Application of surveillance data in evaluation of diagnostic tests for infectious salmon anemia

Carol A. McClure[1], K. Larry Hammell[1,*], Henrik Stryhn[1], Ian R. Dohoo[1], Leighanne J. Hawkins[2,3]

[1]Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, Prince Edward Island C1A 4P3, Canada

[2]New Brunswick Department of Agriculture, Fisheries and Aquaculture, 107 Mount Pleasant Road, St. George, New Brunswick E5C 3K5, Canada

[3]*Present address:* Maritime Veterinary Services, 246 Main Street, PO Box 1105, St. George, New Brunswick E5C 3S9, Canada

ABSTRACT: Infectious salmon anemia (ISA) is a viral disease in farmed Atlantic salmon *Salmo salar*, characterized by lethargy, anorexia, anemia and death. Test methods used for regulatory decisions to remove infected cages include the indirect fluorescent antibody test (IFAT), the reverse transcription–polymerase chain reaction test (RT–PCR), and virus isolation (VI). However, no thorough evaluation of these diagnostic tests has been carried out on field samples. The objective of this study was to evaluate the sensitivity and specificity of ISA diagnostic tests as individual tests and in combinations, using data collected by the provincial government surveillance program. Because a 'gold standard' reference test for ISA was not available, cage status was based on clinical disease records. A pool of fish from negative farms that had never had clinical ISA and a pool of fish from positive cages that were experiencing an outbreak of clinical ISA were obtained and assumed to be uninfected and infected respectively. A total of 1071 fish were used in this study. Depending on the test's cut-off value, the sensitivity and specificity for histopathology ranged from 30 to 73 and 72 to 99% respectively. IFAT had sensitivities and specificities in the range of 64 to 83 and 96 to 100% respectively. For the RT-PCR, sensitivity and specificity were 93 and 98% respectively. Test performances were also evaluated in series and in parallel combinations. Sensitivities are maximized when tests are evaluated in parallel, and ranged from 75 to 98%. Specificities are maximized when the tests are evaluated in series, and ranged from 99 to 100%. Current surveillance testing protocols should be reviewed to capitalize on this newly available information on test characteristics.

KEY WORDS: ISA · Infectious salmon anemia · Diagnostic test · Surveillance · Sensitivity · Specificity · Parallel · Series

INTRODUCTION

Infectious salmon anemia (ISA) virus has caused disease in farmed Atlantic salmon *Salmo salar* in New Brunswick (NB) since 1996 (O'Halloran et al. 1999). This severe disease, characterized by lethargy, anorexia, anemia, internal organ damage and death (Thorud & Djupvik 1988, Byrne et al. 1998), has occurred sporadically throughout New Brunswick fish farms in the Bay of Fundy, and in 1998 approximately 22 of the 83 salmon farms were completely depopulated for

control purposes (O'Halloran et al. 1999). Costly control methods used on New Brunswick Atlantic salmon farms include a surveillance program, early slaughter of fish from test-positive cages, and indemnity programs.

Current industry control programs require ISA testing on moribund fish at least every 6 wk for every farm. Such surveillance results in mandatory slaughter of all fish in a cage if there have been at least 2 positive tests on at least 2 fish and fish in the cage have clinical signs of ISA. There are several commercial diagnostic tests,

including (1) virus isolation (VI) (Dannevig et al. 1995, Bouchard et al. 1999), (2) indirect fluorescent antibody technique (IFAT) (Falk et al. 1998), (3) reverse transcription–polymerase chain reaction (RT–PCR) (Mjaaland et al. 1997), and (4) histopathology on fish tissues (Evensen et al. 1991, Simko et al. 2000). Performance characteristics of these tests are unknown, and test results from the same fish are often inconsistent. Although the ISA diagnostic tests have not been evaluated, results are used to make sizeable monetary decisions.

There are many different parameters to consider when describing test performance, but 2 of the most important are sensitivity and specificity (Noordhuizen et al. 2001). Given that an animal is diseased, diagnostic sensitivity is the proportion of animals that test positive. Given that an animal is not diseased, diagnostic specificity is the proportion that test negative. The computations of these 2 parameters in the simplest situation are computed in Table 1. Often diagnostic tests are used in combination on the same animal to improve the chances of a correct diagnosis, and the results can be interpreted in 1 of 2 ways. When the tests are combined in parallel, the interpretation is positive if 1 or both of the individual tests are positive and negative only when both of the individual tests are negative. When the tests are combined in series, the interpretation is positive only if both of the individual tests are positive and otherwise the interpretation is negative.

Analytical sensitivity differs from diagnostic sensitivity (described above) as it measures the lowest level of detectable pathogen the assay can identify as the pathogen (Jacobson 1998). Many studies have been performed to measure analytical sensitivities for important salmon pathogens including bacterial kidney disease *Renibacterium salmoninarum* (Sakai et al. 1989), infectious hematopoietic necrosis virus (Medina et al. 1992), viral hemorrhagic septicemia virus, and infectious pancreatic necrosis virus (McAllister 1997). Analytical specificity measures an assay's ability to correctly identify a specific pathogen, minimizing the test's cross-reactivity with other pathogens. Studies measuring the diagnostic sensitivities and specificities

of important salmon diseases are very limited (Opitz et al. 2000). A brief evaluation of ISA diagnostic tests from these surveillance data has been reported previously (McClure et al. 2003), but not in the context of evaluating the surveillance program. An efficient surveillance program should be based on a quantitative assessment of test performance in order to make prudent decisions regarding cage removal because of disease.

Because performance reliabilities for each of the diagnostic tests were unknown, many tests were performed simultaneously on tissue from the same fish from 1998 to 2000 by the provincial government as part of the early surveillance program. These results were made available for evaluation of the diagnostic tests. The objectives of the present study were 2-fold. The first objective was to determine the sensitivity and specificity of as many ISA diagnostic tests as possible. The second to evaluate the use of these tests in the current surveillance diagnostic protocol.

MATERIALS AND METHODS

The data. In total, 30 255 test results were available from 8167 specimens of *Salmo salar* from New Brunswick. However, much of the data were not included here because the disease status of each fish's cage was available only from April 1999 to January 2000. All fish that had known diseases other than ISA were removed from the data set. For the purpose of calculating sensitivity and specificity, our gold standards for disease status were (1) ISA negative fish came from farms (sites) that had no outbreak of ISA in any cages during the period, (2) ISA positive fish came from cages that were experiencing clinical disease (defined by >0.05% mortalities d^{-1} at the time of sampling) which was attributed to ISA by regulatory officials.

Further, some of the tests and laboratories were dropped from the analysis because the numbers of samples were too small for statistical analysis. The tests and laboratories that were included in the study included histopathology and VI tests at the Atlantic Veterinary College Diagnostic Laboratory and Aquatic Diagnostic Services (AVC) in Charlottetown, Prince Edward Island; the IFAT test performed at the New Brunswick Department of Agriculture, Fisheries and Aquaculture laboratory (DAFA) in Blacks Harbour, New Brunswick; and the IFAT, RT–PCR, and VI tests performed at the Research and Productivity Council laboratory (RPC) in Fredericton, New Brunswick.

All test results were dichotomous or ordinal. Histopathology was reported on a scale of negative, suspect and positive according to the interpretation of the fish pathologist's report of the histopathologic lesions

Table 1. Estimation of sensitivity (Se) and specificity (Sp) of a diagnostic test using a 2 × 2 table. Sensitivity = a/a + c; specificity = d/b + d. n: total number of specimens tested

| Test result | True disease status | | Total for test result |
	Diseased	Not diseased	
Positive	a	b	a+b
Negative	c	d	c+d
Total for disease status	a + c	b + d	a + b + c + d = n

(Evensen et al. 1991, Simko et al. 2000). For sensitivity and specificity, histopathology data was analyzed in 2 different ways, first with the suspect cases considered positive, and second with the suspect cases considered negative. The IFAT results were reported as negative, 1+, 2+, 3+, or 4+ based on increasing fluorescence intensity of immunologic staining of the virus in kidney tissues as read by a trained technician (Falk et al. 1998). The IFAT results were analyzed using 2 different cut points, first using 1+, 2+, 3+, or 4+ as a positive result (IFAT 1), and second using negative and 1+ as a negative result and 2+, 3+, and 4+ as a positive result (IFAT 2). The RT–PCR which is based on identifying viral RNA in the kidney (Mjaaland et al. 1997), and VI which is based on growing virus in an appropriate cell line (Dannevig et al. 1995, Bouchard et al. 1999) have dichotomous results reported as positive or negative. Because of the expense of the VI test, pools of up to 5 fish were tested as 1 sample, and all fish in the pool were assigned a positive result if the pool were positive. Some fish were tested by VI individually (not in a pool), and the results noted. The resulting data was used in the determination of sensitivity and specificity.

Estimate of sensitivity and specificity for individual tests. Sensitivities (Se), specificities (Sp), and corresponding 95% confidence intervals (CI) were calculated in 2 different ways. Firstly, test sensitivity, specificity, and 95% CI (exact, based on the binomial distribution [Newcombe 1998]) were calculated from a 2×2 table for all test results and the gold standard classifications described above. For the VI tests, sensitivities, specificities and 95% confidence intervals were calculated from the 2×2 table using the results from fish tested individually.

Secondly, to adjust for clustering of test results within a farm, Se and Sp were calculated using a series of steps for all tests except the VI (because of the small number of fish tested). Separate analyses were carried out for the gold standard-positive and for the gold standard-negative fish.

Step 1: A logistic regression model can be used to describe the probability of a positive or a negative test result conditional on the disease state of the fish. Because fish were sampled from different farms and the results from fish from the same farm may not be independent from each other, a random effects logistic regression model with the farm as the random effect, the test result as the outcome, and maximum likelihood estimation of the model's parameters was used (Rabe-Hesketh et al. 2002).

Step 2: A random effects model yields 'subject-specific' estimates, but 'population-averaged' estimates which give the average sensitivity across farms are more appropriate to describe test characteristics. To derive the 'population-averaged' estimates of Se, the logit of the sensitivity (y) was calculated by the formula:

$$y = \beta_0 \div \sqrt{1 + 0.346 + \sigma^2} \qquad (1)$$

in which β_0 was the constant from the random effects logistic regression model for the ISA positive population and σ^2 was the estimated variance of farm random effects (Zeger et al. 1988). The logit of the specificity was calculated similarly.

Step 3: The logit of the 'population-averaged' estimates of the Se was converted to estimates of Se using the formula Se = $e^y/(1+e^y)$ with y from Step 2 for the ISA positive population. Specificity was calculated using Sp = $1 - e^y/(1 + e^y)$, with y from Step 2 for the ISA negative population. Confidence intervals for sensitivity and specificity were calculated with the same formulas when substituting the limits of its CI for the constant, β_0.

Step 4: To assess the degree to which the samples at the same farm were dependent, the intraclass correlation coefficient (ICC) was calculated as $\sigma^2/(\sigma^2 + 3.29)$ (Snijders & Bosker 1999).

Step 5: To assess how the tests would perform if used on a new farm in New Brunswick, 90% prediction intervals for the farm's sensitivity and specificity were computed by the formulas in Step 3 and letting

$$y = \beta_0 \pm 1.645 \sqrt{SE(\beta_0)^2 + \sigma^2} \qquad (2)$$

where $SE(\beta_0)$ is the standard error of β_0.

Prediction intervals are wider than normal confidence intervals because they include additional variation (σ) due to the random effects. The prediction intervals were relaxed to 90% because of the small number of farms in the study.

Estimate of sensitivity and specificity using 2 tests in combination. The sensitivity and specificity for each combined pair of tests was calculated by 2 methods, directly from the actual data which accounts for any dependence between the 2 tests, and indirectly from the individual tests' sensitivity and specificity assuming independence between the 2 tests. If the tests were dependent, test correlations were calculated to evaluate the degree of dependence.

Direct estimate: For the direct estimate, every diagnostic test was paired with each of the other diagnostic tests. The individual test results were evaluated in both series, and parallel, forming a new test result for the combined tests. These new results were then tabulated with the disease status of each fish in a 2×2 table, and sensitivity and specificity were estimated directly.

Indirect estimate assuming independence: Assuming independence between the 2 tests, the sensitivity and specificity of using 2 tests in parallel and in series were estimated using these formulas (Noordhuizen et al. 2001): for parallel interpretation, sensitivity is Se$_p$ = [1 – (1 – Se$_1$) (1 – Se$_2$)], where Se$_p$ indicates par-

allel interpreted sensitivity and Se_1 and Se_2 indicate the sensitivity of the individual tests as calculated above. Specificity is $Sp_p = Sp_1Sp_2$, where Sp_p indicates parallel interpreted specificity and Sp_1 and Sp_2 indicate the specificity of the individual tests. For series interpretation, sensitivity is $Se_s = Se_1Se_2$ and specificity is $Sp_s = [1 - (1 - Sp_1)(1 - Sp_2)]$.

Evaluating dependence between tests: If 2 tests evaluated in combination are independent, the direct estimate of the combined test properties should be the same as the indirect estimate if the same samples are used in each estimate. Because the 2 tests evaluated in combination may not be independent, the covariance can be estimated as a measure of their dependence. Independence of 2 tests, when fish were divided by their true disease status, was assessed by Fisher's exact-test (Gardner et al. 2000). If the p values for the Fisher's exact test were significant at $p \leq 0.10$, the covariances of the test pairs were calculated.

The sensitivity covariance between the 2 tests used in combination is calculated as $\gamma_{Se} = p_{11} - Se_1Se_2$, where p_{11} is the probability of both tests being positive when the fish is diseased, and Se_1 and Se_2 are the sensitivities of the 2 tests that are calculated from only those samples the 2 tests have in common (Gardner et al. 2000). The specificity covariance between the 2 tests used in combination is calculated as $\gamma_{Sp} = p_{00} - Sp_1Sp_2$, where

p_{00} is the probability of both tests being negative when the fish is not diseased, and Sp_1 and Sp_2 are the specificities of the 2 tests calculated from only those samples the 2 tests have in common.

The upper and lower limits for possible values for the covariances were calculated according to Gardner et al. (2000). The covariances are directly affected by the values of Se and Sp, so the strength of the dependence was normalized by γ/maximal possible value of γ, as determined by the upper and lower limits.

RESULTS

The data

The final data set contained 3721 test results from 1071 fish (807 negative and 264 positive). These fish came from 238 different cages and from 23 different farms.

Estimate of sensitivity and specificity for individual tests

Sensitivities and specificities with their associated confidence intervals for each test analyzed without (combined estimate) and with (population estimate)

Table 2. Estimated sensitivities (Se) and specificities (Sp) of 6 infectious salmon anemia diagnostic tests with different threshold levels used by New Brunswick Atlantic salmon farms. Positive, negative: suspects considered positive and negative respectively; n: no. of fish tested; na: not analyzed. IFAT: indirect fluorescent antibody test; RT–PCR: reverse transcription–polymerase chain reaction test; VI: virus isolation. CI: confidence interval; ICC: interclass correlation coefficient. AVC: Atlantic Veterinary College; DAFA: New Brunswick Department of Agriculture, Fisheries, and Aquaculture; RPC: Research and Productivity Council

Test		n	Parameter	Combined estimate (CI)	Population estimate (CI)	Random effect estimated ICC	p-value	90% predicted interval
Histopathology (AVC)								
Positive		159	Se	0.730 (0.653–0.797)	0.730 (0.655–0.793)	0.00	1.000	0.668–0.783
		483	Sp	0.725 (0.682–0.764)	0.721 (0.646–0.794)	0.07	0.000	0.528–0.866
Negative		159	Se	0.302 (0.232–0.380)	0.299 (0.219–0.393)	0.34	0.000	0.047–0.753
		483	Sp	0.994 (0.982–0.999)	0.994 (0.981–0.998)	0.0	1.000	0.984–0.998
IFAT (DAFA)	1	225	Se	0.791 (0.732–0.842)	0.794 (0.693–0.869)	0.14	0.011	0.545–0.940
		646	Sp	0.955 (0.936–0.970)	0.957 (0.924–0.976)	0.11	0.027	0.893–0.989
	2	225	Se	0.644 (0.578–0.707)	0.644 (0.580–0.704)	0.00	1.000	0.590–0.695
		646	Sp	0.999 (0.991–1.000)	0.998 (0.989–1.000)	0.00	1.000	0.992–1.000
IFAT (RPC)	1	52	Se	0.827 (0.697–0.918)	0.827 (0.700–0.907)	0.00	1.000	0.723–0.898
		421	Sp	0.983 (0.966–0.993)	0.980 (0.912–0.996)	0.31	0.070	0.903–0.999
	2	52	Se	0.731 (0.590–0.844)	0.736 (0.567–0.856)	0.05	0.313	0.524–0.882
		421	Sp	0.998 (0.987–1.000)	0.998 (0.983–1.000)	0.00	1.000	0.988–1.000
RT–PCR (RPC)		215	Se	0.926 (0.882–0.957)	0.932 (0.862–0.967)	0.10	0.103	0.828–0.981
		733	Sp	0.981 (0.968–0.990)	0.967 (0.910–0.988)	0.48	0.000	0.845–1.000
VI (AVC)		0	Se	No samples	na	na	na	na
		21	Sp	1.000 (0.839–1.000)	na	na	na	na
VI (RPC)		3	Se	0.667 (0.094–0.992)	na	na	na	na
		69	Sp	0.986 (0.922–1.000)	na	na	na	na

Table 3. Estimated sensitivities, Se (act: actual; indep: independent) of infectious salmon anemia diagnostic tests evaluated in parallel and in series. No. positive samples: number of samples used in calculation of actual sensitivities; p-value: Fisher's exact-test, value for significance of dependence between tests. Further details as in Table 1

Test 1		Test 2	No. positive samples	$Se_{parallel}$ act	indep	Se_{series} act	indep	p-value	Covariance of Se	Cov_{Se}/max
Histopathology (AVC)										
Positive		IFAT 1 (DAFA)	130	0.923	0.944	0.608	0.577	0.146		
		IFAT 2 (DAFA)	130	0.869	0.904	0.515	0.470	0.061	0.038	0.213
		IFAT 1 (RPC)	52	0.904	0.953	0.635	0.604	0.100	0.046	0.375
		IFAT 2 (RPC)	52	0.865	0.927	0.577	0.534	0.081	0.057	0.297
		RT–PCR	135	0.978	0.995	0.719	0.716	0.021	0.016	1.000
Negative		IFAT 1 (DAFA)	130	0.877	0.854	0.232	0.238	0.351		
		IFAT 2 (DAFA)	130	0.746	0.752	0.215	0.194	0.551		
		IFAT 1 (RPC)	52	0.827	0.879	0.192	0.250	0.178		
		IFAT 2 (RPC)	52	0.751	0.812	0.173	0.221	0.254		
		RT–PCR (RPC)	135	0.978	0.987	0.326	0.296	0.551		
IFAT (DAFA) 1		IFAT 1 (RPC)	23	0.870	0.964	0.783	0.654	0.009	0.100	0.697
		IFAT 2 (RPC)	23	0.870	0.944	0.696	0.578	0.040	0.085	0.662
		RT–PCR (RPC)	176	0.949	0.996	0.761	0.776	<0.001	0.035	0.608
	2	IFAT 1 (RPC)	23	0.870	0.938	0.652	0.533	0.067	0.078	0.641
		IFAT 2 (RPC)	23	0.826	0.904	0.609	0.471	0.045	0.095	0.521
		RT–PCR (RPC)	176	0.926	0.993	0.631	0.632	<0.001	0.047	1.000
IFAT (RPC)	1	RT–PCR (RPC)	52	0.962	0.997	0.827	0.811	0.027	0.032	1.000
	2	RT–PCR (RPC)	52	0.962	0.995	0.731	0.717	0.069	0.028	1.000

the random effect of the farm are shown in Table 2. In general, the sensitivity and specificity for histopathology ranged from 30 to 73 and 72 to 99% respectively, depending on the cut-off value. IFAT had sensitivities and specificities in the range of 64 to 83 and 96 to 100% respectively. For RT–PCR sensitivity and specificity were 93 and 98% respectively, and for VI sensitivity was 67% and specificity was in the range of 99 to 100%. The small number of fish that were tested by VI made it impossible to evaluate the random effects of site with this test. Accounting for the within-farm correlation altered the estimates for the other tests only slightly. There was significant within-farm correlation for the sensitivity for histopathology for suspects considered negative and the DAFA IFAT 1 (cut-off between negative, and 1+) as well as the specificity of histopathology with suspects considered positive, DAFA IFAT 1, and the RT–PCR.

Estimate of sensitivity and specificity using 2 tests in combination

Sensitivities and specificities calculated in series and parallel with their associated conditional covariance for each pair of tests are shown in Tables 3 and 4 respectively. The test combinations with the virology tests from 2 different laboratories were not included in the analysis because of the limited number of samples these 2 tests had in

Table 4. Estimated specificities (Sp) for infectious salmon anemia diagnostic tests evaluated in parallel and in series. Further details as in Tables 1 & 3

Test 1		Test 2	No. negative samples	$Sp_{parallel}$ act	indep	Sp_{series} act	indep
Histopathology (AVC)							
Positive		IFAT 1 (DAFA)	341	0.686	0.692	0.988	0.988
		IFAT 2 (DAFA)	341	0.724	0.724	1.000	1.000
		IFAT 1 (RPC)	386	0.700	0.713	0.995	0.995
		IFAT 2 (RPC)	386	0.712	0.724	0.997	0.999
		RT–PCR (RPC)	478	0.713	0.671	0.998	0.980
Negative		IFAT 1 (DAFA)	341	0.938	0.949	1.000	1.000
		IFAT 2 (DAFA)	341	0.988	0.993	1.000	1.000
		IFAT 1 (RPC)	386	0.974	0.977	1.000	1.000
		IFAT 2 (RPC)	386	0.990	0.992	1.000	1.000
		RT–PCR (RPC)	478	0.983	0.920	1.000	1.000
IFAT (DAFA)	1	IFAT 1 (RPC)	268	0.944	0.939	0.996	0.999
		IFAT 2 (RPC)	268	0.963	0.953	1.000	1.000
		RT–PCR (RPC)	574	0.930	0.884	1.000	0.997
	2	IFAT 1 (RPC)	268	0.974	0.982	1.000	1.000
		IFAT 2 (RPC)	268	0.996	0.997	1.000	1.000
		RT–PCR (RPC)	574	0.977	0.925	1.000	1.000
IFAT (RPC)	1	RT–PCR (RPC)	420	0.957	0.910	1.000	0.999
	2	RT–PCR (RPC)	420	0.971	0.924	1.000	1.000

common with the other tests. Sensitivities were maximized when tests are evaluated in parallel, and ranged from 75 to 98% for the direct calculation from the actual data. Specificities were maximized when the tests are evaluated in series, and ranged from 99 to 100%.

Also included in Tables 3 & 4 are the sensitivities and specificities calculated in series and parallel assuming the tests are independent. These numbers were calculated from the individually calculated sensitivities and specificities presented in Table 2.

Except for the combinations of tests with significant dependence, all sensitivities and specificities changed only slightly compared when calculated with the actual data the 2 tests had in common. There were significantly dependent tests (p ≤ 0.05) for the gold standard positive fish for 7 of the 18 test combinations, as shown in Table 3. The covariances were all positive and ranged from 0.016 to 0.100. The sensitivity γ/maximal possible value of γ ranged from 21 to 100%, or complete dependence. There were no significantly dependent tests for the gold standard negative fish.

DISCUSSION

The farmed Atlantic salmon industry in New Brunswick is currently dealing with a diagnostic testing dilemma. The surveillance program tests many moribund fish from all farms in New Brunswick. If a cage is falsely diagnosed as negative for ISA, viral loads may increase and potentially spread to other cages or to neighboring farms. If a cage is falsely diagnosed as positive with ISA, the fish are harvested early, resulting in high economic losses due to large volumes of non-market size fish going to market. As both of these scenarios are unacceptable, the identification of a diagnostic testing program with high sensitivity and specificity is imperative.

Evaluation of individual tests

The results of our study showed the highest sensitivities and specificities in RT–PCR tests performed by the RPC laboratory. The RT–PCR test results are usually returned within a few days. Unfortunately the expense of this test (>35 US$ per fish) may limit its adoption by the industry. The quickest and least expensive test available is the IFAT. Unfortunately, this test's sensitivity is at best 83%, leading to 17% of the truly positive fish testing as falsely negative. Histopathology did not perform very well as an ISA diagnostic test, but this test does have 2 advantages: it is relatively inexpensive and there is the potential to diagnosis a concurrent disease. Performance evalua-

tion of VI was difficult in this data set because most results were reported as pools, and the final evaluation was made only on fish that were tested individually. While the specificity of the VI is excellent, the sensitivity was poor for RPC's VI and not evaluated for AVC due to an insufficient number of samples. An advantage of VI is that a positive result indicates there is live virus in the sample. However, poor sensitivity, high expense and long incubation periods restrict the use of this test (Dannevig et al. 1995).

Interestingly, the sensitivity of the IFAT2 from the RPC laboratory was much higher than the sensitivity of the IFAT2 from the DAFA laboratory (>9%). This inconsistency is an example of how different laboratories can reach different results for the same sample tested by the same type of test. The most likely reason for this discrepancy is the reading of the IFAT slides by a technician. There is a different technician in each laboratory, and their interpretation of the number of fluorescent viral particles on a slide is subjective. Other possible reasons for the discrepancy are the use of different reagents and protocols as well as variations between laboratories, including temperature, light, and equipment.

Sample size

If the sample size were increased by increasing the number of fish tested and number of tests run on each fish, then the precision of the sensitivity and specificity for each test would be improved. Unfortunately, the individual test sensitivities and specificities were calculated using an historic data set in which the number of fish tested and the number of tests run on each fish were already determined.

Because these diagnostic tests are being used to identify infected cages for early harvest, cage-level sensitivity and specificity are actually more important than individual fish sensitivity and specificity of the tests. As the number of fish tested increases per cage, the cage-level sensitivity increases (Dohoo et al. 2003). However, it was not possible to calculate cage-level sensitivity and specificity because the number of fish tested per cage was not fixed but ranged from 1 to 22 fish. The importance of cage-level sensitivities and specificities should not be ignored, and are currently being investigated.

Bias associated with data selection

Although we have estimated the sensitivities and specificities of these diagnostic tests, a critical review of the methods is important. Defining disease status on

samples from perfectly healthy sites and highly diseased cages introduces bias that will cause tests to appear to perform better than they would if applied to all fish (Brenner & Gefeller 1997). Fish that have just been infected and are not showing any signs of disease may not test positive with the available tests, resulting in a loss of sensitivity. Because the data was reduced significantly to identify obviously diseased and disease-free fish, test performance results in this study appear better than they would have had the tests been assessed using the entire spectrum of disease states naturally present in production populations.

Conversely, the sensitivity of the VI test may have been falsely lowered. The DAFA laboratory pooled tissue samples from 1 to 5 fish. Fish in a pool usually came from the same cage, with the number of fish in the pool varying with the severity of disease in the cage. If there were 5 fish in a pool, the cage probably had high mortalities and advanced disease. If there was only 1 fish in a pool, there was probably only that 1 dead fish in the cage; therefore, the chance of the fish in that cage having advanced clinical illness was lower. Such fish may be infected, but may not have abundant live virus to easily create cytopathic effect on the cell culture, the endpoint of the VI test.

Accounting for clustering of within-farm results

The random effects model was used to account for the fact that fish from 1 farm are potentially more alike than fish from different farms. Using Steps 2 and 3 (see 'Materials and methods'), 'population averaged' estimates of sensitivity and specificity were estimated from the results of the random-effects logistic regression model. These estimates would be more accurate than the combined estimates in tests where there was significant within farm correlation. In addition, the model provides prediction intervals (Step 5) for how the test would perform if it were used on a new farm in New Brunswick, or a farm not included in this study. When variation between farms is present, these intervals are wider than the confidence intervals because they incorporate farm-to-farm variation. Possible reasons for extra variation between farms include genetic differences, geographic location, age of fish at the farm, and management factors (feed, handling, sea lice burdens, hygiene, etc.). A hypothetical scenario might involve a strain of Atlantic salmon with improved resistance to ISA virus such that these fish rarely replicate enough virus to test positive with the IFAT test. This would result in an increase in falsely negative tests for fish from farms raising this strain of salmon, and could explain the wide 90% predicted interval for the IFAT tests. A hypothetical geographical

example might be that dead fish that come from more remote farms are not processed as quickly as dead fish from local farms. As dead fish decompose, the integrity of the viral RNA might be jeopardized. Fish from these farms are more likely to show false negatives with the RT–PCR test as a result of the increased time to processing, as reflected by the very wide 90% predicted interval for the sensitivity for the RT–PCR test. The population estimate for the RT–PCR is 0.932, and the 90% predicted interval ranges from 0.828 to 0.981.

Evaluating independence of test results

A common assumption when interpreting tests in combination is that the 2 tests are conditionally independent (Gardner et al. 2000). In this study, we were able to interpret the combinations both by assuming conditional independence, and by the more accurate interpretation in which we accounted for conditional dependence. Conditional dependence of test sensitivities occurs when the second test has a different sensitivity for diseased fish that test positive in the first test compared to those that test negative in the first test. Conditional dependence of test specificities occurs when the second test has a different specificity for nondiseased fish that test negative in the first test compared to those that test positive in the first test. The measures of dependence between 2 tests' sensitivities and specificities are measured by the covariances of the sensitivities and specificities and are gauged by the covariance/maximal covariance of the 2 tests' sensitivity and specificity. If interpreting 2 tests in parallel when the 2 tests have a high sensitivity covariance/maximal covariance, then using the second test will not yield much new information and the resulting sensitivity will be little more than just using the 1 test with the better sensitivity. The same is true for specificities when interpreting in series. It should be noted that 2 tests with a covariance/maximal covariance for sensitivities (or specificities) equal to 1 are only in complete agreement for positive (negative) fish if the sensitivities (or specificities) of the 2 tests are the same.

The best tests to use in combination may not always be those with the best individual sensitivities and specificities, but could possibly be the tests that have the lowest sensitivity and specificity covariance/maximal covariance. One example of this effect is the use of histopathology (with suspects being considered negative) and the DAFA IFAT 2. The sensitivity covariance/maximal covariance was very low at 0.133 and the sensitivity of the tests in parallel jumped from 0.302 for the histopathology and 0.644 for the IFAT to 0.746 for the combined sensitivity.

A positive dependence should be expected in tests that measure the same biological process (Gardner et al. 2000), as can be seen when combining the IFAT with the positive cut-off between negative and 1+ from DAFA and RPC. Both tests measure the level of virus present in kidney imprints on glass slides. The covariance of the sensitivity was 0.100, with a sensitivity covariance/maximal covariance of 0.697, which was the highest positive dependence measured in this study that was not equal to 1. Were these tests independent, we would expect that sensitivity in parallel would be 96%, but because they are very dependent, the sensitivity in parallel is only 87%. This can be compared to the previous example in which histopathology was used in conjunction with IFAT. Positive diagnosis for ISA using histopathology is based on microscopic lesions in the fish tissues, while the IFAT test is based on the presence of virus in the kidney. The biologic processes are different, thus the tests are less likely to be dependent.

Estimates for sensitivities and specificities for ISA diagnostic tests are helpful in choosing which test will be most likely to return a true result. However, each test measures something different about the disease. Virus isolation measures the live virus (Dannevig et al. 1995, Bouchard et al. 1999), RT–PCR measures viral RNA (Mjaaland et al. 1997), IFAT measures viral antigen (Falk et al. 1998), and histopathology assesses lesions (Evensen et al. 1991, Simko et al. 2000). If RT–PCR is positive, there is most probably ISA viral RNA in the fish, but this does not necessarily indicate that the fish is clinically ill or actively shedding virus. Until we are capable of predicting the clinical disease and increased mortalities of the fish in a cage using diagnostic tests, test results should be interpreted cautiously.

Evaluation of surveillance program using performance of combined tests

Currently, the New Brunswick Department of Agriculture, Fisheries, and Aquaculture (NBDAFA) uses a 2 test approach for its ISA surveillance program. Every 6 wk, a sample of suitable moribund or dead fish from every farm is tested for ISA. The whole fish are brought directly to the provincial laboratory on ice and a necropsy is completed, or samples are taken on site, placed on ice, and brought to the laboratory. Samples are taken for all the routine tests available. Kidney imprints are tested by IFAT at the DAFA laboratory immediately, and all other tissue samples are stored for future testing. If the IFAT result is negative, the fish are considered ISA negative and no further testing is made. If the IFAT result is positive, the stored tissue is sent to the RPC laboratory for RT-PCR testing. If this test result is positive, then the fish is considered to be positive; if it is negative, then the fish is considered to be negative. The results of this study suggest that this series interpretation has a sensitivity of 0.76 and a specificity of 1.00. Therefore, using this testing protocol approximately 24% of truly positive fish are falsely being labeled as negative for ISA.

An alternative approach would be to choose a testing strategy that maximizes sensitivity and specificity. The simplest approach would be to test all fish with the RT–PCR test. This test has a high sensitivity (0.93) and a high specificity (0.98). However, due to its price, it is poorly suited for use on all fish tested in the surveillance program. Of the combined tests, the best performing pair of tests was the IFAT and RT–PCR with the tests interpreted in parallel. Unfortunately, this yielded only slightly improved sensitivities and specificities compared to the RT–PCR alone, and would still require testing all fish with the pricey RT–PCR test. Another option is using both of the IFAT tests in parallel. If the DAFA IFAT is interpreted with a positive cut-off between 1+ and 2+ (DAFA IFAT 2) and the RPC IFAT is interpreted with a positive cut-off between negative and 1+ (RPC IFAT 1), then the resulting sensitivities and specificities would be 0.87 and 0.97, respectively. This testing strategy, although not perfect, would offer improved sensitivity at a reasonable cost. Before testing strategies are changed, an in-depth study using a larger disease-positive pool, and an economic analysis using the number of fish tested by the DAFA surveillance program and the cost of the individual tests should be completed.

LITERATURE CITED

Bouchard D, Keleher W, Opitz HM, Blake S, Edwards KC, Nicholson BL (1999) Isolation of infectious salmon anemia virus (ISAV) from Atlantic salmon in New Brunswick, Canada. Dis Aquat Org 35:131–137

Brenner H, Gefeller O (1997) Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 16:981–991

Byrne PJ, MacPhee DD, Ostland VE, Johnson G, Ferguson HW (1998) Haemorrhagic kidney syndrome of Atlantic salmon, *Salmo salar* L. J Fish Dis 21:81–91

Dannevig BH, Falk K, Namork E (1995) Isolation of the causal virus of infectious salmon anaemia (ISA) in a long-term cell line from Atlantic salmon head kidney. J Gen Virol 76: 1353–1359

Dohoo IR, Martin SW, Stryhn H (2003) Veterinary epidemiologic research. Atlantic Veterinary College, Charlottetown, Prince Edward Island

Evensen O, Thorud KE, Olsen YA (1991) A morphological study of the gross and light microscopic lesions of infectious anaemia in Atlantic salmon (*Salmo salar*). Res Vet Sci 51:215–222

Falk K, Namork E, Dannevig BH (1998) Characterization and applications of a monoclonal antibody against infectious salmon anaemia virus. Dis Aquat Org 34:77–85

Gardner IA, Stryhn H, Lind P, Collins MT (2000) Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Prev Vet Med 45:107–122

Jacobson RH (1998) Validation of serological assays for diagnosis of infectious diseases. Rev Sci Tech Off Int Epizoot 17:469–526

McAllister PE (1997) Susceptibility of 12 lineages of chinook salmon embryo cells (CHSE-214) to four viruses from salmonid fish: implications for clinical assay sensitivity. J Aquat Anim Health 9:291–294

McClure CA, Hammell KL, Dohoo IR, Stryhn H, Hawkins LJ (2003) Evaluation of infectious salmon anemia diagnostic tests. In: Miller O, Cipriano RC (tech coords) International response to infectious salmon anemia: prevention, control and eradication; proceedings of a symposium, 3–4 September 2002, New Orleans, LA. Tech Bull 1902: US Department of Agriculture, Animal and Plant Health Inspection Service, US Department of Interior, US Geological Survey, US Department of Commerce, National Marine Fisheries Service, Washington, DC, p 69–73

Medina DJ, Chang PW, Bradley TM, Yeh MT, Sadasiv EC (1992) Diagnosis of infectious hematopoietic necrosis virus in Atlantic salmon *Salmo salar* by enzyme-linked immunosorbent assay. Dis Aquat Org 13:147–150

Mjaaland S, Rimstad E, Falk K, Dannevig BH (1997) Genomic characterization of the virus causing infectious salmon anemia in Atlantic salmon (*Salmo salar* L.): an orthomyxo-like virus in a teleost. J Virol 71:7681–7686

Newcombe RG (1998) Two-sided confidence for the single proportion: comparison of seven methods. Stat Med 17:857–872

Noordhuizen JPTM, Frankena K, Thrusfield MV, Graat EAM (2001) Application of quantitative methods in veterinary epidemiology, 2nd edn. Wageningen Press, Wageningen

O'Halloran JLF, L'Aventure JP, Groman DB, Reid AM (1999) Infectious salmon anemia in Atlantic salmon. Can Vet J 40:351–352

Opitz HM, Bouchard D, Anderson E, Blake S, Nicholson B, Keleher W (2000) A comparison of methods for the detection of experimentally induced subclinical infectious salmon anaemia in Atlantic salmon. Bull Eur Assoc Fish Pathol 200:12–22

Rabe-Hesketh S, Skrondal A, Pickles A (2002) Estimation of generalized linear mixed models. Stata J 2:1–21

Sakai M, Atsuta S, Kobayashi M (1989) Comparison of methods used to detect *Renibacterium salmoninarum*, the causative agent of bacterial kidney disease. J Aquat Anim Health 1:21–24

Simko E, Brown LL, MacKinnon AM, Byrne PJ, Ostland VE, Ferguson HW (2000) Experimental infection of Atlantic salmon, *Salmo salar* L., with infectious salmon anaemia virus: a histopathological study. J Fish Dis 23:27–32

Snijders TAB, Bosker RJ (1999) Multilevel analysis. Sage Publications, London

Thorud K, Djupvik HO (1988) Infectious anaemia in Atlantic salmon (*Salmo salar* L.). Bull Eur Assoc Fish Pathol 8:109–111

Zeger SL, Liang KY, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44:1049–1060