

## Evaluation of Nucleic Acid Amplification Tests as Reference Tests for *Chlamydia trachomatis* Infections in Asymptomatic Men

ROBERT E. JOHNSON,<sup>1\*</sup> TIMOTHY A. GREEN,<sup>2</sup> JULIUS SCHACHTER,<sup>3</sup> ROBERT B. JONES,<sup>4</sup>  
EDWARD W. HOOK III,<sup>5</sup> CAROLYN M. BLACK,<sup>2</sup> DAVID H. MARTIN,<sup>6</sup>  
MICHAEL E. ST. LOUIS,<sup>1</sup> AND WALTER E. STAMM<sup>7</sup>

National Center for HIV, STD, and TB Prevention<sup>1</sup> and National Center for Infectious Diseases,<sup>2</sup> Centers for Disease Control and Prevention, Atlanta, Georgia; Department of Laboratory Medicine, University of California, San Francisco, California<sup>3</sup>; School of Medicine, Indiana University, Indianapolis, Indiana<sup>4</sup>; Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama<sup>5</sup>; School of Medicine, Louisiana State University, New Orleans, Louisiana<sup>6</sup>; and Department of Medicine, University of Washington, Seattle, Washington<sup>7</sup>

Received 9 March 2000/Returned for modification 28 June 2000/Accepted 21 September 2000

**Urine ligase chain reaction (LCR) and PCR tests and urethral swab culture were compared for their abilities to detect *Chlamydia trachomatis* infection in 3,639 asymptomatic men by using one-, two-, and three-test reference standards. Frozen urine at four of five participating centers was also tested by a transcription-mediated amplification assay which was used as a reference test. LCR increased the yield of positive results by 27% and PCR increased the yield of positive results by 26% over the yield of positive results by culture ( $n = 295$ ). LCR and PCR sensitivities were similar, ranging from 80.4 to 93.5%, depending on the reference standard. Culture sensitivity was substantially less. A multiple-test standard yielded LCR, PCR, and culture specificities of 99.6%, with or without discrepant analysis. Test performance varied among centers partly due to different interpretations of the testing protocols. The study confirms that urine LCR and PCR for the detection of *C. trachomatis* have substantially improved sensitivities over that of urethral swab culture for testing of asymptomatic men, enabling screening of this important target group. These tests, perhaps in combination, are also candidate reference tests for the conduct of test evaluation studies.**

*Chlamydia trachomatis* is the most prevalent bacterial sexually transmitted organism in the United States. The marketing of nucleic acid amplification tests (NAATs) that use urine rather than swab specimens dramatically expands options for screening, including the targeting of asymptotically infected men (2, 17, 21). However, NAATs for the detection of *C. trachomatis* infection in asymptomatic men have been evaluated primarily in small studies (4, 11, 20, 22, 23) and head to head only in a study that used a urine enzyme immunoassay as the reference test (20).

Tissue culture, the historical reference standard by which other tests for the detection of *C. trachomatis* infection have been evaluated, has a high specificity but is poorly standardized (2, 15, 18), and its sensitivity, based on experience with cervical swab specimens, is 80 to 85% at best and is probably substantially lower in many laboratories (2, 18, 19). New tests, especially if they are more sensitive than culture, will detect *C. trachomatis* in subjects who have false-negative cultures, leading to serious underestimation of the specificity of the new test if culture is the sole reference standard. To correct for false-negative culture results, most evaluators have used a discrepant analysis technique in which specimens that are culture negative but positive by the new test are further examined by at least one additional test (2). Subjects are classified as infected if culture or any two nonculture tests are positive. This approach may overestimate sensitivity and specificity because use of the additional reference tests depends on the result of the

test under evaluation (7, 8, 10, 12, 13, 16; W. C. Miller, Editorial, Clin. Infect. Dis. 27:1186–1193, 1998; J. Schachter, Editorial, J. Infect. Dis. 27:1181–1185, 1998). An alternative approach proposed recently (1; Miller, Editorial) is to combine culture with a second reference test that is a NAAT and that is applied whenever culture is negative. The resulting reference standard is more sensitive than culture alone and is not susceptible to the bias introduced by discrepant analysis. An additional concern has been that the tests that have been most often used for discrepant analysis have used the same amplification and detection formats as the NAAT evaluated and have not been approved by the U. S. Food and Drug Administration (9, 13; Miller, Editorial).

In this paper we report the results of a five-center evaluation of commercial ligase chain reaction (LCR) and PCR tests performed with urine and of culture performed with urethral swab specimens collected from asymptomatic male subjects for the detection of *C. trachomatis*. Four of the five centers also performed a transcription-mediated amplification (TMA) test with urine.

### MATERIALS AND METHODS

**Study population.** Male patients who presented to a sexually transmitted disease clinic in each of five participating centers, who did not have symptoms of urethritis, and who had not taken antibiotics in the previous 30 days were eligible for enrollment in the study. Study clinicians informed eligible patients about the study and enrolled those who consented to participate. The study was approved by the institutional review boards at each research center and the Centers for Disease Control and Prevention. Enrollment began in October 1995 and ended in August 1997, when enrollment had reached the target sample size of 270 culture-positive patients. If patients were enrolled more than once, only the results from the initial enrollment are included in this report.

**Collection of specimens.** Following the history and physical examination, the clinician collected a urethral swab specimen for a Gram-stained smear and, in

\* Corresponding author. Mailing address: Division of STD Prevention, M.S. E02, Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta, GA 30333. Phone: (404) 639-1894. Fax: (404) 639-8610. E-mail: rej1@cdc.gov.

some clinics, culture for *Neisseria gonorrhoeae*. Clinicians then obtained a second urethral swab for culture for *C. trachomatis* by inserting the swab 2 to 3 cm into the urethral meatus with a rotary motion. Patients were then instructed to void and to save the first 30 ml of urine in marked urine collection cups. The urine was held and transported at 4°C. LCR and PCR testing occurred within 4 days of specimen collection, or for LCR, the specimen was frozen at -20°C or less and was tested within 60 days.

**Laboratory methods.** Research laboratories at each center tested urine specimens by the AMPLICOR microwell plate PCR test (Roche Diagnostic Systems, Branchburg, N.J.) and the LCx LCR test (Abbott Laboratories, Abbott Park, Ill.). The urine was processed and the tests were performed according to the manufacturers' protocols, except that repeated LCR tests for subjects who had equivocal results on initial LCR testing were performed with specimens that had been frozen at -70°C for up to several weeks. An internal control for inhibition has been added to the COBAS version of the AMPLICOR PCR test, but the control was not available for this study. Frozen urine from sub samples of subjects enrolled at four of the five clinics was also tested by a commercial TMA test (Amplified CT; Gen-Probe, Inc., San Diego, Calif.) by following the manufacturer's protocol as part of a separate study to evaluate the TMA test, which was not available at the beginning of this study. Subjects for the subsamples were selected in the order in which they had been selected for this study. Since the delayed testing of frozen urine by TMA could affect the performance of TMA relative to those of LCR and PCR, TMA was used in the study only as an additional reference test.

Research centers used their own research protocols for tissue culture isolation of *C. trachomatis*. Swabs were transported to the laboratory in sucrose phosphate chlamydia transport medium or Eagle's minimal essential medium in Earle salts, each containing fetal calf serum and antibiotics. The specimens were held at 4°C for a maximum of 18 to 72 h, depending on the center, or were frozen at -70°C. Culture medium was inoculated onto cycloheximide- and/or DEAE-dextran-treated McCoy cells in 96-well microtiter plates or 1-dram vials. The cultures were incubated at 35 or 37°C for 48 to 72 h. After incubation, the cells were fixed with methanol or ethanol and were stained with locally prepared or commercial major outer membrane protein-specific fluorescein-conjugated antibody reagents. One center used a lipopolysaccharide-specific antibody stain as well. Three of the centers performed a blind passage if no chlamydia inclusions were noted at 48 to 72 h.

**Reference standards.** We compared estimates of sensitivity and specificity by using the standards described below to classify patients as infected (Table 1).

(i) **Single-test reference standards.** To compare the performances of LCR and PCR head to head, we classified subjects as infected if the culture result was positive. To assess NAATs as single-test reference standards, we also estimated the performances of the remaining tests using LCR and PCR, in turn, as single-test reference standards.

(ii) **Multiple-test (composite) reference standards without discrepant analysis.** By considering each test in turn as the evaluated test, subjects were classified as infected if (i) the results of both of the remaining tests were positive (Both<sup>+</sup> standard) or (ii) the result of either of the remaining tests was positive (Either<sup>+</sup> standard). The Both<sup>+</sup> standard increases the specificity of the reference standard, while it reduces its sensitivity. The Either<sup>+</sup> standard, recently advocated as an alternative to discrepant analysis (1; Miller, Editorial), maximizes the sensitivity of the reference standard, while it reduces its specificity. For subjects for whom TMA test results were available, we extended the Either<sup>+</sup> standard further by classifying subjects as infected if the result of either nonevaluated test was positive or the result of the TMA test was positive (Either<sup>+</sup>TMA<sup>+</sup> standard), thereby minimizing misclassification of truly infected subjects and thus making our estimate of specificity as accurate as possible with three reference tests. The trade-off is that more truly uninfected subjects may be misclassified, resulting in underestimation of the sensitivity of the evaluated test.

(iii) **Multiple-test (composite) reference standards with discrepant analysis.** Subjects were classified as infected if, in turn (i) the culture result was positive or the results of LCR and PCR were positive (DA-Culture standard) or (ii) the results of any two of culture, LCR, and PCR were positive (DA-Any-2<sup>+</sup> standard). These discrepant analysis approaches have been used for nearly all evaluations of NAATs for the detection of *C. trachomatis*. For subjects for whom TMA test results were available, we also classified subjects as infected if the results of either of the tests not under evaluation at that time were positive or the results of the evaluated test and the TMA test were positive (DA-TMA standard).

**Statistical methods.** For each of the reference standards used, sensitivity and specificity were calculated by standard formulae (5). The statistical significance of center-to-center variations in sensitivity and specificity estimates was assessed by the Pearson chi-square test of homogeneity (14). To evaluate center-to-center variations in the difference between sensitivity estimates for pairs of tests, we assigned each specimen with concordant results by both tests a score of zero and each specimen with discordant results by the two tests a score of ±1, depending on which of the two tests had positive results. For each center, the mean of these scores is equal to the difference in sensitivity estimates. The statistical significance of the variation in the mean score among centers was assessed by the analysis of variance F test (14).

TABLE 1. Reference standards

Reference test that must be positive to classify subject as infected (otherwise subject is classified as uninfected)	Test evaluated
<b>Standards not using discrepant analysis</b>	
<b>Single-test standards</b>	
Culture <sup>+</sup> .....	LCR and PCR
LCR <sup>+</sup> .....	Culture and PCR
PCR <sup>+</sup> .....	Culture and LCR
<b>Both<sup>+</sup> standards</b>	
Culture <sup>+</sup> and LCR <sup>+</sup> .....	PCR
Culture <sup>+</sup> and PCR <sup>+</sup> .....	LCR
LCR <sup>+</sup> and PCR <sup>+</sup> .....	Culture
<b>Either<sup>+</sup> standards</b>	
Culture <sup>+</sup> or LCR <sup>+</sup> .....	PCR
Culture <sup>+</sup> or PCR <sup>+</sup> .....	LCR
LCR <sup>+</sup> or PCR <sup>+</sup> .....	Culture
<b>Either<sup>+</sup>/TMA<sup>+</sup> standards</b>	
Culture <sup>+</sup> , LCR <sup>+</sup> , or TMA <sup>+</sup> .....	PCR
Culture <sup>+</sup> , PCR <sup>+</sup> , or TMA <sup>+</sup> .....	LCR
LCR <sup>+</sup> , PCR <sup>+</sup> , or TMA <sup>+</sup> .....	Culture
<b>Standards using discrepant analysis</b>	
DA-Culture standard: culture <sup>+</sup> or LCR <sup>+</sup> and PCR <sup>+</sup> .....	Culture, LCR, and PCR
DA-Any 2 <sup>+</sup> standard: Culture <sup>+</sup> and LCR <sup>+</sup> , culture <sup>+</sup> and PCR <sup>+</sup> , or LCR <sup>+</sup> and PCR <sup>+</sup> .....	Culture, LCR, and PCR
DA-TMA standards	
Culture <sup>+</sup> or LCR <sup>+</sup> , or PCR <sup>+</sup> and TMA <sup>+</sup> .....	PCR
Culture <sup>+</sup> or PCR <sup>+</sup> , or LCR <sup>+</sup> and TMA <sup>+</sup> .....	LCR
LCR <sup>+</sup> or PCR <sup>+</sup> , or culture <sup>+</sup> and TMA <sup>+</sup> .....	Culture

RESULTS

**Enrollment and overall test results.** The number of men enrolled in the study was 3,663, and data for 3,639 (99.3%) men with satisfactory results by culture, LCR, and PCR were included in this analysis (Table 2). The number of subjects who tested culture positive by center ranged from 26 to 97, with a total of 295. The proportion of subjects who were culture positive ranged from 3.2 to 16.2%. Overall, 8.1% of the subjects were culture positive, whereas 10.3% were LCR positive and 10.2% were PCR positive, increases in rates of positivity of 27 and 26%, respectively.

**Head-to-head comparison of sensitivities and specificities of LCR, PCR, and culture using single-test reference standards.** The sensitivities of LCR and PCR, estimated by using culture as an independent reference standard, were similar (Table 3). For LCR, center-specific sensitivity estimates (data not shown) were distributed between 75.8 and 96.2% (P = 0.026), and for PCR, they were distributed between 78.9 and 100.0% (P = 0.069). However, the center-specific estimates of LCR and PCR sensitivities varied together, so that differences between their sensitivities were never greater than 4.2 percentage points (P = 0.868).

Overall, the specificities of LCR and PCR, estimated by using culture as the reference standard, were also similar (Table 3). Center-specific specificity estimates for LCR were dis-

TABLE 2. Number of subjects by result profile for *C. trachomatis* detection tests and research center

Test result profile <sup>a</sup>			No. of subjects at the following research center:					
Culture	LCR	PCR	A	B	C	D	E	All
P	P	P	22	83	57	38	25	225
P	P	N	1	3	15	5	0	24
P	N	P	2	3	18	3	1	27
P	N	N	4	8	5	2	0	19
N	P	P	10	48	11	9	15	93
N	P	N	5	10	1	13	5	34
N	N	P	3	9	10	3	2	27
N	N	N	638	930	470	375	777	3,190
Total			685	1,094	587	448	825	3,639

<sup>a</sup>P, positive; N, negative

tributed between 94.2 and 97.7% ( $P < 0.001$ ), and for PCR, they were distributed between 94.3 and 98.0% ( $P < 0.001$ ). Center-specific differences between LCR and PCR specificities varied significantly ( $P < 0.001$ ), with a maximum difference of 2.5 percentage points.

The sensitivity of LCR for the combined centers exceeded that of culture by 17.8 percentage points when we used PCR as the reference standard (Table 3). Similarly, the sensitivity of PCR exceeded that of culture sensitivity by 18.4 percentage points when we used LCR as the reference standard. However, the center-specific culture sensitivity estimates (data not shown) were distributed between 60.1 and 78.1% when PCR was used as the standard ( $P = 0.017$ ) and between 55.6 and 85.7% when LCR was used as the standard ( $P < 0.001$ ). Moreover, the center-specific estimates of the sensitivities of LCR and culture and the sensitivities of PCR and culture differed by between -7.3 and 32.6 percentage points ( $P < 0.001$ ) and between -4.8 and 33.3 percentage points ( $P < 0.001$ ), respectively.

Overall, the specificities of culture, the specificity of which is usually assumed to be 100%, were only 98.6 and 98.7% when LCR and PCR were used as single-test reference standards, respectively (Table 3). Each estimate varied significantly among centers ( $P < 0.001$ ), primarily because a single center had values less than 96% (data not shown). The specificity of

culture exceeded that of PCR by only 0.3 percentage points overall, and center-specific differences did not vary significantly ( $P = 0.580$ ). The differences in the specificities of culture and LCR varied significantly across centers ( $P = 0.001$ ) but were less than 0.8 percentage points except for a single center, for which the difference was 2.8 percentage points in favor of culture.

**Estimates of sensitivity and specificity obtained using multiple-test reference standards without discrepant analysis.** The addition of a second reference test and use of the requirement that both reference tests must be positive to classify subjects as truly infected (Both<sup>+</sup> standard) increased the sensitivity estimates and decreased the specificity estimates for each test (Table 3). Classification of subjects as truly infected if the result of either reference test was positive (Either<sup>+</sup> standard) had the reverse effect. However, the sensitivities of PCR and LCR were always similar to each other and substantially higher than the sensitivity of culture, regardless of the reference standard used for the comparison, whereas the specificities of PCR and LCR were similar to or somewhat lower than that of culture across the different reference standards (Table 3).

For the four centers that conducted TMA testing, the addition of TMA as a third reference test and the classification of subjects as truly infected if the results of any of the three tests were positive (Either<sup>+</sup>/TMA<sup>+</sup> standard) further decreased the

TABLE 3. Estimates of culture, LCR, and PCR sensitivities and specificities for detection of *C. trachomatis* infection by reference standard

Standard	No. of subjects infected	Sensitivity (%)			No. of subjects not infected	Specificity (%)		
		Culture	LCR	PCR		Culture	LCR	PCR
<b>Single-test standard</b>								
Culture <sup>+</sup>	295		84.4	85.4	3,344		96.2	96.4
LCR <sup>+</sup>	376	66.2		84.6	3,263	98.6		98.3
PCR <sup>+</sup>	372	67.7	85.5		3,267	98.7	98.2	
<b>Multiple-test standard without discrepant analysis<sup>a</sup></b>								
Both <sup>+</sup>	252	70.8	89.3	90.4	3,387	97.9	95.5	95.7
Either <sup>+</sup>	422	64.2	82.4	81.8	3,217	99.4	98.9	99.2
Either <sup>+</sup> / <sup>b</sup> (tested by TMA)	314	63.4	81.2	84.8	2,524	99.4	99.3	99.1
Either <sup>+</sup> /TMA <sup>+</sup> ( <sup>b</sup> tested by TMA)	326	62.9	80.4	84.0	2,512	99.6	99.6	99.6
<b>Multiple-test standard with discrepant analysis</b>								
DA-Culture	388	76.0	88.1	88.9	3,251		99.0	99.2
DA-Any-2 <sup>+</sup>	369	74.8	92.7	93.5	3,270	99.4	99.0	99.2
DA-TMA	321	63.9	81.7	85.4	2,517	99.6	99.6	99.6

<sup>a</sup> For this reference standard, which specimens are classified as infected or uninfected depends on which candidate test is being evaluated. The reported number is the median across the three candidate tests.

<sup>b</sup> For purposes of comparison, Either<sup>+</sup> results were recalculated by using the results only from centers which performed the TMA test.

sensitivity estimates (Table 3), but by only 0.5 percentage points for culture and 0.8 percentage points for LCR and PCR. Substitution of the Either<sup>+</sup>/TMA<sup>+</sup> standard for the Either<sup>+</sup> standard further increased the specificity estimates by 0.2 percentage points for culture, 0.3 percentage points for LCR, and 0.5 percentage points for PCR. The resulting Either<sup>+</sup>/TMA<sup>+</sup> specificity estimates were 99.6% for all three tests (Table 3). Center-specific specificity estimates based on the Either<sup>+</sup>/TMA<sup>+</sup> standard varied significantly across centers (range, 99.1 to 100.0% [ $P = 0.027$ ]) for LCR but not for PCR or culture (data not shown).

**Estimates of sensitivity and specificity obtained using discrepant analysis.** The estimate of culture sensitivity obtained with the DA-Culture standard exceeded the estimates obtained with any other reference standard (Table 3). The estimates of the sensitivities of LCR and PCR obtained with the DA-Culture standard exceeded the estimates obtained with the single-test and Either<sup>+</sup> reference standards (2.6 to 7.1 percentage points), but by less than was the case for estimates of culture sensitivity (8.3 to 11.8 percentage points). The estimates of the sensitivities of LCR and PCR obtained with the DA-Culture standard were less than the estimates obtained with the remaining non-DA standard, the Both<sup>+</sup> standard. Thus, substitution of the DA-Culture standard for non-DA standards reduced the magnitude by which LCR and PCR sensitivity estimates exceeded the culture sensitivity estimate.

The estimates of the sensitivities of culture, LCR, and PCR obtained with the DA-Any-2<sup>+</sup> standard exceeded the estimates obtained with any of the reference standards that did not use discrepant analysis (Table 3). These increases in estimates of the sensitivity of culture (4.0 to 10.6 percentage points) were similar in magnitude to the increases in estimates of the sensitivities of LCR and PCR (3.1 to 11.7 percentage points). Consequently, substitution of the DA-Any-2<sup>+</sup> standard for standards that did not use discrepant analysis had a smaller effect on the differences between the estimated sensitivities of LCR and culture or of the sensitivities of PCR and culture than substitution of the DA-Culture standard.

The estimates of the specificities of LCR and PCR obtained with the DA-Culture and DA-Any-2<sup>+</sup> standards only slightly (<0.02 percentage points) exceeded the estimates obtained with the Either<sup>+</sup> standard (Table 3). The estimates of the specificities of LCR and PCR obtained with the DA-TMA standard exceeded the estimates obtained with the Either<sup>+</sup>/TMA<sup>+</sup> standard by an even smaller amount.

**Differences in number of infections detected.** We did not collect additional urethral swab specimens for an NAAT because our subjects had no symptoms of urethritis. Comparison of the sensitivities of a urine NAAT and urethral swab culture by using the alternate urine NAAT as the reference standard, as we have done in this analysis, may favor the NAAT since the NAATs are performed with the same specimen. Nevertheless, 342 subjects were LCR positive and also culture and/or PCR positive, exceeding the 295 subjects who were culture positive by 16%. Similarly, 345 subjects were PCR positive and also culture and/or LCR positive, exceeding the 295 subjects who were culture positive by 17%. We assume that subjects with positive results by both LCR and PCR are likely to be truly positive.

## DISCUSSION

Other studies have found LCR and PCR to be more sensitive than culture for the detection of *C. trachomatis* infection in women and men undifferentiated with regard to symptoms of urethritis (2, 17, 21). In this multicenter study of asymptomatic

men, the sensitivities and specificities of LCR and PCR were similar to each other. The NAATs detected between 16 and 27% more *C. trachomatis* infections than culture, depending on the proportion of positive tests that are considered truly positive. The specificities of all three tests were 99.6% when we used the alternative tests as independent reference standards. Although the specificity estimates are not directly comparable, the similarity of the results indicates that all three tests have high degrees of specificity.

The sensitivities and specificities of LCR and PCR varied substantially across centers. Culture performance also varied across centers, but independently of NAAT performance. Variations in culture performance were expected because of the difficulty in standardizing culture methods. However, given the greater degree of standardization of LCR and PCR, we expected little variation by center for these tests. During data analysis, center C, which had the lowest LCR sensitivity (76%), became aware that it differed from the other centers in aspirating centrifuged urine from the bottom up rather than the top down due to ambiguity in the testing protocol. A change of the aspiration method to conform to that used by the other centers was associated with an increase in the sensitivity of LCR. (D. H. Martin and C. Cammarata, Thirteenth Meet. Int. Soc. Sexually Transmitted Diseases Research, Abstr. 385, 1999). When the data for center C are excluded from the analysis, the overall sensitivities of LCR and PCR were both 88.5% when culture was used as the reference standard, and the yields of positive tests were 46 and 38% greater for LCR and PCR, respectively, than for culture. Multicenter studies may be especially useful for the identification of intercenter variations in test performance and the factors responsible for the variation.

The LCR and PCR specificity estimates of approximately 96%, based on a culture reference standard, are underestimates due to falsely negative cultures. We considered the use of NAATs in single-, two-, and three-test reference standards that were independent of the evaluated test primarily to avoid underestimation of test specificity without introducing the bias associated with discrepant analysis. The availability of TMA test results allowed us to consider a three-test standard, even though three reference tests might impose substantial burden and the TMA test has received limited evaluation with samples from asymptomatic men. The estimates of culture specificity, which is usually considered to be 100% because inclusions stained with specific antibody have a unique appearance, increased from less than 99.0 to 99.4% and 99.6% when single-test LCR or PCR was replaced with the two- and three-test NAATs with the Either<sup>+</sup> and Either<sup>+</sup>/TMA<sup>+</sup> reference standards, respectively. Even if these standards erroneously classify some subjects as infected, estimates of specificity will not be overestimated as long as the reference tests are independent of the evaluated test for truly uninfected subjects (12, 13; Miller, Editorial). By contrast, such errors contribute a positive bias when discrepant analysis is used, because the results for only the evaluated test-positive subjects can be misclassified by the additional (resolver) reference test (9).

The increases in the estimated specificities of culture, LCR, and PCR from 96 to 99.6% that we observed when we substituted the Either<sup>+</sup> standard and then the Either<sup>+</sup>/TMA<sup>+</sup> standard for single-test standards are important. A prevalence of 2% may represent the lower end of the range in which screening of populations of men for *C. trachomatis* infection would be recommended (3). At a prevalence of 2% and a sensitivity of 85%, the positive predictive value would be 33% if the specificity was 96.4%, as we estimated for PCR using culture alone as the standard. The positive predictive value would be 66% if

the specificity of PCR was 99.1%, as we estimated using the Either<sup>+</sup> standard with culture and LCR as the reference tests. When TMA is added as a third reference test, the estimated specificity of PCR increased to 99.6%; the positive predictive value would be 81%, perhaps acceptable for screening. Thus, culture and an NAAT might not be sufficiently sensitive when an Either<sup>+</sup> reference standard is used. However, two NAATs might be adequate, since our estimate of culture specificity was 99.4% when we used the Either<sup>+</sup> standard with LCR and PCR as composite reference tests, and LCR sensitivity was exceptionally low at one center.

With regard to test sensitivity, the lower estimates that we obtained with the Either<sup>+</sup> or Either<sup>+</sup>/TMA<sup>+</sup> standards compared to those that we obtained with the single-test or Both<sup>+</sup> standards may have occurred because the additional subjects classified as infected by use of the former standards had infections that were more difficult to detect. However, the Either<sup>+</sup> and Either<sup>+</sup>/TMA<sup>+</sup> estimates may also have been lower because those standards misclassified more uninfected subjects. Even limited misclassification of uninfected subjects can cause a substantial underestimation of test sensitivity (6). Due to the resulting ambiguity, test evaluators might compare sensitivity estimates obtained with single- and multiple-test standards to determine the magnitude of any difference.

The DA-Culture standard yielded the highest estimate of culture sensitivity. The DA-Culture standard sensitivity estimates may be positively biased due to discrepant analysis, but the bias may be greatest for culture, because all culture-positive results are accepted as truly positive. Consequently, the DA-Culture standard may underestimate the difference between the sensitivities of nonculture tests and culture. The DA-Any-2<sup>+</sup> standard yielded the highest estimates of LCR and PCR sensitivities. Any bias in LCR and PCR performance estimates contributed by discrepant analysis is greater with the DA-Any-2<sup>+</sup> standard than with the DA-Culture or DA-TMA standard because with the last two standards only the results for patients with discrepant positive results by the evaluated test are subjected to discrepant analysis, whereas with the DA-Any-2<sup>+</sup> standard, the results for patients with discrepant positive results by the reference test are also subjected to discrepant analysis (13).

In contrast to the foregoing case for sensitivity, our estimates of specificity obtained by discrepant analysis were only slightly greater than the estimates that we obtained with the Either<sup>+</sup> and Either<sup>+</sup>/TMA<sup>+</sup> standards. This observation is consistent with those from other studies reporting only a small bias with discrepant analysis estimates of specificity when prevalence is low and when the reference standard has a high degree of specificity (1, 6).

The LCR and PCR tests of urine for detection of *C. trachomatis* offer great potential for the screening of men, an important reservoir of infection for women. The NAATs offer ease of collection and transport, patient acceptability, and high degrees of sensitivity and specificity. However, the multicenter design of our study also permitted us to demonstrate that although the tests are highly standardized, these tests do vary in performance across populations and laboratories. Strict attention to training and quality control and an ongoing program to monitor test performance under real-world conditions will be essential for the more widespread use of NAATs, including

their use as composite reference standards in place of culture and discrepant analysis.

#### ACKNOWLEDGMENT

This work was funded by the Centers for Disease Control and Prevention, Atlanta, Ga. (cooperative agreement RPA 455).

#### REFERENCES

- Alonzo, T. A. and M. S. Pepe. 1999. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Stat. Med.* **18**:2987–3003.
- Black, C. M. 1997. Current methods of laboratory diagnosis of *Chlamydia trachomatis* infections. *Clin. Microbiol. Rev.* **10**:160–184.
- Centers for Disease Control and Prevention. 1998. HIV prevention through early detection and treatment of other sexually transmitted diseases—United States. Recommendations of the Advisory Committee for HIV and STD Prevention. *Morb. Mortal. Wkly. Rep.* **47**(No. RR-12):1–24.
- Chernesky, M. A., H. Lee, J. Schachter, J. D. Burczak, W. E. Stamm, W. M. McCormack, and T. C. Quinn. 1994. Diagnosis of *Chlamydia trachomatis* urethral infection in symptomatic and asymptomatic men by testing first-void urine in a ligase chain reaction assay. *J. Infect. Dis.* **170**:1308–1311.
- Fleiss, J. L. 1981. Statistical methods for rates and proportions, 2nd ed. John Wiley & Sons, Inc., New York, N.Y.
- Green, T. A., C. M. Black, and R. E. Johnson. 1998. Evaluation of bias in diagnostic-test sensitivity and specificity estimates computed by discrepant analysis. *J. Clin. Microbiol.* **36**:375–381.
- Hadgu, A. 1996. The discrepancy in discrepant analysis. *Lancet.* **348**:592–593.
- Hadgu, A. 1997. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. *Stat. Med.* **16**:1391–1399.
- Hadgu, A. 1999. Discrepant analysis: a biased and an unscientific method for estimating test sensitivity and specificity. *J. Clin. Epidemiol.* **52**:1231–1237.
- Irwig, L., A. N. A. Tosteson, C. Gatsonis, J. Lau, G. Colditz, T. C. Chalmers, and F. Mosteller. 1994. Guidelines for meta-analyses evaluating diagnostic tests. *Ann. Intern. Med.* **120**:667–676.
- Jaschek, G., C. A. Gaydos, L. E. Welsh, and T. C. Quinn. 1993. Direct detection of *Chlamydia trachomatis* in urine specimens from symptomatic and asymptomatic men by using a rapid polymerase chain reaction assay. *J. Clin. Microbiol.* **31**:1209–1212.
- McAdam, A. J. 2000. Discrepant analysis: how can we test a test? *J. Clin. Microbiol.* **38**:2027–2029.
- Miller, W. C. 1998. Bias in discrepant analysis: when two wrongs don't make a right. *J. Clin. Epidemiol.* **51**:219–231.
- Moore, D. S., and G. P. McCabe. 1989. Introduction to the practice of statistics. W. H. Freeman & Co., New York, N.Y.
- Pate, M. S., and E. W. Hook. 1995. Laboratory to laboratory variation in *Chlamydia trachomatis* culture practices. *Sex. Transm. Dis.* **22**:322–326.
- Ransohoff, D. F. and A. R. Feinstein. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N. Engl. J. Med.* **299**:926–930.
- Schachter, J. 1997. DFA, EIA, PCR, LCR and other technologies: what tests should be used for diagnosis of chlamydia infections?. *Immunol. Investig.* **26**:157–161.
- Schachter, J., W. E. Stamm, M. A. Chernesky, E. W. Hook, R. B. Jones, F. N. Judson, J. A. Kellogg, B. LeBar, P. A. Mardh, and W. M. McCormack. 1992. Nonculture tests for genital tract chlamydial infection. What does the package insert mean, and will it mean the same thing tomorrow? *Sex. Transm. Dis.* **19**:243–244.
- Stary, A. 1997. Chlamydia screening: which sample for which technique? *Genitourin. Med.* **73**:99–102.
- Stary, A., S. Tomazic-Alen, B. Choueiri, J. Burczak, K. Steyrer, and H. Lee. 1996. Comparison of DNA amplification methods for the detection of *Chlamydia trachomatis* in first-void urine from asymptomatic military recruits. *Sex. Transm. Dis.* **23**:97–102.
- Taylor-Robinson, D. 1997. Evaluation and comparison of tests to diagnose *Chlamydia trachomatis* genital infections. *Hum. Reprod.* **12**:113–120.
- Toye, B., R. W. Peeling, P. Jessamine, P. Claman, and I. Gemmill. 1996. Diagnosis of *Chlamydia trachomatis* infections in asymptomatic men and women by PCR assay. *J. Clin. Microbiol.* **34**:1396–1400.
- Van der Pol, B., T. C. Quinn, C. A. Gaydos, K. Crotchfelt, J. Schachter, J. Moncada, D. Jungkind, D. H. Martin, B. Turner, C. Peyton, and R. B. Jones. 2000. Multicenter evaluation of the AMPLICOR and automated COBAS AMPLICOR CT/NG tests for detection of *Chlamydia trachomatis*. *J. Clin. Microbiol.* **38**:1105–1112.