J Vet Clin 18(4) : 329-333(2001)

Diagnostic Accuracy of Urease and Polymerase Chain Reaction to Detect *Helicobacter* Species Infection in Dogs

Son-Il Pak* and Tae-Ho Oh[1]

**College of Veterinary Medicine and Biomedical Science, Colorado State University.*
College of Veterinary Medicine, Kyungpook National University.

Abstract : Evaluation on the diagnostic performances of urease test and polymerase chain reaction (PCR) for detection of *Helicobacter* species infection in dogs has rarely been performed in research with site-specific situations, although assessing diagnostic tests is an essential part prior to its practical use in a variety of clinical settings. The clinical value of a diagnostic test may be misjudged and comparisons between different tests may yield misleading conclusions when high within-patient correlations are present. We applied a conceptually simple statistical approach to estimate the sensitivity and specificity of urease test and PCR for detection of *Helicobacter* species infection in dogs. This approach assumes that responses from three different sampling sites within an animal are correlated where unit for statistical analysis is the site rather than the animal. The sensitivity and specificity of urease test was 0.74 (95% confidence interval, 0.64-0.84) and 0.87 (95% CI, 0.67-1.00), respectively. For PCR, the sensitivity was 0.95 (95% CI, 0.89-1.00) and specificity 0.90 (95% CI, 0.70-1.00). Two tests were almost equally specific. Urease test, however, has a lower diagnostic accuracy and thus should only be used after careful validation in terms of sensitivity.

Key words **:** *Helicobacter*, Diagnosis, Sensitivity, Specificity. Dog

Introduction

Most species of the genus *Helicobacter* are efficient colonizers of mammalian stomachs within a restricted host range[19]. In particular almost all adult dogs are infected with *Helicobacter* species based on evaluation wither with the urease test, gastric histology or direct staining of gastric mucus[5,12], although several epidemiological characteristics such as the route of infection and specific prevalence rate related to age of the animal are not exactly known.

The accuracy of a diagnostic test for predicting the presence or absence of a disease is often evaluated by estimating its sensitivity and specificity with respect to a gold standard in making the diagnosis. Sensitivity is the probability that a test will be positive when true diagnosis is positive, and specificity is the probability that a test will be negative given the true diagnosis is negative. Their complements, 1-sensitivity and 1-specificity, are the false negative and false positive error rates associated with the test.

Typically, a widely recognized assumption for experiments where patients are the experimental units of analysis is that responses are independently distributed, and therefore the theoretical sensitivity and specificity of a test are calculated directly from the probability model based on the assumption of independence of observations. Fleiss[6] discussed the problem of estimating sensitivity and specificity when the observations are independent.

However, for tests that measure similar biologic phenomena such as serum antibody responses to infectious agents, it is logical to expect that test results will be dependent, conditional on an animals true status. Vacek[24] and Gardner *et al*[8] described statistical approaches to investigate dependence among tests. Lachenbruch[16] and Schulzer[23] described the situation where a new test is given to each subject several times.

In other situations for site-specific observations where the experimental unit of analysis is the site the sensitivity and specificity are the probability that a site within a patient with the condition will be classified by the test as being with the condition. In addition the results for sites from the same subject can be highly correlated depending on the magnitude of the correlation and the number of sites per patient sampled. Ignoring the correlation between sites tends to yield significant underestimation of the true variance of sensitivity and specificity. Therefore, site-specific data should be analyzed with a statistical methodology that accounts for the dependence of within-patient observations.

Hujoel *et al*[14] discussed the use of the correlated binary models of Bahadur[2] for obtaining standard errors of sensitivity and specificity estimates when observations are correlated. Kupper and Haseman[15] proposed correlated binomial model to assess the strength of possible correlation in the same litter for certain toxicological experiments.

The increased use of diagnostic tests to determine a condition in animal and human medicine has raised concerns about the diagnostic accuracy of the results of these procedures. The purpose of this study was to estimate the diagnostic accuracy of urease test and polymerase chain reaction (PCR) for site-specific *Helicobater* species infection in dogs.

[1]Corresponding author.
E-mail : thoh@knu.ac.kr

Materials and Methods

Study animal

A total of 78 dogs were recruited for the study with owners informed consent in three local hospitals, Colorado, USA. About half of the subjects (52.9%) were from a random sample of the outpatient and the others were selected from the in-patient. They were all did not associated with gastric illnesses. Nine dogs were excluded in the analysis because of missing data or unreliable results of testing. The median age of dogs was 2.3 years, ranging from 1.5 to 3.2 years. The presence or absence of gastric *Helicobacter* spp. was ascertained in dogs by evaluating gastric biopsies for urease activity, histopathology, and PCR.

Gastric biopsy

Biopsies of the stomach were obtained from anesthetized dogs with a pediatric endoscope. Endoscopic biopsies were procured from the pyloric antrum (incisura to pyloric sphincter), the body (greater curvature) and the cardia. Three biopsies were taken from each site, two for urease testing and light microscopy one for PCR.

Urease test

Urease activity was evaluated as previously described[21]. The specimens were inserted into Christensens Urea 2% agar for 24 h and observed for 24 h for a change in the color of the indicator medium. A change from orange-red to bright pink was considered a positive result.

PCR

Gastric biopsies collected endoscopically were frozen at -80°C. PCR was performed using primers with *Helicobacter* genus-specific primers directed against 16S rDNA. DNA extraction and PCR procedures were followed by the method performed by other researchers[5,7].

Histological examination

Samples stained with hematoxylin and eosin (HE) and modified Steiners stain for histopathology were evaluated for the number of organisms, degree of inflammation, and the presence of lymphoid follicles. The criteria for histological diagnoses of gastric biopsies were described in detail elsewhere[12]. Only a pathologist evaluated all tissue sections. For microscopy, the fragments were smeared onto a glass slide, heat-fixed, stained with 40% carbolfuchsin for 5 min, and examined under oil immersion on the basis of size and spiral, rod-shaped morphology. In this study the results from histopathologic findings were considered as a reference test.

Statistical analysis

In situation that site-specific responses within a patient are correlated several methods are available. Among these approaches we compared four estimation methods: binomial estimator (BE)[16], ratio estimator (RE)[22], correlation estimator (CE)[4], and weighted estimator (WE)[17]. A Fortran computer program was used to estimate the sensitivity and specificity for correlated responses.

Table 1. Test results of urease activity and PCR

Test *	Gold test	Test result	
		Positive	Negative
Urease	Pathology findings (+)	82	29
	Pathology findings (-)	4	26
PCR	Pathology findings (+)	105	6
	Pathology findings (-)	3	27

*For urease test, 82 out of 111 sites are true positive results and 26 out of 30 are true negative results. For PCR, 105 out of 111 sites are true positive results and 27 out of 30 are true negative results.

Results

The test results between the two diagnostic tests and a histopathologic findings that ascertains true disease status were summarized in Table 1, which is assumed responses on sites within a patient were assumed to be independent observations. The sample-sizes in both groups are not coincide, since in some patients all sites were not subjected to the both test.

Table 2 shows a comparison of the estimates of sensitivity and specificity, standard error and 95% confidence interval (CI). There were no great differences in the estimates of sensitivity and specificity among four methods. The standard errors of the estimators are similar except that of the BE. The BE showed narrower 95% CI than the other estimators since it ignores the correlation between sites within a patient.

Discussion

Gastric *Helicobacter* infection among humans and domestic animals is common, and the species have been reported in dogs[12], cats[20], mice[10], swine[11], and cattle[3]. The infection has been associated with chronic gastritis, peptic ulcers, and cancers[18] but no clear indication of clinical importance for practitioners. In animal population there is no established gold standard to diagnose *Helicobacter*-associated infection in terms of perfect sensitivity and specificity; relatively specific test lacks sensitivity.

Estimation of sensitivity and specificity is a simple matter when the true diagnostic status can be determined[6]. Unfortunately, this is often impractical or impossible, and sensitivity and specificity are estimated by comparing a new test with a reference test, which also has error rates associated with it. If one disregards the error rates of the reference test and calculates the error rates for the new test in the usual

Table 2. Estimates of the sensitivity and specificity with standard error (SE) among four methods

Method*	Sensitivity	SE	95% CI#	Specificity	SE	95% CI#
			Urease test			
BE	0.74	0.042	0.66 0.82	0.87	0.062	0.75 0.99
RE	0.74	0.050	0.64 0.84	0.87	0.102	0.67 1.00
CE	0.74	0.050	0.64 0.84	0.87	0.098	0.68 1.00
WE	0.74	0.050	0.64 0.84	0.87	0.102	0.67 1.00
			PCR			
BE	0.95	0.021	0.90 0.99	0.90	0.055	0.79 1.00
RE	0.95	0.030	0.89 1.00	0.90	0.100	0.70 1.00
CE	0.95	0.030	0.89 1.00	0.90	0.095	0.71 1.00
WE	0.95	0.030	0.89 1.00	0.90	0.100	0.70 1.00

*BE, binomial estimator; RE, ratio estimator; CE, within-patient correlation estimator; WE, weighted estimator.
confidence interval.

manner, the estimates will be biased. When the error rates of the reference test are known, however, appropriate estimates for the error rates of the new test can be obtained as described by Gart and Buck[9]. Hui and Walter[13] have shown that if two diagnostic tests with unknown error rates are simultaneously applied to individuals from two populations with differing disease prevalences, the maximum likelihood estimates for the error rate of both tests, as well as the prevalence rates for the two populations, can be obtained.

The primary reason using a correlated model is that the variances of estimates can differ considerably. In this study the estimates of two diagnostic tests showed almost the same among the four methods applied. The standard error of the sensitivity and specificity assuming responses on sites within a patient were to be independent observations would be 0.042 or $\sqrt{(0.74\times0.26)/111}$ and 0.062 or $\sqrt{(0.867\times0.133)/30}$, respectively. The confidence interval of BE, however, was narrower than those of the other 3 methods. This indicates that BE can considerably reduce the variance of the estimates when the within-patient correlation exists, resulting in inflation of type I error rate. The specificity of urease test for BE was 0.062 whereas 0.102 for RE. Thus 95% confidence interval using BE would yield an interval with only 80% actual coverage (z-score = $1.96\times0.062/0.102 = 1.19$) not 95%.

In this study, several statistical methods have been evaluated for the estimation of the sensitivity and specificity of the site-specific diagnostic test results are derived from multiple sites within individual animals. The statistical methods used in this paper are described in detail elsewhere[1]. Briefly, when the observations are independent, the appropriate model is the binomial model:

$$P(X=x) = nCx\ p^x q^{n-x}$$

Where nCx represnts the number of possible combinations of n animals taken x at a time, x denotes the number of animals with true positive or true negative test results, n is the number of diseased or non-diseased animals, p is the sensitivity or specificity parameter, and q=1-p. An estimate of p can be obtained by dividing the true number of true positive or true negative test results x by the number of diseased or non-diseased sites n. The variance of the estimated sensitivity or specificity is pq/n. The BE ignores the correlation between sites within each animal. Rao and Scott[22] proposed a simple way of estimating the variance of correlated binary data. The overall proportion in RE is equal to BE, but RE uses a concept of design effect and sample size in estimation due to clustering. The RE method assigns a patient with k sites the same weight as k patients with 1 site each although the correlation of sites within patients implies that a patient with k sites contributes less information.

The CE method incorporates the correlation of the responses within each animal. The within-patient correlation coefficient can be obtained in analysis of variance table by treating the outcome of each animal, coded as 0 or 1, as a continuous variable. Lee and Dubin[17] proposed the weighted estimator to overcome the drawbacks of RE method, which assign equal weight regardless the number of sites. Ahn[1] recommended the choice of estimators depending on the values of the number of patients, sensitivity and specificity estimate, and the estimate of the within-patient correlation.

We can expand the concept of conditional dependence between tests into two-tests with two populations. For example, if both tests are based on a particular antibody reaction, something when inhibits the reaction or causes a false reaction for one of the tests may have a similar effect on the other. It is important to get accurate estimates of sensitivity and specificity of a diagnostic test to avoid misinterpretation of test results.

Conclusion

Because diagnostic tests are not fully accurate, often site-specific correlation exists within an animal diagnostic accu-

racy estimates should be used with caution because of underestimation of variance estimates. This paper showed several statistical methods, which allow the calculation of sensitivity and specificity for site-specificity observations when reference test results are available on an entire study samples. We illustrated that the assumption of test independence in site-specific situation will result in an underestimation of the error rates of the new test if it is positively correlated with the reference test.

Reference

1. Ahn C. An evaluation of methods for diagnostic sensitivity for correlated observations. Technical Report, Division of Clinical Epidemiology, University of Texas Health Science Center, 1996.
2. Bahadur RR. A representation of the joint distribution of responses to n dichotomous items. In: Soloman H (ed.), Studies in item analysis and prediction. Stanford University Press, Stanford, 1961.
3. De Groote D, van Doorn LJ, Duscatelle R, Verschuuren A, Tilmant K, Quint WG, Jalava K, Vandamme P. Phylogenetic characterization of *Candidatus Helicobacter bovis*, a new gastric *helicobacter* in cattle. Int J Syst Bacteriol 1999; 49: 1707-1715.
4. Donner A, Klar N. Confidence interval construction for effect measures arising from cluster randomization trials. J Clin Epidemiol 1993; 46: 123-131.
5. Eaton KA, Dewhirst FE, Paster BJ, Tzellas N, Coleman BE, Paola J, Sherding N. Prevalence and varieties of *Helicobacter* species in dogs from sources and pet dogs: Animal public health implications. J Clin Microbiol 1996; 34: 3165-3170.
6. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. Wiley-Intersciences, New Yrok, 1981.
7. Fox JG, Dewhirst FE, Shen Z, Feng Y, Taylor NS, Paster BJ, Ericson RL, Lau CN, Correa P, Araya JC, Roa I. Hepatic Helicobacter species identified in bile and gallbladder tissue from Chileans with chronic cholecystitis. Gastroenterology 1998; 114: 755-763.
8. Gardner IA, Stryhn H, Lind P, Collins MT. Conditional dependence between tests affects the diagnosis and surveillance of animal disease. Prev Vet Med 2000; 45: 107-122.
9. Gart JJ, Buck AA. Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. Am J Epidemiol 1966; 83: 593-602.
10. Goto K, Ohashi H, Ebukuro S, Itoh K, Tohma Y, Takakura A, Wakana S, Ito M, Itoh T. Isolation and characterization of *Helicobacter* species from the stomach of the house musk shrew (*Suncus murinus*) with chronic gastritis. Curr Microbiol 1998; 37: 44-51.
11. Grasso GM, Ripabelli G, Sammarco ML, Ruberto A, Iannitto G. Prevalence of *Helicobacter*-like organisms in porcine gastric mucosa: A study of swine slaughtered in Italy. Comp Immun Microbiol Infect Dis 1996; 19: 213-217.
12. Happonen I, Saari S, Castren L, Tyni O, Hännien M-L, Westermarck E. Occurrence and typographical mapping of gastric *Helicobacter*-like organisms and their association with histological changes in apparently halthy dogs and cats. J Vet Med 1996; 43: 305-315.
13. Hui SL, Walter SD. Estimating the error rates diagnostic tests. Biometrics 1980; 36: 167-171.
14. Hujoel PP, Moultron LH, Loesche WJ. Estimation of sensitivity and specificity of site-specific diagnostic tests. J Periodont Res 1990; 25: 193-196.
15. Kupper LL, Haseman KJ. The use of a correlated binomial model for the analysis of certain toxicological experiments. Biometrics 1978; 34: 69-76.
16. Lachenbruch PA. Multiple reading procedures: the performance of diagnostic tests. Stat Med 1988; 7: 549-557,.
17. Lee E, Dubin N. Estimation and sample size considerations for clustered binary responses. Stat Med 1994; 13: 1241-1252.
18. McGee DJ, Mobley HL. Mechanisms of *Helicobacter pylori* infection: bacterial factors. Curr Top Microbiol Immunol 1999; 241: 155-180.
19. Otto G, Hazell SH, Fox JG, Howlett CR, Murphy JC, ORourke JL, Lee A. Animal and public health implications of gastric colonization of cats by *Helicobacter*-like organisms. J Clin Microbiol 1994; 32: 1043-1049.
20. Paster BJ, Lee A, Fox JG, Dewhirst FE, Tordoff LA, Fraser GJ, ORourke JL, Taylor NS, Ferrero R. Phylogeny of *Helicobacter felis* spp. Nov., *Helicoabcter mustelae*, and related bacteria. Int J Syst Bacteriol 1991; 41: 31-38.
21. Radin MJ, Eaton KA, Krakowka S, Morgan DR, Lee A, Otto G, Fox JG. *Helicobacter pylori* gastric infection in gnotobiotic beagle dogs. Infect Immun 1990; 58: 2606-2612.
22. Rao JNK, Scott A. A simple method for the analysis of clustered binary data. Biometrics 1992; 48: 577-585.
23. Schulzer M, Anderson DR, Drance S. Sensitivity and specificity of a diagnostic test determined by repeated observations in the absence of an external standard. J Clin Epidemiol 1991; 44: 1167-1179.
24. Vacek PM. The effects of conditional dependence on the evaluation of diagnostic tests. Biometrics 1985; 41: 959-968.

개에서 *Helicobacter*균 감염을 검출하기 위한 urease 검사와 PCR 검사의 진단적 정확도

박선일* · 오태호[1]

*College of Veterinary Medicine and Biomedical Science, Colorado State University

경북대학교 수의과대학

요 약 : 새로 개발되거나 혹은 기존의 어떤 진단검사를 다양한 임상상황에 적용하기 위해서는 먼저 이들 검사법의 진단적 정확도를 추정하는 연구가 반드시 선행되어야 한다. 진단의 정확도에 대한 추정치를 모른다면 검사결과를 해석하는 것이 불가능하기 때문이다. 특히 동일한 개체에서 감염부위별로 3개 이상의 시료를 얻어 진단검사를 적용하는 경우 각 시료의 검사결과는 독립적인 측정시료가 아니라 개체내에서 연관성이 매우 높은 종속적인 시료에 해당한다. 즉 동일한 개체에서 얻은 시료일수록 검사결과에서 유사한 반응을 보이며 이 경우 분석의 단위는 각각의 개체가 아니라 검사부위가 되는데 이는 의학연구에서 매우 흔하다. 본 연구에서는 *Helicobacter* 균에 의한 감염을 검출하기 위하여 동일한 개로부터 위의 해부학적 구조상 pyloric antrum, body 및 cardia의 생검시료에 대하여 urease 검사와 PCR 검사를 적용하여 각 검사의 진단적 정확도를 추정하였다. urease 검사의 민감도와 특이도는 0.74 (95% 신뢰구간: 0.64-0.84)와 0.87 (95% 신뢰구간: 0.67-1.00)이었으며 PCR 검사의 민감도와 특이도는 0.95 (95% 신뢰구간: 0.89-1.00)와 0.90 (95% 신뢰구간: 0.70-1.00)로 두 검사의 특이도는 높은 것으로 나타났다. 그러나 PCR 검사에 비하여 urease 검사의 경우 가음성 (false negative)의 가능성이 높기 때문에 진단결과에 대한 신중한 해석이 요구된다.