Analysis of the Disagreement Between Automated Bioluminescence-Based and Culture Methods for Detecting Significant Bacteriuria, with Proposals for Standardizing Evaluations of Bacteriuria Detection Methods

WRIGHT W. NICHOLS,* GORDON D. W. CURTIS, AND H. H. JOHNSTON

Regional Public Health Laboratory and Department of Microbiology, John Radcliffe Hospital, Oxford, OX3 9DU, United Kingdom

Received 13 October 1981/Accepted 24 December 1981

A fully automated method for detecting significant bacteriuria is described which uses firefly luciferin and luciferase to detect bacterial ATP in urine. The automated method was calibrated and evaluated, using 308 urine specimens, against two reference culture methods. We obtained a specificity of 0.79 and sensitivity of 0.75 using a quantitative pour plate reference test and a specificity of 0.79 and a sensitivity of 0.90 using a semiquantitative standard loop reference test. The majority of specimens negative by the automated test but positive by the pour plate reference test were specimens which grew several bacterial species. We suggest that such disagreement was most likely for urine containing around 10^5 colony-forming units per ml (the culture threshold of positivity) and that these specimens were ones contaminated by urethral or vaginal flora. We propose standard procedures for calibrating and evaluating rapid or automated methods for the detection of significant bacteriuria and have analyzed our results using these procedures. We recommend that identical analyses should be reported for other evaluations of bacteriuria detection methods.

In the absence of clear symptoms, the most effective way of distinguishing between significant bacteriuria, implying urinary tract infection, and urethral or vaginal contamination of urine is by quantitative bacteriology of carefully collected urine specimens (14). In attempts to speed up this process, the firefly bioluminescent reaction, suggested as a means of counting bacteria by Chappelle and Levin (6), has been examined by several workers (1, 7, 13, 23, 27).

Conventional quantitative culture results do not agree fully either with manual bioluminescence-based methods (1, 3, 7, 23, 27) or with automated bioluminescence-based methods (8, 13, 20). This paper analyzes reasons for the discrepancy between the automated methods and the quantitative culture method of detecting significant bacteriuria.

As yet there is no agreement on the operation parameters necessary for a rapid or automated method to be suitable for routine use in a diagnostic bacteriology laboratory. This has led to the use of a variety of different reference methods for detecting bacteriuria, for example, pour plate quantitative colony counts (27), "Droplette" quantitative colony counts (4a, 8), surface quantitative colony counts (11), semiquantitative standard loop colony counts (1, 2, 4a, 7, 8, 10, 12, 17, 20), filter paper strip methods (4a, 25), and a spiral plating method (24a). Moreover, these quantitative or semiquantitative procedures have been defined as positive on the basis of different colony count criteria: 1 × 10^5 colony-forming units (CFU)/ml (1, 7, 8, 10, 12, 20, 23, 27), 7 × 10^6 CFU/ml (11), or different counts depending on qualitative factors such as the identity of the organism(s) grown or the presence or absence of leukocytes (2, 24–25). It is thus extremely difficult to compare the results of one rapid or automated method with another. Such comparisons are essential either for improving existing automated methods or for judging the applicability of current rapid or automated methods to the screening of urine specimens in a particular laboratory. We therefore propose that the four basic analyses used in this paper should be applied to the calibration and evaluation of other rapid or automated methods of detecting significant bacteriuria.

MATERIALS AND METHODS

Reagents. All solutions were in sterile, pyrogen-free distilled water (water for injections was from British Pharmacopoeia).

ATP solutions were prepared and stored frozen as described previously (22). Firefly luciferin-luciferase reagent, nucleotide releasant (somatic), and apyrase were used as previously described (8). Apyrase, nucleotide releasant (somatic), and luciferin-luciferase solu-
tions were filtered (membrane filters; pore size, 0.8 μm) before use to prevent blockage of reagent lines. Tris base (Trizma) was from Sigma Chemical Co., St. Louis, Mo., and Tween 20 was from Koch-Light Laboratories.

**Specimens.** Urine specimens were taken randomly from those submitted to this laboratory for routine bacteriological examination, except that those preserved with boric acid were excluded as previously described (8).

**Organisms.** *Escherichia coli* (NCTC 10418), *Staphylococcus aureus* (NCTC 6571), *Pseudomonas aeruginosa* (NCTC 10662), and a laboratory strain of *Proteus mirabilis* were grown at 37°C for 16 h in 5 ml of brain heart infusion (Oxoid Ltd., Basingstoke, England). For determinations of the ATP contents of these organisms, each was diluted with a further 5 ml of brain heart infusion, incubated at 37°C for 1.7 h, and then cooled. The resultant dilutions for continuous-flow assay were in ice-cold water. Viable counts of each dilution were made using the pour plate method described below.

**Reference methods of detecting significant bacteriuria.** Two culture methods were used for detecting significant bacteriuria. A standard bacteriological loop was used to streak 0.002 ml of urine on blood agar and MacConkey agar plates (18) which were incubated, in air, overnight at 37°C. The specimen was considered positive if 200 or more colonies were observed on either plate. This was carried out by the technical staff for every urine specimen sent to the laboratory. The second culture method was a pour plate method for which 10⁻², 10⁻³, 10⁻⁴, and 10⁻⁶ dilutions in sterile water were made. A 1-ml volume of each dilution was mixed with 15 ml of Columbia agar (Oxoid) held at 40°C. The mixture was poured immediately into a petri dish, allowed to cool, and incubated, in air, at 37°C overnight. The colonies in the appropriate plate were counted and related to the original urine specimen. The pour plate method was considered positive if we observed the equivalent of ≥10⁶ CFU/ml of the original urine specimen.

**Equipment.** Continuous flow equipment (Fig. 1) was as described previously (8), with alterations to the mixing of sample and luciferin-luciferase as described in the next section.

**Automated method of detecting significant bacteriuria.** Figure 1 shows the apparatus used. The firefly luciferin-luciferase reagent was mixed with the sample outside the flow cell in a small, four-turn mixing coil (internal volume, 85 μl; diameter of each turn, 9.7 mm) (Technicon Instruments Corp. Ltd., Basingstoke, England). The resulting reaction mixture was fed to a helical flow coil (Technicon Instruments) held with the axis vertical, in front of the photomultiplier tube of the Chem Glow photometer (American Instrument Co. [Travenol Laboratories, Inc.], Silver Spring, Md.) used previously (8, 22). Acetate was used as the anion in the Tris buffer since it inhibits luciferase less than chloride, nitrate, or phosphate (21). Tris-acetate (pH 8.08) was used to counter the acidity of urine specimens and bring them within the range pH 7.5 to pH 8.0 while minimizing the concentration of added acetate (12a). However, for the measurements of ATP contents of bacteria, the buffer was 444 mM Tris-acetate (pH 7.75).

The peak light response of the luciferin-luciferase and photometer detection system was recalibrated each day, or within the day for each fresh solution of luciferin-luciferase reagent, by replacing the apyrase reagent with water and sampling ATP standards (0, 1.0, 2.5, 5.0, and 10.0 nM) in duplicate. The peak output (in millivolts) of the photometer (which resulted from a combination of mixing, luciferase reaction kinetics, and arrival of sample at the flow cell) was analyzed in terms of chemiluminescence units as described previously (22). Calibration curves were drawn of chemiluminescence units against ATP concentration and used to convert peak photometer readings obtained with urine specimens to apparent ATP concentrations. For example, one such calibration curve followed the least-squares regression line: \( y = 0.39x - 0.23 \), where \( y \) was the apparent nanomolar ATP concentration and \( x \) was the peak photometer reading (in chemiluminescence units).

**Calibration of the automated method of detecting significant bacteriuria.** Results from the automated method consisted of an apparent bacterial ATP concentration for each specimen. To define a threshold of positivity for the results of the automated method, it was calibrated empirically against each reference method (23, 27). We rejected the alternative of using an estimated mean bacterial content of ATP to decide theoretically a threshold of positivity (1, 7) since this would not allow for uncontrolled factors such as luciferase inhibition by urine (22) and acetate (21). ATP derived from anaerobic or carboxyphilic bacteria, and ATP from somatic cells (H. H. Johnston and G. D. W. Curtis, Int. Symp. Appl. Biololuminescence Chemiluminescence, 1979, p. 446-447).

The results of the empirical calibration procedure are shown in Fig. 2. We have arbitrarily taken 21% false-positives (expressed as the proportion of reference test negatives) as acceptable for the automated method to be used routinely, and this determined the automated method’s threshold of positivity (20). If the pour plate method was used as the reference, the positivity threshold was 2.0 nM ATP; if the standard loop method was used as the reference, the positivity threshold was 2.4 nM ATP. The method could then be compared with other methods on the basis of the resulting false-negative rate (expressed as a percentage of the reference test positives) as was carried out previously (20).

**Cell counts.** Leukocytes, erythrocytes, and squamous epithelial cells in urine were counted microscopically using a Fuchs-Rosenthal counting chamber.

**Judgement of contamination of a specimen with urethral or vaginal flora.** A urine specimen was deemed to be contaminated (i) if it contained squamous epithelial cells at a concentration greater than 10⁶/ml, (ii) if any diptheroid organisms were cultured from the specimen, or (iii) if two or more organisms were grown from the urine.

**RESULTS**

As in a previous study (8), of those specimens showing greater than 10⁵ CFU/ml, pure growths of *E. coli* (25 specimens) and mixed cultures (27 specimens) predominated. Table 1 shows the agreement of the automated method with culture methods compared to four other bioluminescence-based automated methods. All automated methods agreed more closely with the standard
loop method than with the pour plate method. When results were normalized so that a false-positive rate of 21% was obtained, false-negative rates varied between 6 and 40%, with a value of 10 or 25% for the study described here depending on which reference method was used. The samples positive by the pour plate method and negative by the automated test were, in large part, also negative by the standard loop method. This is also demonstrated in Table 1; when those urine samples for which standard loop and pour plate results differed were excluded from the analysis, then the false-negative rate of the automated test decreased substantially, from 25 to 9%.

The culture results for the specimens falsely negative by the automated method were as follows. Against the pour plate reference, of 18 specimens, 15 showed more than one species (containing a total of $1.0 \times 10^2$ through $6.0 \times 10^6$ CFU/ml) and 3 showed pure growths of *Proteus* sp., *Streptococcus* sp., and *Staphylococcus* sp. ($3.1 \times 10^5$, $6.6 \times 10^5$, and $7.9 \times 10^5$ CFU/ml, respectively). The above three specimens showing pure growths were also falsely negative by the automated method with respect to the standard loop reference method. A further specimen containing one species (*E. coli* at $6.0 \times 10^6$ CFU/ml) was falsely negative with respect only to the standard loop reference method. However, the 15 specimens which showed mixed growths were negative by the standard loop method as well as by the automated method. This accounted for the closer agreement (Table 1) between the automated method and the standard loop reference method than between the automated method and the pour plate reference method.

Growth of several organisms from a voided
FIG. 2. Variation in false-positive and false-negative rates with the automated test's threshold of positivity. Symbols: calculated using the pour plate reference test: false-positive rate (•) and false-negative rate (○). Symbols: calculated using the standard loop reference test: false-positive rate (○) and false-negative rate (•).

Table 2. Results of the automated bioluminescence-based methods of detecting significant bacteriuria compared to culture.

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of specimens</th>
<th>No. of false-positive</th>
<th>% of false-positive</th>
<th>No. of false-negative</th>
<th>% of false-negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>Unpublished</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>24</td>
<td>15</td>
</tr>
</tbody>
</table>

Values in parentheses are those after exclusion of specimens for which the standard loop reference and pour plate reference results disagreed.

The false-positive and false-negative rates were calculated as described in the text.

VOL. 15, 1982

Downloaded from http://jcm.asm.org on August 14, 2017 by guest
The results of lower sensitivity demonstrates that coefficients of variation for measurements of apparent bacterial ATP concentrations in urine ranged from 9 through 33%. The results of examining bacterial suspensions by the automated method are shown in Fig. 5. The lower sensitivity limit of the method varied between $10^5$ and $10^6$ CFU/ml. Thus, in attempting to match an instrument positivity threshold with $10^5$ CFU/ml in urine, the instrument was being used at its limit of sensitivity. We suggest that this contributed to the high coefficients of variation reported in Table 2.

Table 3 shows that bacterial ATP was not significantly destroyed by the treatments used in the automated method. The ATP contents of selected bacteria, measured using the automated method, agreed well with those obtained by Chappelle and Levin (6).

DISCUSSION

Disagreement between the bioluminescence-based method and the quantitative culture method of detecting significant bacteriuria. The specimens positive by the quantitative culture (pour plate) method of detecting significant bacteriuria but falsely negative by the automated method were principally those contaminated by urethral or vaginal flora (Fig. 3). Such contaminated specimens were also the major source of falsely negative results with other automated or rapid methods of detecting significant bacteriuria (2, 4a, 12, 24a, 25).

The automated and reference methods of detecting significant bacteriuria were based on different parameters for estimating microbial biomass (bacterial ATP content and ability to grow, respectively). Furthermore, point estimates of apparent bacterial ATP concentrations in urine (Table 2) and of bacterial colony-forming units per unit volume of suspension (19) are subject to significant variability. Therefore we propose that some disagreement between the automated and culture methods was inevitable. In addition, we suggest that when any method is compared to culture using results in the conventional dichotomous form (+/-), then the disagreement will be observed principally with those specimens whose culture result is at, or near, the chosen reference threshold of positivity. We believe that this is the reason that grossly contaminated urine specimens constituted the majority of falsely negative results, as it is generally these specimens which have bacterial
counts in the region of $10^5$ CFU/ml (5, 26). This reasoning is supported by the observations that the exclusion from the study of specimens with urethral or vaginal contamination resulted both in lower numbers of specimens with culture results near $10^5$ CFU/ml (Fig. 3) and in a reduction of the proportion of false-negatives (legend to Fig. 3). On the other hand, the disagreement between the quantitative (pour plate) and semi-quantitative (standard loop) culture results (Table 1) was, we suggest, due to systematic underestimation of the colony count on those standard loop plates showing several bacterial species but not on those showing pure growth.

Although we believe that some disagreement is inevitable between two methods of detecting significant bacteriuria based on different microbial properties, nevertheless specific reasons can be given for the large disagreements reported here (Table 1). First, the automated instrument is currently working at its limit of sensitivity (Fig. 5). Second, variable concentrations of luciferase inhibitors remain even though their effects have been reduced by dilution (20, 22). Third, different species of bacteria contain differing quantities of ATP per colony-forming unit (Table 3) (6), whereas the reference test was based solely on the number of colony-forming units per unit volume of specimen. From these points we suggest that the agreement between the automated method and culture will increase if the sensitivity of bacterial detection by the automated method is increased. This is feasible since a 6- to 10-fold decrease in the sensitivity of measurements of ATP concentrations is observed on changing from discrete analysis to continuous flow analysis (unpublished observations). Furthermore, the fourfold dilution of specimen during sampling would be unnecessary if luciferase inhibitors could be effectively removed within the automated method (for example, by continuous dialysis or by automated centrifugation and resuspension). An overall sensitivity increase of 25- to 40-fold is thus theoretically achievable.

Despite the problems discussed above, we believe that the automated method may be used to screen midstream urine specimens collected from carefully prepared, asymptomatic patients, i.e., those specimens which are positive by the automated test being cultured. On the other hand, specimens from symptomatic patients still require full cultural examination.

**Calibration and evaluation of rapid and automated methods of detecting significant bacteriuria.** As stated above, the analyses of the automated bacteriuria detection method described in this paper were designed to make the results fully comparable both with those from other laboratories and with the absolute requirements of any diagnostic test. The procedures which we recommend for further similar studies are described below.

Full calibration curves should be reported of false-positive and false-negative rates as functions of the rapid or automated method’s threshold of positivity (4a, 23; Fig. 2). False-positive and false-negative values must be calculated according to the accepted convention (4, 9), taking the rapid or automated method to yield
the test result and the reference method to give the "true" result. This recommendation is the most important, and it applies regardless of the chosen reference test. In general, when the results of a new diagnostic test which gives results as values of a continuous variable are compared to the results of a dichotomous (±) reference test, false-positive and false-negative levels are inversely related (4). Thus, in new rapid or automated tests for detecting significant bacteriuria, false-positive and false-negative levels vary inversely as the positivity threshold of the new test is varied (4a, 23; Fig. 2). Such calibration curves allow one to examine false-negative levels at a particular false-negative level, or vice versa, regardless of the automated or rapid method. We have arbitrarily taken 21% false-positives to be acceptable (20; Table 1); others have used 10% false-positives (24a). On the other hand, it is unclear how, for example, 34% false-positives with 7% false-negatives (10) may be compared to 7% false-positives with 27% false-negatives (1) without the above information.

For the reference method, aerobic pour plate colony counts using nutrient agar should be made, and 10^5 CFU/ml in the original urine specimen should be considered the threshold of positivity (14, 15, 26). For results from different laboratories, the above calibration curves may only be compared if the reference methods are identical because the reference method may affect the results of calibration. For example, for several automated methods, different levels of agreement were found with a quantitative pour plate reference method and with a semiquantitative standard loop reference method (Table 1).

The frequency distribution of bacterial counts should be plotted to allow monitoring of the extent to which the 10^-1 CFU/ml criterion is valid (14, 15, 26; Fig. 3). The 10^-1 CFU/ml reference threshold of positivity was proposed as a discriminator, for asymptomatic patients, between significant bacteriuria, implying an underlying urinary tract infection, and bacteriuria due to urethral or vaginal contamination (14, 15). This purely quantitative criterion is of high predictive value (of the order of 95%) only for catheter or midstream specimens obtained from carefully prepared patients (16, 26). Under these conditions the 10^-1 CFU/ml value falls in the trough of a bimodal frequency distribution of specimen colony counts (5, 14). This recommendation is important because when our automated method was compared with the pour plate reference method, different results were obtained depending on the proportion of urethrally or vaginally contaminated specimens (which accounted for the bulk of the cultural results in the region of 10^3 CFU/ml [Fig. 3]).

Where possible the plot should be reported of the frequency distribution of the variable obtained from the rapid or automated method. If the rapid or automated method yields results in terms of a continuous variable, then for an ideal diagnostic test, the frequency distribution of this variable in the population should also be bimodal (4), where the two modes represent infection and noninfection. This recommendation allows monitoring of the extent to which a new rapid or automated test fits this criterion of an ideal diagnostic test, independently of a comparison.
with bacterial colony counts (Fig. 4).

The advantages of adopting these recommendations in the design of studies to calibrate and evaluate new rapid or automated methods of detecting significant bacteriuria are as follows: results would be directly comparable between different laboratories; results for different methods or instruments, or both, would be directly comparable; and the procedures may be used as a basis for the definition of required instrument performance criteria (although these may vary between individual laboratories).

ACKNOWLEDGMENTS

This work was supported by an Equipment Research and Development grant, reference 55/79 R/E1048/240, from the British Department of Health and Social Security. We thank R. Jay (Technicon Instruments Corp. Ltd.) for glassware design, E. Anderson and S. Musgrave for technical assistance, and D. F. J. Brown and J. Scholefield for copies of manuscripts in press. R. T. Mayon-White, R. G. Mitchell, and M. P. E. Slack are thanked for critical discussions.

LITERATURE CITED


