

Relative Accuracy of Nucleic Acid Amplification Tests and Culture in Detecting *Chlamydia* in Asymptomatic Men

HONG CHENG,^{1*} MAURIZIO MACALUSO,¹ STEN H. VERMUND,¹ AND EDWARD W. HOOK III²

*Department of Epidemiology and International Health¹ and Department of Medicine,²
University of Alabama at Birmingham, Birmingham, Alabama 35294*

Received 5 February 2001/Returned for modification 7 July 2001/Accepted 19 August 2001

Published estimates of the sensitivity and specificity of PCR and ligase chain reaction (LCR) for detecting *Chlamydia trachomatis* are potentially biased because of study design limitations (confirmation of test results was limited to subjects who were PCR or LCR positive but culture negative). Relative measures of test accuracy are less prone to bias in incomplete study designs. We estimated the relative sensitivity (RSN) and relative false-positive rate (RFP) for PCR and LCR versus cell culture among 1,138 asymptomatic men and evaluated the potential bias of RSN and RFP estimates. PCR and LCR testing in urine were compared to culture of urethral specimens. Discordant results (PCR or LCR positive, but culture negative) were confirmed by using a sequence including the other DNA amplification test, direct fluorescent antibody testing, and a DNA amplification test to detect chlamydial major outer membrane protein. The RSN estimates for PCR and LCR were 1.45 (95% confidence interval [CI] = 1.3 to 1.7) and 1.49 (95% CI = 1.3 to 1.7), respectively, indicating that both methods are more sensitive than culture. Very few false-positive results were found, indicating that the specificity levels of PCR, LCR, and culture are high. The potential bias in RSN and RFP estimates were <5 and <20%, respectively. The estimation of bias is based on the most likely and probably conservative parameter settings. If the sensitivity of culture is between 60 and 65%, then the true sensitivity of PCR and LCR is between 90 and 97%. Our findings indicate that PCR and LCR are significantly more sensitive than culture, while the three tests have similar specificities.

Chlamydia trachomatis infection is the most common bacterial sexually transmitted disease (STD) worldwide, with more than 4 million new cases estimated to occur annually in the United States alone (2, 21). Although *C. trachomatis* infection can be treated with antibiotics, control of the disease has been impeded, in part because symptoms of the infection are often absent or insufficient to lead to treatment among many infected women and men (2, 5, 6, 10, 11, 15, 17, 18). The large group of asymptotically infected persons is not only at risk of serious long-term sequelae but also sustains transmission within communities.

Screening for chlamydia is a critical component of the control strategy recommended by the Centers for Disease Control and Prevention (2). Establishment and maintenance of successful large-scale chlamydia control and screening programs could be facilitated by the availability of highly sensitive and specific tests, particularly if specimens for testing could be collected without invasive procedures.

Application of two recently developed tests based on amplification of organism-specific DNA sequences—PCR and ligase chain reaction (LCR)—to first-void urine samples, rather than urethral or cervical swab specimens, is becoming a preferred method for detecting infection among asymptomatic patients (5, 6, 10, 11, 15, 17). Previous data suggest that PCR and LCR tests may be more sensitive than the other currently available chlamydia tests and probably have very high specificity (5, 6,

10, 11, 15, 17). The methods used to evaluate the accuracy of PCR and LCR, however, have been the subject of intense controversy (1, 10, 11, 13, 17).

Much of this controversy is related to the fact that the studies that have estimated the sensitivity and specificity of PCR and LCR have discriminated between true-positive results and false-positive results by using cell culture (often with additional tests) to confirm the positive results of the new tests, without, however, confirming the negative results (and thus potentially missing false-negative results). Discrepant analysis is a procedure in which the confirmation of test results is further restricted to specimens that yield a positive result by the new test, while they are negative according to another, presumably less-sensitive method such as cell culture. Discrepant analysis has been used in some of the largest studies that have evaluated PCR and LCR tests for chlamydia (1, 5–11, 13–17, 19). Studies with incomplete confirmation procedures, however, are prone to missing false-negative results when the results of all the tests performed are negative. Some false-positive results may also be missed in discrepant analysis because not all positive results are subject to the confirmation procedure. As a consequence, the sensitivity of the new tests tends to be overestimated because the denominator (the true positive) tends to be underestimated. Also, the false-positive rates of the new tests tend to be underestimated since the denominator (the true negative) is inflated by an unknown number of false-negative results (1). The problem is further complicated by the potential for interdependence among test results. The degree to which measures of accuracy are biased is not only a function in the true accuracy of all the tests involved but also depends on the degree of interdependence among their results (7).

* Corresponding author. Mailing address: Department of Epidemiology and International Health, School of Public Health, University of Alabama at Birmingham, 1530 3rd Ave., S, Birmingham, AL 35294. Phone: (205) 975-8679. Fax: (205) 975-7058. E-mail: hcheng@ms.soph.uab.edu.

While absolute measures of accuracy are likely to be biased in the contexts described above, estimates of the relative sensitivity (*RSN*, the ratio of two sensitivities) and the relative false-positive rate (*RFP*, the ratio of two false-positive rates) do not require denominator information and thus are less likely to be influenced by the interdependence of test results, considerably alleviating the concern about bias (3, 4).

In this study, we estimated the relative accuracy (*RSN* and *RFP*) of PCR and LCR on urine specimens compared with the accuracy of cell culture on urethral swab specimens obtained from asymptomatic men attending a metropolitan STD clinic. The potential bias in the *RSN* and *RFP* estimates was evaluated by examining the influence of alternative assumptions about the accuracy and interdependence of the tests involved and about the true disease prevalence.

MATERIALS AND METHODS

Study group. Men attending the Jefferson County Department of Health STD clinic in Birmingham, Ala., who did not have symptoms of urethritis and had not taken antibiotics in the previous 30 days were eligible for enrollment.

Collection of specimens. After informed consent was obtained, a nurse clinician collected each subject's history and carried out a standardized, limited physical examination, collecting a urethral swab specimen for a Gram stain and culture for *Neisseria gonorrhoeae*, followed by a second swab specimen for *C. trachomatis* culture. The specimens were collected by using Dacron-tipped, steel shaft swabs by inserting the swab into the urethral meatus with a rotary motion. To enhance the likelihood of collecting adequate material, the first swab was inserted about 1 cm and the second swab was inserted 2 to 3 cm into the urethra. Swabs used for specimen collection were pretested and bulk purchased to ensure that there were no inhibitory substances that might interfere with culture performance. Immediately after specimen collection, the swabs were placed in vials containing 1.5 ml of 0.2 M sucrose-phosphate buffer transport medium containing 2% fetal bovine serum and antibiotics. The samples were maintained at 4°C at the collection site and transported within 18 h to the laboratory, where they could be stored for as long as 72 h in a -70°C freezer prior to inoculation of the cultures.

After collection of swab specimens, participants were then instructed to void, saving the first 30 ml of urine in a marked urine collection cup. Urine was not frozen before testing but maintained at 4°C and transported to the laboratory on ice.

Laboratory methods. Urine specimens were tested by the Amplicor Microwell Plate PCR test (Roche Diagnostic Systems, Branchburg, N.J.) and the LCx LCR test (Abbott Laboratories, Abbott Park, Ill.) according to the manufacturers' protocols.

Urethral swab specimens were cultured, after thawing, on DEAE-dextran-treated McCoy cells in 96-well microtiter plates. Three 100- μ l aliquots of transport medium were inoculated into three microtiter wells (300 μ l [20%] of transport medium, total volume): two wells on a "primary" plate and one on a "pass" plate for subsequent passage if primary inoculations were negative. All specimens were centrifuged at 800 \times g at 37°C for 1 h after the inoculation and then incubated in 5% carbon dioxide (CO₂) for 30 min. After this, the medium was aspirated from each well, and 200 μ l of cycloheximide medium was added. The microtiter plates were then incubated at 37°C in 5% CO₂ for 48 to 72 h. After this incubation, medium was aspirated from the primary plate, and the cells were fixed to the plate by using methanol. After 3 min of methanol fixation, the methanol was aspirated, and each well was stained by using commercially available immunofluorescence stains. One duplicate well on this plate was stained by using monoclonal antibody detection reagents directed at major outer membrane protein (MOMP; Syva Microtrak, Palo Alto, Calif.), and one was stained by using commercially available anti-lipopolysaccharide (LPS) reagents (Kallstad, Chaska, Minn.). Microtiter plates were then read for the presence of chlamydial inclusions with a Zeiss inverted fluorescent microscope, and inclusions were graded on a continuous scale of 1 to 4. If no inclusions were detected on the primary plate, cells from the pass plate were transferred to a secondary culture plate treated in the same manner as the primary plate and incubated for an additional 48 to 72 h. After the second incubation, the pass plate was stained with the LPS reagent. The transport medium of selected samples was also used for the direct fluorescent antibody assay (DFA) when the results of PCR, LCR, and culture were discordant (see below).

Resolution of discordant results. The same procedure was employed to evaluate either PCR or LCR (henceforth called the "new test"). If the new test was positive on a urine specimen and culture was positive on a urethral specimen from the same subject, no further resolution was pursued. Otherwise, discordant results between the new test and negative cultures were resolved by using the other DNA amplification test (i.e., LCR was used to confirm PCR and vice versa), DFA, and the MOMP test in sequence, stopping at the first positive result or after all tests were negative (Fig. 1). Discordant results were classified as positive if any of the sequentially performed reference tests was positive. The discordant results were classified as false positive if all reference tests were negative. The protocol of study was reviewed and approved by the Institutional Review Board for Human Use of the University of Alabama at Birmingham.

Statistical methods. (i) **Estimation of RSN and RFP.** In discrepant analysis, equation 1 (Eq 1) and Eq 2 are used to estimate the sensitivity (*SN*) and specificity (*SP*) of the new test, respectively (see data layout in Fig. 2). The estimation

$$\hat{SN} = \frac{a + b'}{a + b' + c} \quad (1)$$

$$\hat{SP} = \frac{d}{b'' + d} \quad (2)$$

process usually implies the assumptions that the specificity of culture is 100% and that both the sensitivity and the specificity of the confirmation procedure are 100%.

Although the two assumptions are necessary conditions for the validity of the estimates, they are not sufficient. In fact, Eq 1 and Eq 2 yield biased estimates unless the sensitivity of both the new test and culture is 100% (i.e., false-negative results are impossible). The direction and size of the bias depend on the number of diagnostic errors in each cell.

Because the specificity of PCR, LCR, and culture is likely to be very high, the bias caused by unconfirmed false-positive results when both the new test and culture are positive (cell *a*) may be worth ignoring. In discrepant analyses, results that are positive by the new test and negative by culture are specifically targeted by the confirmation procedure. Thus, the residual error in this cell (cell *b*) is probably negligible.

Results that are positive by culture only (cell *c*) include an unknown number of false-positive culture results, which may not be negligible, depending on the specificity of culture. False-positive culture results in cell *c* would lead to underestimating the sensitivity of the new test (because they would be erroneously interpreted as false-negative results of the new test). Also, false-positive culture results in cell *c* would lead to overestimating the specificity of the new test (because they would be erroneously removed from the number of true-negative results). Confirmation of this subset of results is desirable but has been omitted in most studies of PCR and LCR for chlamydia detection (7-9, 14, 16). In the present investigation, we applied the confirmation procedure to all discordant results, including those classified in cell *c*.

Results that are negative for both culture and the new test (cell *d*) include an unknown number of false-negative results. These errors cannot be detected by the confirmation procedure and are not counted in the denominator of Eq 1, leading to overestimation of the sensitivity of the new test. Specificity can also be overestimated because the false-negative results included in cell *d* are erroneously counted as true negatives in Eq 2.

Using similar considerations, Green et al. found that the validity of sensitivity estimates of LCR depends critically on whether culture specificity is equal to 100%, while specificity estimates are less prone to bias (7). When the specificity of culture is reduced slightly from 100 to 99.6%, the bias in estimates of LCR or PCR sensitivity can be larger than five percentage points (7). We show here that relative accuracy estimates are considerably more robust.

The estimation of *RSN* and *RFP* uses numerator information only and is less prone to the bias resulting from the exclusion of concordant negative results from the confirmation procedure (3, 4). The *RSN*, *RFP*, and 95% confidence intervals (95% CI) of $\ln RSN$ and $\ln RFP$ are calculated by using the following equations (see Table 1 for data layout):

$$\hat{RSN} = \frac{SN_1}{SN_2} = \frac{a' + b'}{a'' + c'} \quad (3)$$

$$\hat{RFP} = \frac{1 - SP_1}{1 - SP_2} = \frac{a'' + b''}{a'' + c''} \quad (4)$$

$$95\% \text{ CI}(\ln \hat{R}) = \ln \hat{R} \pm 1.96 \sqrt{V\hat{ar}(\ln \hat{R})} \quad (5)$$

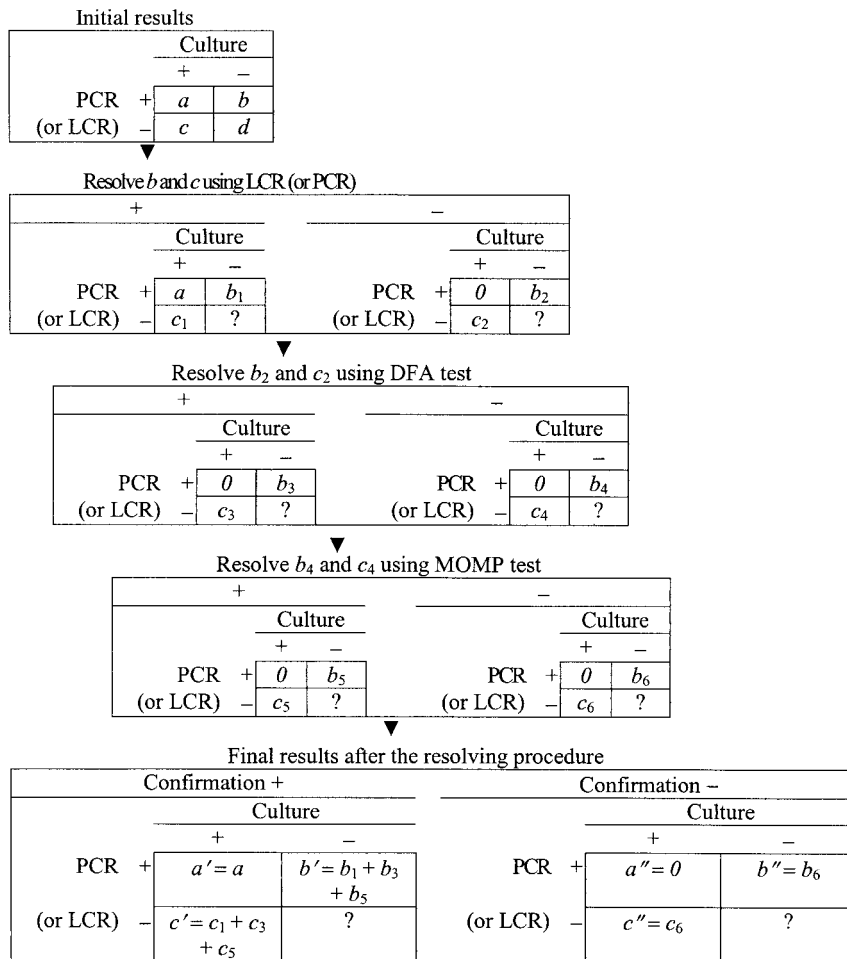


FIG. 1. Confirmation procedure in discrepant analysis.

In the equations above and in the following text, the subscripts for *SN* and *SP* indicate the type of tests (1 for PCR or LCR, 2 for culture). The \hat{R} in Eq 5 can be replaced by either \hat{R}_{SN} or \hat{R}_{FP} , while $\hat{Var}(\ln \hat{R})$ is

$$\frac{b' + c'}{(a' + b')(a' + c')}$$

for $\ln \hat{R}_{SN}$ and

$$\frac{b'' + c''}{(a'' + b'')(a'' + c'')}$$

for $\ln \hat{R}_{FP}$.

If the confirmation procedure is 100% accurate, estimates from Eq 3 and Eq 4 are unbiased. For data in this study, the *RSN* and *RFP* of the new test versus culture were estimated based on the confirmed discordant results and the unconfirmed concordant positive results. That is, in Eq 3 and Eq 4, *a'* was replaced with *a* and *a''* was assumed to be zero (Fig. 1). *RSN* and *RFP* estimates were still potentially biased because none of the confirmation tests for the resolution of discordant results is perfectly accurate and, probably to a lower degree, potential for bias could arise from excluding results in cell *a* from the confirmation procedure.

(ii) **Evaluation of bias in the *RSN* and *RFP* estimates.** Bias in the estimates of *RSN*, *RFP*, and *SN* of PCR and LCR was evaluated when only discordant results are resolved by using an imperfect confirmation procedure using a set of mathematical expressions including all parameters that influence accuracy (see Appendix). To assess the potential for bias, the *RSN*, *RFP*, and *SN* (of PCR and LCR) estimates (Eq A1 to Eq A12) for a given set of parameters were compared

with their corresponding theoretical values. The percent bias of each relative accuracy estimate was computed as $100 \times (\hat{R} - R)/R$, where \hat{R} is the *RSN*, *RFP*, *SN*, or *SP* estimate and *R* is the theoretical value of *RSN*, *RFP*, *SN*, or *SP*. The range of the potential bias was obtained by calculating the bias in *RSN*, *RFP*, *SN*, and *SP* estimates under the alternative assumptions that the tests were conditionally independent or highly interdependent (to the maximum degree allowed by the parameter settings).

We assumed that the sensitivity of the new test (*SN*₁), of culture (*SN*₂), and of

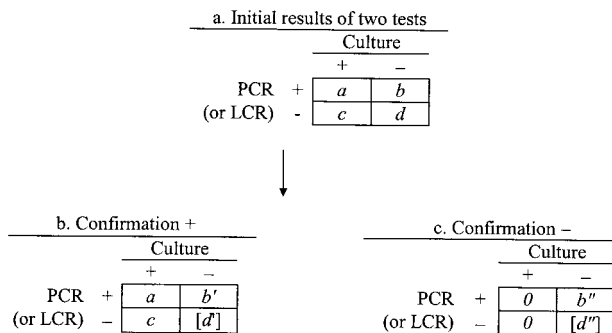


FIG. 2. Data layout of initial and confirmed results in discrepant analysis. Brackets indicate unknown value.

the confirmation procedure (reference, SN_r) were all greater than 55%, that the corresponding false-positive rates (FP_1 , FP_2 , and FP_3) were all less than 5%, and that the disease prevalence ranged from 2 to 15%. Specifically, the parameter values used in the evaluation study were generated by letting the sensitivity and specificity of the relevant tests vary within the following ranges: the SN and SP of PCR, LCR, or MOMP were 80 to 100% and 95 to 100%, respectively; the SN and SP of cell culture were 55 to 100% and 95 to 100%, respectively; and the SN and SP of DFA were 60 to 100% and 95 to 100%, respectively.

The sensitivity of the confirmation procedure was calculated according to Eq 8 by using the sensitivity of each test involved in the resolution. The specificity was calculated according to Eq 9 by using the specificity of each test involved in the resolution.

$$P(T_r | D) = \sum_i P(T_i | D) - \sum_{i \neq j} P(T_i \cap T_j | D) + P(\cap_i T_i | D), \quad (8)$$

$$P(\bar{T}_r | \bar{D}) = P(\cap_i \bar{T}_i | \bar{D}), \quad (9)$$

where $i = 1, 2, 3$ and indicates LCR (or PCR), DFA, and MOMP, respectively, if all three tests were involved in the confirmation procedure. If only two tests, e.g., LCR and DFA, were used in the confirmation tests, then $i = 1, 2$. For example, the minimum value of the combined sensitivity of LCR and DFA was $(0.8 + 0.6 - 0.48) \times 100\% = 92\%$ and the corresponding specificity was $(0.95 \times 0.95) \times 100\% = 90.25\%$ under the assumption that the accuracy of these two tests was independent conditional on disease status. Alternatively, the minimum value of the combined sensitivity of LCR and DFA was $(0.8 + 0.6 - 0.6) \times 100\% = 80\%$ and the corresponding specificity was $0.95 \times 100\% = 95\%$, under the assumption that the accuracy of these two tests was maximally interdependent conditional on disease status.

Bias was evaluated under the circumstances that (i) all the test accuracy parameters were mutually independent conditional on disease status, (ii) test sensitivities were maximally interdependent and specificities were mutually independent conditional on disease status, and (iii) all test accuracy parameters were highly interdependent. In the calculation of RFP estimates, we added 0.000001 to cell probabilities to avoid zero marginal probabilities when the test accuracy was maximally interdependent.

Interpretation of RSN and RFP estimates in terms of sensitivity and specificity. Based on the estimates of RSN and RFP , the corresponding sensitivity and specificity of PCR and LCR were calculated as $SN_1 = RSN \times SN_2$ and $SP_1 = 1 - RFP_1 = 1 - RFP \times FP_2$, respectively, for specified levels of the sensitivity and specificity accuracy of culture (7).

RESULTS

Enrollment of subjects began in October 1995 and ended in August 1997. A total of 1,138 asymptomatic men were enrolled in this study; 1,136 subjects were tested for the comparison of PCR and culture, 1,134 subjects were tested for the comparison of LCR and culture (Fig. 3 and 4), and 1,132 subjects were tested by all three methods.

RSN and RFP estimates of PCR or LCR versus cell culture. The RSN estimate of PCR was 1.4 (95% CI = 1.3 to 1.7), and the RFP estimate was 4.0 (95% CI = 0.5 to 36), after the discordant results were resolved by using the LCR, DFA, and MOMP methods sequentially (Fig. 3d). The RSN estimate of LCR was 1.5 (95% CI = 1.3 to 1.7), after the discordant results were resolved by using the PCR, DFA, and MOMP methods sequentially (Fig. 4d). Because no false-positive LCR results were found, the RFP of LCR was zero, and its variance could not be estimated.

Evaluation of bias in RSN estimates. Figures 5 to 12 display selected results of the systematic evaluation of the potential for bias in relative measures of accuracy. Each figure displays the percent bias in the RSN estimate as a function of two of the seven relevant parameters, holding the other five constant at specified, plausible levels. Figures 5 to 8 present the results of analyses based on the commonly held assumption that test

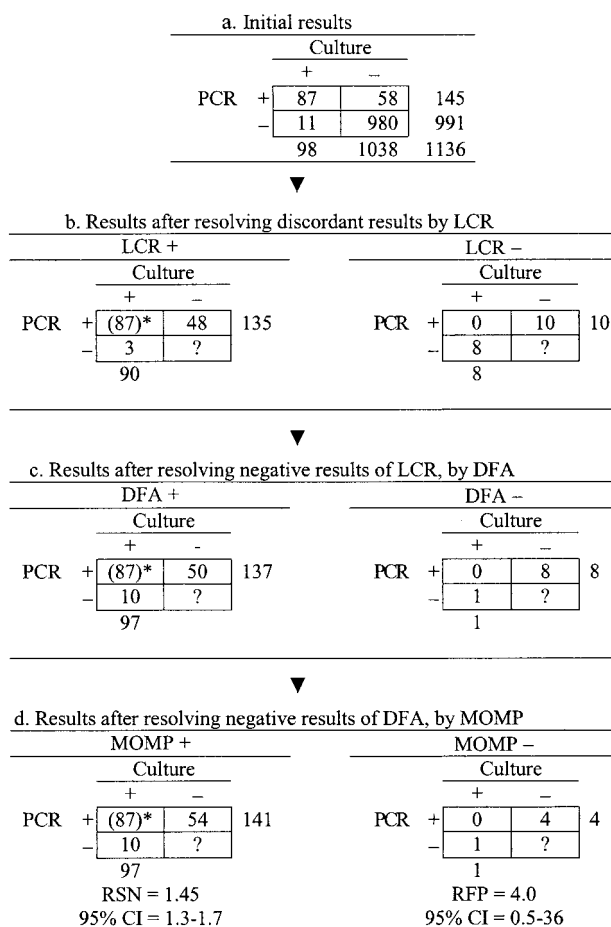


FIG. 3. Confirmation procedure in the discrepant analysis of PCR-culture comparison. The number 87 in parentheses was positive for both PCR and cell culture and was classified as positive without resolution.

results are independent from each other conditionally with regard to disease status, whereas Fig. 9 to 12 are based on the assumption that test sensitivities are positively associated, while specificities are mutually independent. If we assume that the sensitivities and specificities of all the tests were independent, the percent bias was most often less than 5% if the sensitivity of the confirmation procedure (SN_r) was $\geq 85\%$ (Fig. 5 to 8). The bias of the RSN estimate was increasingly negative (i.e., toward underestimating the RSN) as the true sensitivity of the new test increased and as the sensitivity of culture decreased, holding other parameters constant (Fig. 5). The specificity of the new test and of culture seemed to have a stronger influence on the sign and size of the bias, which was increasingly negative as the true specificity of the new test increased and as the specificity of culture decreased (Fig. 6 and 7). The bias also was increasingly negative as the true sensitivity of the confirmation procedure decreased but did not vary appreciably with the specificity of the procedure, if other parameters are held constant (Fig. 7). Finally, the bias did not vary appreciably over a relatively wide range of prevalence rates (Fig. 8). Overall, in these analyses the bias was less than 4% if the difference between SN_1 (sensitivity of the new test)

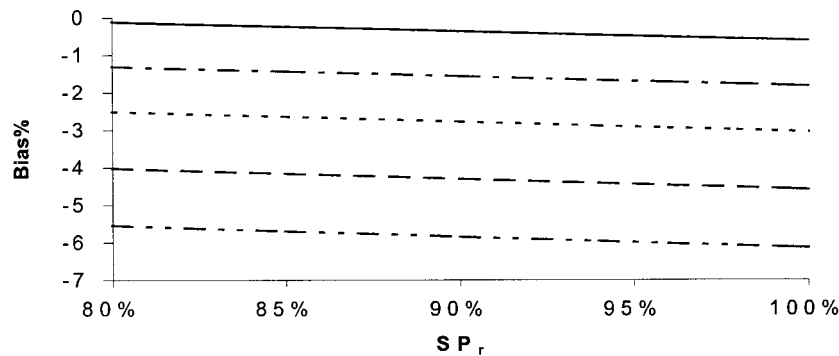


FIG. 6. Percent bias of *RSN* estimates, independence. $SN_1 = 90\%$, $SP_1 = 99\%$, $SN_2 = 60\%$, $SP_2 = 99.5\%$, and prevalence = 10%. SN_r : 80%; 85%; 90%; - - - - 94%; —, 98%.

independent, the bias in *RFP* estimates was similar to the case when all the test accuracy were independent. The bias was less than 2% when SN_r was higher than or equal to the higher values of SN_1 and SN_2 .

If the sensitivities and specificities of all the tests were maximally interdependent, the bias was generally larger than 50%, indicating that discrepant analysis is not a suitable design for situations in which false-positive errors may be correlated.

Interpretation of *RSN* estimates in terms of sensitivity and comparison of the bias of absolute and relative estimates. The estimates of *RSN* calculated in this study were applied to alternative theoretical values of the sensitivity of culture to evaluate the possible range of sensitivity values of PCR and LCR (Table 1). This analysis showed that if the sensitivity of the culture methods used in this study was as low as 60%, the corresponding sensitivity levels of PCR and LCR would be between 80 and 90%, whereas for values of culture sensitivity close to 70% the sensitivity of PCR and LCR would be virtually 100%.

Table 2 compares the direction and size of the bias associated with *RSN* estimates with the corresponding bias associated with absolute estimates of sensitivity for selected combinations of parameters. Whereas the absolute estimates of sensitivity were overestimated by as much as 25%, the *RSN* estimates were only slightly underestimated (<10%, but most

often <5%). Furthermore, *RSN* estimates were conservative (biased toward to the null) in many circumstances.

DISCUSSION

The accuracy of PCR and LCR has been the subject of intense debate. Most of the reported sensitivity estimates of PCR or LCR on urine specimens range from 86 to 96%, and specificity estimates are usually higher than 99.5%. The validity of these estimates, however, has been criticized with criticisms focused on the process of "discrepant analysis" which leads to selective confirmation of initial test results. Cell culture, the traditional standard for diagnosing *C. trachomatis* infection and for evaluating the accuracy of new tests, clearly detects fewer than 90% of infections and, as more sensitive methods for chlamydial detection are developed, is probably no longer a suitable standard (1, 5, 6, 10, 11, 13, 15, 17). Using additional tests to resolve discordant results between PCR (LCR)-positive and cell culture-negative (discrepant analysis) has been advocated by Schachter et al. (16) and criticized by Hadgu et al. (8, 9) and Green et al. (7, 11).

In the present study, we addressed the main criticism about selective confirmation of test results in discrepant analysis by (i) applying the confirmation procedure to all the discordant results between PCR or LCR and culture, including both cul-

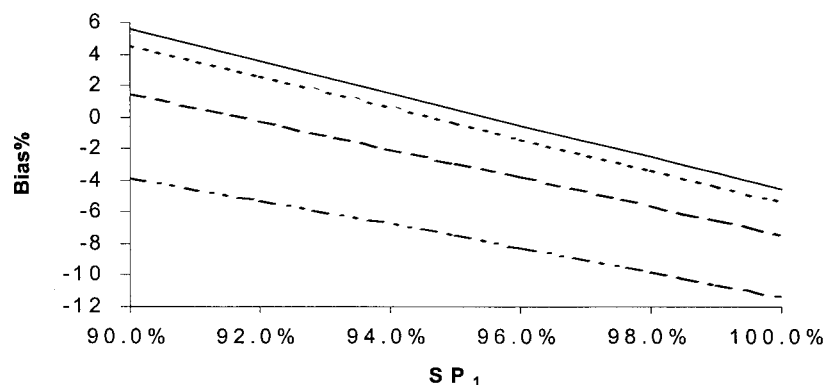


FIG. 7. Percent bias of *RSN* estimates, independence. $SN_r = 85\%$, $SP_r = 90\%$, $SN_1 = 90\%$, $SN_2 = 60\%$, and prevalence = 10%. SP_2 : 95%; 98%; 99.5%; —, 100%.

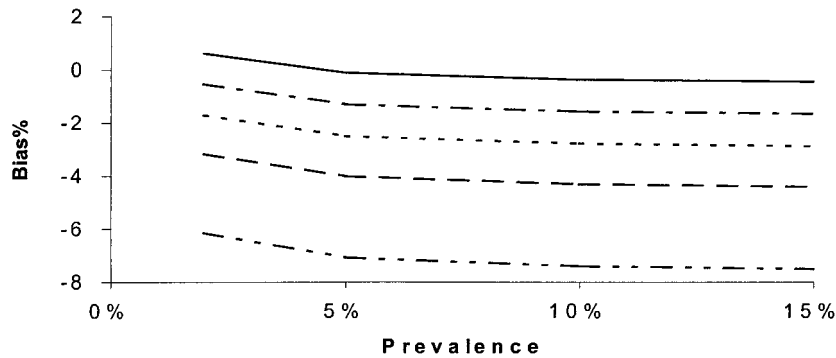


FIG. 8. Percent bias of *RSN* estimates, independence. $SP_r = 90\%$, $SN_1 = 90\%$, $SP_1 = 99\%$, $SN_2 = 60\%$, and $SP_2 = 99.5\%$. SN_r : 70%; 75%; 85%; 90%; - - - - 94%; ———, 98%.

ture-negative, PCR (or LCR)-positive results and culture-positive, PCR (or LCR)-negative results; (ii) estimating *RSN* and *RFP* to reduce the bias due to partial confirmation of the denominators of sensitivity and specificity; and (iii) estimating the residual bias associated with relative estimates of accuracy under a range of plausible assumptions about the true value of the unknown parameters.

We found that *RSN* estimates of PCR or LCR versus cell cultures range from 1.4 to 1.5 and are significantly higher than the null value of 1, suggesting that the sensitivity of PCR and LCR is substantially larger than that of cell culture. The confirmation procedure identified few false-positive results overall. The *RFP* estimates varied from 1.2 to 8, with CIs that were always wide and included the null value of 1. Thus, although the point estimates suggest that the specificities of PCR and LCR are lower than that of cell culture, these results are inconclusive and are also compatible with equivalent specificities. The difference in sensitivity between PCR and LCR seems small, since their *RSN* estimates were very similar. These results are consistent with the findings of previous studies (13).

The average sensitivity of culture on cervical swab specimens was estimated to be about 80% in expert laboratories (5, 12). Culture sensitivity may be lower among asymptomatic men because of the difficulty in obtaining satisfactory urethral specimens, as well as the potentially lower concentrations of *C.*

trachomatis (5, 6). If the value 1.5 is a reasonable estimate for the *RSN* of PCR or LCR and if the sensitivity of the culture test is 60 to 65%, the sensitivity of LCR and PCR is 90 to 97% (Table 4). This range of estimates is consistent with the findings of previous studies (5, 6).

Since the accuracy of the confirmation procedure was not perfect and the tests evaluated are probably not mutually independent, even *RSN* and *RFP* estimates, which are less prone to bias than absolute estimates of sensitivity and specificity, may be distorted. Because of the great concern about the potential for bias in the published estimates of the sensitivity and specificity of LCR and PCR (1, 5, 6, 10, 11, 13, 15, 17), we evaluated the direction and size of the potential bias by using a relatively simple mathematical model and assuming plausible ranges for all relevant parameters. In addition, we considered the possibility that diagnostic errors might not be mutually independent. Interdependence of test sensitivities is biologically plausible because all tests depend on the presence of whole or partial chlamydia organisms. In contrast, false-positive rates are likely to be independent of each other, because different mechanisms would lead to false-positive results in the tests evaluated: reduced culture specificity on urethral specimens can result from cross-contamination of specimens or misclassification due to the presence of cell artifacts that resemble inclusions, while false-positive PCR and LCR results

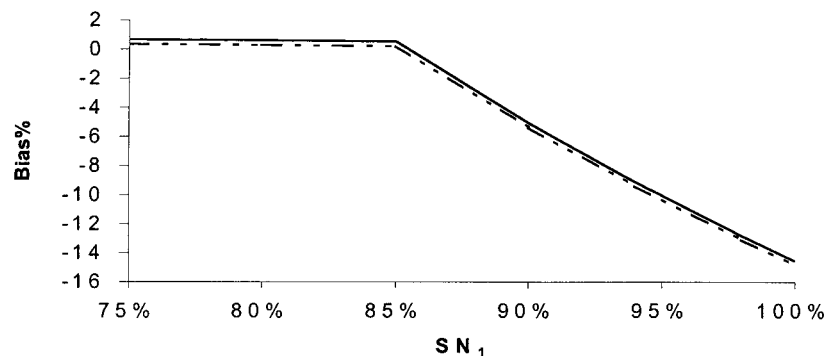


FIG. 9. Percent bias of *RSN* estimates according to SN_1 . SNs are maximum dependent, and SPs are independent. $SN_r = 90\%$, $SP_r = 90\%$, $SP_1 = 99\%$, $SN_2 = 60\%$, and prevalence = 10%. SN_2 : 55%; ———, 85%.

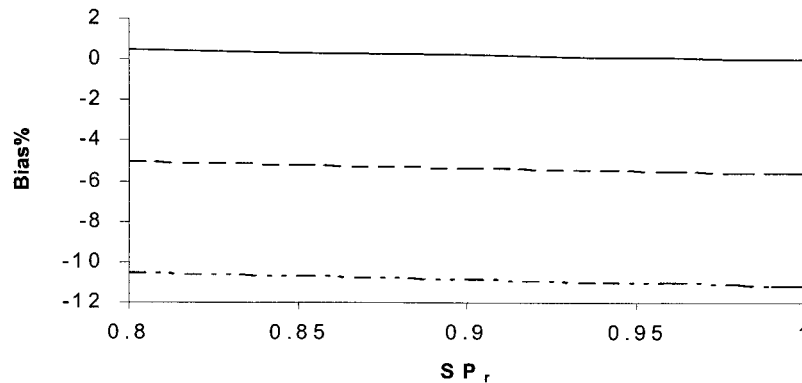


FIG. 10. Percent bias of *RSN* estimates according to SP_r . SNs are maximum dependent, and SPs are independent. $SP_1 = 90\%$, $SP_1 = 99\%$, $SN_2 = 60\%$, $SP_2 = 99.5\%$, and prevalence = 10%. SN_r : - - - - - , 80%; - · - · - · - , 85%; —, 90 to 100%.

are presumably most often due to carryover contamination of urine specimens.

In a realistic scenario (sensitivity of PCR, LCR, and confirmatory procedure, $\geq 80\%$; sensitivity of culture, $< 80\%$; specificity of all tests, $> 95\%$; prevalence, 10%, with moderate interdependence of false-negative errors), the *RSN* estimates presented here underestimate the true values by ca. 5%. Thus, the true value of the *RSN* is about 1.5 (i.e., 1.45/0.95 for PCR). For the same range of parameter values, the *RFP* estimates are likely to have been overestimated by 15 to 20%. The results suggest that the specificities of PCR and LCR are slightly lower than culture specificity (Table 4). If the true culture specificity is 99 to 99.5%, PCR or LCR specificity would be 95 to 97%. However, conclusions are much harder to draw about the specificity levels because of the imprecision of the estimates, and the data are also compatible with no difference in specificity among PCR, LCR, and culture.

Hadgu et al. (8, 9) and Green et al. (7) found that discrepant analysis leads to overestimating the absolute sensitivity of PCR and LCR and that the bias is large (e.g., $> 10\%$) under a broad array of circumstances. Our calculations suggest that, whereas absolute sensitivity estimates would be biased by as much as 25%, the *RSN* estimates are only slightly underestimated (most often by $< 5\%$). Because of the conservative nature of the bias

of *RSN* estimates, the findings of this study support the conclusion that DNA amplification tests are considerably more sensitive than culture.

The bias of absolute specificity estimates is usually smaller than the bias of the *RFP* estimates when test results are mutually independent. The absolute specificity is generally underestimated and the *RFP* tends to be overestimated. When test results are interdependent, however, bias in both estimates is usually very large. Thus, our data indicate that the use of discrepant analysis is unlikely to yield acceptable results (and should not be employed) in situations where false-positive test results may be correlated.

A potential limitation of the comparisons evaluated in this study is that the sampling procedures employed to obtain the specimens for DNA amplification tests (first-void urine) were different from the procedures used to obtain specimens for cell culture (urethral swabs). Thus, the measures of absolute or relative accuracy refer not just to the laboratory assay (DNA amplification versus culture) but, more properly, to the combination of sampling procedure and laboratory assay. If urethral specimens were likely to contain more *C. trachomatis* organisms, the association of culture and swab would bias the comparisons in favor of culture performance. Alternatively, if the invasiveness of the procedure for obtaining urethral spec-

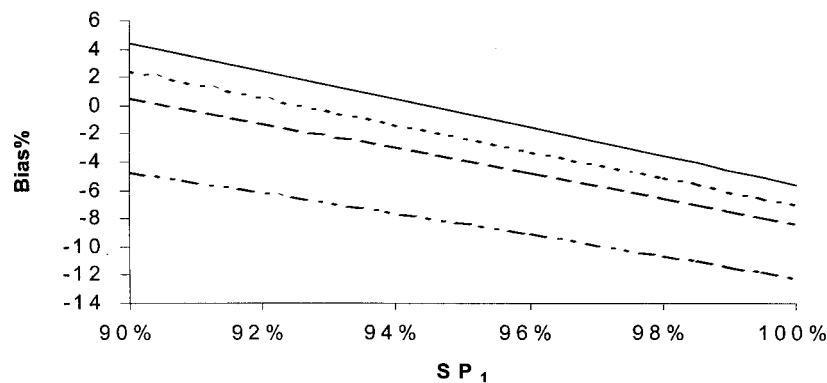


FIG. 11. Percent bias of *RSN* estimates according to SP_1 . SNs are maximum dependent, and SPs are independent. $SN_r = 85\%$, $SP_r = 90\%$, $SN_1 = 90\%$, $SN_2 = 60\%$, and prevalence = 10%. SP_2 : - - - - - , 95%; - · - · - · - , 98%; ·····, 99%; —, 100%.

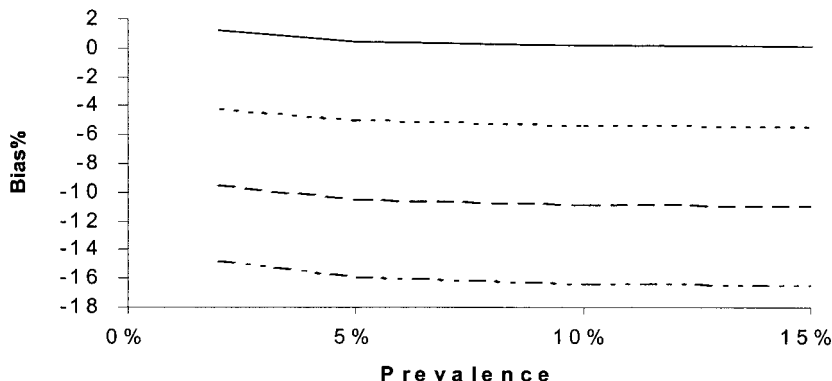


FIG. 12. Percent bias of *RSN* estimates according to prevalence. *SNs* are maximum dependent, and *SPs* are independent. $SN_r = 90\%$, $SP_r = 90\%$, $SN_1 = 90\%$, $SN_2 = 60\%$, and prevalence = 10%. SP_2 : ·····, 95%; ······, 98%; ······, 99%; —, 100%.

imens led to less-satisfactory samples with fewer organisms, culture performance would be impaired, biasing the comparisons in favor of DNA amplification tests. Furthermore, the comparative research design required that multiple samples be taken from the same individual. It is possible that the requirement to obtain two urethral swabs before the first-void urine reduced the number of chlamydia organisms shed by the urethra into the urine, increasing the likelihood of false-negative PCR or LCR results. Thus, the sensitivity of PCR and LCR could have been higher if only a urine sample (as would happen in a screening program) had been taken. The order in which specimens were taken in a study comparing two methods for diagnosing genital human papillomavirus infection was evaluated by Vermund et al. and showed slight influence on the diagnostic accuracy (20). On the other hand, to the extent

that one is interested in evaluating the performance of a procedure that can be broadly applied to asymptomatic men, compared to a procedure that can only be applied within the clinic setting, the comparisons made in this analysis are appropriate.

In summary, Hadgu and Green's concerns (7-9) about bias in the estimates of test sensitivity and specificity are valid and should be carefully evaluated as new tests are developed. The *RSN*, *RFP*, and bias estimates in this study suggest that they do not dramatically distort calculation of the accuracy of PCR and LCR. We concluded that the sensitivity of PCR and LCR is significantly greater than the sensitivity of culture for screening asymptomatic men and that the specificities of these tests are very similar.

APPENDIX

When a confirmation procedure is applied to the discordant results of the two compared tests, *RSN* can be estimated using Eq A1 and *RFP* can be estimated by using Eq A2, as follows:

$$R\hat{S}N = \frac{P(T_1 | D, T_r)P(D) + P(T_1 \cap T_2 | D, \bar{T}_r)P(D) + P(T_1 | \bar{D}, T_r)P(\bar{D}) + P(T_1 \cap T_2 | \bar{D}, \bar{T}_r)P(\bar{D})}{P(T_2 | D, T_r)P(D) + P(T_1 \cap T_2 | D, \bar{T}_r)P(D) + P(T_2 | \bar{D}, T_r)P(\bar{D}) + P(T_1 \cap T_2 | \bar{D}, \bar{T}_r)P(\bar{D})} \tag{A1}$$

and

$$R\hat{F}P = \frac{P(T_1 \cap \bar{T}_2 | \bar{D}, \bar{T}_r)P(\bar{D}) + P(T_1 \cap \bar{T}_2 | D, \bar{T}_r)P(D)}{P(\bar{T}_1 \cap T_2 | \bar{D}, \bar{T}_r)P(\bar{D}) + P(\bar{T}_1 \cap T_2 | D, \bar{T}_r)P(D)} \tag{A2}$$

In Eq A1 and Eq A2, T_1 , T_2 , and T_r stand for positive results of test 1 (in the present study, either PCR or LCR), test 2 (culture), and the confirmation (or reference) procedure (the sequence of the other DNA amplification test, DFA, and MOMP test employed to verify discordant results); \bar{T}_1 , \bar{T}_2 , and \bar{T}_r stand for the corresponding negative results; D denotes the presence of disease; and $P(D)$ is the disease prevalence. $P(T_1 | D, T_r)$ is the probability that test 1 is positive, depending on the presence of disease and positive confirmation results. For example, if the confirmation procedure and test 1 are indepen-

TABLE 1. Sensitivity of LCR and PCR derived from estimates of the sensitivity of culture and of *RSN*

<i>RSN</i>	Sensitivity (%)	
	Culture	PCR or LCR
1.50	67.0	100.0
	65.0	97.5
	60.0	90.0
	55.0	82.5
1.45	70.0	100.0
	65.0	94.3
	62.0	89.9
	60.0	87.0
	55.0	79.8
1.40	71.4	100.0
	70.0	98.0
	65.0	91.0
	60.0	84.0
	55.0	77.0
1.30	76.9	100.0
	75.4	98.0
	73.1	95.0
	70.0	91.0
	65.0	84.5
	60.0	78.0
	55.0	71.5

equals $\min\{[SN_1 - \min(SN_r, SN_1)], [SN_2 - \min(SN_r, SN_2)]\}P(D) + SP_r(1 - SP_1)(1 - SP_2)P(\bar{D})$ and

$$R\hat{F}P = \frac{SN_1 - \min(SN_r, SN_1) + SP_r(1 - SP_1)P(\bar{D}) - G}{SN_1 - \min(SN_r, SN_1) + SP_r(1 - SP_2)P(\bar{D}) - G} \tag{A8}$$

By using the notation above, absolute sensitivity and specificity can be estimated based on the data layout in Fig. 1, 3, and 4.

When test results are mutually independent conditional on the presence of disease, the potentially biased estimates of absolute sensitivity and specificity PCR and LCR can be calculated by using Eq A9 and Eq A10, respectively, as follows:

$$S\hat{N}_1 = \frac{x + y}{x + y + z} \tag{A9}$$

where x equals $SN_1SN_2P(D) + (1 - SP_1)(1 - SP_2)P(\bar{D})$, y equals $SN_1(1 - SN_2)SN_rP(D) + (1 - SP_1)SP_2(1 - SP_r)P(\bar{D})$, and z equals $(1 - SN_1)SN_2SN_rP(D) + SP_1(1 - SP_2)(1 - SP_r)P(\bar{D})$ and

$$S\hat{P}_1 = 1 - \frac{SN_1(1 - SN_2)(1 - SN_r)P(D) + (1 - SP_1)SP_2SP_rP(D)}{\{(1 - SN_r)P(D) - SN_1SN_2(1 - SN_r)P(D) + SP_rP(\bar{D}) - SP_r(1 - SP_1)(1 - SP_2)P(\bar{D})\}} \tag{A10}$$

When sensitivities are maximally interdependent but specificities are mutually independent, the SN_1 and SP_2 of PCR and LCR can be estimated by using Eq A11 and Eq A12, respectively, as follows:

$$S\hat{N}_1 = \frac{E + G}{E + F + G} \tag{A11}$$

and

$$S\hat{P}_1 = \frac{SN_1 - \min(SN_r, SN_1) + SP_r(1 - SP_1)P(\bar{D}) - G}{(1 - SN_r)P(D) + SP_rP(\bar{D}) - G} \tag{A12}$$

where $E, F,$ and G are the same as used in Eq A7 and Eq A8.

REFERENCES

1. **Black, C. M.** 1997. Current methods of laboratory diagnosis of *Chlamydia trachomatis*. Clin. Microbiol. Rev. **10**:160-184.
2. **Centers for Disease Control and Prevention.** 1993. Recommendations for the prevention and management of *Chlamydia trachomatis* infections. Morb. Mortal. Wkly. Rep. **42**(RR-12):1-39.
3. **Cheng, H., and M. Macaluso.** 1997. Comparison of the accuracy of two tests with a confirmation procedure limited to positive results. Epidemiology **8**:104-106.
4. **Cheng, H., M. Macaluso, and M. Hardin.** 2000. Validity and coverage of estimates of relative accuracy. Ann. Epidemiol. **10**:251-260.
5. **Chernesky, M. A., D. Jang, H. Lee, J. D. Burczak, H. Hu, J. Sellors, S. J. Tomazic-Allen, and J. B. Mahony.** 1994. Diagnosis of *Chlamydia trachomatis* infections in men and women by testing first-void urine by ligase chain reaction. J. Infect. Dis. **32**:2682-2685.
6. **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.
7. **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**:2540-2543.
8. **Hadgu, A.** 1997. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. Stat. Med. **16**:1391-1399.
9. **Hadgu, A.** 1996. The discrepancy in discrepant analysis. Lancet **348**:592-593.
10. **Hillis, S., C. Black, J. Newhall, C. Walsh, and S. Groseclose.** 1995. New opportunities for *Chlamydia* prevention: applications of science to public health practice. Sex. Transm. Dis. **3**:197-202.
11. **Johnson, E. T., T. A. Green, J. Schachter, R. B. Jones, E. W. Hook III, C. M. Black, D. H. Martin, M. E. St. Louis, and W. E. Stamm.** 2000. Evaluation of nucleic acid amplification tests as reference tests for *Chlamydia trachomatis* infections in asymptomatic men. J. Clin. Microbiol. **38**:4382-4386.
12. **Pate, M. S., and E. W. Hook III.** 1995. Laboratory to laboratory variation in *Chlamydia trachomatis* culture practices. Sex. Transm. Dis. **22**:322-326.
13. **Schachter, J.** 1997. DFA, EIA, PCR, LCR and other technologies: what tests should be used for diagnosis of *Chlamydia* infection? Immunol. Investig. **26**:157-161.
14. **Schachter, J., et al.** 1998. Discrepant analysis and screening for *Chlamydia trachomatis*. Lancet **351**:217-218.
15. **Schachter, J., W. E. Stamm, T. C. Quinn, W. W. Andrews, J. D. Burczak, and H. H. Lee.** 1994. Ligase chain reaction to detect *Chlamydia trachomatis* infection of the cervix. J. Clin. Microbiol. **32**:2540-2543.
16. **Schachter, J., et al.** 1996. Discrepant analysis and screening for *Chlamydia trachomatis*. Lancet **348**:1308-1309.
17. **Sculnick, M., R. Chua, A. E. Simor, D. E. Low, H. E. Khosid, S. Fraser, E. Lyons, E. A. Legere, and D. A. Kitching.** 1994. Use of the polymerase chain reaction for the detection of *Chlamydia trachomatis* from endocervical and urine specimens in an asymptomatic low-prevalence population of women. Diagn. Microbiol. Infect. Dis. **20**:195-201.
18. **Stamm, W. E., and K. K. Holmes.** 1990. *Chlamydia trachomatis* infections of adults, p. 181-194. In K. K. Holmes, P.-A. Mardh, P. F. Sparling, P. J. Wiesner, et al. (ed). Sexually transmitted diseases, part V: sexually transmitted agents. McGraw-Hill Information Services Company, New York, N.Y.
19. **Taylor-Robinson, D.** 1997. Evaluation and comparison of tests to diagnose *Chlamydia trachomatis* genital infections. Hum. Reprod. **12**:113-120.
20. **Vermund, S. H., M. H. Schiffman, G. L. Goldberg, D. B. Ritter, A. Weltman, and R. D. Burk.** 1989. Molecular diagnosis of genital human papillomavirus infection: comparison of two methods used to collect exfoliated cervical cells. Am. J. Obstet. Gynecol. **160**:304-308.
21. **World Health Organization.** 1995. An overview of selected curable sexually transmitted diseases. Global Program on AIDS. World Health Organization, Geneva, Switzerland.