JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 1991, p. 1730–1731
0095-1137/91/081730-02$02.00/0

Vol. 29, No. 8

Reproducibility of a Scoring System for Gram Stain Diagnosis of Bacterial Vaginosis

M. RIDUAN JOESOEF,[1]* SHARON L. HILLIER,[2] SUHARNO JOSODIWONDO,[3] AND MICHAEL LINNAN[4]

Division of STD/HIV Prevention, Center for Prevention Services,[1] and International Health Program Office,[4] Centers for Disease Control, Atlanta, Georgia 30333; Department of Obstetrics and Gynecology, University of Washington, Seattle, Washington 98195[2]; and Department of Microbiology, School of Medicine, University of Indonesia, Jakarta, Indonesia[3]

Received 11 February 1991/Accepted 13 May 1991

A total of 225 pairs of duplicate Gram-stained slides from three hospitals in Jakarta were evaluated independently by a local (University of Indonesia, Jakarta) and a referral (University of Washington, Seattle) laboratory by the new scoring criteria proposed by Nugent et al. The correlation coefficients of the duplicate Gram stain scores ranged from 0.65 to 0.83. The kappa statistics for the bacterial vaginosis category (no, score of 0 to 6; yes, score of 7 to 10) ranged from 0.62 to 0.77. These findings confirm the good to excellent interobserver reliability of the new scoring system and the importance of slide preparation.

Several studies have demonstrated that Gram stains of vaginal fluid correlate well with the clinical diagnosis of bacterial vaginosis (1, 3, 6). Recently, a multicenter study (5) evaluated the intercenter reliability of Gram-stained smears and found a moderate reliability using criteria described by Spiegel et al. (6) and good to excellent reliability for a new scoring system proposed by Nugent et al. (5). Our study, in Jakarta, Indonesia, is the first independent attempt to evaluate the interobserver reliability of the proposed new scoring system by using data generated under field conditions.

The specimens for this study were obtained from pregnant women who attended clinics for prenatal care at 16 to 30 weeks of gestation in three hospitals in Jakarta. A total of 257 vaginal smear specimens were randomly selected from December 1989 through May 1990 for preparation of duplicate slides (i.e., two slides were prepared from each patient specimen). One slide was sent to the local microbiology laboratory (Department of Microbiology, University of Indonesia, Jakarta) and the other slide was sent to the referral laboratory (Department of Obstetrics and Gynecology, University of Washington, Seattle).

The vaginal smears were obtained from the vaginal sidewall by using a moistened cotton swab. A technician rolled the swab over clean glass slides, fixed the slides with heat, and sent the slides to the local and referral laboratories for staining (using safranin as the counterstain) and interpretation. Each Gram-stained slide was evaluated under oil immersion (1,000×).

At the beginning of the study an Indonesian microbiologist (S.J.) was trained for 3 days in the referral laboratory at the University of Washington, Seattle. He was the only person in the local laboratory who stained and interpreted the Gram-stained slides, and he was unaware of the clinical condition of the patient. The results of evaluation from one laboratory were not known by members of the other laboratory.

Of the 257 pairs of Gram-stained slides, 32 slides could not be evaluated because either the amount of material was insufficient for assessment or the identification numbers and/or dates of specimen collection were illegible (i.e., pairs could not be matched). The remaining 225 pairs of slides were analyzed; 108 slides were prepared at hospital A, 69 slides were prepared at hospital B, and 48 slides were prepared at hospital C.

Gram stain interpretation was done by the new method proposed by Nugent et al. (5). Briefly, in the new scoring system, three morphotypes were used to create a total score of 0 to 10. These three morphotypes (the most reliably recognized morphotypes) are large gram-positive rods (*Lactobacillus*), small gram-negative or gram-variable rods (*Bacteroides* or *Gardnerella*), and curved gram-negative to gram-variable rods (*Mobiluncus* spp.). The total scores were computed by adding the weighted quantitation (0 to 4+) of the three morphotypes. A score of 7 to 10 was considered to indicate bacterial vaginosis and a score of 0 to 6 was considered to indicate no bacterial vaginosis.

Different statistical analyses were carried out for the two types of Gram stain interpretations: quantitative (raw score) and categorical (bacterial vaginosis versus no bacterial vaginosis). Spearman's rank correlations of the scores were computed. For the categorical data, we computed the percentages of agreement and the kappa statistics and used the McNemar test to evaluate the differences in the proportion of samples that were classified as being positive for bacterial vaginosis.

Spearman's rank correlation coefficients of the duplicate Gram stain scores ranged from 0.65 in hospital C to 0.83 in hospital B (Table 1). All of the correlation coefficients are statistically significant. According to the criteria for kappa (2) (excellent, value of >0.75; good to fair, 0.75 to 0.40; and poor, <0.40), the findings indicate that the interobserver reliabilities were excellent for two hospitals (A and B) but only good to fair for hospital C.

The percentage of agreement for the presence or absence of bacterial vaginosis (no, score of 0 to 6; yes, score of 7 to 10) ranged from 87.5% in hospital C to 90.7% in hospital A (Table 1). Similarly, the kappa statistics ranged from 0.62 in hospital C to 0.77 in hospital B. The differences between the referral and local laboratories in the proportions of samples diagnosed as positive for bacterial vaginosis were not statistically significant.

These findings indicate that the new scoring method for Gram stain diagnosis of bacterial vaginosis is reliable and that a microbiologist can be trained in a relatively short time to reproducibly evaluate the smears. Although the presence

* Corresponding author.

TABLE 1. Correlation coefficient, percentage of agreement, and kappa statistics for Gram-stained duplicates

Hospital	Correlation coefficient[a]	% Agreement[b]	Kappa	Proportion difference[c]	Total no. in sample
A	0.81	90.7	0.75	0.0	108
B	0.83	89.9	0.77	0.07	69
C	0.65	87.5	0.62	0.0	48
A, B, and C	0.78	89.8	0.75	0.02	225

[a] For all values, $P < 0.0001$.
[b] Percent that agreed on diagnosis of bacterial vaginosis (no, score of 0 to 6; yes, score of 7 to 10).
[c] Difference in the proportions classified as bacterial vaginosis between referral and local labs. For all values, $P > 0.10$.

of clinical signs (discharge, elevated pH, clue cells, and amine odor) is the standard method used for the diagnosis of bacterial vaginosis, the clinical diagnosis poses a reliability problem in large, multicenter studies which involve more than one clinician or in developing countries where such clinical expertise is scarce. In either situation, a simple, easily standardized, and reliable method of diagnosis is needed. This study is the first to confirm that the Gram stain method using the new scoring system is reliable under field conditions in a developing country.

Our findings are consistent with those of the study of Nugent et al. (5). The coefficients of correlation (0.82 in the study of Nugent et al. and 0.78 in this study) are similar. Since the original study by Spiegel et al. (6) in 1983, two other studies have evaluated the reliability of Gram-stained vaginal smears for diagnosing bacterial vaginosis and have found good agreement (4, 5).

Although all the Gram-stained slides were evaluated by one local microbiologist, the interobserver reliability was significantly different for the slides prepared in hospitals A and B (correlation coefficients of 0.81 and 0.83, respectively) and slides in hospital C (correlation coefficient of 0.65). These findings indicate the importance of careful slide preparation.

This work was supported by a grant no. 497-0253 from the U.S. Agency for International Development Mission in Jakarta.

Budi Utomo, Gulardi Wiknjosastro, and Nyoman Kandung assisted in this study.

REFERENCES

1. **Eschenbach, D. A., S. L. Hillier, C. Critchlow, C. Steven, T. DeRouen, and K. K. Holmes.** 1988. Diagnosis and clinical manifestation of bacterial vaginosis. Am. J. Obstet. Gynecol. **158**:819–828.
2. **Fleiss, J. L. (ed.).** 1981. The measurement of interrater agreement, p. 212–236. Statistical methods for rates and proportions. John Wiley & Sons, New York.
3. **Krohn, M. A., S. L. Hillier, and D. A. Eschenbach.** 1989. Comparison of methods for diagnosing bacterial vaginosis among pregnant women. J. Clin. Microbiol. **27**:1266–1271.
4. **Mazzulli, T., A. E. Simor, and D. E. Low.** 1990. Reproducibility of interpretation of Gram-stained vaginal smear for the diagnosis of bacterial vaginosis. J. Clin. Microbiol. **28**:1506–1508.
5. **Nugent, R. P., M. A. Krohn, and S. L. Hillier.** 1991. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of Gram stain interpretation. J. Clin. Microbiol. **29**:297–301.
6. **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.