Comparison of results of three commercial heartworm antigen test kits in dogs with low heartworm burdens

Clarke E. Atkins, DVM, DACVIM

Objective—To compare results of 3 commercial heartworm antigen test kits performed on serum samples from dogs infected with low numbers of adult female heartworms.

Design—Blinded laboratory evaluation.

Sample Population—Serum samples from dogs (n = 208) proven at necropsy to be infected with 1 to 4 adult female heartworms and from dogs (32) without heartworms.

Procedure—Samples were sequentially tested with each test kit, following the manufacturers’ instructions, by licensed veterinary technicians in private practice who were not aware of infection status of the dogs. The order of test kit evaluations was randomly chosen. For each test kit, sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were evaluated.

Results—All tests yielded some false-negative results, and there were significant differences among tests in regard to ability to detect low heartworm burdens. Sensitivity of the test kits ranged from 78 to 84%. For all test kits, sensitivity increased as number of female heartworms increased. All 3 test kits had high specificity (97%).

Conclusions and Clinical Relevance—Results indicated that sensitivity of the 3 commercially available heartworm antigen test kits ranged from 78 to 84% when used to test serum samples from dogs with low heartworm burdens, and that sensitivity varied among test kits. For all 3 test kits, specificity was 97%. All 3 test kits yielded false-positive and false-negative results for some dogs with low heartworm burdens. (J Am Vet Med Assoc 2003;222:1221–1223)
interpret the tests as they would in their own practices, but without consultation among themselves. In response to questions, the technicians were found to be less familiar with the Solo Step and CHAT test kits than with the Snap test kit. Therefore, to avoid potential bias, technicians were given a chance to become familiar with the use of each test kit prior to testing of study samples, and all technicians stated that they were comfortable with test kit procedure before testing for the present study was begun. Each technician was assigned 40 serum samples and instructed to analyze each of the 40 samples 3 times, using each of the 3 test kits in succession. Because samples were rerandomized and relabeled between testing sessions, the technicians did not know if they were evaluating the same or different samples in each successive session. Technicians were provided with a data sheet with the code for each sample and recorded the sample code number on each test kit device to ensure precision. Results were recorded as positive or negative, and technicians signed the data sheet at the end of each test session. It was predetermined that samples would not be reanalyzed in the event of discrepant results. All testing was performed on the same day.

The order of test kit use was determined by random draw. The Snap test kit was used to analyze all 240 samples first, followed by the CHAT and then the Solo Step test kit. Immediately prior to each test session, the technicians were required to read the test kit instructions. This was followed by a demonstration of the test kit by a technician familiar with all 3 test kits. Each technician was then required to test at least 1 positive control and 1 negative control sample with each test kit. Technicians were allowed to repeat this process with each test kit until they were comfortable with its use.

Sensitivity, specificity, accuracy, and positive and negative predictive values and their 95% confidence limits were calculated as described. The Cochran Q test followed by the exact McNemar test with binomial probabilities was used to compare estimates of sensitivity and accuracy among the 3 test kits. Bonferroni corrections were applied when doing pairwise comparisons. Confidence limits were calculated by use of exact binomial probabilities. Values of \( P < 0.05 \) were considered to be significant.

**Results**

Each of the 3 test kits yielded a single false-positive result (Table 1); therefore, specificity was 97% for all 3. Each test kit identified a different sample as false positive.

Sensitivity of the Snap test kit was significantly (\( P = 0.01 \)) higher than sensitivity of the other 2 test kits (Table 1). For each test kit, sensitivity increased as the number of infecting adult female heartworms increased (Table 2). For each heartworm burden (ie, 1, 2, 3, or 4 adult female heartworms identified at necropsy), sensitivity was not significantly different among test kits. When results for samples from dogs infected with 1 or 2 adult female heartworms were combined, sensitivities of the Snap, Solo Step, and CHAT test kits were 75, 73, and 68%, respectively, with sensitivity of the Snap test kit being significantly higher than that of the other 2 test kits.

Accuracy was >80% for all 3 test kits (Table 1). Accuracy of the Snap test kit was significantly higher than accuracy of the other 2 test kits.

Positive predictive values (ie, probability of heartworm infection for a dog with a positive test result) were identical for the 3 test kits (99% confidence limits, 97 to 100%). Negative predictive values (ie, probability that a dog with a negative test result would be free from heartworm disease) were lower, reflecting the relatively small percentage of samples from uninfected dogs that were evaluated. Negative predictive values were 41% (95% confidence limits, 30 to 53%) for the CHAT test kit, 48% (36 to 61%) for the Snap test kit, and 42% (31 to 54%) for the Solo Step test kit. Negative predictive value of the Snap test kit was significantly higher than values for the other 2 test kits.

**Discussion**

Previous studies have suggested that the current generation of heartworm antigen test kits is specific. This was confirmed in the present study in which specificity of each test kit was 97%.

Sensitivity of the 3 test kits evaluated in the present study was generally good, ranging from 78 to 84%. It should be emphasized that the study was designed to evaluate the sensitivity of these test kits in dogs with low heartworm burdens (≤4 adult female heartworms), and that 133 of the 240 serum samples were from dogs with only 1, 2, or 3 adult female heartworms identified at necropsy. Thus, conditions of this study created a particular challenge for the test kits. As heartworm burden increased, the sensitivities of the test kits improved.

### Table 1—Results of 3 commercial heartworm antigen tests performed on serum samples from 208 dogs with low heartworm burdens (≤4 adult female heartworms identified at necropsy) and 32 dogs not infected with heartworms

<table>
<thead>
<tr>
<th>No. of samples</th>
<th>True positive</th>
<th>True negative</th>
<th>False positive</th>
<th>False negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test kit</strong></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
<td>Accuracy (%)</td>
<td></td>
</tr>
<tr>
<td>CHAT</td>
<td>163 (72–84)</td>
<td>97 (84–100)</td>
<td>81 (71–82)</td>
<td></td>
</tr>
<tr>
<td>Snap</td>
<td>175 (76–89)*</td>
<td>97 (84–100)</td>
<td>82 (71–82)</td>
<td></td>
</tr>
<tr>
<td>Solo Step</td>
<td>165 (73–85)</td>
<td>97 (84–100)</td>
<td>82 (71–82)</td>
<td></td>
</tr>
</tbody>
</table>

*Value was significantly \( P < 0.05 \) different from values for the other 2 test kits.
In the present study, the Snap test kit had significantly higher sensitivity, accuracy, and negative predictive value than did the CHAT and Solo Step test kits. Although all 3 kits are membrane-format antigen tests, test technology differs. Results of the present study, in conjunction with results of a previous study,7 would seem to suggest that the membrane ELISA format used in the Snap test kit, although slightly more complicated to perform, is superior to the lateral flow immunoassay methodology used by the other 2 test kits.

Recent publications4,5 have reported results of similar studies of the sensitivity of heartworm antigen tests performed by commercial laboratories and of heartworm test kits. McCall et al.4 showed that commercial laboratories performed by commercial laboratories and of heartworm studies of the sensitivity of heartworm antigen tests with a single female heartworm, but were otherwise comparable in sensitivity to tests performed in commercial laboratories. Authors of that study evaluated 1 kit (Solo Step test kit) and the immediate predecessor6 of another test kit used in the present study (Snap test kit) and found that sensitivities were 76% in dogs infected with 1 female heartworm and 95% and 85%, respectively, in dogs infected with 2 female heartworms. These values were higher than sensitivities found in the present study, whereas in a similar study7 involving serum samples from naturally infected dogs, sensitivities of the Solo Step and Snap test kits were 56 and 65%, respectively, for dogs infected with 1 or 2 female heartworms, compared with sensitivities of 73 and 75% in the present study. The reasons for these different sensitivity and specificity values are unclear, but these differences may reflect the fact that tests were performed on different samples by different technicians. It is also possible that the higher sensitivities obtained more recently represent improvements in technology.

Results of this study demonstrate that the sensitivity of the current generation of heartworm antigen tests, even under these rigorous conditions, is overall quite good and that, as expected, the sensitivity of these tests improved as heartworm burdens increased. In fact, for the 2 test kits that have been evaluated previously (Solo Step and Snap), sensitivity estimates in the present study, using serum samples from dogs infected with a maximum of 4 adult female heartworms, were higher than estimates in a previous study3 in which samples from dogs with higher heartworm burdens (≤10 female heartworms) were used. This most likely represents improvements in technology as the test kits continue to evolve. It is notable that the improved sensitivity has apparently not been associated with a decrease in specificity.

Clinically relevant findings of the present study include verification of the high sensitivity of currently available heartworm antigen test kits, even in dogs with low heartworm burdens, and identification of differences in sensitivity among kits. Nevertheless, results indicate that with all 3 test kits, results may be falsely negative in some dogs with low heartworm burdens. On the other hand, results confirmed the high specificity of these test kits, indicating that false-positive results are unlikely. However, because each test kit mistakenly provided a positive test result for 1 uninfected dog, repeating tests for dogs with positive results, particularly dogs without clinical signs of heartworm disease, would seem to be judicious.

In the present study, negative predictive values (ie, probability that a dog with a negative test result would be free from heartworm disease) for all 3 test kits were <50%. This low value reflects the low heartworm burdens in infected dogs from which samples were collected and the low percentage of samples from uninfected dogs. Therefore, these values are likely lower than the values that can be expected in a practice situation. On the other hand, positive predictive values (ie, probability of heartworm infection for a dog with a positive test result) and accuracy (ie, percentage of samples that yielded correct results) were likely higher than values that can be expected in a practice situation for the same reason that negative predictive values were low. The relationship between prevalence and positive and negative predictive values has been shown previously.1

Results of the present study must be interpreted in light of the experimental design. The performance of heartworm antigen test kits varies, depending on infection status (ie, proportion of infected vs uninfected dogs, heartworm burden, heartworm sex ratio, and presence of immature or dying heartworms) of the dogs being studied.1 Additionally, serum testing may not exactly mimic results obtained with blood, given the potential for residual blood components to diminish the ability to properly interpret results.2 Finally, the significant differences in sensitivity and accuracy among test kits in the present study might not be clinically important at lower prevalence rates, and results of the present study cannot be expected to exactly mimic those found in any given practice because of the variability of population characteristics.

References