Reliability of the Whiff Test in Clinical Practice

Andreas Cohrssen, MD, Matthew Anderson, MD, MSc, Alina Merrill, DO, and Diane McKee, MD, MSc

Vaginitis is among the most common reasons for gynecological consultation in primary care. Although the work-up of vaginal symptoms is well described in the literature, women often go without a diagnosis,[1,2] and a recent study using cultures as a gold standard found that clinician diagnoses were not very accurate.[3]

Bacterial vaginosis (BV) is the most common cause of vaginitis in patients presenting to health care providers. The diagnosis of BV rests on four criteria, one of which is the "whiff test."[4] The whiff test is performed by mixing a sample of vaginal discharge with potassium hydroxide and smelling the sample for a characteristic fishy odor.[5]

Although the whiff test is one of the most common clinical tests in primary care, its reliability has never been assessed. We undertook this study to determine if the whiff test was a reliable diagnostic maneuver as measured by interobserver variability.

Participants, Methods, and Results

The study was conducted at 3 academic urban family practice clinics (clinics A, B, and C) serving primarily working-class communities in New York City. Each time that a clinician collected a specimen of vaginal discharge for the evaluation of a symptomatic patient, the sample was considered eligible for the study. The clinician collecting the sample (clinician 1) identified another clinician (clinician 2) who happened to be available at the time and passed along several drops of the discharge to him/her. Both clinicians separately performed whiff tests on the sample and noted the results on half of a pre-numbered perforated card. Samples were coded as "definitely positive" or "not definitely positive." The patient was managed according to the assessment of clinician 1, and neither clinician communicated their results to the other at any time.

The whiff test is performed routinely at all sites. We assumed all clinicians were competent to perform the test and provided no training or standardization before the study. The clinicians involved were all attending physicians except for one family nurse practitioner and one resident in family medicine (an author, AM). The Institutional Review Boards at Beth Israel Hospital and Montefiore Medical Center considered the study exempt.

Fifty-two samples were collected. The overall raw concordance between observers was 85% and the κ value was 0.68 (see Table 1). Values for the 3 individual clinics were as follows: clinic A, κ 0.47 (17 patients); clinic B, κ 0.70 (20 patients); clinic C, κ 0.86 (15 patients).

Comment

A κ value of 0.68 is generally interpreted as showing moderate agreement between 2 observers and is not an uncommon value for diagnostic tests. A recent review article on the κ statistic[6] cites κ values of 0.56 for the detection of jugular venous distention, 0.75 for the diagnosis of alcoholism from the CAGE questionnaire, and 0.82 for the straight leg raise test. Thus, our data confirm that the whiff test provides useable clinical information.

The κ values for the 3 clinics were quite divergent. Our study was not designed to evaluate these differences or their significance. However, the data raise the intriguing possibility that clinical practice might in some sites be sub-optimal and a target for improvement.

Several reasons may account for disagreement between observers. Among the test-related factors might be the use of KOH bottles of differing potency, any delay in performance of the test, use of insufficient quantity of discharge by one observer, or interference with the test by use of absorbent material (such as a cotton swab). Among observer-

Submitted 10 February 2005; revised 7 April 2005; accepted 12 April 2005.

From the Institute for Urban Family Health, Beth Israel Medical Center, New York, NY (AC, AM); and Department of Family and Social Medicine, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, NY (MA, DM).

Conflict of interest: none declared.

Corresponding author: Matt Anderson, MD, Montefiore Family Health Center, 360 E. 193rd St., Bronx, NY 10458 (e-mail: andersonma@aol.com).

Table 1. Interobserver Variability of the Whiff Test*

	Clinician 2		
	Test Positive	Test Negative	
Clinician 1			
Test positive	16	6	22
Test not definitely positive	2	28	30
	18	34	52

* Raw concordance: 44/52 = 85%; κ value: 0.68.

dependent factors might be the degree of skill in performing the test and the ability to smell. The degree of ventilation and distance from the sample during the test may also have altered results between observers. We did not collect data on these various factors. It is probable, however, that under more controlled circumstances, following a specified protocol and with specifically trained clinicians, that the whiff test might perform better.

This study examined the performance of the whiff test in actual clinical practice and not as performed in a research setting. Even under these less than ideal circumstances the whiff test appears to be a moderately reliable clinical tool.

This study was undertaken as part of the New York City Research and Improvement Network, a practice-based research network, and we thank our many colleagues who participated. Drs. Arthur Blank and Clyde Schechter provided invaluable statistical advice.

References

1. Schaaf VM, Perez-Stable EJ, Borchardt K. The limited value of symptoms and signs in the diagnosis of vaginal infections. Arch Intern Med 1990;150:1929–33.
2. Berg AO, Heidrich FE, Fihn SD, et al. Establishing the cause of genitourinary symptoms in women in a family practice. Comparison of clinical examination and comprehensive microbiology. JAMA 1984;251:620–5.
3. Allen-Davis JT, Beck A, Parker R, Ellis JL, Polley D. Assessment of vulvovaginal complaints: accuracy of telephone triage and in-office diagnosis. Obstet Gynecol 2002;99:18–22.
4. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. Am J Med 1983;74:14–22.
5. Gardner HL, Dukes CD. Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified "nonspecific" vaginitis. Am J Obstet Gynecol 1955;69:962–76.
6. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). CMAJ 2004;171:1369–73.