Abstract: Serodiagnostic test results do not always predict the status of an animal as might be expected. When few false-negative and few false- positive test results are reported for a test (high test sensitivity and specificity), the assumption is that the test is a very accurate predictor of infection/disease status. This assumption is correct if disease prevalence is high. However, when disease prevalence decreases to, for instance, 0.1% such as may be seen after several years of a vaccination campaign, a test having sensitivity of 99% and specificity of 99% becomes a poor predictor of infected animals. In this scenario, a positive test result will be wrong 91% of the time. A negative test result, however, virtually always will correctly identify non infection animals when prevalence of infection remains low.

How well do serodiagnostic tests predict the infection or disease status of cats?

Richard H. Jacobson, PhD

Serodiagnostic test results do not always predict the disease status of an animal as might be expected. When few false-negative and few false-positive test results are reported for a test (high test sensitivity and specificity), the assumption is that the test is a very accurate predictor of infection/disease status. This assumption is correct if disease prevalence is high. However, when disease prevalence decreases to, for instance, 0.1% such as may be seen after several years of a vaccination campaign, a test having sensitivity of 99% and specificity of 99% becomes a poor predictor of infected animals. In this scenario, a positive test result will be wrong 91% of the time. A negative test result, however, virtually always will correctly identify noninfected animals when prevalence of infection remains low.

In veterinary medicine, the differential diagnosis is usually influenced substantially by laboratory test results. In telephone consultations with veterinarians, questions are often raised about the reliability of diagnostic test results and the inferences that can be made from them. When a test result is reported by a laboratory as positive at a given titer, can one always conclude that the animal is infected with the agent in question? When laboratories or companies provide an interpretation statement for test results but give no indication of the likelihood that the results may be errant, does a positive test result always reflect an infected/diseased animal and a negative test result signify that the animal is not infected? When a manufacturer claims, for instance, 97% specificity and 97% sensitivity for an in-house test kit, can it be surmised that an incorrect result will be obtained only about 3 of 100 times? The answer to all of these questions is no.

To further confuse the interpretation of serologic test results, the terminology used to describe the test's expected performance or interpretation of test results may have several meanings. The terms sensitivity and specificity are often used erroneously to predict whether a positive or negative test result will reflect disease status of the animal.

Table 1—Description of a hypothetical population of 1,000 animals of known infection status, in terms of test result sensitivity and specificity

Infection status	Serodiagnostic test result	
	No. of observations	Classification
100 Infected/diseased animals	90 tested Positive	90% Sensitive
	10 tested Negative	False negative
900 Noninfected animals	810 tested Negative	90% Specific
	90 tested Positive	False positive

Sensitivity—The proportion of infected/diseased animals that test positive; Specificity—The proportion of uninfected animals that test negative.

They do not. Sensitivity and specificity (Table 1) are intrinsic characteristics of the test and reflect the degree of meticulousness with which the test is done. The prevalence of infection/disease also must be known or estimated before one can determine the probability that a test result accurately reflects the infection/disease status of the animal. But even when these factors are known, the statistical terminology that describes the test may frustrate those not versed in statistical inferences, leading to further misunderstandings in interpreting the test's performance.

The veterinarian knows intuitively that, for a cat with obvious signs of FeLV infection, a positive test result confirms the diagnosis. Alternatively, for a healthy cat having a history that is completely inconsistent with exposure to FeLV, the veterinarian intuitively begins to question a positive test result and usually seeks confirmation of the diagnosis through additional testing. Although intuition gained through practice is helpful, it would be desirable to have a simple scientific basis for concluding what we feel to be intuitively correct. Accordingly, the veterinarian needs to know, in direct and unambiguous terms, the capacity of a test result to predict the infection or disease status of the animal. Therefore, the implications of a given serotest result will be reviewed herein, using statistics applied in a simple and intuitive manner, and some precautions in regard to interpreting test results will be offered.

Criteria for evaluation of serodiagnostic test results—If a perfect test could be devised, all infected

		Sensitivity->	99%								70%							
		Specificity->	99%	98%	95%	90%	80%	70%	50%	25%	99%	98%	95%	90%	80%	70%	50%	25%
D i s e a s e	p r e v a l e n c e	0.1%	9	5	2	1	0.5	0.3	0.2	0.1	7	3	1	0.7	0.4	0.2	0.1	0.1
		1%	50	33	17	9	5	3	2	1	41	26	12	7	3	2	1	1
		5%	84	72	51	34	21	15	9	7	78	65	42	27	18	11	7	5
		10%	92	84	68	52	35	27	18	13	88	80	61	44	28	21	13	9
		20%	96	93	83	71	55	45	33	25	95	90	78	63	47	37	26	19
		30%	98	96	89	81	68	59	46	36	97	94	86	75	60	50	38	29
		40%	98	97	93	87	77	69	57	47	98	96	90	82	70	61	48	38
		50%	99	98	95	91	83	77	66	57	99	97	93	88	78	70	58	48

*Instructions: Estimate the disease prevalence of the population of animals from which the patient is derived and determine sensitivity and specificity of the test from published information or data from laboratory/company. Find the prevalence figure and read across to the probability value under the test's specificity. Note: Sensitivities of only 99% and 70% are shown because sensitivity has a limited effect on the ability of a positive test result to identify an infected animal. Interpolate for other sensitivities: for example, if the test's sensitivity is 85% (not shown), the specificity is 95%, and the disease prevalence is estimated at 5%, the chance that the test is accurate is about 47% (halfway between 51 and 42%).

animals would test positive (100% sensitivity) and all noninfected animals would test negative (100% specificity). Because of limitations in current diagnostic test technology, errors introduced by technicians, and biological variables among animals that are the subjects of the tests, it has been impossible to devise the perfect test. We must, therefore, be satisfied with test results that merely predict the infection status of the animal, knowing that there will be several sources of error in the prediction.[a] The question then is how to assess a test result for its reliability (its degree of "correctness," also known as the predictive value of the test). Three essential criteria need to be assessed when determining the reliability of a test result:

(1) What is the estimated prevalence of the infection/disease for a population of animals that represent the animal in question? For instance, one would expect multiple-cat households to have a higher prevalence of FeLV infection than do single-cat households.

(2) What is the sensitivity and specificity of the test? Were these values properly determined experimentally? Were the samples, used in determining these criteria, obtained from animals that had proven infection status? To what extent did the technician contribute to reduced test sensitivity and specificity? Information on these factors is usually available from the company producing the test kit or the laboratory conducting the tests.

(3) What is the reproducibility of the test when comparing results between animals and between different runs of the assay? Is the test "robust" (ie, does it give the same results even when the technician was relatively careless) or does it require exquisite care to obtain the same result in 2 successive runs of the test?

[a]Estimates of diagnostic test reliability (predictive values of test results) are given as single values. Statistically, there is an element of error associated with each point estimate that ordinarily would be accounted for by calculating confidence intervals (ci). We have chosen to not include the ci to make the illustrations less confusing. The reader should thus be aware that the probability values in the tables are approximations of the given point estimates.

Positive test result: The effect of prevalence of infection/disease on its interpretation—Of the 3 aforementioned elements, probably the most neglected and misunderstood is the effect of infection/disease prevalence on the reliability of a test result. If the infection/disease prevalence is high, even a fairly inefficient diagnostic test will have a reasonable likelihood of accurately predicting the status of the animal. For instance, if 40% of the animals are infected in the population, and the test has sensitivity of only 70% and specificity of 80%, a positive test result is still likely to detect about 70% of infected animals (Table 2). Should infection/disease prevalence decrease, for instance, to 0.1% as a result of an effective vaccination campaign, we intuitively know that a positive test result is more suspect for an animal from such a population than when prevalence is high. Indeed, even a very improved test (99% sensitivity and 99% specificity) will give a greater number of errant than correct results when the prevalence is very low. In fact, only about 1 of 10 positive test results will accurately predict an infected animal: actually, 91% of those positive test results will be erroneously positive (Table 2). Although a more specific and sensitive test was used in the latter case, a decrease in disease prevalence had a marked detrimental effect on the ability of a positive test result to predict infected animals. So, when infection/disease prevalence is low, confirmatory testing is helpful in clarifying the differential diagnosis.

The reasons for the effect of decreasing infection/disease prevalence on reliability of the positive test result may not be clear; an example may clarify the point. For purposes of this example, assume that the test has been properly developed by the laboratory or manufacturer and that the technician introduced no errors in conducting the test. How reliable is a positive test result if the estimated prevalence of infection/disease in the population is 1% and the test kit has a known sensitivity and specificity of 90%? For a population of 1,000 cats that are representative of the cat in question:

(1) Of 10 infected cats (1% of 1,000 = prev-

A	B	C	D	E	F	G	H	I	
			Infected			Noninfected			
Size of population	Estimated infection or disease prevalence	Sensitivity of test	No. of animals	No. that test positive (true pos)	Specificity of test	No. of animals	No. that test positive (false pos)	Total no. of positive test results	Reliability of positive test result
			A × B	C × D		A − D	(100% − F) × G	E + H	(E ÷ I) × 100
1,000	1%	90%	10	9	90%	990	99	108	8.3%
1,000	30%	90%	300	270	90%	700	70	340	79%
1,000	1%	98.3%	10	9.8	98.3%	990	16.8	26.6	37%
1,000	30%	98.3%	300	275	98.3%	700	11.9	287	96%
1,000*									

*Use this row to insert values to determine specific reliability of a positive test result.
Steps in calculation: (1) Arbitrarily choose a population size of 1,000 animals and insert in the table above.
(2) Estimate the infection/disease prevalence in the population that represents the animal you are testing and insert in the table above.
(3) Insert sensitivity and specificity of test (provided by company or published in literature).
(4) Calculate reliability (known as predictive value to statisticians) of a positive test result.

alence), 9 will test positive (because the test's sensitivity is 90%).

(2) Of 990 noninfected cats, 891 will test negative (90% specificity) leaving 99 noninfected cats that will have erroneously positive test results.

(3) Although only 10 cats of 1,000 are actually infected, the test gave positive results for 108 cats (9 positive test results from infected cats and 99 erroneously positive tests from noninfected cats).

(4) Thus, only 8.3% (9/108) of positive test results will correctly identify an infected cat. Obviously, 91.7% of the positive test results will be erroneous (false positive).

The calculated results (Table 3; lines 1 and 2) reveal the considerable impact of increasing infection/disease prevalence on the capacity of a positive test result to predict an infected cat. When the infection/disease prevalences of 1 and 30% are compared, the probability that a positive test result will identify an infected cat increases from 8.3% to 79.4%. Table 3 provides examples of calculated test reliability and is intended principally for those who wish to follow the logic and calculations in determining the probability with which a positive test result will predict infection status (the predictive value of a positive test result).

Negative test result: The effect of prevalence of infection/disease on its interpretation—Using the same scenario as outlined previously, a decrease in prevalence of infection/disease from 30% to 1% has a minimal effect upon the capacity of a negative test result to detect cats that are not infected (Table 4). In our example:

(1) Of 10 infected cats, 9 will test positive (90% sensitivity), leaving 1 infected cat that will test erroneously negative.

(2) Of the 990 noninfected cats, 891 will test negative (90% specificity).

(3) Although 990 cats of 1,000 are actually noninfected, the test result was negative for 892 cats (891 negative test results from noninfected cats and 1 erroneously negative test result from an infected cat).

(4) Thus, 99.9% (891/892) of the negative test results will correctly identify noninfected cats.

When the estimated infection/disease prevalence is 30%, a negative test result still detected 95% of the noninfected cats (Table 5). When the prevalence of infection/disease is >5% for a test having a sensitivity of 99.9%, even if test specificity decreases to 20%, >99% of negative test results will be accurate. A negative test result is thus more resistant to the effect of decreasing infection/disease prevalence and is reliable in predicting that a cat does not have infection/disease. A worksheet with instructions for its use is provided (Table 5) to calculate the predictive values of a negative test result.

Hidden factors that may affect test sensitivity—The proportion of infected/diseased animals that test positive is the test's sensitivity. In the aforementioned examples, the reliability of a test result was not only a function of disease prevalence, but also was greatly dependent on the sensitivity and, to a lesser extent, the specificity of the test.

Unfortunately, the way in which sensitivity and specificity are determined is variable because it depends on the test developer's choice of standards. In development of a test, a criterion is chosen, against which the serotest results are evaluated. The most criterion standard is a large group of sera from animals of absolutely proven infection status that represent the general population of animals. Calculation of test sensitivity, using such a standard, is the most accurate method for determining diagnostic sensitivity.

During test development, it is often difficult to obtain enough serum samples representing animals with proven infection status. Test developers may select sera that have been evaluated by use of another laboratory test. When a test is developed, for which the criterion is another test result (not the proven infection status of the animals), the calculated sensitivity is known as relative sensitivity. Because the test under development and the test

Sensitivity->		99%								70%							
Specificity->		99%	98%	95%	90%	80%	70%	50%	25%	99%	98%	95%	90%	80%	70%	50%	25%
P r e v a l e n c e	0.1%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
D i s e a s e	1%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99
	5%	100	100	100	100	100	100	100	100	98	98	98	98	98	98	97	94
	10%	100	100	100	100	100	100	100	100	97	97	97	96	96	95	94	88
	20%	100	100	100	100	100	100	99	99	93	93	93	92	91	90	87	77
	30%	100	100	100	100	99	99	99	98	89	88	88	88	86	84	80	66
	40%	99	99	99	99	99	99	99	97	83	83	83	82	80	78	71	56
	50%	99	99	99	99	99	99	98	96	77	77	76	75	73	70	63	45

*Instructions: Estimate the infection/disease prevalence of the population of animals from which the patient is derived and determine sensitivity and specificity of the test from published information or data from laboratory/company. Find the prevalence figure and read across to the probability value in the column under the test's specificity. Note: Sensitivities of only 99 and 70% are shown because sensitivity has a limited affect on the ability of a negative test result to correctly identify a noninfected animal.

Table 5—Intuitive method for determining the reliability (predictive value) of a negative test result

A	B	C	D	E	F	G	H	I
			Infected			Uninfected		
Size of population	Estimated infection or disease prevalence	Sensitivity of test	No. of animals	No. that test negative (false neg)	Specificity of test	No. of animals	No. that test negative (true neg)	Total no. of negative test results	Reliability of negative test result
			$A \times B$	$(100\% - C) \times D$		$A - D$	$F \times G$	$E + H$	$(E \div I) \times 100$
1,000	1%	90%	10	1	90%	990	891	892	99.9%
1,000	30%	90%	300	30	90%	700	630	660	95.4%
1,000	1%	98.3%	10	<1	98.3%	990	990	990	100%
1,000	30%	98.3%	300	5.1	98.3%	700	688	693	99.3%
1,000*								

*Use this row to insert values to determine specific reliability of a negative test result.

Steps in calculation: (1) Arbitrarily choose a population size of 1000 animals.
(2) Estimate the infection/disease prevalence in the population that represents the animal you are testing and insert in the table above.
(3) Insert sensitivity and specificity of test (provided by company or published in literature).
(4) Calculate reliability (known as predictive value to statisticians) of a positive test result.

used for comparison have intrinsic sources of error, by definition, the stated sensitivity of the test under development will be more errant. The amount of error introduced is directly related to the accuracy of the reference test. Confidence in test results is contingent on an assurance that the criterion was the best possible approximation/proof of the infection status of the animals in question.

Unfortunately, test kit manufacturers and laboratories don't always report how test sensitivity was determined. Serologic test for *Borrelia burgdorferi* infection in animals are a case in point. Extreme difficulty has been encountered in attempts to identify animals of proven infection status so the diagnostic sensitivity of laboratory tests for that disease is actually unknown. Some commercial laboratories offer assays that have been developed, assuming that the differentiation (cut-off) between positive and negative test results mimics what occurs in human beings and thus, choose an arbitrary cutoff titer of 1:64. There is no way to estimate whether such test results are correct; in fact, results of such tests are an expensive, usually uneducated guess.

Hidden factors that may affect test specificity—The proportion of noninfected animals that test negative is the test's specificity. When the infection/disease prevalence is low, the reliability of a negative test result is excellent even when the test

specificity is only 25% (Table 4). As test sensitivity decreases, a negative test result is still a good predictor that an animal is not infected, as long as the infection/disease prevalence is ≤10% (Table 4).

In our experience, the principal contributor to false-positive results for in-house ELISA is not the reagents in the test kit; rather it is technician error. If the reaction vessels are not washed sufficiently between steps, color develops, leading to false-positive reactions. We are often asked by veterinarians for confirmation of a positive FeLV ELISA result that they obtained by in-house testing. We frequently find no evidence of FeLV on retesting, which confirms observations of others.[1]

Predicting the FeLV status of cats by use of the IFA test and ELISA—To predict the reliability of a positive or negative test result, we have seen that diagnostic sensitivity, specificity, and disease prevalence must be known. Sensitivity and specificity data are usually available from published literature. For instance, in studies conducted over the past 18 years, Hardy and Zuckerman[1] have determined that for 348 samples tested by viral isolation (their criterion for comparison) and by the immunofluorescent antibody (IFA) test:

(1) The IFA test was 98.3% sensitive and 98.3% specific.

(2) On the basis of 1% prevalence of infection/disease, a positive IFA test result, will predict

37% of infected cats (calculated using Table 5; lines 3 and 4). Sixty-three percent of positive IFA test results will be falsely positive.

(3) If the infection/disease prevalence is 30%, such as in a multiple-cat household in which FeLV is known to exist, positive IFA test results would accurately reflect the FeLV status of the cat in 96% of samples tested (Table 3).

(4) The predictive values for a negative IFA test result are 100% and 99.3% when the disease prevalence is 1% and 30%, respectively (Table 3; line 4); again, a negative test result is a strong indicator that the cat does not have FeLV infection.

Several early comparison studies have indicated considerable discrepancy between the IFA test and ELISA.[2-7] Later studies, probably reflecting improved test kits and more experience in their use, have suggested greater concordance between ELISA and the IFA test.[8-11] Generally, the studies using FeLV commercial test kits have indicated that the ELISA false-negative rate is low, resulting in a sensitivity that exceeds 98% when based on the IFA test as the standard. The ELISA false-positive rate (with the IFA test or virus isolation as the criterion for comparison) has been higher. Assume for purposes of this study, that in a veterinarian's office, false-positive results are obtained among 10% of all samples tested. The reliability of ELISA is then calculated as follows:

(1) The ELISA is 99% sensitive and 90% specific.

(2) On the basis of 1% disease prevalence, a positive test result will accurately predict 9.1% of infected cats. Of 10 positive test results, 9 will be errant.

(3) If disease prevalence is 30%, the predictive value of a positive test result increases to 81%.

(4) The reliability of a negative ELISA result is virtually 100% when FeLV prevalence is either 1% or 30%.

After reviewing other reports (in this issue) on FeLV prevalence among various populations of cats, and the efficacy of diagnostic tests for detection of FeLV infection, Tables 3 and 5 can be used to calculate the capacity of positive and negative test results to reliably predict the infection status of cats.

Conclusion—When evaluating a serodiagnostic test result, the veterinarian should first consider whether the cat is at high risk (from a high prevalence group) or low risk (from a low prevalence group) for the condition under consideration. If the test result is positive, a different interpretation is required depending on the level of risk. A positive test result for a cat in a low-risk group has a much greater probability of being errant than does the same test result for a cat in a high-risk group. For a test reported by the manufacturer to have high sensitivity (99%) and specificity (99%), a positive test result may be very misleading if the technician using the test is sloppy. For such a test, even if the technician is meticulous, a positive test result for a low-risk cat (from a group having a low infection/disease prevalence of about 0.1%) should cause the practitioner to seriously question that positive test result; it is most likely errant. Conversely, negative test results are good prognosticators of noninfected cats even if the sensitivity and specificity of the test are not good. As infection/disease prevalence for FeLV testing decreases because of vaccination and test-and-removal programs, a positive ELISA result should be confirmed by retesting, using ELISA or IFA testing.

References
1. Hardy WD Jr, Zuckerman EE. Development of the immunofluorescence antibody test for detection of feline leukemia virus infections in cats. *J Am Vet Med Assoc* 1991;199:1327–1335.
2. Hardy WD Jr. The feline leukemia virus. *J Am Anim Hosp Assoc* 1981;17:951–980.
3. Jarrett O, Golder MC, Weijer K. A comparison of three methods of feline leukaemia virus diagnosis. *Vet Rec* 1982; 110:325–328.
4. Kahn DE, Mia AS, Tierney MM. Field evaluation of Leukassay F*, an FeLV detection test kit. *Feline Pract* 1980; 10:41–45.
5. Lutz H, Pedersen N, Harris CW, et al. Detection of feline leukemia virus infection. *Feline Pract* 1980;10:13–23.
6. Waits MR, Reggiardo C, Meininger AC. Comparison of the ELISA and FA test for the detection of FeLV. *Southwest Vet* 1982;34:208–209.
7. Hirsch VM, Search GP, Bellamy JEC. Comparison of ELISA and immunofluorescence assays for detection of FeLV antigens in blood of cats. *J Am Anim Hosp Assoc* 1982;18:933–938.
8. Lopez NA, Jacobson RH, Scarlett JM, et al. Sensitivity and specificity of blood test kits for feline leukemia virus antigen. *J Am Vet Med Assoc* 1989;195:747–751.
9. Hawks DM, Legendre AM, Rohrbach B. Comparison of four test kits for feline leukemia virus antigen. *J Am Vet Med Assoc* 1991;199:1373–1377.
10. Swango LJ. Evaluation of test results for feline leukemia virus diagnostic tests available for in-office use by veterinarians. *J Am Vet Med Assoc* 1991;199:1386–1389.
11. Lopez NA, Jacobson RH. False-positive reactions associated with anti-mouse activity in serotests for feline leukemia virus antigen. *J Am Vet Med Assoc* 1989;195:741–746.