrepresentative samples. Second, evaluations of sensitivity and specificity done in the laboratory or even in an experimental setting usually overestimate the real performance of the test in the field. Third, technical modifications to improve test sensitivity may result in a loss of specificity because of the inverse relationship between these values. Finally, a single cutoff value for a quantitative diagnostic test is not necessarily optimal for all testing situations. The choice of an appropriate cutoff should depend on many factors, including the reason for the testing, whether results are to be interpreted at a group or individual level, the costs of classification errors, and the specific goals of the user.

In the field, the limitations of sensitivity and specificity as measures of test performance are clear. For example, most practitioners do not have the luxury of determining the true status of all animals being tested. Rather, practitioners obtain a test result and then interpret it by asking questions such as how likely it is that an animal or sample is truly positive (or truly negative) given a positive (or negative) test result. Such questions equate to the predictive value of a positive or negative test result, respectively. The answers depend not only on the sensitivity and specificity of the test but also on the prevalence of the agent or residue in the population. By use of Bayes' theorem, predictive values can be calculated from sensitivity, specificity, and prevalence (13). Most pathogens and residues of interest probably occur on dairies at low prevalence, and, hence, the positive predictive value tends to be low for most diagnostic tests. For example, a test that is 90% sensitive and 90% specific has a positive predictive value of 50% when the prevalence is 10%, but the positive predictive value decreases to 8.3% when the prevalence is 1% (Figure 1). If the test were 100% specific, which might be true for bacterial and viral isolation techniques in most diagnostic laboratories, the positive predictive value would be 100%, regardless of prevalence.

For conditions of low prevalence, two tests can also be used in combination to achieve the goal of having a test system that has good predictive value for positive samples without compromising predictive value for negative samples. To achieve this goal, samples are screened with a rapid, inexpensive assay of high sensitivity (>99%). Positive samples are retested with a

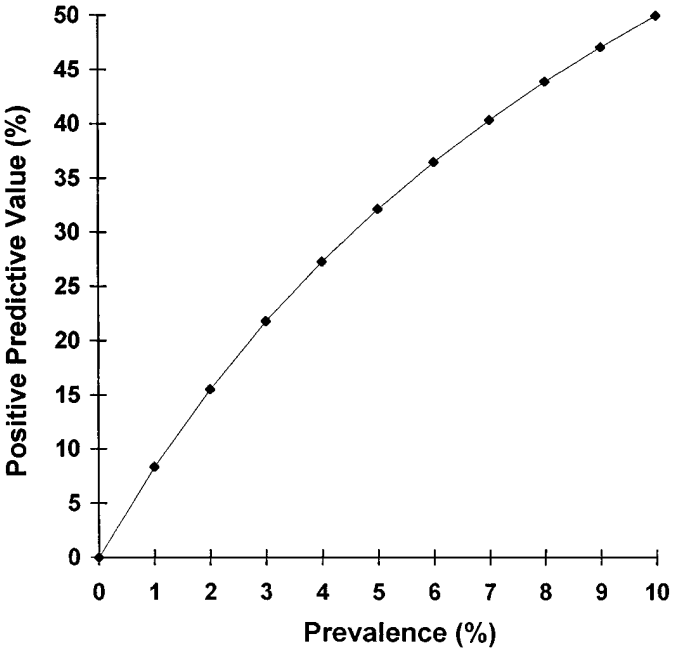

Figure 1. Effect of prevalence on the positive predictive value for a test with sensitivity of 90% and specificity of 90%. Predictive value positive = (sensitivity × prevalence)/((sensitivity × prevalence) + (1 − specificity) × (1 − prevalence)).

second highly specific (>99%) and more expensive assay; only samples that are positive on both tests are considered to be positive.

Herd Versus Individual Tests

In the previous sections, I emphasized individual test interpretation and the need for tests of high specificity when prevalence is low. Sometimes the question is whether the herd (or age or production group within a herd) is affected by the agent or residue. This question can be addressed by collecting a pooled sample from the group or by sampling individual animals and interpreting group results. Typically, milk is used for the former, and serum and milk are used for the latter. One limitation of a bulk tank sample is that it does not represent a pool from all cows in the herd, and the contributions of high producing cows and those in early to midlactation are overrepresented relative to low producing and late lactation cows. Consequently, the sensitivity of a test for a pooled sample of milk from several cows usually is less than for an individual cow sample if the same cutoff point is used to interpret the test result.

When multiple samples of milk or blood are collected, and the herd is the unit of interest, herd level sensitivity (probability that the herd test yields a positive when the herd is truly positive) and herd level specificity (probability that the herd test yields a negative when the herd is truly negative) are the measures analogous to their individual counterparts. Their values depend on the sensitivity and specificity of the individual test, the prevalence of the agent or residue in positive herds, the number tested, and the number of positives that are required in the group before the herd can be designated positive (14). These herd level measures are almost always unknown, but herd level sensitivity and specificity are inversely related as they are at the individual level. One consequence of the use of imperfectly specific tests is that the more tests that are conducted, the greater the probability is of obtaining false-positive results for the herd. For example, if 100 cows from a negative herd are tested with a test that is 95% specific, 5 positive results would be expected. If 5 or fewer positives were used to classify the herd as positive, the herd would be erroneously considered positive.

Antibiotic Residue Assays for Milk

Seventeen assays were performance tested for monitoring β-lactam residues by the Association of Official Analytical Chemists, International and were accepted by the FDA Center for Veterinary Medicine for monitoring raw, bulk, commingled milk. Initial development of the assays was laboratory-based and, because of consumer concerns, emphasized analytical sensitivity such that assays can detect concentrations of drugs in milk samples even when below the safe or tolerance limit or safe concentration established by FDA. The tests have a sigmoid dose-response curve: as drug concentration increases, a corresponding increase occurs in the percentage of positive samples until a concentration plateau is reached above which all samples are positive (5).

When these assays were used by a number of investigators on field samples, specificity was inadequate for tests on individual cows, and there were some field reports of positive bulk tank tests when antibiotics had not been used. False-positive results were reported in samples from cows with experimentally induced endotoxic mastitis and naturally occurring mastitis (9, 15, 19, 20, 21). Despite almost universal use for monitoring milk from tanker trucks in the US, the predictive value of positive and negative assay results is not known. Although such tests have not been adequately validated, the consciousness of the dairy industry for the need for responsible antibiotic use appears to have been greatly height-

ened by introduction of monitoring of tanker truck milk.

Assays for monitoring raw, commingled milk for chloramphenicol (one test) and sulfonamides (one test) have been accepted by FDA and by the National Conference on Interstate Milk Shipments for Appendix N, Pasteurized Milk Ordinance testing. States participate in monitoring milk for tetracycline, aminoglycosides, and gentamicin using both accepted and tests that have not been accepted nationally.

Serologic and Milk Assays for *Salmonella dublin*

The limited sensitivity of methods and the need for repeat culture of milk and feces to identify *Salmonella* carriers in dairy herds has led researchers to develop mixed or group-specific ELISA assays (12,16,17) for various *Salmonella* serogroups (B, C1, C2, C3, D1, and E1). These tests can help to identify *Salmonella* carriers, determine which herds have been exposed previously to *Salmonella*, and be used to estimate the herd and individual animal prevalence of infection or exposure (18). Although such tests have not been evaluated adequately for their sensitivity and specificity in the US, the ELISA is probably more sensitive and less specific than culture.

At the herd level, field experience indicates that an unexposed herd has very few ELISA-positive animals and only when tested at low optical density values. Herds with persistent salmonellosis usually have a high proportion of positive animals at high optical density values. The true status of herds that have several positive animals near the cutoff for the assay has not been determined. Vaccination, which is commonly practiced in dairy herds in California, and a recent history of clinical salmonellosis complicate serologic interpretation. Repeated sampling is considered necessary when the goal is to identify and cull persistently seropositive (suspect carrier) cows (18). Although titers are generally higher in infected than in vaccinated, uninfected cows, distinguishing one group from the other is difficult without the repeated culture of milk and feces.

For herds that are positive for the group D serotype (most frequently *S. dublin*), serotesting of all dairy animals over 6 mo of age has been recommended, as has culling of animals that are persistently positive after 60 to 90 d. When ongoing attempts are made to remove seropositive cows, quarantine of purchased females and testing for *S. dublin* also should be implemented. This strategy, combined with other recognized control procedures, may prove to be beneficial in reducing the prevalence of infection (16).

SAMPLE SIZES FOR USE IN ON-FARM TESTING

After limitations of available tests and possible testing strategies have been defined, the number of needed samples must be determined. For simplicity, I have restricted the discussion to samples from animals rather than the environment and address three common questions. First, if the goal is to certify freedom from a particular agent or residue, all animals should be tested with a test that is 100% sensitive, and all test results should be negative. Second, to determine whether a pathogen or agent is present in animals at or above a specified prevalence, the desired confidence level in the result and the expected prevalence in the herd need to be specified (1,10). In general, more samples are required as the prevalence decreases and as the desired confidence in the result increases. For example, to be 95% confident of detecting at least 1 positive animal in a large herd with 10% prevalence (assuming that the test is 100% sensitive and specific) requires 30 samples. A corollary of this calculation would be whether zero risk can be inferred if no positives are detected in 30 randomly selected samples from a herd. Such an inference is not warranted and can only be made if the entire herd were tested with a perfectly sensitive test. The upper 95% confidence limit for prevalence when no positives are found in n samples is approximately 3/n (10) or, in this case, 10%. Sample size formulas can be modified to adjust for test sensitivity values (14). In this example, if a test is known to be only 50% sensitive, then the required sample size would need to be doubled to 60 because only one-half of the truly positive animals would be detected. Third to determine how many samples are needed to obtain a precise estimate of prevalence, three factors that affect this calculation must be considered: prevalence in the population, the allowable error limits for the prevalence estimate, and the confidence in the estimate. When prevalence is not known, it is often advisable to evaluate the worst case scenario, which occurs when prevalence equals 50%. For example, if the prevalence in the herd were 50% and a veterinarian wanted to sample enough animals to be 95% confident that the estimate based on a sample from the herd would be between 40 and 60% (i.e., error limits were 10%), 96 animals would need to be sampled. When prevalence in the population is close to either 0 or 100%, much smaller samples are necessary. Difficulties that are associated with interpretation of prevalence estimates based on small samples have recently been described (2).

CONCLUSIONS

On-farm programs for monitoring antibiotic residues are costly to implement. The estimated an-

nual cost of testing to detect β-lactam antibiotic residues in tanker truck milk in the US is between $8 and $35 million (11). Extension of testing to other drugs, chemicals, and pathogens in bulk tank milk and other media stretch the resources of the dairy industry. The lack of field-validated tests for most of the chemical and infectious agents of concern makes it difficult to ensure that stated goals of HACCP programs are consistently achieved. Much basic and field research is needed to develop and validate these tests, which can only be achieved through a substantial infusion of new research funding. Validation of tests is essential for selection of the most appropriate testing strategies, to estimate predictive values and for appropriate test interpretation, and to ensure that testing programs operate as efficiently as possible. Consideration of requirements for sample size is also an important part of development of on-farm testing programs.

REFERENCES

1 Cannon, R., and R. T. Roe. 1982. Livestock Disease Surveys: a Field Manual for Veterinarians. Australian Govt. Publ. Serv., Canberra, A.C.T., Australia.

2 Carpenter, T. E., and I. A. Gardner. 1996. Simulation modeling to determine herd-level predictive values and sensitivity based on individual-animal test sensitivity and specificity and sample size. Prev. Vet. Med. 27:57.

3 Centers for Disease Control and Prevention. 1991. Page 340 *in* CDC Public Health Service: Healthy People 2000. Dep. Human Health Serv. Publ. 91-50212. Ctr. Disease Control Prevention, Atlanta, GA.

4 Centers for Epidemiology and Animal Health. 1996. National Animal Monitoring System's (NAHMS) Dairy '96. Page 15 *in* Part 1: Reference of 1996 Dairy Management Practices, Natl. Anim. Health Monitoring Syst. Publ. N200.496.

5 Charm, S. E. 1994. Determination of non-actionable positives associated with antibiotic tests. Dairy Food Environ. Sanit.14: 151.

6 Coleman, W. W. 1995. Animal food safety and dairy regulations, now and in the future: from farm to fork, a state perspective. J. Dairy Sci. 78:1204.

7 Cullor, J. S. 1995. Common pathogens that cause foodborne disease: can they be controlled on the dairy? Vet. Med. 90:185.

8 Cullor, J. S. 1995. Implementing the HACCP program on your clients' dairies. Vet. Med. 90:290.

9 Cullor, J. S., A. L. Van Eenennaam, I. Gardner, L. Perani, J. Dellinger, W. L. Smith, T. Thompson, M. A. Payne, L. Jensen, and W. M. Guterbock. 1994. Performance of various tests used to screen for antibiotic residues in milk samples from individual animals. J. AOAC 77:862.

10 Di Giacomo, R. F., and T. D. Koepsell. 1986. Sampling for detection of infection or disease in animal populations. JAVMA 189:22.

11 Gardner, I. A., J. S. Cullor, F. D. Galey, W. Sischo, M. Salman, B. Slenning, H. N. Erb, and J. W. Tyler. Alternatives for validation of diagnostic assays used to detect antibiotic residues in milk. JAVMA 209:1.

12 Hoorfar, J., P. Lind, and V. Bitsch. 1995. Evaluation of an O antigen enzyme-linked immunosorbent assay for screening of milk samples for *Salmonella dublin* infection in dairy herds. Can. J. Vet. Res. 59:142.

13 Martin, S. W. 1988. The interpretation of laboratory results. Vet. Clin. North Am. Food Anim. Pract. 4:61.

14 Martin, S. W., M. Shoukri, and M. A. Thorburn. 1992. Evaluating the health status of herds based on test applied to individuals. Prev. Vet. Med. 14:33.

15 Okada, Y. 1986. Bacterial inhibitors in milk produced by cows treated with no antibiotics. J. Jpn. Vet. Med. Assoc. 39:97.

16 Smith, B. P., D. G. Oliver, P. Singh, G. Dilling, P. A. Martin, B. P. Ram, L. S. Jang, N. Sharkov, J. S. Orsburn, and K. Jackett. 1993. Detection of *Salmonella dublin* mammary gland infection in carrier cows, using an enzyme-linked immunosorbent assay for antibody in milk or serum. Am. J. Vet. Res. 50:1352.

17 Smith, B. P., G. W. Dilling, J. K. House, H. Konrad, and N. Moore. 1995. Enzyme-linked immunosorbent assay for *Salmonella* serology using lipopolysaccharide antigen. J. Vet. Diagn. Invest. 7:481.

18 Smith B. P., and J. K. House. 1996. Use of serology as part of a salmonella control program. Page 493 *in* Proc. Am. Coll. Vet. Intern. Med. Forum, San Antonio, TX. Am. Coll. Vet. Intern. Med., Lakewood, CO.

19 Sischo, W. M., and C. M. Burns. 1993. Field trial of four cowside antibiotic-residue screening tests. JAVMA 202:1249.

20 Tyler, J. W., J. S. Cullor, R. J. Erskine, W. L. Smith, J. Dellinger, and K. McClure. 1992. Milk antimicrobial drug residue assay results in cattle with experimental, endotoxin-induced mastitis. JAVMA 201:1378.

21 Van Eenennaam, A. L., J. S. Cullor, L. Perani, I. A. Gardner, W. L. Smith, J. Dellinger, W. M. Guterbock, and L. Jensen. 1993. Evaluation of milk antibiotic residue screening tests in cattle with naturally occurring clinical mastitis. J. Dairy Sci. 1993:76: 3041.

22 Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39:561.