Reliability of three bulk-tank antimicrobial residue detection assays used to test individual milk samples from cows with mild clinical mastitis

Suzanne N. Gibbons-Burgener, DVM, PhD; John B. Kaneene, DVM, MPH, PhD;
James W. Lloyd, DVM, PhD; Joseph F. Leykam, BS; Ronald J. Erskine, DVM, PhD

Objective—To determine the likelihood of false-positive results when testing milk samples from individual cows by use of 3 commercially available assays (Penzyme Milk Test and the SNAP β-lactam and Delvo-SP assays) labeled for use with commingled milk.

Sample Population—Milk samples from 111 cows with mild clinical mastitis.

Procedure—Cows were randomly assigned to the control (no antimicrobials) or intramammary treatment group. Posttreatment milk samples were collected at the first milking after the labeled withholding period or an equivalent time for controls, randomly ordered, and tested twice by use of each assay and once by use of high-performance liquid chromatography. Sensitivity, specificity, and positive and negative predictive values were determined for each assay. Concordance of results for the same sample was assessed for each assay by calculating κ.

Results—Sensitivities of the Delvo-SP and SNAP β-lactam assays were > 90%, whereas the sensitivity of the Penzyme Milk Test was 60%. Positive predictive values (range, 39.29 to 73.68%) were poor for all 3 assays. Concordance of test results was excellent for the SNAP β-lactam and Delvo-SP assays (κ = 0.846 and 0.813, respectively) but was less for the Penzyme Milk Test (κ = 0.545).

Conclusions and Clinical Relevance—Because of the low positive predictive values, these 3 assays may not be useful for detecting violative antimicrobial residues in individual milk samples from cows treated for mild clinical mastitis. However, repeatability of each assay was considered good to excellent. (*Am J Vet Res* 2001;62:1716–1720)

Dairy farmers, veterinarians, dairy manufacturers, and researchers believe it is highly desirable to have at least 1 quick and reliable test for the detection of antimicrobials in milk. Prior to the approval of new drugs for use in lactating cows, pharmaceutical companies must demonstrate that an assay or method exists for detection of their drug in marketable milk; however, these tests are typically applied to commingled milk from a group or herd of cows. The ability to test milk samples from individual cows for residues is essential for the determination of labeled withholding periods. Fortunately, there are FDA-approved medications with withholding periods for treatment of mastitis, the most common disease in cows. However, results of earlier studies[1,2] indicate that farmers depend more on residue testing than labeled withholding periods when deciding to withhold milk from a treated cow. The practice of testing individual cow milk samples off- or on-site is widespread and promoted by the dairy industry. Because there are no rapid assays intended for testing milk from individual cows, it has become commonplace to use approved commingled milk testing assays for this purpose.

Results of numerous studies[3-7] indicate that current assays used for on-farm testing of milk from individual cows for drug residues often yield false-positive results. Thus, caution is warranted in the use of these assays. Although the specificity and sensitivity of these assays have been established for testing commingled milk under controlled laboratory conditions, it is not known whether results are accurate under field conditions. To date, no studies have validated the use of such tests for assessing individual cows' milk. Michigan dairy farmers and milk-handlers most commonly use 3 commercially available tests designed for assessing commingled milk to test individual milk samples for antimicrobial residues. We hypothesized that these assays would yield 10% false-positive results when used for testing milk from cows that were treated for mild clinical mastitis. To test this hypothesis, we included only cows with naturally occurring mastitis, randomized mastitis treatment, excluded cows treated with antimicrobials in an extralabel manner, and used quantitative gold standards to evaluate assay reliability. The objective of the study reported here was to determine the likelihood of false-positive results when testing milk samples from individual cows by use of 3 commercially available assays labeled for use with commingled milk.

Materials and Methods

Study design—A longitudinal experimental study of cows with mild clinical mastitis was conducted to evaluate the reliability of 3 commercially available tests (the Penzyme Milk Test[a] and the SNAP β-lactam[b] and Delvo-SP[c] assays) when used to assess individual cow milk samples for antimicrobial residues.

Received Sep 13, 2000.
Accepted Mar 12, 2001.
From the Population Medicine Center (Gibbons-Burgener, Kaneene, Lloyd) and Department of Large Animal Clinical Sciences (Gibbons-Burgener, Kaneene, Lloyd, Erskine), College of Veterinary Medicine, the Department of Agricultural Economics, College of Agriculture and Natural Resources (Lloyd), and the Department of Biochemistry, College of Natural Sciences (Leykam), Michigan State University, East Lansing, MI 48824. Dr. Gibbons-Burgener's present address is the Wisconsin Veterinary Diagnostic Laboratory, University of Wisconsin-Madison, 6101 Mineral Point Rd, Madison, WI 53705.
Supported by the American Veterinary Medical Foundation, Michigan Animal Initiative Research Fund, and the Population Medicine Center.
Address correspondence to Dr. Kaneene.

Sample size required for estimating a single proportion (ie, false-positive results/all results) is less than that required for evaluating potential associations with risk factors. By estimating that 10% of the tests would yield a positive result and allowing a 5% margin of error ($\alpha = 0.05$), the required sample size for estimating the likelihood of false-positive results was 138 tests. A test was defined as 1 analysis of each posttreatment sample by use of each assay.

Case definition—A case of mild clinical mastitis was defined as a cow with visibly abnormal milk from an affected quarter; an affected quarter may have been enlarged or reddened. Cows were specifically excluded from the study if they had received antimicrobial treatment for any reason within the 30 days preceding entry to the study, were previously entered in the study (a repeat case), had a concurrent illness requiring antimicrobial treatment, or had severe mastitis that required systemic administration (IV, IM, or SQ) of antimicrobials. Dairy farmers or veterinarians initially identified affected cows. Eight farms participated in the study. Of 111 cows initially enrolled in the study, 92 remained through the posttreatment sample collection (83% case retention rate).

Treatment groups—After diagnosing mild clinical mastitis and collecting the pretreatment milk sample, the dairy farmer entered the cow's identification number in chronologic sequence on the provided data sheet on which the randomly assigned treatment group was indicated. Forty five of the 92 cows (48.9%) remaining in the study were assigned to the treatment group, whereas 47 (51.1%) were assigned to the control group. Cows assigned to the antimicrobial treatment group were treated according to label directions with an FDA-approved **intramammary (IMM)** antimicrobial therapy selected by either the producer or veterinarian. Twenty-six (57.8%) received pirlimycin, 9 (20%) received hetacillin, and 10 (22.2%) received cephapirin. Secondary nonantimicrobial medications (eg, oxytocin, flunixin meglumine) were also administered to some cows. Cows assigned to the control treatment group received either no treatment, nonantimicrobial medications, or an IMM infusion with saline (0.9% NaCl) solution.

Sample collection—Two samples (pre- and posttreatment) were collected from each cow by the dairy farmer. The pretreatment sample was collected after mastitis was diagnosed but before treatment was initiated. For cows in the antimicrobial treatment group, the posttreatment sample was collected the first time cows were milked following completion of the labeled withholding period. The timing for collection of the posttreatment sample from a cow in the control group was determined by using the same withholding period as the most recently treated cow. Administration of drugs other than antimicrobials may have required a variety of actual withholding periods prior to shipment of milk from the farm. Producers observed these recommended withholding periods and instructions prior to including milk in the bulk tank.

Before collection of samples, foremilk from each quarter was discarded. An 80-ml composite milk sample comprising approximately 20 ml of milk from each quarter was then collected by hand and stored at 8 C. Samples were retrieved and transferred to the laboratory within 4 to 48 hours of collection. In the laboratory, four 5- to 8-ml aliquots were transferred to plastic vials and stored at –70 C for antimicrobial residue analyses. We believed the use of –70 C instead of refrigeration or –10 C to store these aliquots would reduce the risk of antimicrobial degradation.[d]

Antimicrobial residue analyses—Milk samples were thawed in an ice-water bath and vortexed briefly. Approximately every 3 months, pretreatment samples were randomly ordered in batches and tested once with each assay. Likewise, each posttreatment sample was randomized twice and tested twice. Thus, to yield the sample size necessary to estimate the likelihood of false-positive results, we ran at least 138 tests for each assay with posttreatment samples. Except for the use of individual and thawed milk samples, the 3 commercial assays were performed according to each manufacturer's recommendations. Each assay qualitatively detects antimicrobial residues by use of a different mechanism.[8] The Delvo-SP assay assesses microbial growth inhibition, whereas the SNAP β-lactam assay identifies antimicrobials by use of enzyme-linked receptor binding and the Penzyme Milk Test by use of an enzymatic colorimetric technique. The same person (SNGB) visually interpreted assay results while blinded to treatment group.

Antimicrobial residues were also measured once in each sample by use of **high-performance liquid chromatography (HPLC)**. As suggested by results of several previous studies,[8-10] HPLC is the gold standard to identify and quantify antimicrobial residues in milk samples. The specific extraction and detection methods have been described.[11,e] By comparing the concentration of each antimicrobial determined by use of HPLC with the FDA-tolerance level and assay detection limits (**Appendix**), we were able to determine whether each assay should have detected the residue and whether a given residue was considered violative (ie, greater than the FDA-tolerance level). Because the SNAP β-lactam assay and Penzyme Milk Test are not indicated for the detection of pirlimycin, if a sample containing pirlimycin residues yielded positive results by use of either of these assays, we considered these false-positive results.

Statistical analyses—The reliability of each of the residue detection assays was expressed in terms of sensitivity, specificity, **positive predictive value (PPV)**, and **negative predictive value (NPV)**, using the following equations:

$$\text{Sensitivity} = (\text{No. of positive results by use of both HPLC and assay/No. of positive results by use of HPLC}) \times 100$$

$$\text{Specificity} = (\text{No. of negative results by use of both HPLC and assay/No. of negative results by use of HPLC}) \times 100$$

$$\text{PPV} = (\text{No. of positive results by use of both HPLC and assay/ No. of positive results by use of assay}) \times 100$$

$$\text{NPV} = (\text{No. of negative results by use of both HPLC and assay/ No. of negative results by use of assay}) \times 100$$

Reliability statistics were calculated first by use of the specific assay's detection limits and then by use of FDA-established tolerance levels for each of the antimicrobials.

Concordance of results of the first and second tests for the posttreatment samples was assessed by calculating κ for each of the 3 commercial assays, using the formula[12]:

$$\kappa = (P_o - P_e)/(1 - P_e)$$

where P_o = the observed probability of concordance between 2 tests and P_e = the expected probability of concordance between 2 tests.

Results

None of the pretreatment samples were found to have antimicrobial residues that would have hindered the interpretation of posttreatment assay results. We were unable to interpret the posttreatment assay results for samples from 3 cows. In addition, occasionally a milk sample would produce no visual result on a given assay; these tests were not included in the evaluation of

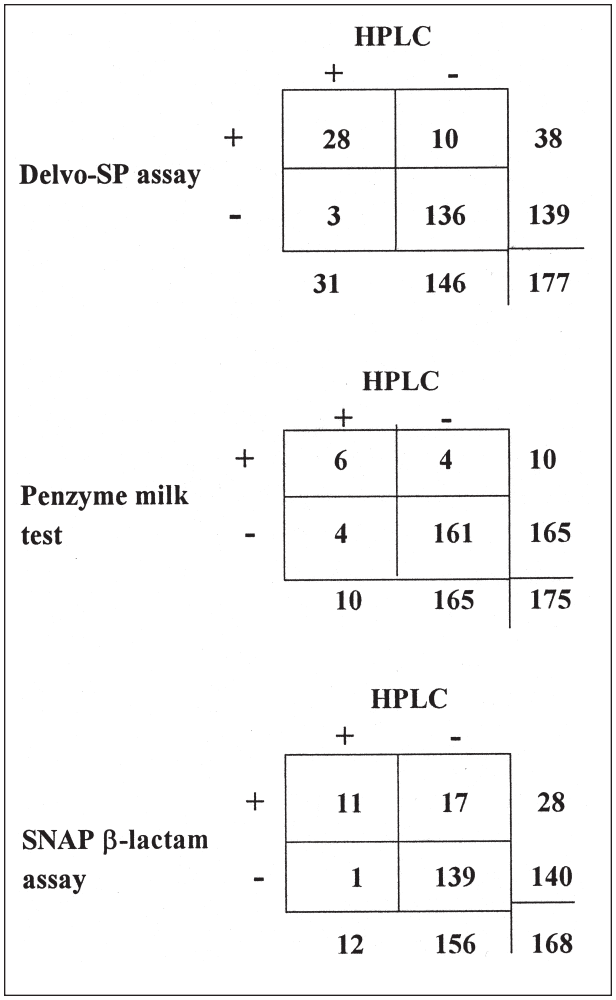

Figure 1—2 X 2 charts comparing results of 3 commercial assays for detection of antimicrobial residues (ampicillin, cephapirin, or pirlimycin) in individual milk samples with results of high-performance liquid chromatography (HPLC). A positive result was defined as a concentration greater than or equal to the detection limit of each assay.

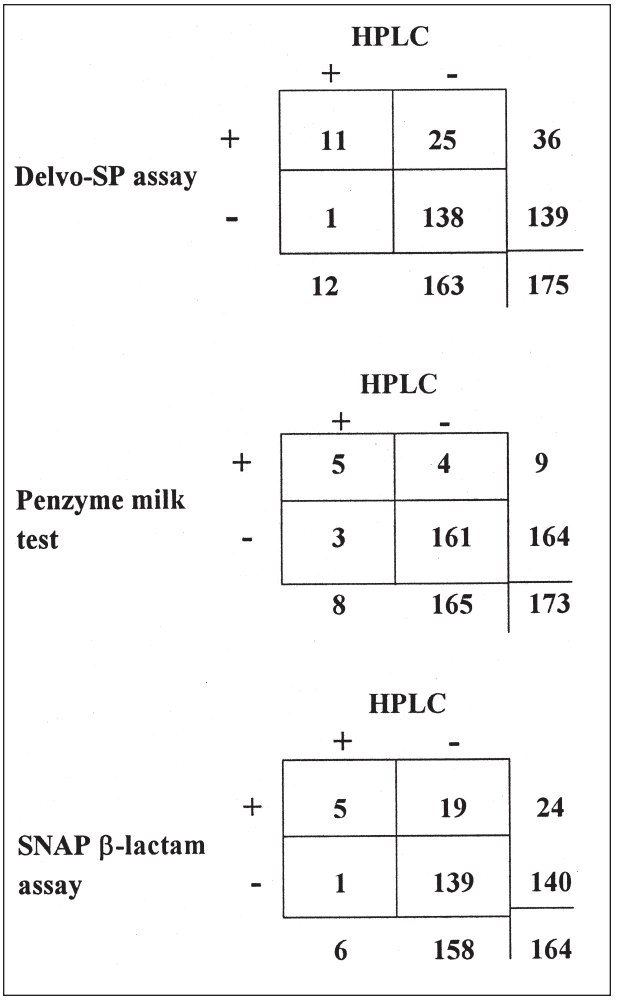

Figure 2—2 X 2 charts comparing results of 3 commercial assays for detection of antimicrobial residues (ampicillin, cephapirin, or pirlimycin) in individual milk samples with results of HPLC. A positive result was defined as a concentration greater than the FDA-established tolerance level for each antimicrobial.

Table 1—Characteristics of 3 commercially available assays for detection of antimicrobial residues at greater than or equal to the minimum detection limit of each assay in individual milk samples

Assay (No. of tests*)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
SNAP β-lactam (168)	91.67 (61.52–99.79)	89.10 (83.12–93.52)	39.29 (21.50–59.42)	99.29 (96.08–99.98)
Penzyme (175)	60.0 (26.24–87.84)	97.58 (93.91–99.34)	60.0 (26.24–87.84)	97.58 (93.91–99.34)
Delvo-SP (177)	90.32 (74.25–97.96)	93.15 (87.76–96.67)	73.68 (56.90–86.60)	97.84 (93.82–99.55)

Data are reported as proportion (95% confidence interval).
*Test was defined as 1 analysis of each sample by use of each assay; samples were analyzed in duplicate.
PPV = Positive predictive value. NPV = Negative predictive value.

each assay. Consequently, of the potential 178 tests performed for each assay, results of 175, 168, and 177 were analyzed for evaluation of the Penzyme Milk Test and SNAP β-lactam and Delvo-SP assays, respectively.

Results of assays were evaluated, using both the detection limit of each assay and FDA-tolerance levels. Determination of the frequency distributions for these results indicated a low number of positive results regardless of assay used (**Fig 1 and 2**). Milk samples from 23 cows had concentrations of antimicrobial residues that were detectable by at least 1 of the commercial assays (**Table 1**). However, samples for only 6 of those 23 cows had violative concentrations. The sensitivities of the Penzyme Milk Test and SNAP β-lactam and Delvo-SP assays were 62.5, 83.33, and 91.67%, respectively, for detection of samples containing violative residues (**Table 2**).

Concordance of results of duplicate tests run on the same posttreatment samples was excellent for the SNAP β-lactam ($\kappa = 0.846$) and Delvo-SP ($\kappa = 0.813$) assays. Concordance was good for results of the Penzyme Milk Test ($\kappa = 0.545$).

Table 2—Characteristics of 3 commercially available assays for detection of antimicrobial residues at greater than the FDA-established tolerance level in individual milk samples

Assay (No. of tests*)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
SNAP β-lactam (164)	83.33 (35.88–99.58)	87.97 (81.86–92.60)	20.83 (7.13–42.15)	99.29 (96.08–99.98)
Penzyme (173)	62.5 (24.49–91.48)	97.58 (93.91–99.34)	55.56 (21.20–86.30)	98.17 (94.74–99.62)
Delvo-SP (175)	91.67 (61.52–99.79)	84.66 (78.20–89.82)	30.56 (16.35–48.11)	99.28 (96.06–99.98)

See Table 1 for key.

Discussion

Unlike the protocol used in some earlier studies,[3,13,14] this study was designed to simulate the collection of milk for residue testing as commonly performed on farms. Samples were collected from cows treated for mastitis; samples were not spiked with known quantities of antimicrobials. Instead, milk from these affected cows had been withheld from the bulk tank as a result of treatment with antimicrobials, and posttreatment samples were only collected after appropriate withholding periods were observed. As a consequence of the present study's design, only farms with managers or owners willing and capable of following the study protocol were included. These farms had a low incidence of mild clinical mastitis, thus limiting the number of accessible cases. However, we were able to collect enough samples to achieve the minimum sample size necessary to provide adequate precision in the calculation of reliability statistics.

Pretreatment milk samples were also collected, but testing of these samples was done only to ensure the absence of residues that could interfere with interpretation of the posttreatment test results. Because clinical mastitis was initially diagnosed in all cows entered in the study, and because abnormal milk is a contraindication for use of the Penzyme Milk Test or Delvo-SP assay, it was inappropriate to use pretreatment test results to evaluate assay reliability. In addition, it is rare that milk is tested on-site for residues prior to treatment. By blindly testing the posttreatment samples twice and comparing results between tests by use of the κ statistic, we found strong evidence that outcome repeatability was good to excellent for all 3 assays. This finding is consistent with the report of a discussion forum,[10] in which the use of 2 assays in series was recommended rather than simply repeating the same assay on samples that yielded presumptive positive results. However, in the present study, the assays could not be evaluated in series, because each assay detected different antimicrobials with different detection limits. Interpretation of results of such assays performed in series may be difficult. This was particularly evident in that only the Delvo-SP assay was reported to detect pirlimycin residues. As recommended by the discussion forum,[10] it would be most beneficial to use a sensitive assay initially, followed by a specific second assay for those samples with positive results. The 3 assays we evaluated had similar detection limits for ampicillin and cephapirin and, thus, provided little improvement of specificity to distinguish false-positive results.

Sensitivities were similar between the SNAP β-lactam and Delvo-SP assays. The Penzyme Milk Test yielded more false-negative results, which, combined with a low prevalence of detectable residues, had a profound effect on sensitivity. Specificities were comparable among assays. Predictive values are better indicators of what a dairy farmer may encounter when deciding to discard or sell the milk from a tested cow. The PPV is the likelihood that a positive assay result truly identifies a sample with an antimicrobial residue concentration greater than or equal to the detection limit of the assay or greater than the FDA-established tolerance level. Each of the assays we evaluated had PPV less than expected given the good specificities. This may have been a result of the low prevalence of detectable residues in the sample population. The PPV of the SNAP β-lactam assay was low, which may have been attributable in part to an undocumented cross-reactivity with pirlimycin residues. Six of the 17 false-positive results were recorded for samples with pirlimycin residues detected by use of HPLC. We propose that the enzyme conjugate of the β-lactam assay may bind not only β-lactams but also pirlimycin. If the assay had been labeled for detection of pirlimycin residues, PPV would have increased from 39 to 61%. Other potential factors for false-positive or -negative results, such as treatment received, were considered. Negative predictive values for each assay were excellent, because false-negative results were rare. Users of these assays should feel confident in a negative result if the appropriate assay is performed to test for residues of the administered antimicrobial agent.

We found that 6 of the 45 cows treated with an approved antimicrobial according to label directions had violative antimicrobial residues even after the withholding period was observed. However, the FDA-established tolerance levels are set for commingled or bulk-tank milk samples; milk from an individual cow would be diluted in such samples. Because these assays only measure residues qualitatively, the exact concentration of antimicrobials in milk cannot be determined from a positive assay result. The conservative approach for handling an individual milk sample that yields a positive result by use of 1 of the 3 assays evaluated in the present study is to discard the milk and retest at a later time. However, if we had followed this practice, milk from 17 cows may have been unnecessarily discarded. Because of the low PPV of these assays and because prevalence of violative antimicrobial residues may be low in a given population of cows, the usefulness of any of these 3 assays in deciding the fate of milk from individual cows receiving treatment for mastitis is highly questionable, particularly if a producer wishes to minimize the quantity of milk discarded unnecessarily.

[a]Penzyme milk test, Cultor Food Science Group, New York, NY.
[b]SNAP β-lactam assay, IDEXX Laboratories Inc, Westbrook, Me.

[c]Delvo-SP assay, Gist Brocades Food Ingredients Inc, Menomonee Falls, Wis.

[d]Walker R, Microbiology Laboratory, Animal Health Diagnostic Laboratory, Michigan State University, East Lansing, Mich: Personal communication, 1997.

[e]Gibbons-Burgener SN. Identification and quantification of ampicillin, cephapirin and pirlimycin in cows' milk using high performance liquid chromatography and fluorescence detection. In: *An epidemiological study of antimicrobial residues detected in Michigan cows' milk*. PhD dissertation, Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, Mich, 2000.

Appendix

Tolerance levels established by the FDA for specific antimicrobials in marketable milk and visual minimum detection limits of 3 commercial assays designed to measure antimicrobial residues in commingled milk samples

Variable	Ampicillin*	Cephapirin	Pirlimycin
Tolerance level (ppb)	10	20	400
Detection limit (ppb)			
SNAP β-lactam	4–6	2	NA
Penzyme	4–6	4–8	NA
Delvo-SP	4	5	50–200

*Ampicillin is the immediate product of hetacillin metabolism.
ppb = Parts per billion. NA = Not available.

References

1. 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.

2. Gibbons-Burgener SN, Kaneene JB, Lloyd JW, et al. Evaluation of certification in the Milk and Dairy Beef Quality Assurance Program and associated factors on the risk of having violative antibiotic residues in milk from dairy farms in Michigan. *Am J Vet Res* 1999;60:1312–1316.

3. Andrew SM, Frobish RA, Paape MJ, et al. Evaluation of selected antibiotic residue screening tests for milk from individual cows and examination of factors that affect the probability of false-positive outcomes. *J Dairy Sci* 1997;80:3050–3057.

4. Cullor JS. Tests for identifying antibiotic residues in milk: how well do they work? *Vet Med* 1992;87:1235–1241.

5. Cullor JS, van Eenennaam A, Gardner I, et al. Performance of various tests used to screen antibiotic residues in milk samples from individual animals. *J AOAC Int* 1994;77:862–870.

6. Sischo WM, Burns CM. Field trial of four cowside antibiotic-residue screening tests. *J Am Vet Med Assoc* 1993;202:1249–1254.

7. Van Eenennaam AL, Cullor JS, Perani L, et al. Evaluation of milk antibiotic residue screening tests in cattle with naturally occurring mastitis. *J Dairy Sci* 1993;76:3041–3053.

8. Mitchell JM, Griffiths MW, McEwen SA, et al. Antimicrobial drug residues in milk and meat: causes, concerns, prevalence, regulations, tests, and test performance. *J Food Protect* 1998;61:742–756.

9. Anderson KL, Moats WA, Rushing JE, et al. Detection of milk antibiotic residues by use of screening tests and liquid chromatography after intramammary administration of amoxicillin or penicillin G in cows with clinical mastitis. *Am J Vet Res* 1998;59:1096–1100.

10. Gardener 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.

11. Moats WA, Romanowski RD. Multiresidue determination of β-lactam antibiotics in milk and tissues with the aid of high-performance liquid chromatographic fractionation for clean up. *J Chromatogr A* 1998;812:237–247.

12. Rosner BA. Hypothesis testing: categorical data. In: Kugushev A, ed. *Fundamentals of biostatistics*. 4th ed. London: International Thomson Publishing, 1995;423–426.

13. Harik-Khan R, Moats WA. Identification and measurement of β-lactam antibiotic residues in milk: integration of screening kits with liquid chromatography. *J AOAC Int* 1995;78:978–986.

14. Halbert LW, Erskine RJ, Bartlett PC, et al. Incidence of false-positive results for assays used to detect antibiotics in milk. *J Food Protect* 1996;59:886–888.