Journal of Veterinary Cardiology (2009) **11**, S33—S40

ELSEVIER

Journal of Veterinary Cardiology
ESVC

www.elsevier.com/locate/jvc

Measures of accuracy and performance of diagnostic tests

Kenneth J. Drobatz, DVM, MSCE

Department of Clinical Studies, School of Veterinary Medicine, University of Pennsylvania, 3900 Delancey Street, Philadelphia, PA 19104-6010, USA

Received 12 November 2008; received in revised form 2 January 2009; accepted 10 March 2009

KEYWORDS
Diagnostic test;
Sensitivity;
Specificity;
Predictive values;
Likelihood ratio;
Receiver operating curve;
Confidence interval

Abstract Diagnostic tests are integral to the practice of veterinary cardiology, any other specialty, and general veterinary medicine. Developing and understanding diagnostic tests is one of the cornerstones of clinical research. This manuscript describes the diagnostic test properties including sensitivity, specificity, predictive value, likelihood ratio, receiver operating characteristic curve.
Methods: Review of practical book chapters and standard statistics manuscripts.
Results: Diagnostics such as sensitivity, specificity, predictive value, likelihood ratio, and receiver operating characteristic curve are described and illustrated.
Conclusion: Basic understanding of how diagnostic tests are developed and interpreted is essential in reviewing clinical scientific papers and understanding evidence based medicine.

Introduction

Evaluations of biomarkers for cardiac disease have recently been reported in veterinary clinical medicine. Several investigators have studied N-terminal pro-B-type natriuretic peptide (NT-proBNP) and N-terminal pro-A-type natriuretic peptide (NT-proANP) as markers to detect cardiac disease.[1—4] Their objective is to use these molecules as diagnostic tests for the presence of heart disease. Why would they want to develop these diagnostic tests when there are the gold standards of thoracic radiography and echocardiography? The reasons are several fold and include cost, invasiveness, risk, turn-around time, and efficiency, to name a few. If the gold standard test is readily available, economical, non-invasive, efficient, causes minimal risk to the patient, requires minimal expertise and is easily performed, then there is no logical reason to develop another diagnostic test for the disease that is investigated. However, this is rarely the case and clinicians are constantly exploring new ways to diagnose disease with tests that fit the above criteria. In the case of the cardiac biomarkers, clinicians are hoping to develop tests that detect cardiac disease with

E-mail address: drobatz@vet.upenn.edu

1760-2734/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jvc.2009.03.004

a simple blood sample rather than having to subject every suspect patient to the more costly, less efficient, expertise-heavy, and sometimes risky (in unstable patients) thoracic radiography and echocardiography. Given the economic climate and the rising costs of veterinary medicine, developing good diagnostic tests can be extremely beneficial for the patients, their owners, veterinarians and their staff. From a strictly diagnostic perspective, a good diagnostic test would never miss an animal with the disease and would always correctly identify animals with the disease being tested for. This review will discuss the diagnostic test parameters so that the readers can more easily understand the development of a diagnostic test, its properties and whether it can be beneficial in their practice.

Hypothetical example

Veterinary clinicians are often presented with cats in respiratory distress. Dyspneic cats are often unstable and even a small amount of handling or restraint can prove fatal. Unfortunately, the underlying cause of respiratory distress is seldom obvious and diagnostic tests are required to identify the etiology, and more specifically to differentiate heart failure and primary respiratory disease. While both thoracic radiography and echocardiography help rule in or rule out cardiac disease, these diagnostic procedures require substantial and prolonged animal restraint that could compromise or even result in the death of some severely dyspneic cats. Clearly, a simple test that is economical, easy to perform, safe, relatively non-invasive, and could definitively rule in or rule out cardiac disease would be ideal. Veterinary cardiologists and emergency clinicians recognize this diagnostic dilemma and are investigating it. As a hypothetical example of developing a diagnostic test to delineate these two conditions using a simple blood test, the following data have been developed. This test will be referred to as the Sick Cat Cardiac Hormone (SCCH). Assuming that the assay for this hormone has already been validated, it now needs to be evaluated in its ability to distinguish cats with cardiac disease causing respiratory distress versus cats with primary respiratory disease as the cause of the problem. Ideally, this hormone measurement would be positive in all cats that truly have cardiac disease (sensitivity) and would be negative in all cats without heart disease as the cause for the respiratory distress (specificity). It would also be ideal if the animal tests positive for the disease, then it is highly likely to have the disease (positive predictive value) and if the animal tests negative for the disease, then it is highly unlikely to have the disease (negative predictive value). Additionally, the test should have good discriminating properties so that when faced with an animal with a positive test, it is much more likely to be positive in a cat with cardiac disease compared to a cat without cardiac disease (likelihood ratio positive). Conversely, when faced with a cat with a negative test, the test would have a much higher probability of being negative in a cat with cardiac disease compared to a cat without cardiac disease (likelihood ratio negative). Each of these diagnostic test properties is reviewed in detail below.

Sensitivity and specificity

The sensitivity and specificity are the more commonly recognized properties of diagnostic tests. These are most commonly utilized when the diagnostic results are dichotomous (e.g., yes or no, present or not present, etc.). For example, some clinicians anecdotally believe that auscultation of a heart murmur in cats with respiratory distress is an excellent indicator of cardiac disease as the underlying cause. Rather than using anecdotal impression, one could evaluate whether this is true by prospectively evaluating cats with respiratory distress and specifically making a point of auscultating for a cardiac murmur. Then determine whether they truly have cardiac disease as the cause of the respiratory distress by utilizing the gold standard of thoracic radiography and echocardiography. Because detecting a murmur can be subjective, at least two clinicians should auscultate the hearts of these cats as well. Both clinicians should be ''blind'' to the other's conclusion regarding the presence of a murmur (one should not know whether the other detected a murmur or not so that they are not influenced by each other). Additionally, the gold standards (thoracic radiography and echocardiography) should be performed after the clinicians' auscultations and be applied to all cats in the study (for this example, 100 consecutive cats are evaluated). Additionally, the clinicians interpreting the thoracic radiographs and echocardiograms should not know whether or not a murmur had been detected. For ease of discussion, let's assume that both clinicians are in agreement about the presence of a murmur or not in all the cats. The results are presented below.

In this hypothetical example there were 60 cats with cardiac disease as the cause of the respiratory distress and 40 cats that had primary respiratory disease. Of the 60 cats with cardiac disease, 50