Food Animal Economics

Economic evaluation of risks to producers who use milk residue testing programs

Barrett Durand Slenning, DVM, MPVM, and Ian Andrew Gardner, BVSc, PhD

Objective—To evaluate the decision to test for milk antimicrobial residues in milk from dairy cows treated with procaine penicillin G (PPG).

Design—Economic-decision analysis after stochastic simulation.

Sample Population—1,000 computer-simulated cows/model.

Procedure—Meta-analysis of the Food Animal Residue Avoidance Databank was used to generate PPG disappearance curves for cows given single PPG treatments, IM, of 6,600 U/kg (3,000 U/lb) of body weight or 26,400 U/kg (12,000 U/lb), and multiple treatments at 26,400 U/kg (12,000 U/lb), IM. These curves were entered into 1,000-replication stochastic pharmacokinetic models, generating population-level milk PPG profiles for each treatment group for each day after treatment, which were subjected to economic-decision analyses of feasibility of residue testing. The model was evaluated for changes in herd size, proportion of herd available for testing, milk production, test price, test sensitivity/specificity, and withdrawal periods.

Results—For both single-treatment groups, a 2-day withdrawal period avoided violative residues. However, nearly two thirds of the cows risked false identification for violative residues. For the multiple-treated group, nearly 40% had violative residues after a 5-day withdrawal period, and an additional 10 to 15% risked false identification for violative residues. Economic analysis yielded a decision against testing; mean cost was $2 (ie, 5% more than the mean cost of not testing).

Clinical Implications—Complex dynamics of current milk residue tests discourage practitioners from recommending procedures to clients. In general, increases in herd size, milk production, proportion of a herd available for testing, or milk price will increase the value of testing. Increasing test sensitivity decreases its desirability to producers. (*J Am Vet Med Assoc* 1997;211:419–427)

Public concern about food safety, in general, and agricultural chemicals in food, in particular, has been growing.[1-7,a] Whether the reasons for this concern are warranted is irrelevant; an industry that does not respond to public concerns will lose public confidence and will likely have nonoptimal regulations imposed on it. Hence, agriculture must respond to public concerns about food safety through public education and quality control of production. For animal agriculture, 1 high-profile quality control method involves residue avoidance programs.[8-11]

The US dairy industry has long promoted quality control and residue avoidance. Prior to the 1990s, the National Mastitis Council developed a 5-point plan for mastitis prevention.[11] Furthermore, many processors offered premiums to producers to enhance quality. In 1991, the Milk and Dairy Beef Residue Prevention Protocol was created with the intention of focusing on events within control of management that are associated with an increased risk of residues.[11-14] The resulting protocol emphasizes 3 dairy management areas that directly minimize risks of residues: management of a healthy herd; control of drug use; and residue testing.

The protocol has not, however, been an unqualified success. In 1993, a sample of 240 California dairies revealed that there was poor awareness or implementation of the protocol.[15] The protocol was described by producers as an invasion by government. In addition, veterinarians perceived themselves doing the regulators' jobs, so they did not promote the program. Finally, producers and veterinarians voiced concern about the quality of the residue tests.

Antimicrobial residue tests are imperfect. False-positive results may be caused by natural milk components[16-19] or because of the nature of the test used.[20-24] Unfortunately, the means used to approve tests are insufficient to allow intelligent choices of appropriate tests. Tests are required to show a minimum sensitivity (the ability to identify violative milk samples) of 90% but are not required to pass a realistic requirement for specificity (the ability to identify nonviolative milk samples).[25] This is paradoxic, because most samples submitted for testing will be nonviolative when producers adhere to recommended withdrawal periods. Second, the subjects chosen for test evaluation do not represent in-field specimens; samples are commingled raw milk, not separate samples from antibiotic-treated cows. Finally, tests often detect residues at concentrations that are less than established safe levels, making the likelihood of false-positive results high.

Consequently, producers and veterinarians have been provided with inadequate information on test quality and performance, making selection of the best test kit problematic. In addition, uncertainty about actual

From the Population Medicine Program, Department of Food Animal and Equine Medicine, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606 (Slenning), and the Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA 95616 (Gardner).

test performance, and the way it is affected by the health practices, milk production, and economic situation of specific dairies, eliminates the potential to reliably determine the value of a testing program or to compare testing programs.[26] The purpose of our study was to evaluate the interactions of cow-health management and herd characteristics with decisions about residue testing and to measure their impacts on test system costs.

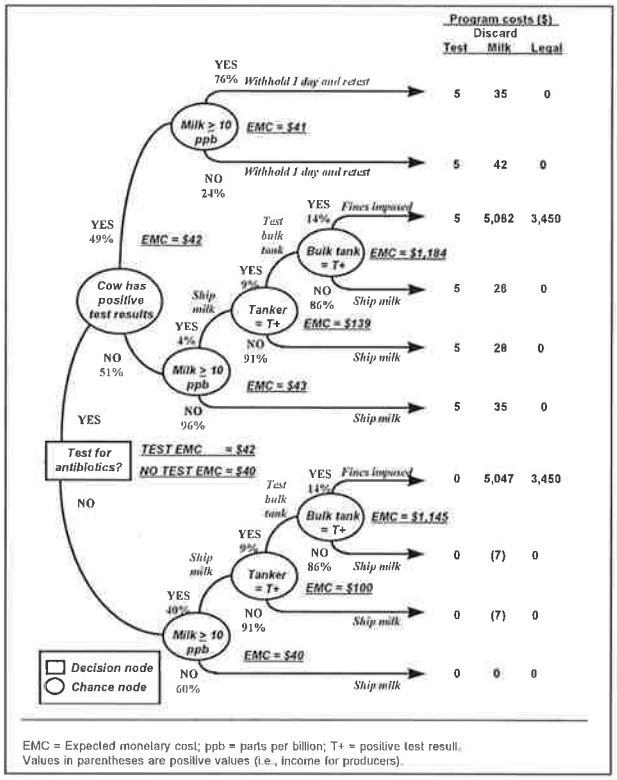
Materials and Methods

The **Food Animal Residue Avoidance Databank (FARAD)**[27] was searched for records pertaining to specific antimicrobials in lactating dairy cows. The description of the methods and results of this initial work has been reported previously.[b] The search was narrowed to include only **procaine penicillin G (PPG)** given IM and in which the PPG concentration was measured in milk. The study was used as the unit of interest. The FARAD data were evaluated for information on body weight of cows, dosage rates, milk production, and measures of variation for these variables. Meta-analysis[28] was used to develop statistical models of PPG disappearance after treatment administered according to 1 of 3 treatment protocols. The **first protocol (1X)**, representing the US label dosage,[b] was a single treatment of 6,600 U of PPG/kg (3,000 U/lb) of body weight, IM. Because the effectiveness of treatment at such dosages has been questioned,[29-34] the **second protocol (4X)** examined was a single dose administered at 4X the label dosage (26,400 U/kg [12,000 U/lb]). Third, to investigate most cows treated multiple times, the **final protocol (MULTI)** examined cows receiving > 2 treatments at 4X the label dosage. Final linear predictive equations of drug disappearance by treatment and number of exposure events were developed. Through general linear model analysis, estimates for the disappearance coefficients and their cross-study variability were determined.

Nineteen studies from the FARAD database met the following selection criteria: adequate data, as determined on the basis of evaluation of the aforementioned information, and analysis that yielded residuals within 2 SD of the overall mean values determined for curve slope. Tests for significance in curve slopes and point estimates for y intercepts were performed, using commercially available statistical programs.[c]

A stochastic model of PPG disappearance in dairy cows was built. The specific model algorithm was an adaptation of a previous model[35,36] built by use of commercially available spreadsheets.[d] The model developed 1,000 stochastic replications for each of the 3 treatment protocols (1X, 4X, MULTI) and yielded estimates, over time, of the proportion of cows having milk PPG concentrations at or greater than legal limits (≥ 10.0 ppb[37]), greater than minimum detectable limits but less than legal limits (5.0 to 9.9 ppb[37]), and less than minimum detectable limits (< 5.0 ppb[37]). The relative size of each of these categories defined time-dependent risks of cows that correctly had test-positive results for violative concentrations of PPG (true positives), incorrectly had false-negative results for violative concentrations of PPG (false negatives), were correctly identified as having negative results for violative concentrations of PPG (true negatives), and incorrectly were identified as having positive results for violative concentrations of PPG (false positives). Given the lack of difference in outcomes between the 1X and 4X groups, plus the fact that they were unlikely to be used in dairy cows, the remainder of the study was limited to cows in the MULTI group.

For our study, we used the premise that, under the current US system,[25,38] a sample of milk may be submitted



EMC = Expected monetary cost; ppb = parts per billion; T+ = positive test result.
Values in parentheses are positive values (i.e., income for producers).

Figure 1—Schematic diagram of the decision tree for antimicrobial residue testing of milk samples.

for antimicrobial residue testing up to 3 times. First, a producer tests a sample from a specific cow at the end of her withdrawal period (a decision we modeled in our study). Second, milk is sampled at the tanker-truck level. Third, should the tanker-truck sample have positive results, bulk-tank samples from each dairy that contributed to that tanker truck are tested. For our study, we assumed the same test was used for all 3 samples. The test was performed only once per sample, and other confirmatory tests were not used. This procedure is allowed by the US rule.[38] The effects of allowing the same test to be used as a confirmatory test on bulk-tank samples initially testing positive for violative residues was examined, because this is a common practice.[39]

We calculated population-level distributions of risk over time and applied them to a herd of known size, economic structure, and milk production. Entry values and their justification for use were determined (Appendix). These distributions determined probabilities of samples having violative residues on a given test day and the likelihood of tests yielding correct and incorrect results, using estimates of test sensitivity and specificity.

The value of milk, estimates of testing costs, regulatory fines and penalties, and costs of milk discarded because it was incorrectly identified as containing violative residues were used to generate expected monetary values (**expected monetary costs [EMC]**), which were used in the analysis of choosing a testing scheme over a strategy of not testing. The method used to generate the EMC was decision-diagram analysis, a widely used technique that accounts for uncertainty inherent in decisions.[35,40-43] The decision diagram (Fig 1) examined the choice between milk-residue testing of cows in the MULTI group on day 5 after treatment versus simply entering the same cows into the milking herd without testing. When a cow was tested and had a positive test result, the producer would withhold milk for 1 additional day and

schedule that cow for another test on the next day. The cost structure of the cows with a positive test result included a charge for running the test and the value of the milk discarded during withholding (5 days × 31.4 L/d × $0.22/L = $34.54, which was rounded to $35). For true-positive cows, the discarded milk (approx $7) was not charged to the program, because this milk would have been discarded anyway; however, for false-positive cows, the loss from inappropriately discarded milk was included.

We assumed that when a producer opted not to use the residue test, the cow's milk would be placed into the dairy's bulk tank for pickup. The section of the decision tree that traces this event mirrored that of the false-negative path under the testing option.

On the basis of standard US milk handling protocols,[11,39] when a tanker-truck sample has a positive result for PPG residues, the bulk-tank samples representing the dairies that contributed to that tanker would each be evaluated. A positive result from a bulk-tank sample would trigger regulatory actions and penalties, represented by the farm being charged for the cost of all the milk in the tanker truck, added-on cost recoveries, fines, and special shipping charges (in this case amounting to more than $8,500) until the farm had negative test results in accordance with the rules established in the Pasteurized Milk Ordinance.[25] When a cow with violative residues was allowed to contribute to the bulk tank and the milk from that tank was not identified as having violative residues, the model added in the bonus of $7 for milk sold that should have been discarded, which decreased the cost from $35 to $28.

Sensitivity of the decision-diagram analysis to changes in selected herd and test variables was accomplished by holding other inputs constant and allowing the selected variable to change over a wide range of values. Six variables were chosen: herd size, milk production, test price, withdrawal time, proportion of the herd at end of withdrawal period, and test sensitivity/specificity. The impact on absolute and relative changes in the EMC of the opposing options was evaluated graphically.

Results

Final equations depicting PPG disappearance curves were as follows:

for 1×, ln(PPG concentration in milk) = 1.269 + [–0.1375 (days after treatment)]; for 4×, ln(PPG concentration in milk) = 2.604 + [–0.1375 (days after treatment)]; and for MULTI, ln(PPG concentration in milk) = 2.604 + [–0.0523 (days after treatment)].

All were significant ($P < 0.05$) and explained the majority of variation in the curves (adjusted R^2 for MULTI was the worst at 0.877). The Y intercepts (estimates of theoretical time-zero milk concentrations) for 4× and MULTI were not significantly different, so values were combined for the simulations. Similarly, the disappearance slopes for the single-treatment protocols were not significantly different and were combined for the simulations.

Outputs were determined for 1,000 stochastic simulations of each treatment protocol (Table 1). The daily frequency of cows, by treatment protocol, was calculated in each of 4 PPG milk concentration categories. The concentration categories of 0.0 ppb and 0.1 to 4.9 ppb signified cows that had nonviolative residues and would have amounts that were less than the test's minimum detectable limits. The category of 5.0 to 9.9 ppb was the group of cows that still had nonviolative residues but had amounts that were greater than the test's minimum detectable limits. Finally, the category of ≥ 10.0 ppb depicted those cows that had violative residues and amounts that were greater than minimum detectable limits. For cows in the 1× and 4× groups, the period evaluated after treatment was only 3 days, because all cows cleared PPG from their milk in that time.

A single solution of the decision-diagram analysis was depicted (Fig 1). On the basis of the FARAD meta-analysis and the stochastic simulations, on day 5 after the MULTI treatment protocol, 49% of the cows would be identified as having violative residues. However, of those cows with positive test results, nearly 1 in 4 would be false positives. Of those cows with negative test results for PPG, 96% would be true negatives (nonviolative), and 4% would be false negatives, as determined on the basis of the stochastic modeling. Given the distribution of estimated PPG concentrations in milk of these cows and their relative contribution to the tanker truck, the milk processor was likely to identify 9% of these tanker-truck loads as having

Table 1—Stochastic simulation outputs for 3 treatment protocols

Treatment protocol*	Concentration category (ppb)	Days after treatment									
		1	2	3	4	5	6	7	8	9	10
1×	0.0	0	0	1000	–	–	–	–	–	–	–
	0.1–4.9	0	336	0	–	–	–	–	–	–	–
	5.0–9.9	0	664	0	–	–	–	–	–	–	–
	≥ 10.0	1000	0	0	–	–	–	–	–	–	–
4×	0.0	0	0	1000	–	–	–	–	–	–	–
	0.1–4.9	0	337	0	–	–	–	–	–	–	–
	5.0–9.9	0	623	0	–	–	–	–	–	–	–
	≥ 10.0	1000	0	0	–	–	–	–	–	–	–
MULTI	0.0	0	0	0	22	160	331	457	662	714	771
	0.1–4.9	0	0	22	187	265	331	287	144	138	126
	5.0–9.9	0	0	61	122	170	82	62	69	45	42
	≥ 10.0	1000	1000	917	669	405	256	194	125	103	61

*Treatment protocols of procaine penicillin-G were designed as follows: 1×, 6,600 U/kg (3,000 U/lb), IM, given once; 4×, 26,400 U/kg (12,000 U/lb), IM, given once; MULTI, 26,400 U/kg (12,000 U/lb), IM, given >2 times. ppb = parts per billion.

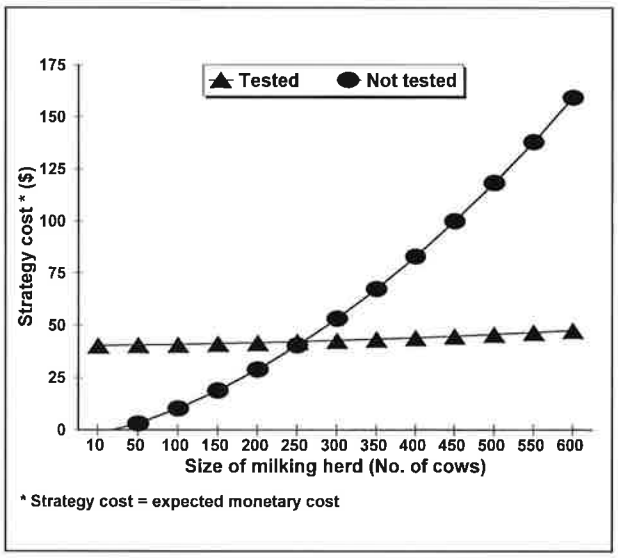


Figure 2—Sensitivity analysis for changes in size of milking herd and its effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

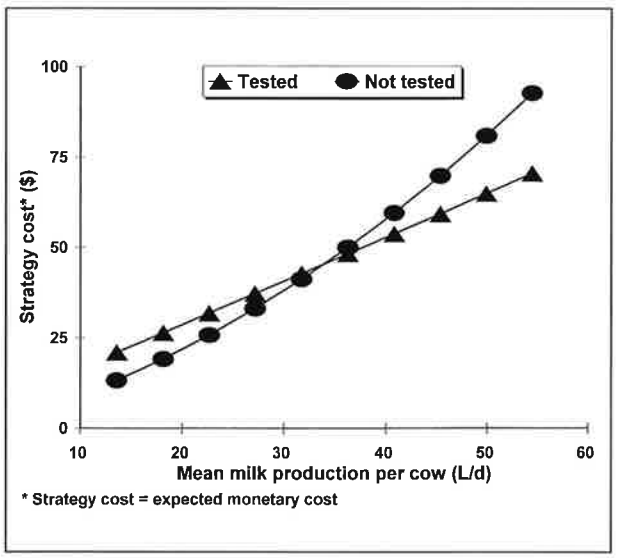


Figure 3—Sensitivity analysis for changes in mean milk production per cow and its effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

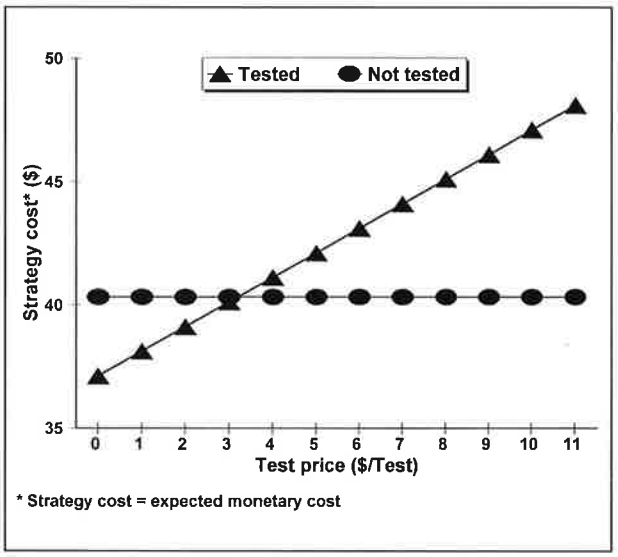


Figure 4—Sensitivity analysis for changes in price of test kit and its effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

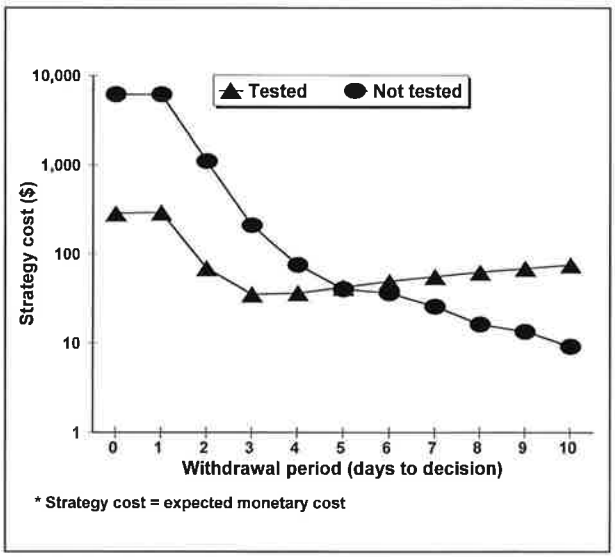


Figure 5—Sensitivity analysis for changes in duration of withdrawal period and its effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

violative residues. On testing bulk-tank samples from each farm, the stochastic simulation suggested a 14% likelihood that samples from a dairy's bulk tank would be identified as the violative component.

For the option of not testing, the simulation model predicted similar likelihoods. In this case, as the proportion of cows to be tested was low, the similarity even carried through to the probabilities of true positives, false positives, true negatives, and false negatives. In other simulations (not shown) in which the treatment dosages were higher or treatments were given for a greater duration, withdrawal periods were shorter, or proportion of cows that were at the end of a withdrawal period was higher, this similarity in probabilities was not detected.

Overall, when the likelihood of events was balanced with their associated costs, producers, on average, would have a cost of $42 every time a cow was chosen to be tested under these conditions (EMC of test). The EMC for not testing was $40, suggesting that a producer in these conditions who opts not to test cows prior to selling their milk would typically have a $40 cost for each cow. Under the conditions of this decision, the optimal path would be the option to not test, because it yielded a cost that was $2 less than the test option.

The sensitivity of the $2 (5%) difference between

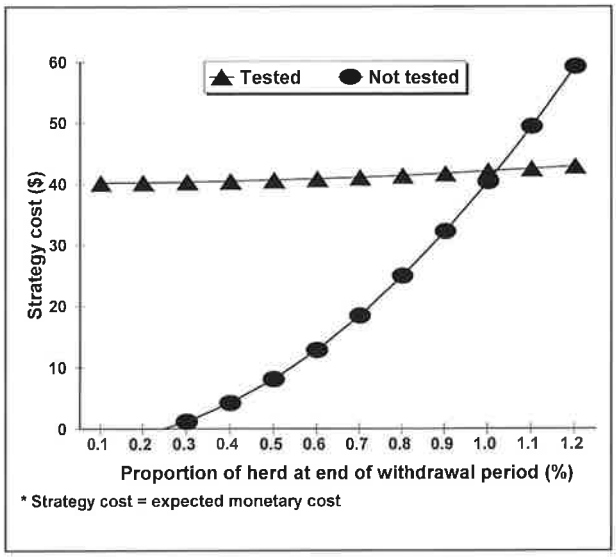


Figure 6—Sensitivity analysis for changes in the percentage of a herd available for testing at the end of a withdrawal period and its effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

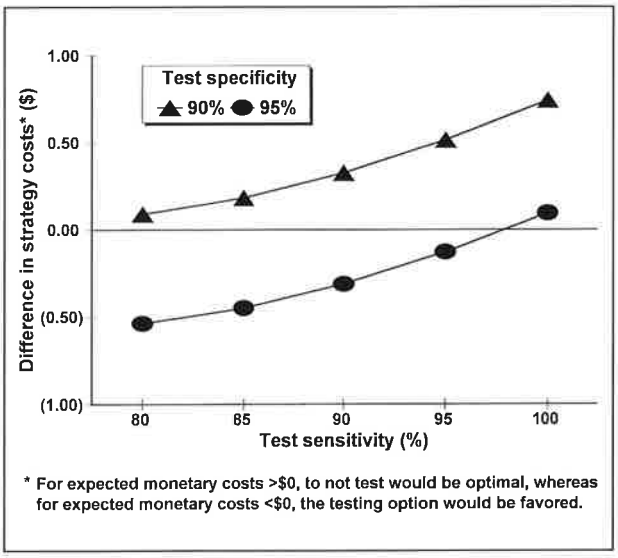


Figure 7—Sensitivity analysis for changes in test sensitivity and specificity and the effects on the decision to submit milk samples from specific cows for antimicrobial residue testing.

Table 2—Decision diagram probabilities and expected monetary costs (EMC) to a producer when a processor uses the same test to confirm positive results for screening tests on milk samples

		Milk ≥ 10 ppb			Tanker-truck sample tests positive			Bulk-tank sample tests positive		
Option	Node EMC	Yes	No	EMC	Yes	No	EMC	Yes	No	EMC
Testing	$41	0.04	0.96	$39	0.02	0.98	$36	0.03	0.97	$25
Not testing	($1)	0.40	0.60	($1)	0.02	0.98	($3)	0.03	0.97	$212

Parentheses indicate a negative cost (ie, positive income for the producer).
ppb = parts per billion.

options was evaluated for changes in cost structure as well as changes in characteristics of the test and herd. In general, the option to not test was favored until 1 of the following breakpoints was reached: herd size exceeded 257 cows (Fig 2), daily milk production exceeded 34 L/cow (Fig 3), test price was < $3.25/sample (Fig 4), withdrawal period was < 5 days (Fig 5), percentage of herd at risk (proportion of herd at end of selected withdrawal period) exceeded 1.0% (Fig 6), or test specificity exceeded 95% (Fig 7).

For the aforementioned sensitivity analyses, we assumed that confirmation of test results was not performed by the processor. A recommendation has been made that processors should repeat the same test on any samples with positive results.[39] The likely efficiency of such a system, given the same test attributes as described previously, would result in a 9% reduction in true positives, compared with the use of the screening test alone, and also would result in a decrease of 81% in false positives. These outcomes were mirrored in the absolute changes of the false negatives and true negatives, respectively. Using a double-testing protocol (ie, screening and confirmatory tests), the option to not test was favored by a $42 margin (Table 2).

Discussion

In this study, we affirmed that recommended withdrawal periods would be sufficient to prevent adulterated milk from entering the food chain after a cow is given a single treatment with PPG. This was true for a labeled withdrawal time of 48 hours, regardless of whether the drug was administered once at the label dosage or once at 4X the label dosage. However, approximately two-thirds of the cows were at risk of having positive test results because of the test's ability to identify PPG concentrations at half the FDA-approved tolerance limit. For cattle receiving multiple treatments of 4X the label dosage, we chose a withdrawal period of 5 days, because this constituted an extra-label use, and, thus, a withdrawal period of more than twice the label recommendation was selected. However, even with this extended withdrawal period, the model indicated that 40% of these cows had violative residues, and an additional 10 to 15% were at risk of having false-positive results even though they had concentrations that were less than tolerance limits. The fact that the initial decision solution we used favored the strategy of not testing treated cows prior to reentry to the milking string supported the findings of Montgomery[15] in which

nearly 9 of 10 surveyed dairies reported they did not test milk from cows prior to allowing them back in the milking string.

One aspect of a decision, which is often not appreciated, is the risk inherent in that decision. Although, on average, the option to not test yielded the lower cost, it also was associated with a risk of delivering the highest costs. This trade-off, lower mean expected costs with the potential for higher costs at the extreme, is not uncommon in business choices. Methods exist to measure a decision maker's relative acceptance of, or aversion to, risk, and to accommodate these characteristics while minimizing risk in an investment system.[44,45] The major analytical method to minimize risk at a given reward level is portfolio analysis, an optimization technique that mixes various investment options, similar to ration-balancing linear programs. Given that the difference between the 2 strategies under the conditions of this examination was minimal ($42 – $40 = $2), this decision point was likely to be unstable; small changes in some of the underlying variables could tip the optimal strategy choice towards the testing option. For that reason, it was deemed appropriate to perform extensive sensitivity testing on the model.

The effects of changes in herd size, milk production, test price, and percentage of herd at the end of a withdrawal period made intuitive sense. For instance, it was logical that even a perfect test could be priced too high for the value of its information. Hence, intuitively there should be a price threshold at which the test becomes unacceptable. In addition, when the increasing amounts of milk placed at risk by not using a testing protocol were considered, we expected that there would be a threshold for herd size, milk production, and percentage of herd to be tested. Likewise, it seemed reasonable that each variable would have a point of indifference (a value at which the decision maker is indifferent to choosing either option). However, intuition is not sufficient for making decisions. Generating actual points of indifference, as was done in this study, are true aids in decision making.

Furthermore, intuition fails when the relationships to be evaluated have complex, nonlinear dynamics. For example, changes in duration of the withdrawal period impacted the relative value of the testing and not testing strategies. Both strategies had extremely high per-cow costs when the withdrawal period was < 3 days, attributable to the increased likelihood of shipping milk from cows with violative residues, being identified as the dairy responsible for the violative residues, and paying the fines and costs for such a violation. The point of indifference was on day 5. An interesting dynamic existed between the EMC of the testing strategy and withdrawal time. For the testing option, the EMC hit a low point (approx $35) on day 3 and then rose to $76 by day 10. The EMC for the not testing strategy did not display this change with time; it continually decreased throughout the evaluated period.

Similarly, the interaction of test sensitivity and specificity did not follow an intuitive pattern. There was a difference between strategy-specific EMC values; a positive difference indicated the EMC for the not testing option was less than the EMC for the testing option (and, was therefore, optimal). Conversely, a negative difference would have indicated that the EMC for testing was less than the EMC for not testing. The testing option was not favored until specificity was ≥ 95%. Paradoxically, improvements in sensitivity actually decreased the test's value to producers, because an increase in test sensitivity works primarily to increase the processor's likelihood of identifying samples from tanker trucks and bulk tanks with antimicrobial residues whether the concentrations of those residues are actually violative or are merely greater than the test's minimum detectable limit.

These sensitivity analyses also highlighted potential conflicts that arise with a residue testing program. For instance, the overall health status of the herd is an important factor in determining the relative value of a testing option (Fig 6). In general, the fewer the number of cows at risk of having violative residues in their milk, the more likely that a testing program's cost will exceed that of an option to not test. In addition, increasing the number of cows given multiple extra-label treatments in a herd increases the importance of testing (Table 1). Programmatically, this suggests that a producer with good health management might choose to not participate in a testing program. This is not different from other disease control programs; the lower the prevalence, the less confidence and value that can be ascribed to testing.[46,47] Because dairy veterinarians, extension agents, and sanitarians are constantly striving to improve the health of a herd, this paradox between improving overall herd-health and the subsequent increase in risk exposure as a result of testing will grow in importance and will have to be addressed in the regulations to avoid conflict and lack of compliance.

Another conflict identified through these sensitivity analyses was related to the proper establishment of withdrawal periods (Fig 5). Even given the imperfect performance of a test, the withdrawal period can be optimized, because there is an economic trade-off between short withdrawal periods and testing. At 1 extreme, when a producer chooses to use short withdrawal periods, it is financially inappropriate not to test, because the likelihood of shipping milk with violative residues and having to pay for the consequences is great. However, at the other extreme, should a producer wait until the last cow is guaranteed to have cleared the drug from her system, the cost of discarded milk would be excessive. Hence, the optimal point of testing will be between these extremes and will balance the risk of shipping milk with violative residues as a result of the use of short withdrawal periods against the direct cost of discarding milk that would result from the use of long withdrawal periods. This requires, however, decision makers to rely on information other than simple drug half-lives and published withdrawal

times, a focus that has been discounted in the literature,[48] although it is still commonly recommended.

The final area of residue testing conflict illustrated in our study concerned the interaction of test sensitivity and specificity with the relative value of the testing strategy. Our analyses (Fig 7) should raise concerns to regulators and producers. First, from the viewpoint of regulators, the fact that test specificity must exceed 90% to be economically viable would suggest that the current system of test evaluation is inadequate and that there are strong economic reasons for a producer to choose against testing. The proxy for specificity used in test evaluations is benchmarked against milk with a zero-concentration of residues, not against milk with residues at or less than tolerance limits. Hence, the ability to identify nonviolative samples (ie, true specificity), has been reported to be ≤ 80% for some sample types.[25] At such specificities, the analysis reported here could not support the choice of testing cows, because the costs associated with false-positive results outweighed the regulatory penalties and fines in our model. From the view of producers, the tendency of the curves to favor the option to not test as sensitivity increases (slope of the curve moves up and to the right), raises concerns with the regulatory concentration on test sensitivity. At the low prevalence that would exist in a well-managed herd, increasing sensitivity would actually lessen the value of a testing program, because increasing sensitivity tends to lower specificity and simply increases the likelihood that a processor will identify tanker-truck contents as having violative residues (ie, positive results), irrespective of whether the milk truly has violative residues.

One way to help decrease the false-positive rate is for processors to retest all samples that had positive results, using the same test. The attraction for processors would be that they would not need to invest in a second, expensive, analytical test. Using the test from this study as the example test, we applied the results to > 1 million samples. The example assumed the 2 tests were independent, which is not the case. However, the assumption of independence yielded the highest possible performance (ie, best improvement possible over a single-test system) of the system. The greater the correlation of the tests, the less performance enhancement is possible; if the tests were perfectly correlated, there would not be a gain from the second test, and the only result would be an increase in system cost of $5 for each retested sample. The example used an initial violative-residue prevalence of 0.063%, the rate reported by the FDA for 1995.[16] Overall, the test-retest system decreased the number of true-positive results from 573 to 521 (9%) for the single-test system versus the test-retest system, respectively, and decreased the number of false-positive results from 189,880 to 36,077 (81%). Although this represents a net gain for producers, it is questionable whether a system that decreased the ability to truly identify samples that have violative residues would be acceptable to regulators or the public.

The test-retest decision-diagram analysis outcomes were determined. For purposes of comparison, the remainder of the decision tree was unchanged for our example. In reality, it is likely that an increased surveillance intensity would include increased fines (reflecting the likelihood that a sample with a positive result has a greater chance of truly having violative residues) and would require that at least some of the increased testing and records costs be passed on to producers. The changes in system efficiency resulting from a test-retest protocol resulted in an EMC for the testing option of a $41 cost, but an EMC for the option to not test resulted in a $1 gain. Hence, whereas the single-testing protocol for the processor yielded a $2 difference in favor of not testing, the test-retest protocol yielded a $42 difference, also in favor of the option to not test. Adapting the test-retest protocol, then, would increase the relative economic advantage of the option to not test. Hence, rational producers, facing a processor who uses a double-testing method, should be even less inclined to test specific cows than when a single-testing protocol is used. If the program's goal is to minimize the likelihood that producers would add contaminated milk to the food chain, then a double-testing program would presumably undercut its own goal.

The model used in our study, similar to all models, has weaknesses. The pharmacokinetic relationships were determined on the basis of data that primarily investigated drug disappearance in healthy cattle. The data in most pharmacokinetic studies confound dose rates with sampling period. In many studies, the higher the dosing rate of a drug, the longer the sampling time frame. This can bias the calculation of half-lives, inappropriately inflating the disappearance times at higher doses. We made the choice to use the input study as the unit of interest by choosing to refrain from weighting the studies on the basis of the number of animals in each study. We, therefore, had the potential to increase apparent variability in the population, because a study with only 2 or 3 subjects would carry the same impact as would a study comprising 20 or 30 animals. The amount of inappropriate variability this might have introduced was not measured.

For our model, we assumed there was not a regulatory memory (eg, a report of a violation does not increase the costs of the next violation). In reality, this is not the case; therefore, our model underestimated the true costs of repeat violations. Our model also assumed milk that was not put into the market channels was completely lost as an income source. Many dairies use such milk as feed for calves, and processors can use such milk for nonfood industrial purposes. Hence, this study likely overestimated the costs for withholding milk from the markets.

Evaluating these analyses, the dangers of oversimplification become apparent. Our common tools used to evaluate the merit of a program from test efficiency measures, mean drug disappearance rates, and including other measures of mean milk production and herd size, all attempt to provide simple answers by factor-

ing out variability. Our attempts to simplify the network of effects that determine the efficiency of a residue-testing program only produce simplistic answers that ignore many interactions inherent in the complex, dynamic events surrounding the decisions at each step. Using stochastic modeling and economic-decision analysis would allow an analyst or decision maker to capture those interactions and account for variability. Not surprisingly, they can lead to complex answers, yet important trends and factors can be elucidated.

Even with relatively inefficient tests, testing programs can be optimized. A factor confounding the problem, however, is that official, yet seemingly inappropriate, measures of test efficiency make program choices an uncertain affair. This results in a national program that is driven by the trend of constantly improving test technology. A situation then develops in which producers and practitioners will always be chasing an ever-moving target limit for detectable residues, forcing the use of increasingly sensitive tests. As indicated in this study, the paradox arises in that these technologically advanced tests are becoming more costly, making less economic sense for this use.

Appendix

Input values for the antimicrobial residue testing decision analytic model

Item or characteristic	Value	Units	Justification for use/ clarification of range available
Test sensitivity	0.910	NA[1]	Federal Drug Administration minimum requirement for sensitivity = 90%[25]
Test specificity	0.810	NA	Minimum not regulated; assumed to be less than sensitivity
Test cost per unit	5.00	$	Informal survey yielded on-farm test prices between $2 and $8
Minimum detection limit	5.0	ppb[1]	Milk and Dairy Beef Quality Assurance Program lists mode for minimum detection limit at 5 ppb[37]
Size of milking herd	250	cows	Approx. US mean herd size = 267 cows, by cow population distribution[49]
Mean daily milk production	31.4	L/cow	Equivalent to approximately 9,600 kg (21,120 lb) of milk during a 305-day lactation; above-average herd[50]
Value of milk	0.22	$/L	Rounded to lower end of published US milk price after accounting for fees and transport[50]
Antibiotic used			Procaine Penicillin G, IM, at dosage of 26,400 U/Kg (12,000 U/lb), multiple times, on basis of clinical experience of senior author as a common treatment protocol
Legal tolerance limit	10.0	ppb	Milk and dairy beef residue prevention protocol[37]
Percentage of herd at end of withdrawal period	1.0	%	Arbitrary
Interval after treatment to test	5.0	days	Arbitrary at 2.5 × the label recommendation for withdrawal period

[1]NA = not applicable; ppb = parts per billion.

[a]Muller M. *Antibiotic tissue residues detected in california dairy cattle: 1988-1990.* Dissertation, Department of Epidemiology and Preventive Medicine, University of California, Davis, 1991.

[b]Slenning BD, Burstein H. Stochastic modeling of pharmacokinetics to measure risk and efficacy of antimicrobic use in dairy cattle (abstr). In: Martin SW, ed. *Proceedings, 6th Symposium International Society of Veterinary Epidemiology and Economics*. Ottawa, Calif: ISVEE, 1991;639.

[c]Statistical Analysis System, SAS Institute Inc, Cary, NC; Statistix, Analytical Software, St Paul, Minn.

[d]Lotus 1-2-3, Release 2.4, Cambridge, Mass.

References

1. Bevill RF. Factors influencing the occurrence of drug residues in animal tissues after the use of antimicrobial agents in animal feeds. *J Am Vet Med Assoc* 1984;185:1124–1126.
2. Egan J, Frayne EJ, O'Connor F. The control of antibiotic residues in liquid milk. *Irish Vet News* 1988;10:30-31, 33.
3. Fuhrmann T. Overview of residue concerns of the dairy industry. *J Am Vet Med Assoc* 1991;198:836–838.
4. Ingersoll B. Dairy dilemma—milk is found tainted with a range of drugs farmers give cattle. *Wall Street Journal* Dec 29, 1989;1.
5. Kaneene JB, Ahl AS. Drug residues in dairy cattle industry: epidemiological evaluation of factors influencing their occurrence. *J Dairy Sci* 1987;70:2176–2180.
6. Riviere JE. Pharmacologic principles of residue avoidance for veterinary practitioners. *J Am Vet Med Assoc* 1991;198:809–816.
7. Vautier HE, Postigo CB. Bovine mastitis and antibiotic residues in milk: risks to public health. *World Anim Rev* 1986;60:41–43.
8. Kaneene JB, Willeberg P. Influence of management factors on the occurrence of antibiotic residues in milk: a case-control study in Michigan dairy herds, with examples of suspected information bias. *Acta Vet Scand Suppl* 1988;84:473–476.
9. McEwen SA, Black WD, Meek AH. Antibiotic residue prevention methods, farm management, and occurrence of antibiotic residues in milk. *J Dairy Sci* 1991;74:2128–2137.
10. McEwen SA, Meek AH, Black WD. A dairy farm survey of antibiotic treatment practices, residue control methods and associations with inhibitors in milk. *J Food Protect* 1991;54:454–459.
11. Sundlof SF, Hentschel AF, Fuhrmann T, et al. Milk and dairy beef ten-point quality assurance program. *Agri-Practice* 1991;12(6):5–12.
12. Clarke RH. The hazard analysis critical control point (HACCP) approach to slaughter quality control in red meat abattoirs, in *Proceedings*.10th Symp World Assoc Vet Food Hygienists 1989;194–198.
13. National Advisory Committee on Microbiological Criteria for Foods. *HACCP principles for food production*. Washington, DC: US Food Inspection Service, 1989.
14. Draft principles and application of the Hazard Analysis Critical Control Point (HACCP) system, in *Proceedings*. Report 25th Session Codex Committee Food Hygiene 1990;75–80.
15. Montgomery KR. A survey of California dairies to determine awareness and implementation of quality assurance and residue avoidance plans. In: *1993 USDA epidemiology symposium*. Knoxville: USDA/APHIS.VS 1993.
16. Milk monitoring with antimicrobial drug screening tests. *FDA Vet* 1996;XI:8.
17. Carlsson A, Bjorck L, Persson K. Lactoferrin and lysozyme in milk during acute mastitis and their inhibitory effect in Delvotest P. *J Dairy Sci* 1989;72:3166–3175.
18. Egan J, Meaney WJ. The inhibitory effect of mastitic milk and colostrum on test methods used for antibiotic detection. *Irish J Food Sci Tech* 1984;8:115–120.
19. TerHune TN, Upson DW. Factors affecting the accuracy of the live animal swab test for detecting urine oxytetracycline and predicting oxytetracycline residues in calves. *J Am Vet Med Assoc* 1989;194:918–921.
20. Charm SE, Chi R. Microbial receptor assay for rapid detection and identification of seven families of antimicrobial drugs in milk: collaborative study. *J Assoc Off Anal Chem* 1988;71:304–316.
21. Haapoja A, Korkeala H. Antimicrobial residues in milk. Comparison of different agar diffusion methods. *Acta Vet Scand* 1984;25:250–259.
22. Jones GM, Seymour EH. Cowside antibiotic residue testing. *J Dairy Sci* 1988;71:1691–1699.
23. Seymour EH, Jones GM, McGilliard ML. Comparisons of on-farm screening tests for detection of antibiotic residues. *J Dairy Sci* 1988;71:539–544.

24. Tritschler JP II, Duby RT, Oliver SP, et al. Microbiological screening tests to detect antibiotic residues in cull dairy cows. *J Food Protection* 1987;50:97–102.

25. Gardner IA, Cullor JS, Galey FD, et al. Alternatives for validation of diagnostic assays used to detect antibiotic residues in milk. *J Am Vet Med Assoc* 1996;209:46–52.

26. Slenning BD. A proposal towards developing a system of test efficiency indices aimed at classifying test quality when tests are used over populations. *Kenya Vet* 1994;18:562–565.

27. Sundlof SF, Riviere JE, Craigmill AL, et al. Computerized food-animal residue-avoidance data bank for veterinarians. *J Am Vet Med Assoc* 1986;188:73–76.

28. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. London: Academic Press Ltd, 1985.

29. Black WD. The use of antimicrobial drugs in agriculture. *Can J Physiol Pharmacol* 1984;62:1044–1048.

30. Booth JM. Intramammary antibiotic preparations and their withholding times. *Vet Rec* 1986;118:34–35.

31. Food and Drug Administration. *FDA and the veterinarian*. Washington, DC: HHS Publications (FDA), 1989;89–6046.

32. Kaneene JB, Coe PH, Smith JH, et al. Drug residues in milk after intrauterine injection of oxytetracycline, lincomycin-spectinomycin, and povidone-iodine in cows with metritis. *Am J Vet Res* 1986;47:1363–1365.

33. Krainock RJ. Prolonged milk residue in 2 cows after subcutaneous injections of penicillin at an extra-label dose. *J Am Vet Med Assoc* 1991;198:862–863.

34. Seymour EH, Jones GM, McGilliard ML. Persistence of residues in milk following antibiotic treatment of dairy cattle. *J Dairy Sci* 1988;71:2292–2296.

35. Slenning BD. Financial analysis of a clinical trial comparing simple estrus detection against estrus detection following prostaglandin-based appointment breeding in a commercial dairy herd in California, USA. *Prev Vet Med* 1994;18:239–257.

36. Slenning BD, Wheeler MB. Risk evaluation for bovine embryo transfer services using computer simulation and economic decision theory. *Theriogenology* 1989;31:653–673.

37. *Milk and dairy beef residue prevention protocol. 1995 veterinarian manual*. Stratford, Iowa: Agri-Education Inc, 1995.

38. *Milk and dairy beef residue prevention protocol. 1995 producer manual*. Stratford Iowa: Agri-Education Inc, 1995.

39. Cullor JS. Antibiotic residue testing: cowside, bulk tank and tanker load. In: *National Mastitis Council, 1995 regional meeting, Harrisburg, Pa*. Arlington, Va: National Mastitis Council, 1995;7–15.

40. Weinstein MC, Fineberg HV, Elstein AS, et al. *Clinical decision analysis*. Philadelphia: WB Saunders Co, 1980.

41. Pitcher PM, Galligan DT. Decision analysis and economic evaluation of the use of the rapid milk progesterone assay for early detection of pregnancy status of cows. *J Am Vet Med Assoc* 1990;197:1586–1590.

42. White ME, Hollis NE. Decision analysis in bovine practice. *Compend Contin Educ Pract Vet* 1982;4:s426–s430.

43. Collins MT, Morgan IR. Economic decision analysis model of a paratuberculosis test and cull program. *J Am Vet Med Assoc* 1991;199:1724–1729.

44. Morrow WEM, Leman AD, William EM, et al. An economic study of lifetime piglet production for sows allocated to treatments designed to improve parity 2 litter size. *Prev Vet Med* 1990;10:105–118.

45. Galligan DT, Ramberg C, Curtis C, et al. Application of portfolio theory in decision tree analysis. *J Dairy Sci* 1991;74:2138–2144.

46. Tyler JW, Cullor JS. Titers, tests, and truisms: rational interpretation of diagnostic serologic testing. *J Am Vet Med Assoc* 1989;194:1550–1558.

47. Martin SW. The interpretation of laboratory results. In: *Vet clin North America: food animal practice*. Vol 4. Philadelphia: WB Saunders, 1988;61–78.

48. Dix LP, Bai SA, Rogers RA, et al. Pharmacokinetics of digoxin in sheep: limitations of the use of biological half-life for interspecies extrapolation. *Am J Vet Res* 1985;46:470–472.

49. Table 14—Cattle and calves, sales & inventory. In: *1992 census of agriculture*. Washington, DC: USDA, 1992.

50. Markets. *Dairy Herd Manage* 1995;32:36–38.