JOURNAL OF CLINICAL MICROBIOLOGY, July 1990, p. 1506–1508
0095-1137/90/071506-03$02.00/0

Vol. 28, No. 7

Reproducibility of Interpretation of Gram-Stained Vaginal Smears for the Diagnosis of Bacterial Vaginosis

TONY MAZZULLI, ANDREW E. SIMOR,* AND DONALD E. LOW

Department of Microbiology, Mount Sinai Hospital, and University of Toronto, Toronto, Ontario M5G 1X5, Canada

Received 4 January 1990/Accepted 4 April 1990

In the diagnostic microbiology laboratory, interpretation of Gram-stained slides of vaginal swab specimens is used to support the clinical diagnosis of bacterial vaginosis. The reproducibility with which technologists interpret these Gram-stained slides was evaluated by presenting, in coded fashion, 80 original slides and 80 duplicate slides of vaginal swab specimens to three technologists. They each interpreted the original slide twice and the duplicate slide from the same specimen once. Intraobserver and interobserver agreement was assessed by use of the weighted kappa statistic. Semiquantitation of *Lactobacillus* and *Gardnerella* morphotypes and a diagnosis of bacterial vaginosis showed the greatest intraobserver agreement, with kappa values ranging from 0.772 to 1.000. Interobserver agreement was also high for rating *Lactobacillus* morphotypes and clue cells (kappa values between 0.735 and 0.869) but decreased slightly for *Gardnerella* morphotypes and a diagnosis of bacterial vaginosis (kappa values between 0.656 and 0.800). These results indicate that there is good agreement for the interpretation of Gram-stained slides of vaginal swab specimens and that this method alone, without culture, can be used reliably to support the clinical diagnosis of bacterial vaginosis.

Although the term bacterial vaginosis had been used previously (17, 18), in 1984, Weström et al. (20) first defined it and proposed that it be used to replace the term nonspecific vaginitis. Clinically, the diagnosis of bacterial vaginosis is based on several factors including the presence of a characteristic homogeneous, thin, grey vaginal discharge, a vaginal pH of ≥4.5, a positive amide odor test (release of a fishy amine odor when vaginal fluid is mixed with 10% KOH), and the identification of clue cells (vaginal epithelial cells heavily coated with coccobacilli) seen on microscopic examination (1). However, vaginal swab specimens are often sent to the diagnostic microbiology laboratory to identify the microbial cause of bacterial vaginosis. Currently, it is recognized that there are several organisms that may be associated with this condition including *Gardnerella vaginalis*, anaerobic bacteria (*Bacteroides* species and *Mobiluncus* species), and even *Mycoplasma hominis* (3, 18, 20). The Gram stain of vaginal fluid from patients with a clinical diagnosis of bacterial vaginosis has a characteristic appearance. Typically, it will show many small gram-negative organisms resembling *G. vaginalis* in the absence of *Lactobacillus* species (5). A study by Spiegel et al. (17) showed that a Gram stain was consistent with bacterial vaginosis in 25 of 25 women with that clinical diagnosis and in none of 35 women with *Candida* vaginitis or normal examinations. Thus, it has been recommended that a Gram stain alone without culture of vaginal fluid could be used for the diagnosis of bacterial vaginosis (6, 12, 17). Although intraobserver and interobserver variability was assessed by Spiegel et al. (17), this was done using specimens from only 10 patients, with vaginal washings rather than vaginal swabs, and it was not clear what level of expertise (microbiologist or laboratory technologist) the evaluators had. It has been shown that there can be great variability in intraobserver and interobserver interpretation of Gram-stained specimens from other body sites, such as from sputum (2, 19). Since this potential problem has not been fully evaluated for the interpretation of vaginal swab Gram stains and because the Gram stain alone, without culture, is being more frequently used in microbiology laboratories to support the diagnosis of bacterial vaginosis, we examined the reproducibility with which technologists interpret Gram stains of vaginal swabs.

* Corresponding author.

MATERIALS AND METHODS

Specimens. Eighty consecutive vaginal swab specimens received in the microbiology laboratory from different patients were evaluated. Specimens were obtained from the gynecology clinic, family physicians, and the emergency department. All specimens were collected with a cotton-tipped swab and transported to the laboratory in Amies transport medium with charcoal (NCS Diagnostics, Mississauga, Ontario, Canada). Two Gram-stained slides were prepared from each specimen. All slides were coded numerically, with no other identifying features placed on the slides.

Study protocol. Gram-stained slides were presented to three experienced microbiology technologists. A total of 240 slides were presented to each technologist in sets of 20 at a time (80 original slides, 80 duplicate slides of the same specimen, and the 80 original slides presented a second time). All slides were presented with different code numbers, so that no technologist was aware that they might be reading the same slide twice. Neither the amount of time each technologist could take to read a slide nor the number of fields to be examined was specified, but rather, the technologists were asked to interpret the slides as if they were routine clinical specimens.

Smear interpretation. Gram-stained slides were examined under oil immersion (×1,000). Each technologist was asked to quantitate the number of large gram-positive bacilli (*Lactobacillus* morphotypes), small gram-variable bacilli (*Gardnerella* morphotypes), curved gram-negative bacilli (*Mobiluncus* morphotypes), and all other bacteria, yeast cells, or trichomonads on a scale of 1 to 4 according to the method of Spiegel et al. (17). The categories used were 1+ (<1 cell per field), 2+ (1 to 5 cells per field), 3+ (6 to 30 cells per field), and 4+ (>30 cells per field). They were also asked to quantify the number of polymorphonuclear neutrophils (pus cells) using the same rating scale. Clue cells were rated as

TABLE 1. Intraobserver agreement for reading 80 Gram-stained slides of vaginal swab specimens twice

Interpretation	Kappa value for observer:		
	1	2	3
Bacterial morphotype			
Lactobacillus	0.937	0.927	0.849
Gardnerella	0.955	0.944	0.798
Mobiluncus	0.463	0.956	0.897
Cell morphotype			
Pus cells	0.543	0.777	0.678
Clue cells	0.945	0.875	0.836
Diagnosis (bacterial vaginosis)	0.945	0.875	0.772

either present or absent. Finally, all technologists were asked to interpret each slide as being either consistent with bacterial vaginosis or showing no evidence of bacterial vaginosis.

Statistical analysis. The weighted kappa statistic (4) was used to measure agreement about the quantity of the various bacteria and pus cells within and between technologists. The kappa statistic ranges from 0 to 1, with 1 representing perfect agreement and 0 representing no more agreement than would be expected to occur on the basis of chance alone. The weighted kappa statistic gives no credit for agreement equal to that expected by chance alone but gives partial credit for responses that are in close but not perfect agreement. In this study, perfect agreement was given a weight of 1, differences of one category (e.g., 1+ versus 2+) received a weight of 0.889, differences of two categories (e.g., 1+ versus 3+) received a weight of 0.556, and differences of three categories (e.g., 1+ versus 4+) received a weight of 0 (7). In assessing the assessments of the technologists for the presence of clue cells and their overall interpretation of the slides for the presence or absence of bacterial vaginosis, a weight of 1 was given for perfect agreement and a weight of 0 was given for no agreement. Calculated kappa values of ≤ 0.40 are considered to reflect fair to poor reproducibility or agreement, those of ≥ 0.41 and ≤ 0.80 are considered to reflect moderate to substantial agreement, and those of ≥ 0.81 reflect almost perfect agreement (11).

Intraobserver reproducibility was determined by comparing the rating of each technologist of the original slide and their rating of the same slide on a separate reading. Interobserver reproducibility was determined by pairing each rating of a smear by a technologist with that of both the other technologists. Also, the interpretation of the 80 original slides for the diagnosis of bacterial vaginosis made by the three technologists under test conditions was compared with the interpretation of the same 80 slides made by technologists in the routine microbiology laboratory as they were processed under nontest conditions.

RESULTS

Table 1 shows the kappa values for intraobserver agreement for reading the same slide twice. Agreement between the two interpretations was good, with the strength of agreement ranging from moderate-substantial (kappa, ≥ 0.41 and ≤ 0.80) to almost perfect (kappa, ≥ 0.81). Intraobserver rating of pus cells, whose presence suggests vaginitis rather than vaginosis, showed the weakest agreement (kappa ranging from 0.543 to 0.777), while agreement for the rating of lactobacilli and clue cells was the greatest, with kappa values of ≥ 0.81 suggesting almost perfect agreement.

TABLE 2. Intraobserver agreement for reading 80 Gram-stained slides and 80 duplicate slides made from the same vaginal swab specimens

Interpretation	Kappa value for observer:		
	1	2	3
Bacterial morphotype			
Lactobacillus	0.876	0.854	0.755
Gardnerella	0.730	0.781	0.802
Mobiluncus	0.700	0.815	0.794
Cell morphotype			
Pus cells	0.638	0.692	0.500
Clue cells	0.902	0.701	0.953
Diagnosis (bacterial vaginosis)	0.902	0.750	1.000

Table 2 shows intraobserver agreement for reading two separate slides made from the same specimen. The results are similar to those for the interpretation of the same slide twice, with the greatest agreement being for the rating of lactobacilli and a diagnosis of bacterial vaginosis and the least agreement for the rating of pus cells.

Table 3 shows the interobserver agreement for different technologists reading the same slide. These results continued to show almost perfect agreement for the rating of lactobacilli among the three technologists but showed a lower rate of agreement for all other categories, with kappa values in the moderate to substantial category.

A review of the routine records of the microbiology laboratory revealed that 9 (11.2%) of the original 80 slides were given a final interpretation of being consistent with bacterial vaginosis. This compared with a range of 10 to 12 slides (12.5 to 15%) given this diagnosis under the test conditions. Comparing each of the study technologists with the routine microbiology laboratory for the interpretation of slides for the diagnosis of bacterial vaginosis gave kappa values between 0.580 and 0.945.

DISCUSSION

The diagnosis of bacterial vaginosis is no longer a diagnosis of exclusion made in patients with symptoms of vaginitis from whom *Trichomonas* or *Candida* species are not isolated but rather is based on specific clinical criteria (15). The diversity of the bacteria, including *G. vaginalis*, *Mobiluncus*

TABLE 3. Interobserver agreement for reading the same 160 Gram-stained slides of vaginal swab specimens

Interpretation	Kappa value for observer:		
	1 vs 2	1 vs 3	2 vs 3
Bacterial morphotype			
Lactobacillus	0.869	0.830	0.817
Gardnerella	0.782	0.696	0.656
Mobiluncus	0.526	0.683	0.710
Cell morphotype			
Pus cells	0.685	0.587	0.685
Clue cells	0.866	0.735	0.790
Diagnosis (bacterial vaginosis)	0.672	0.800	0.781

spp., and other anaerobic organisms, in these patients has been well recognized (3, 14, 18, 20). Although the diagnosis of bacterial vaginosis is based on clinical criteria, vaginal swab specimens are often sent to the microbiology laboratory for processing to rule out other possible diagnoses and to add support to the clinical diagnosis. Originally, the standard in the laboratory was culture of vaginal specimens for *G. vaginalis* (8, 9). However, with the recognition that bacterial vaginosis is associated with a diverse group of organisms, many of which are difficult, cumbersome, and costly to culture in the laboratory, and since *G. vaginalis* can be found in up to 40 to 50% of otherwise healthy women (10, 16, 18), it has become increasingly more common to process these specimens by Gram stain alone, without culture (13, 17). The correlation between a clinical diagnosis of bacterial vaginosis and a positive Gram-stained smear showing distinctive characteristics has been shown (17). However, the reproducibility with which Gram-stained slides are interpreted by the same observer as well as different observers has not previously been well documented. This study therefore attempted to determine the intraobserver and interobserver variability in the interpretation of Gram-stained slides of vaginal swab specimens. The principal finding of this study was that the reproducibility of the Gram stain interpretations was very good, often showing substantial to near-perfect agreement.

There are several possible reasons for this high level of consistency. First, the technologists may have been more likely to adhere to a quantitation scale under study conditions than in routine daily specimen processing. However, our results indicate that even when comparing the final interpretation of the Gram-stained slides for the diagnosis of bacterial vaginosis made under the test conditions and by the routine microbiology laboratory under nontest conditions, there continued to be a high level of agreement. Prior to commencing the study, a teaching session to review the microbiological characteristics of bacterial vaginosis was carried out for all the technologists in the laboratory. Therefore, the high degree of consistency may reflect the fact that the features of bacterial vaginosis were fresh in the minds of the technologists. Finally, the high level of agreement for intraobserver interpretation may be the result of the technologist's recall or identification of slides previously read despite efforts to code and randomize slides to prevent this. However, this would seem unlikely, since there was a large sample of slides (240 readings) and the agreement remained high for the reading of the original slides and the duplicate slides which would have been seen only once by each technologist.

Since the interpretation of Gram-stained smears of vaginal swab specimens has been shown to correlate with the clinical diagnosis of bacterial vaginosis and because a high level of intraobserver and interobserver agreement has been demonstrated, we believe the Gram-stained smear alone, without culture, can be used to evaluate vaginal swab specimens for bacterial vaginosis.

ACKNOWLEDGMENTS

We thank P. Lo, B. Willey, and C. Porter for technical assistance, S. Bull for statistical analysis, and M. Apanay for secretarial assistance.

LITERATURE CITED

1. **Amsel, R., P. A. Totten, C. A. Spiegel, K. C. Chen, D. Eschenbach, and K. K. Holmes.** 1983. Nonspecific vaginitis: diagnostic criteria and microbial and epidemiological associations. Am. J. Med. **74**:14–22.
2. **Bartlett, R. C., J. Tetrault, J. Evers, J. Officer, and D. Derench.** 1979. Quality assurance of gram-stained direct smears. Am. J. Clin. Pathol. **72**:984–990.
3. **Blackwell, A. L., A. R. Fox, I. Phillips, and D. Barlow.** 1983. Anaerobic vaginosis (non-specific vaginitis): clinical, microbiological and therapeutic findings. Lancet **ii**:1379–1382.
4. **Cohen, J.** 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol. Bull. **70**:213–220.
5. **Dunkelberg, W. E.** 1965. Diagnosis of *Haemophilus vaginalis* vaginitis by gram-stained smears. Am. J. Obstet. Gynecol. **91**:998–1000.
6. **Eschenbach, D. A., S. Hillier, C. Critchlow, C. Stevens, T. DeRouen, and K. K. Holmes.** 1988. Diagnosis and clinical manifestations of bacterial vaginosis. Am. J. Obstet. Gynecol. **158**:819–828.
7. **Fleiss, J. L.** 1981. Statistical methods for rates and proportions, 2nd ed., p. 121–236. John Wiley & Sons, Inc., New York.
8. **Gardner, H. L., and C. D. Dukes.** 1955. *Haemophilus vaginalis* vaginitis. Am. J. Obstet. Gynecol. **69**:962–976.
9. **Gardner, H. L., and C. D. Dukes.** 1959. *Haemophilus vaginalis* vaginitis. Ann. N.Y. Acad. Sci. **83**:280–289.
10. **Jones, B. M., G. R. Kinghorn, and B. I. Duerden.** 1982. An overview of the diagnosis and treatment of *Gardnerella vaginalis* and *Bacteroides* associated vaginitis. Eur. J. Clin. Microbiol. **1**:320–325.
11. **Kramer, M. S., and A. F. Feinstein.** 1981. Clinical biostatistics. LIV. The biostatistics of concordance. Clin. Pharmacol. Ther. **29**:111–123.
12. **Krohn, M. A., S. Hillier, and D. Eschenbach.** 1989. Comparison of methods for diagnosing bacterial vaginosis among pregnant women. J. Clin. Microbiol. **27**:1266–1271.
13. **Milatovic, D., K. Machka, R. V. Brosch, H. J. Wallner, and I. Braveny.** 1982. Comparison of microscopic and cultural findings in the diagnosis of *Gardnerella vaginalis* infection. Eur. J. Clin. Microbiol. **1**:294–297.
14. **Piot, P., E. Van Dyck, P. Godts, and J. Vanderheyden.** 1982. The vaginal microbial flora in non-specific vaginitis. Eur. J. Clin. Microbiol. **1**:301–306.
15. **Sobel, J. D.** 1989. Bacterial vaginosis—an ecologic mystery. Ann. Intern. Med. **111**:551–553.
16. **Spiegel, C. A., R. Amsel, D. Eschenbach, F. Schoenknecht, and K. K. Holmes.** 1980. Anaerobic bacteria in nonspecific vaginitis. N. Engl. J. Med. **303**:601–606.
17. **Spiegel, C. A., R. Amsel, and K. K. Holmes.** 1983. Diagnosis of bacterial vaginosis by direct Gram stain of vaginal fluid. J. Clin. Microbiol. **18**:170–177.
18. **Spiegel, C. A., D. Eschenbach, R. Amsel, and K. K. Holmes.** 1983. Curved anaerobic bacteria in bacterial (nonspecific) vaginosis and their response to antimicrobial therapy. J. Infect. Dis. **148**:817–822.
19. **Valenstein, P. N.** 1988. Semiquantitation of bacteria in sputum Gram stains. J. Clin. Microbiol. **26**:1791–1794.
20. **Weström, L., G. Evaldson, K. K. Holmes, W. van der Meijden, E. Rylander, and B. Fredriksson.** 1984. Taxonomy of vaginosis; bacterial vaginosis—a definition. Scand. J. Urol. Nephrol. Suppl. **86**:259–264.