

Diagnosis of bacterial vaginosis: need for validation of microscopic image area used for scoring bacterial morphotypes

P-G Larsson, B Carlsson, L Fåhræus, T Jakobsson, U Forsum

Sex Transm Infect 2004;**80**:63–67. doi: 10.1136/sti.2003.006106

See end of article for authors' affiliations

Correspondence to:
P-G Larsson, Department of Obstetrics and Gynaecology, Kärnsjukhuset, 541 85 Skövde, Sweden; p-g.larsson@vgregion.se

Accepted for publication 22 August 2003

Background: The diagnosis of bacterial vaginosis (BV) is often made according to Nugent's classification, a scoring system based on bacterial counting of Gram stained slides of vaginal secretion. However as the image area of the microscope field will influence the number of morphotypes seen there is a need to standardise the area.

Methods: A graph intended for recalculation of number of bacterial morphotypes seen by the observer using 1000× magnification from various microscope set-ups was constructed and applied to data sets typical for scoring BV. The graph was used in recalculation of Nugent scores, which were also compared with the Ison/Hay scores to evaluate the consequences for the diagnosis of BV.

Results: The observed image area differed by 300% among the investigated microscope set-ups. In two different data sets, one treatment study and one screening study, a considerable change in the number of women classified as intermediate was seen when the graph was used to standardise the image area. The recalculated numbers were also compared to the Ison/Hay classification. Weighted kappa indexes between the different methods were 0.84, 0.88, and 0.90, indicating that the methods are comparable.

Conclusion: Because of the considerable differences among image areas covered by different microscope set-ups used in Nugent and Ison/Hay scoring, there is a need to standardise the area in order to reach comparable scores reflecting the diagnosis of BV in different laboratories. The differences in the intermediate group will have a considerable effect on the results from both treatment and prevalence studies, even though the kappa indexes indicate very good agreement between the methods used.

The prevalence of bacterial vaginosis (BV) in different populations around the world ranges between 10–30%,¹ making BV the most common health problem affecting women. In BV the vaginal flora are characterised by 1000–10 000-fold increased concentration of anaerobic bacteria such as *Gardnerella vaginalis*, *Mycoplasma*, and *Mobiluncus*. To diagnose BV using the commonly used Amsel's criteria, the patient has to fulfil three out of four different criteria: typical homogeneous discharge, pH above 4.5, positive sniff test, and clue cells seen with microscopy.²

Gram stained smears for the diagnosis of BV were introduced by Spiegel, who investigated different bacterial morphotypes in relation to semiquantitative measures of the vaginal flora, but the presence of clue cells was excluded from the diagnosis. To fulfil the diagnosis of BV "Lactobacillus morphotype" should be depressed (0 to 2+) or absent together with a predominance of "Gardnerella morphotype" (3 or 4+).³ Spiegel's idea of counting bacterial morphotypes on an ordinal scale was incorporated into Nugent's classification, which was introduced in 1991,⁴ and this method is now commonly used. Nugent's scoring has been applied by different investigators with good kappa values.⁵

Another classification was recently suggested by Ison and Hay. They assess the relation between the *Lactobacillus* morphotype and *Gardnerella* morphotype in Gram stained vaginal smears rather than counting the bacteria on an ordinal scale.^{6,7}

The original work of Nugent does not state the area covered by the high power microscopic field. The area covered by the microscopic field does, however, profoundly influence the outcome of scoring. For example, in assessing the histological grades of breast cancer, the mitotic counts per microscopic field are calculated.⁸ The mitotic activity is

assessed in a minimum of 10 fields. Up to nine mitoses per 10 fields gives a score of 1, 10–19 mitoses gives a score of 2, and more than 20 mitoses a score of 3, based on a microscope with a field area of 0.274 mm². This scoring system can then easily be adapted to other microscopes with different field areas using a graph that has been constructed for this purpose and which compensates for differences in microscopic area.⁹ This standardisation enables pathologists to compare data with those from other investigators.

The purpose of this study was to construct a similar graph for use in recalculating Nugent scores for the diagnosis of BV and to evaluate the consequences for Nugent's and Ison/Hay's scoring in two different patient materials, one from a treatment study and one from a prevalence study.

MATERIAL AND METHODS

The diagnosis of BV was based on the Nugent scoring system⁴ (table 1) where the numbers of different bacterial morphotypes in Gram stained smears are counted in high power fields using microscopy with a 1000× magnification. The points achieved from the number of different bacterial morphotypes are added together, with a total score of 0–3 considered normal and a score of 7–10 consistent with BV. A score of 4–6 is classified as intermediate.

The diameter of the image areas covered by six different lens and ocular microscopic set-ups (table 2) was measured using a stage micrometer with a 0.01 mm interval scale. The diameter of the image area of each microscope was calculated using the formula $A = r^2 \times \pi$.

The image area of one microscope set-up (Zeiss FL30 in table 2) was used as the reference for Nugent scoring, since this microscope was used in an international validation of

Table 1 The Nugent scoring system

Score	<i>Lactobacillus</i> morphotype per field	<i>Gardnerella</i> morphotype per field	Curved bacteria (<i>Mobiluncus</i>) per field
0	>30	0	0
1	5–30	<1	1–5
2	1–4	1–4	>5
3	<1	5–30	
4	0	>30	

Nugent scoring by one investigator in which excellent kappa values were achieved.⁵

A graph for the number of bacteria for different scores by field diameter, with the Zeiss FL30 set-up as the reference microscope, was constructed using the difference between the largest and smallest image areas in different microscopes and a table created by adding the same percentage increase in the number of bacteria used as the cut-off point for the Nugent scores (fig 1).

Slides from one ongoing and one recently completed study were analysed, comprising 913 slides from a treatment study for BV¹⁰ and 8985 slides from a trial for screening and treatment of women with BV and with intermediate flora during pregnancy (Fåhraeus, to be published 2004). All slides were Gram stained and scored according to Nugent by one investigator. In all vaginal samples at least four fields were evaluated, and the estimated number of *Lactobacillus* morphotypes and *Gardnerella* morphotypes together with the estimated number of curved rods (*Mobiluncus* morphotype) per high power field were noted as intervals on an ordinal scale (range 0–100 000 bacteria per field). Estimation of numbers of bacteria in intervals was done assuming that the number of bacteria (1–30) noted in a part of a representative microscopic field can be used to estimate the approximate number of bacteria in the whole field.¹¹ The numbers of different bacterial morphotypes were then transformed to intervals in accordance with Nugent. In addition, the presence of other bacterial morphotypes was also recorded. As we have the numbers of bacteria, these can be used to recalculate new scores, after compensation for the image area of the microscope, with new intervals.

In order to classify the slides according to Ison/Hay, the number and morphotypes of the bacteria were used, and most of the slides with a Nugent score above 3 were re-evaluated by one investigator (BC) alone or together with another (PGL) investigator using the Hay/Ison classification. The Hay/Ison classification grade 1 is considered as normal *Lactobacillus* morphotype only; grade 2 as intermediate, reduced *Lactobacillus* morphotype with mixed bacterial morphotypes; and grade 3 as BV, mixed bacterial morphotypes with few or absent *Lactobacillus* morphotypes. Ison and Hay have also observed that in a high power field, slides with very few bacteria will be classified as intermediate according to Nugent, but these slides do not represent a true

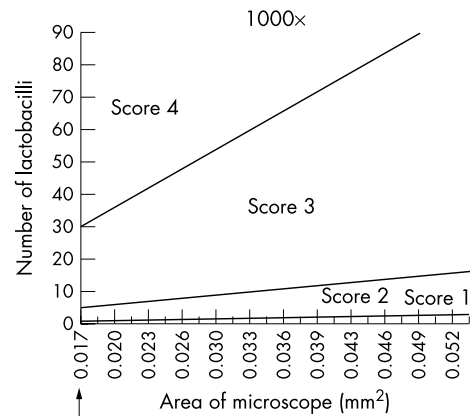


Figure 1 Graph of number of bacteria for different scores by field diameter, with the smallest microscope, with an area of 0.0165 mm², as reference (arrow).

intermediate group in between normal and BV. Therefore, Ison and Hay introduced a new group called grade 0. Furthermore, slides with no or few *Lactobacillus* morphotype or *Gardnerella* morphotype but with larger cocci morphotypes are graded as a new tentative grade 4. However, none of the slides graded 0 and 4 are, according to Amsels composite criteria, from women with a abnormal vaginal status and these two groups (group 0 and 4) are therefore regarded as normal by Ison and Hay.⁷

Statistics

A weighted kappa index was calculated to investigate the agreement between the different methods. A kappa above 0.70 is regarded as good agreement.^{12 13}

Ethics

Both clinical studies were approved by the regional ethics committee.

RESULTS

The diameter of the image area of the six microscope set-ups varied between 0.145 mm and 0.250 mm (table 2), and thus the area varied between 0.0165 and 0.049 mm². The scores

Table 2 The diameter and the area differ considerably for different microscope set-ups

Microscope	Ocular	Lens	Image diameter (mm)	Area (mm ²)	Image area difference with the smallest as reference
Zeiss FL30	10×/18	100×/1.25	0.145	0.0165	100%
Leitz Laborlux D (1)	10×18	100×/0.90	0.21	0.035	210%
Leitz Laborlux D (2)	10×/18	100×/1.32	0.18	0.025	154%
Leica DM LB (1)	10×/25	100×/1.30	0.25	0.049	297%
Leica DM LB (2)	10×/20	100×/1.30	0.19	0.028	172%

Table 3 Assignment of points for numbers of bacterial morphotypes in Nugent scoring using different microscope set-ups

Nugent score points	Zeiss FL30 100% relative image size	Leitz Laborlux (1) 210% relative image size	Leica DM LB (1) 297% relative image size
>30	>30	>50	>90
29-5	5-29	10-49	15-89
4-1	1-4	2-8	3-12
1-0	0-1	0-2	0-3
0	0	0	0

for the different microscope set-ups were adjusted using the Zeiss LF30 microscope set-up that was validated during the international workshop as reference (image area 100%) (table 3). By measuring the diameter of any individual microscope, it is possible to "calibrate" the Nugent scoring system.

The consequences of the calibrations were tested on the two patient materials.

Table 4 shows a comparison between the Nugent scores validated in the international workshop (Zeiss LF30 microscope set-up) and the scores recorded with the Leica DM LB microscope set up, with an area that is 297% larger than that in the Zeiss FL30 microscope set-up used in the treatment study of BV. Slides assigned to the normal group and the BV group did not change Nugent scores. Among the slides originally assigned to the intermediate group (25.4% of the slides) there was a 24% reduction of this group down to 19.5% of the slides after calibration. The weighted kappa value between the two methods was 0.94.

The comparison between validated Nugent scores (Zeiss LF30 microscope set-up) and scores obtained with the Leica DM LB microscope set-up for the material from pregnant women is shown in table 5. When the 1176 slides originally assigned to the intermediate group were recalculated, there was a 39% reduction of the intermediate group, 717 were still classified as intermediate, 49 women were reclassified as BV, and 426 as normal. Most of the slides classified as intermediate using the non-validated microscope set-up were thus reclassified as normal when recalculated to validated Nugent scores. No changes were seen in slides from women

diagnosed as normal or with BV. The weighted kappa value between the two methods was 0.88.

A comparison of slides with Ison/Hay scoring and validated Nugent scores for the two studies is shown in tables 6 and 7. The kappa index in the treatment study was 0.88 (table 6) and in the prevalence study it was 0.90 (table 7). In the treatment study the cure rate increased from 54.6% to 63% and the failure rate increased from 16.7% to 22.0% after calibration of the scores.

DISCUSSION

In all research one must be able to compare data from different investigations. The image area observed in microscopes is often an overlooked variable in various scoring procedures used in clinical research. However, standardisation has been achieved by pathologists in order to compare survival after treatment for breast cancer when the diagnosis is based on mitotic counts per image area seen in the microscope. This kind of standardisation indicates the direction similar efforts in clinical microbiology should take in counting or assessing, on an ordinal scale, numbers of bacteria seen in stained smears.

The differences between Nugent's original scoring validated in the international workshop⁵ and the scores obtained with compensating for image area are not large. The calculated weighted kappa index was 0.94 and 0.88, respectively, in the two clinical materials. Such high kappa indexes demonstrate very good agreement. We found good agreement in the international workshop, with kappa indexes of 0.74-0.87 for 12 of the collaborators investigating vaginal smears using

Table 4 The effect of using the graph in figure 1 to recalculate the cut-off point in the Nugent scores after compensation for the area of the microscope. Patients were subjects in a treatment study for BV

Original Nugent scores	Normal	Intermediate	BV	Total
Normal	422	0	0	422
Intermediate	36	178	18	232
BV	0	0	259	259
Total	458	178	277	913

Table 5 The effect of using the graph in figure 1 to recalculate the cut-off points in the Nugent scores after compensation for the area of the microscope. Slides were from a screening and treatment study of pregnant women with the aim of studying pregnancy outcome among women with BV

Original Nugent scores	Negative	Intermediate	BV	Total
Normal	7066	8	0	7074
Intermediate	426	701	49	1176
BV	2	8	725	735
Total	7494	717	774	8985

Table 6 Comparison between the Ison/Hay classification and the Nugent classification after compensation for the area of the microscope. Patients were subjects in a treatment study for BV

Nugent scores after compensation for area	Ison/Hay's classification			Total
	Normal	Intermediate	BV	
Normal	458	0	0	458
Intermediate	48	91	39	178
BV	0	16	261	277
Total	506	107	300	913

Weighted kappa index 0.88.

Nugent's original classification.⁵ However, complete agreement occurred in only 63% of the slides, and in some slides the score could vary from 0 to 10. Most divergences occurred in the slides classified as intermediate. A contributing factor could be the non-standardised image area, as there was no standardisation of the microscopes used in the study. The contributors came from all parts of the world.

We do not know the area of the microscope that was used in Nugent's original study which was published 20 years ago. Most new microscopes used today are equipped with wide angle oculars and improved lenses, which give a considerably larger microscopic image area. This is why we chose the ZeissLP30 microscope set-up from early 1990s as the reference set up, as the microscope set-up of the new Leica DM LB has a three times larger area than the ZeissLP30.

The Nugent scoring system is excellent in diagnosing either negative or BV positive smears but not intermediate smears. The reason for this could be that this system was based on only five different series of 50 pregnant women—too small a material to find the troublesome slides. The slides with only *Lactobacillus* morphotype or with only *Gardnerella* morphotype are likely to be judged in the same way by various investigators. The controversy occurs in the intermediate group. Three different kinds of slides can be identified as difficult to interpret. Firstly, slides from patients who have BV might be scored as intermediate because there is more than one *Lactobacillus* morphotype in the slide. This will give 2 points for *Lactobacillus* morphotype and the maximum of 4 points for more than 30 *Gardnerella* morphotypes even if the slides contain more than 10⁸ *Gardnerella* morphotypes per field and also contain clue cells. The slide will be assigned a score of 6—that is, intermediate. With an increase in image area in a calibrated microscope set up, more than three *Lactobacillus* morphotype should be the cut-off limit. Secondly, slides from patients with normal flora that contain 300–500 pleomorphic *Lactobacillus* morphotypes per field may be difficult to interpret. Some of these *Lactobacillus* morphotypes are so small that they might be counted as *Gardnerella* morphotypes. With more than 30 such misinterpreted morphotypes, the slide ends up with a total score of 6 and

is regarded as intermediate. With an increase in image area in a calibrated microscope set-up there could be as many as 90 *Gardnerella* morphotypes per field before the slide would be regarded as intermediate. Thirdly, slides with less than 30 *Lactobacillus* morphotypes (1 point) and with more than five *Gardnerella* morphotypes (3 points) will score 4 and be labelled intermediate. This situation could be very common if the woman has recently been treated with antibiotics, particularly clindamycin vaginal cream. Lactobacilli are very sensitive to clindamycin and are considerably decreased during a period after treatment.

The ordinal scale of Nugent scoring for *Lactobacillus* morphotypes and *Gardnerella* morphotypes has the same intervals, 0, <1, 1–4, 5–29, >30. Schmidt and Hansen¹⁴ suggested a different ordinal scale, 0–1, 1–4, 5–15, 16–29, >30, in order to increase the validity of Nugent scoring in primary healthcare populations. However, when Schmidt's intervals were applied to the dataset in the international workshop, the kappa index in the workshop was not increased (unpublished). Both Nugent's and Schmidt's ordinal scales have an interval of 0–30 bacteria per field. In our screening material of 8985 pregnant women, the estimated mean number of bacteria with *Lactobacillus* morphotype was 169 (range 1–1500), while the estimated mean number of bacteria with *Gardnerella* morphotype was 1664, with a range of 5–10 000. It might be that the interval for *Lactobacillus* morphotype should be in the range of 0–150, while the interval for *Gardnerella* morphotypes should be 10 times larger, or 0–1500, to adequately reflect the image seen by the microscopist.

As we found kappa indexes of 0.88 and 0.94 when comparing the original Nugent classification with the Nugent classification modified by area, we have demonstrated very good agreement. However, the consequences in our two different materials are considerable. In the screening material the intermediate group would decrease from 13.1% to 8.0%, as most of the intermediate group would be regarded as normal.

In the treatment study the intermediate group would decrease from 25.4% to 19.5%; two thirds of those reclassified

Table 7 Comparison between the Ison/Hay classification and the Nugent classification after compensation for the area of the microscope. Slides were from a screening and treatment study of pregnant women with the aim of studying pregnancy outcome among women with BV

Nugent scores after compensation for area	Ison/Hay classification			Total
	Normal	Intermediate	BV	
Normal	7494	0	0	7494
Intermediate	202	391	124	717
BV	10	21	743	774
Total	7706	412	867	8985

Weighted kappa index 0.90.

would then be regarded as normal and one third would be reclassified as BV. As this was a treatment study the consequence would be that the cure rate would increase from 54.6% to 62.0% and that the failure rate would increase from 16.7% to 18.5% (data not shown).

In the international workshop,⁵ the Ison/Hay^{6,7} classification which divided slides into three different groups dichotomously, and not on an ordinal scale, had the best kappa index. Interpretation of stained smears of bacteria cannot constitute only a simple counting of the number of morphotypes, as this is a time consuming and futile endeavour if large numbers of bacteria are to be counted or the definition of a specific morphotype is not given or is unclear. The dimensions included in the interpretation of the image seen by the microscopist should be analysed and validated so that classification of the images can be made along what is, in principle, an ordinal scale including intervals of numbers of morphotypes. We therefore compared the compensated Nugent score with the Ison/Hay classification and also found very high kappa indexes of 0.90 and 0.88, respectively, indicating that the two methods are very alike. Nevertheless, the consequences in the treatment study when using the Ison/Hay classification are that the cure rate will increase further to 63.0% and the failure rate will increase to 22.2%. In the screening material the intermediate group will be reduced further, from 717 to 412.

The intermediate flora could be regarded as a flora that is between normal and BV,¹⁵ but this requires that it is a well defined group. There have also been reports that women with intermediate flora have the same increased risk for adverse outcome as women with BV.^{6,16-19}

After standardisation of the microscopic area, it would be possible to have an international consensus regarding new cut offs. This might be the next step in an international agreement concerning standardisation of the diagnosis of BV and intermediate flora. The Ison/Hay criteria seem to constitute the best classification method, as this will allow

the microscopist to synthesise an impression from several microscopic fields rather than a specific number, so that the influence of surface area and bacterial density should be lessened, but Hay/Ison classification it is still rather difficult to teach because it is based on a clinical view rather than on counting bacterial morphotypes in the microscope. It could be that the Ison/Hay criteria expressed as bacterial morphotypes per high power field with a known microscopic area would further improve the scoring for bacterial vaginosis.

ACKNOWLEDGEMENTS

This study did not receive support from any company. The screening study for BV during pregnancy was supported by the Health Research Council in the Southeast of Sweden (FORSS).

CONTRIBUTORS

P-GL, LF, TJ, and UF contributed to the design of the study, whereas BC examined all slides; P-GL re-evaluated some of the slides and, together with LF and UF, wrote the report.

Authors' affiliations

P-G Larsson, Department of Obstetrics and Gynaecology, Kärnsjukhuset, Skövde, Sweden

P-G Larsson, B Carlsson, L Fåhraeus, T Jakobsson, U Forsum, Department of Molecular and Clinical Medicine, Linköpings Universitet, Sweden

No conflicts of interest are stated.

REFERENCES

- Morris M, Nicoll A, Simms I, et al. Bacterial vaginosis: a public health review. *Br J Obstet Gynaecol* 2001;**108**:439-50.
- Amsel R, Totten PA, Spiegel CA, et al. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med* 1983;**74**:14-22.
- Spiegel CA, Amsel R, Holmes KK. Diagnosis of bacterial vaginosis by direct Gram stain of vaginal fluid. *J Clin Microbiol* 1983;**18**:170-7.
- Nugent RP, Krohn MA, Hillier SH. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of Gram stain interpretation. *J Clin Microbiol* 1991;**29**:297-301.
- Forsum U, Jakobsson T, Larsson PG, et al. An international study of the interobserver variation between interpretations of vaginal smear criteria of bacterial vaginosis. *Apmis* 2002;**110**:811-18.
- Hay PE, Lamont RF, Taylor-Robinson D, et al. Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage. *BMJ* 1994;**308**:295-8.
- Ison CA, Hay PE. Validation of a simplified grading of Gram stained vaginal smears for use in genitourinary medicine clinics. *Sex Transm Infect* 2002;**78**:413-15.
- Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. 1. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991;**19**:402-10.
- National Coordinating Group For Breast Screening Pathology. *Pathology reporting in breast cancer screening*, 2nd ed. NHSBSP publications no 3 April, 1995.
- Eriksson K, Carlsson B, Forsum U, et al. A double blind treatment study of with normal vaginal lactobacilli after an open treatment with vaginal clindamycin ovules bacterial vaginosis. *Acta Dermato Vener* 2003;submitted.
- Koch AL. Growth measurement. In: Gerhardt P, Murray RGE, Wood WA, Krieg NR, eds. *Methods for general and molecular bacteriology*. Washington DC: American Society for Microbiology, 1994:248-77.
- Gjorup T, Jensen AM. [The kappa coefficient—a goal for evaluating the reproducibility of nominal and ordinary data]. *Nord Med* 1986;**101**:90-4.
- Rigby AS. Statistical methods in epidemiology. V. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 2000;**22**:339-44.
- Schmidt H, Hansen JG. Diagnosis of bacterial vaginosis by wet mount identification of bacterial morphotypes in vaginal fluid. *Int J STD AIDS* 2000;**11**:150-5.
- Rosenstein IJ, Morgan DJ, Sheehan M, et al. Bacterial vaginosis in pregnancy: distribution of bacterial species in different gram-stain categories of the vaginal flora. *J Med Microbiol* 1996;**45**:120-6.
- Lahi-Camp JM, Rai R, Ison C, et al. Association of bacterial vaginosis with a history of second trimester miscarriage. *Hum Reprod* 1996;**11**:1575-8.
- Larsson PG, Platz-Christensen JJ, Dalaker K, et al. Treatment with 2% clindamycin vaginal cream prior to first trimester surgical abortion to reduce signs of postoperative infection: a prospective, double-blinded, placebo-controlled, multicenter study. *Acta Obstet Gynecol Scand* 2000;**79**:390-6.
- Larsson PG, Carlsson B. Does pre- and postoperative metronidazole treatment lower vaginal cuff infection rate after abdominal hysterectomy among women with bacterial vaginosis? *Infect Dis Obstet Gynecol* 2002;**10**:133-40.
- Oakeshott P, Hay P, Hay S, et al. Association between bacterial vaginosis or chlamydial infection and miscarriage before 16 weeks gestation: prospective community based cohort study. *BMJ* 2002;**325**:1334.

Key messages

- There are problems in diagnosing BV
- There are considerable differences among image areas covered by different microscope set-ups used in Nugent and Ison/Hay scoring
- Using the Nugent scoring the intermediate group is the most difficult
- Differences in the intermediate group will have a considerable effect on the results from both treatment and prevalence studies, even though the kappa indexes indicate very good agreement between the methods used
- The Ison/Hay criteria seem to constitute the best classification method, as this will allow the microscopist to synthesise an impression from several microscopic fields rather than a specific number, so that the influence of surface area and bacterial density should be lessened
- The Hay/Ison classification is still difficult to teach because it is based on a clinical view rather than on counting bacterial morphotypes in the microscope
- It could be that the Ison/Hay criteria expressed as bacterial morphotypes per high power field with a known microscopic area would further improve the scoring for bacterial vaginosis.