Validity of using serological tests for diagnosis of diseases in wild animals

LA. GARDNER *, S. HIETALA ** and W.M. BOYCE ***

Summary: Transposition of diagnostic tests used in domestic livestock species to free-ranging and captive wildlife species has two problems. First, most existing tests have not been adequately validated in domestic livestock. Second, assumptions that a serological test will perform identically in wildlife and livestock species may not be correct, due to differences in pathogenic strains and serovars, host serological responses, and exposure to organisms of similar antigenic structure which produce cross-reacting antibodies. Some assays require species-specific reagents/test components which might not be commercially available, and most assays have not been standardized. The authors outline the principles involved in the evaluation of a serological diagnostic test, and provide examples of how knowledge of test sensitivity and specificity can be used to estimate true prevalence, to determine whether a population is infected, and to facilitate management decisions with regard to animal translocations.


INTRODUCTION

Diagnostic testing has an important role in monitoring and maintaining the health of captive and free-ranging wildlife populations. In wildlife, tests are used to detect exposure to an agent, and to detect and identify the agent involved in an individual animal infection or a herd epidemic. Tests are also used for epidemiological purposes, including estimation of the prevalence of disease, determination of whether populations are infected, study of risk factors for disease, and evaluation of disease transmission risk between wildlife and domestic species, as well as between wildlife and humans.

Of the diagnostic tests used in wildlife, serological testing (the measurement of serum antibody) seems to be the most frequent, as a single sample can be used easily – and at relatively little expense – to test for exposure to multiple infectious agents (2, 14). This approach is being used with increasing frequency, due to the collection and
banking of sera from many wildlife species (12). The usual goal of serological testing is to discriminate between exposed and non-exposed animals. It is not always possible to make this distinction with complete confidence, however, as the distributions of test responses of the two groups typically overlap to some extent (17). Serological responses are usually measured on a continuous scale (i.e. optical densities in an enzyme-linked immunosorbent assay [ELISA]) or on an ordinal scale (i.e. titres of 1:4, 1:8, 1:16, etc.), and a cut-off value is typically set. Above this level animals are considered positive, and below it they are considered negative. A third category, namely ‘suspect’ or ‘intermediate’, is sometimes used for test results with values where the distributions of responses in the exposed and non-exposed populations overlap. If the true exposure or infection status of the animals is known, sensitivity and specificity can be calculated at the selected cut-off. Sensitivity and specificity are inherent properties (measures of validity or accuracy) of a serological test. In contrast, test performance (predictive value) is influenced by the prevalence of antibodies in the population. The prevalence of detectable antibodies, in turn, depends on the rate of infection with age and time, antibody persistence, and the rate of antibody loss. These factors, together with the reproducibility of assays within and between laboratories, determine how accurately the test results reflect the true exposure status of the animal or group of animals.

Most serological tests used in wildlife species – including tests for brucellosis, anaplasmosis and bluetongue – have been directly transposed, without validation, from use in domestic livestock species. This results in two separate but related problems. First, many tests have not been adequately evaluated in livestock species: indeed, few data have been published on the sensitivity and specificity of many of the tests commonly used in veterinary medicine. Second, even if a test has been validated in domestic species, the test may not perform identically in wildlife species due to differences in pathogen strains and serovars, host serological responses, and exposure to organisms with similar antigens which cause the production of cross-reacting antibodies. Some assays also need species-specific reagents/test components which may not be commercially available. The limitation entailed by the lack of antisera to IgG for many exotic species has been overcome, to a certain extent, by use of either protein A or protein G bound to enzymes such as horseradish peroxidase (19). For many species, these complexes can be used as alternatives to secondary antibody in ELISA and Western blotting (4). In addition to these technical issues, quality assessment and standardization of procedures have not been implemented for most existing tests used in wildlife species (3).

Compared with use in livestock species, one advantage of using a serological test in free-ranging wildlife is that interpretation of titres or test results for most infectious diseases is not usually complicated by vaccination. Brucellosis in bison is a notable exception, as interpretation of serological titres might depend on whether strain 19 vaccine has been used (6). This advantage may recede in the future, for many infectious diseases, as vaccination of wildlife becomes a more frequent management tool.

In this paper, the authors outline the principles involved in the evaluation of a serodiagnostic test for a pathogen in wildlife species, and show how data on sensitivity and specificity can be used to estimate true prevalence, to determine whether a population is infected, and to facilitate management decisions with regard to animal translocation.
ESTIMATION OF TEST SENSITIVITY AND SPECIFICITY

To validate a serological test, test results should be compared with the true status of an animal (frequently termed the ‘gold standard’) for the infection or disease of interest. A ‘gold standard’ is a diagnostic method (or combination of methods) which determines absolutely and without error whether an animal has been exposed or infected. For some diseases, no ‘gold standard’ exists. In other cases, use of this perfect diagnostic method is not practical in the field – due to high cost, labour or invasiveness – and therefore other less expensive and more practical methods are used. However, this surrogate measure may be used in diagnostic decision-making only if comparison between the serological test and the surrogate yields results which are sufficiently close to those achieved by comparison with the ‘gold standard’. With brucellosis, for example, bacterial culture is used as the standard, even though it may be imperfectly sensitive. For bluetongue, there is general agreement that virus isolation is the best indicator of current infection, although there is evidence that polymerase chain reaction may be equally sensitive with clinical samples (1). On the basis of comparison with known positive and negative samples, sensitivity and specificity can be calculated. Sensitivity and specificity may be difficult or costly to determine for some diseases, but such characteristics should be known for any test which is widely used and forms a basis for wildlife management decisions.

Sensitivity

‘Sensitivity’ is the probability that a test will correctly identify animals which have been exposed to a given pathogen (true positives). A test with high sensitivity will correctly identify a greater proportion of exposed animals and a lower proportion of false negatives. False-negative serological reactions can occur for any of the following reasons:

- if an animal was recently infected and a test system failed to detect a low concentration of antibodies
- if an animal was immunologically tolerant to infection (producing few or no antibodies)
- if the cut-off for the test was set too high
- if the test was not correctly designed or performed to detect the specific antibody in question.

Buyers of animals and regulatory officials in importing countries, states and regions usually require tests of almost 100% sensitivity to minimize the risk of introducing new pathogens. This line of reasoning could also be applied to wild animals which are being acquired by zoological collections or are being translocated to establish or augment free-ranging populations.

Specificity

‘Specificity’ is the probability that a test will correctly identify animals which have not been exposed to a given pathogen (true negatives). False-positive reactions can be caused by any of the following factors:

- presence of specific and non-specific antibodies following vaccination
- errors in handling and testing in the laboratory
- excessively low assay cut-off values
- cross-reactions with antibodies produced by antigenically-related organisms.
For some infectious diseases (e.g. bluetongue), the latter problem has been alleviated by the use of competitive ELISA assays which use monoclonal antibodies to reduce cross-reactivity between antigenically-related organisms (19). Those who sell animals usually require tests of perfect specificity to maximize the chance of sales. Similarly, for wildlife species, a high test specificity ensures that animals deemed important for maintenance of the genetic diversity of a population are not excluded from reintroduction programmes due to false-positive results.

One effect of using imperfectly specific tests is that multiple testing increases the chance of abnormal results in healthy, non-exposed animals. For example, one may consider a situation where a veterinarian or other wildlife health professional is interested in screening banked sera for ten unrelated bacterial and viral infections, as part of a disease survey. If each test has a specificity of 90% and is performed on a single non-exposed animal, the probability of at least one positive test result is $1 - (0.9)^{10}$ or 0.65. Therefore, if the same ten tests were performed on a population of one hundred non-exposed animals, one would expect sixty-five animals to give a positive result in at least one of the tests. Of course, the investigator would never know that these animals had not been exposed to the agents in question, and might thus erroneously conclude that the population was exposed to some or all of the agents.

If the specificity of each test were higher, however, the probability of a false herd diagnosis would be reduced.

Sensitivity and specificity are generally considered ‘fixed’ values, although for some diseases the amount of the agent present, or the stage of disease (clinical or subclinical) may affect test sensitivity. Specificity may differ by geographical region, due to differences in the distribution of organisms of related antigenic structure which can produce cross-reacting antibodies, or due to differences in host populations (e.g. genetics, diet).

To estimate the validity of tests, sera from animals of known exposure or infection status must be tested. Experimental infections are unlikely to be ethical or practical in wildlife, and therefore it is most appropriate to use field samples from animals of representative age, health status, etc. for these determinations. If the test is to be used primarily to detect subclinical infections, evaluation of clinically-ill animals will tend to overestimate the sensitivity of the test. Non-exposed animals used for specificity estimation should include healthy animals, as well as those affected by other conditions which need to be differentiated from the infection under study. Detailed discussions of issues related to the design of test validation studies are given elsewhere (10, 13).

The following sampling designs can be used to assess the sensitivity and specificity of a test:

a) sampling based on the true exposure status (this approach necessitates the availability of banked samples of known exposure status)

b) sampling based on the serological test results, with follow-up of positive and negative individuals to determine their true exposure status, e.g. by necropsy and culture (this design requires a correction factor for the sampling fraction)

c) cross-sectional sampling, with determination of test results and true exposure status in all sampled animals.
Test results should be evaluated ‘blind’ (i.e. independently of knowledge of the true exposure status) to avoid the potential for biased interpretation. Once results are available, frequency distributions categorized by exposure status can be plotted in a bar graph (Fig. 1). For a specific cut-off value, results can be classed as positive or negative and cross-tabulated as shown in Table I. Sensitivity, specificity, and confidence intervals can then be calculated (8). Ideally, there should be at least one hundred exposed and one hundred non-exposed animals to obtain relatively precise estimates of both sensitivity and specificity.

**FIG. 1**

**Frequency distribution of serological titres for infected and non-infected animals**

Figure 1 shows that the frequency distributions of results for non-exposed and exposed animals overlap in the range 1:4 to 1:16. A cut-off of 1:8 might be appropriate to balance false positives and negatives, but if false positives are not allowed (100% specificity) the cut-off should be increased to 1:32. In this case, sensitivity would decrease to 81.7%. If false negatives are not allowed (100% sensitivity) the cut-off should be moved to 1:4, and specificity would decrease to 83.3%. Changing the cut-off shows the inverse relationship between sensitivity and specificity (10). Selection of an appropriate cut-off value for test interpretation depends on several factors, including the purpose of testing, the relative cost of false positives and false negatives, and the availability of definitive (confirmatory) tests. Although diagnostic laboratories usually report results as positive or negative at a single cut-off, different cut-offs might be appropriate, given different circumstances for testing and consequences of misclassification.
TABLE I
Sensitivity and specificity of results shown in Figure 1

<table>
<thead>
<tr>
<th>True status</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>101</td>
<td>5</td>
</tr>
<tr>
<td>-</td>
<td>3</td>
<td>85</td>
</tr>
</tbody>
</table>

Estimated sensitivity = 101/104 = 97.1 (95% confidence interval: 91.2-99.3)
Estimated specificity = 85/90 = 94.4 (95% confidence interval: 86.9-97.9)

One improvement on reporting sensitivity and specificity at a single cut-off value is to present a receiver-operating characteristic (ROC) plot (20) for the serological test. The ROC graph plots the true-positive proportion (sensitivity) on the y-axis versus the false-positive proportion (1 – specificity) on the x-axis over all possible cut-off values for the test. This method is especially useful for comparing the accuracy of two tests for the same agent. The ROC curve for the data in Figure 1 is shown in Figure 2.

FIG. 2
Receiver-operating characteristic plot for all possible cut-off values for the distributions of serological titres from infected and non-infected animals in Figure 1
PREDICTIVE VALUES

Predictive values are probabilities which provide an answer to the following related questions:
- What proportion of the test-positive animals are truly infected?
- What proportion of the test-negative animals are truly non-infected?

Predictive values depend on sensitivity, specificity and prevalence. Given that sensitivity and specificity estimates are available, predictive value curves can be generated to show how prevalence affects these estimates (Fig. 3). As prevalence decreases, the predictive value of a positive test decreases while the predictive value of a negative test increases.

Despite the use of an accurate test (sensitivity = 97.1%; specificity = 94.4%) in the example described above, most test positives will be non-infected (false positives) if the infection is present in a population at low prevalence.

The following strategies may be used to improve the predictive value of a positive test result when testing for infections of low prevalence:

a) testing ‘high-risk’ groups (e.g. those with clinical signs rather than apparently-healthy animals) or adults vs. juveniles (if there is an age-related risk of exposure)

b) increasing the cut-off value for the test, to increase specificity
c) using a testing technique of higher specificity

d) re-testing samples which gave a positive result in the first test with a confirmatory test of specificity close to 100%, and using series interpretation (only animals which also give a positive result in the second test are considered positive).

**ESTIMATION OF TRUE PREVALENCE FROM TEST RESULTS**

True prevalence can be estimated by using the following formula (15):

\[
\text{true prevalence} = \frac{\text{AP} + \text{sp} - 1}{\text{se} + \text{sp} - 1}
\]

where:

- \( \text{AP} \) = apparent prevalence = \( \frac{\text{no. of positives}}{\text{no. tested}} \)
- \( \text{sp} \) = specificity
- \( \text{se} \) = sensitivity.

Using the same example as in the previous section, a seroprevalence (i.e. apparent prevalence) of 10% obtained using a test with a sensitivity of 97.1% and specificity of 94.4% (cut-off of 1:8 in Figure 1) gives a true prevalence estimate of 4.8%. In some cases, adjustment for sensitivity and specificity yields a negative or zero estimate for true prevalence; such a finding suggests that the population is not infected. If sensitivity and specificity are unknown, however, this formula cannot be applied and the true prevalence of infection in the population cannot be calculated directly. Only the seroprevalence can be reported, and this may under- or overestimate the true prevalence of infection (depending on the relative magnitude of the sensitivity and specificity values).

**INFECTION STATUS OF POPULATIONS**

Often, the primary interest in testing is to determine whether the population has been exposed or is actually harbouring a pathogenic agent, rather than to estimate prevalence per se. This might be important in geographically-distinct populations where questions of animal translocation need to be addressed, or when risk factors for infection are being studied. The following factors affect sensitivity and specificity at the population level (11):

- \( a) \) sensitivity and specificity of the individual test
- \( b) \) number of animals tested
- \( c) \) prevalence of infection in the population
- \( d) \) number of positives used to classify the population as positive.

The lack of data on individual tests means that population sensitivity and specificity, as well as predictive values, are almost always unknown. Wildlife health professionals are often required to interpret population results on the basis of incomplete knowledge.
of individual test performance. The apparent prevalence (0-100%) and the number of seropositive individuals are both considered in determining the status of the population. Usually, if there is a high percentage of test positives, the decision on the exposure status of the population is clear. But what if a low apparent prevalence (e.g. < 20%) is detected when the population is tested? In these situations, knowledge of specificity alone would facilitate judgements about the infection status of the population.

In another example, a population of one hundred deer without a previous history of exposure to bluetongue virus is tested for presence of the virus by agar gel immunodiffusion, and five reactors (test positives) are detected, i.e. seroprevalence is 5%. The underlying question is whether or not the population has been genuinely exposed. Without knowing the characteristics of the test, one approach might be to attempt isolation of the virus or to use polymerase chain reaction methods to detect viral DNA in blood samples from the seropositive deer. If the specificity of the test was precisely known, however, a more rational assessment could be made without the need for further testing. For example, if the test had a specificity of 0.95, the expected number of positives in a non-infected population of one hundred deer would be five, i.e. exactly what was found on testing; if the specificity was < 0.95, the expectation that these were false positives would increase, as the expected number of false positives would be greater than five; if the test had a specificity of 0.999, as many as five positive results would be highly unlikely if the population was truly non-exposed.

Population status is most accurately determined by testing all animals, but this is often unnecessary and not cost-effective, especially if populations are infected at or above moderate prevalence (> 30%). Calculations for numbers of samples needed to detect infection/exposure in populations are given elsewhere (5), but are based on the assumption that tests are perfectly sensitive and specific. Sample sizes needed to detect significant differences in seroprevalence between two populations can be calculated with public domain software, such as Epi Info (7).

**USING TESTS TO MANAGE THE RISK OF INTRODUCING PATHOGENS**

Serological testing can play an important role in reducing the risk of introducing new pathogens, and has been recommended as an important component in veterinary assessment of the health status of source/founder stock (18). The epidemiological considerations can be demonstrated by the example below.

A rancher with a brucellosis-free bison herd is interested in augmenting the herd with twenty new bison from a large population estimated to have a 10% prevalence of brucellosis. The probability of introducing infection (i.e. ≥ 1 infected bison) if n bison are randomly selected from the herd can be calculated using the following formula (9):

\[ 1 - (1 - \text{prevalence})^n. \]

* If twenty bison are selected, the probability of introducing infection is \( 1 - (0.9)^{20} \) or 0.88. Therefore, brucellosis will almost certainly be introduced if the twenty animals are added to the herd without testing. Indeed, the risk of introducing infection is
proportional to the number of introduced bison. If a serological test with a sensitivity of 0.9 and a specificity of 1 is used, however, and only bison which test negative are introduced, what is the probability that disease will be introduced due to false-negative test results? The predictive value of a negative test in these circumstances can be calculated as equal to 0.989. Thus the probability that at least one infected bison will be introduced is equal to $1 - (0.989)^{20}$ or 0.20, i.e. a substantially lower risk. To totally eliminate risk, the bison should be purchased from 'brucellosis-free populations' or, if this is not possible, a test with a sensitivity of 100% should be used.

It should be emphasized that introduction of an infectious agent is a necessary but not a sufficient criterion for transmission of the infection to animals in the recipient herd. The probability of infection becoming established in the recipient herd will be less than or equal to the probability of introducing infection, and depends on management and environmental factors.

To minimize the risk of disease transmission among populations, it is preferable to use data which include histories and serological profiles of the complete population, rather than relying on test results of purchased animals. To unequivocally establish 'freedom' from a particular pathogen, every animal in a population must be tested using a test which is 100% sensitive. This requirement might be realistically met in a captive population, but is unlikely ever to be satisfactorily achieved in free-ranging or semi-captive populations. In an ideal world, translocations of animals might be best restricted to populations of similar health status, as the ramifications of the introduction of an agent can sometimes be serious (11).

**CONCLUSIONS**

Decision-making by those managing the health of wildlife populations would be facilitated if better data were available on the sensitivity and specificity of commonly-used tests. Such data would improve estimates of disease prevalence and would help in assessing the risk of disease transmission from the imported animal, or from local animals to the imports. The risk of transmission between wild animals and livestock populations could also be better evaluated. Recently-developed molecular technologies provide the opportunity to detect infectious agents with improved sensitivity and specificity, but there is still a need to validate these tests, assure appropriate standardization of tests and reagents, and implement quality control procedures. Whether testing provides cost-effective information for management of wildlife health and conservation is another important consideration. Decision analysis (16) using risk estimates (such as those described above) can be used to address this question, but the topic is outside the scope of this article.
VALIDITÉ DE L'UTILISATION SUR LES ANIMAUX SAUVAGES DES ÉPREUVES SÉROLOGIQUES DE DIAGNOSTIC. - I.A. Gardner, S. Hietala et W.M. Boyce.

Résumé : L'utilisation des épreuves de diagnostic destinées au bétail domestique sur des espèces sauvages maintenues en captivité ou vivant en liberté, soulève deux problèmes. En premier lieu, il n'existe pratiquement pas de tests dont l'utilisation ait été validée pour le bétail. Ensuite, une épreuve sérologique n'entraîne pas nécessairement les mêmes réactions chez les espèces domestiques et sauvages ; en effet, il existe des différences parmi les souches et les sérovars pathogènes, parmi les réponses sérologiques et enfin parmi les expositions à des micro-organismes de structure antigénique similaire, mais induisant des réactions sérologiques croisées. Certaines épreuves nécessitent des réactifs ou des composants spécifiques qui ne sont pas toujours disponibles ; de surcroît, la plupart des tests n'ont pas été normalisés. Les auteurs décrivent les principes devant régir l'évaluation d'un test de diagnostic sérologique. Ils expliquent aussi, en l'illustrant d'exemples, que la connaissance précise de la sensibilité et de la spécificité d'un test peut servir à estimer la prévalence réelle, à déterminer si une population est infectée et à faciliter la prise de décision en matière de déplacements d'animaux.


* * *

VALIDEZ DEL EMPLEO DE PRUEBAS SEROLÓGICAS PARA EL DIAGNÓSTICO DE ENFERMEDADES EN ANIMALES SALVAJES. – I.A. Gardner, S. Hietala y W.M. Boyce.

Resumen: La transposición de las pruebas de diagnóstico utilizadas para especies de ganado doméstico a especies salvajes, ya sea en libertad o en cautividad, presenta dos problemas. Primero, la mayoría de las pruebas existentes no han sido validadas de forma adecuada sobre el ganado doméstico. Y segundo, el presupuesto de que una prueba serológica funcionará de manera idéntica en la fauna salvaje y en especies domésticas puede no ser correcto, debido a la existencia de diferencias en las cepas patógenas y serovares, en la respuesta serológica del huésped y en la exposición a microorganismos de estructura antigénica similar que producen anticuerpos capaces de reacciones cruzadas. Algunos ensayos requieren la utilización de ciertos reactivos o componentes específicos de especie que tal vez no estén disponibles en el mercado, y la mayor parte de las pruebas no han sido estandarizadas. Los autores señalan los principios que deben regir la evaluación de una prueba de diagnóstico serológica, y ofrecen ejemplos sobre cómo el conocimiento de la sensibilidad y especificidad de una prueba puede servir para estimar la prevalencia real, para determinar si una población está infectada y para facilitar las decisiones de manejo relativas al traslado de animales.

REFERENCES


