

Evaluation of Bias in Diagnostic-Test Sensitivity and Specificity Estimates Computed by Discrepant Analysis

TIMOTHY A. GREEN,^{1*} CAROLYN M. BLACK,¹ AND ROBERT E. JOHNSON²

*Division of AIDS, STD, and TB Laboratory Research, National Center for Infectious Diseases,¹ and
Division of Sexually Transmitted Diseases Prevention, National Center for HIV, STD, and
TB Prevention,² Centers for Disease Control and Prevention, Atlanta, Georgia 30333*

Received 2 June 1997/Returned for modification 12 August 1997/Accepted 30 October 1997

When a new diagnostic test is potentially more sensitive than the reference test used to classify persons as infected or uninfected, a substantial number of specimens from infected persons may be reference-test negative but new-test positive. Discrepant analysis involves the performance of one or more additional tests with these specimens, reclassification as infected those persons for whom the new-test-positive results are confirmed, and recalculation of the estimates of new-test sensitivity and specificity by using the revised classification. This approach has been criticized because of the bias introduced by the selective use of confirmation testing. Under conditions appropriate for evaluating a nucleic acid amplification (NAA) test for *Chlamydia trachomatis* infection with cell culture as the reference test, we compared the bias in estimates based on the discrepant-analysis classification of persons as infected or uninfected with that in estimates based on the culture classification. We concluded that the bias in estimates of NAA-test specificity based on discrepant analysis is small and generally less than that in estimates based on culture. However, the accuracy of discrepant-analysis-based estimates of NAA-test sensitivity depends critically on whether culture specificity is equal to or is slightly less than 100%, and it is affected by competing biases that are not fully taken into account by discrepant analysis.

New tests for the diagnosis of the presence or absence of various viral and bacterial infections are continually being developed. The accuracy of such a test can be described by its sensitivity, i.e., the probability that a specimen from an infected person tests positive, and its specificity, i.e., the probability that a specimen from an uninfected person tests negative. If a test is evaluated with specimens from persons whose infection status is known with certainty, the proportion of positive test results for specimens from infected persons provides an unbiased estimate of sensitivity, and the proportion of negative test results for specimens from uninfected persons provides an unbiased estimate of specificity.

In most settings, tests must be evaluated with specimens from persons whose infection status cannot be known with certainty. Under these circumstances, a reference test performed with a clinical specimen is used to classify each person as infected or uninfected. To the extent that the reference test has less than perfect sensitivity or specificity, both the estimate of the sensitivity of the test being evaluated given by the proportion of new-test-positive results for specimens with reference-test-positive results and the estimate of specificity given by the proportion of new-test-negative results for specimens with reference-test-negative results may be biased.

The bias resulting from the use of imperfect reference tests has received much attention in the statistical and epidemiological literature (19). From a statistical standpoint, the problem is that the number of degrees of freedom is insufficient to estimate all parameters of interest. Typical statistical approaches require either that constraints be imposed on the parameters to reduce the number being estimated or that the

number of degrees of freedom be increased either by evaluating three or more tests, including the reference test, with a single population or by evaluating both the reference test and the new test with two or more populations in whom the prevalence of infection differs.

The introduction of nucleic acid amplification (NAA) tests to the field of diagnostic testing has highlighted the problems presented by imperfect reference tests. For example, with the diagnosis of *Chlamydia trachomatis* infection, the traditional reference test, cell culture, is believed to have high, perhaps even perfect, specificity but considerably lower sensitivity (3, 16). It is biologically plausible, however, that NAA tests such as PCR or ligase chain reaction (LCR) have much higher sensitivities than culture while retaining very high specificities (6, 11). If this is true, a substantial number of specimens from infected persons may test negative by cell culture but test positive by an NAA test. To improve the accuracy of estimates of NAA-test sensitivity and specificity, many investigators have adopted a practice termed "discrepant analysis," wherein culture-negative, NAA-test-positive specimens undergo one or more additional tests to determine whether the positive NAA test result can be confirmed (2–4, 17, 18). The additional tests typically include a second NAA test containing probes for a target sequence different from the target sequence of the probes used in the test under evaluation. An example of this is a PCR or LCR test that targets the major outer membrane protein (MOMP) gene of *C. trachomatis*. Persons with confirmed NAA-test-positive results are reclassified as infected, and estimates of NAA-test sensitivity and specificity are recalculated by using the revised classification.

Discrepant analysis has been criticized because of the bias introduced by the selective nature of the confirmation testing (7, 8). To provide a framework for evaluating the accuracy of published estimates of NAA-test sensitivity and specificity, we compared the bias in estimates based on the discrepant-anal-

* Corresponding author. Mailing address: Centers for Disease Control and Prevention, 1600 Clifton Road, N.E. (MS-A12), Atlanta, GA 30333. Phone: (404) 639-4460. Fax: (404) 639-4664. E-mail: tag1@cdc.gov.

TABLE 1. Estimates of LCR sensitivity and specificity by using both culture and discrepant analysis to classify persons as infected or uninfected^a

LCR result	Culture-positive result	Culture-negative result
Positive	<i>a</i>	$b_+ + b_-^b$
Negative	<i>c</i>	<i>d</i>

^a Culture-based estimates of sensitivity and specificity are calculated as follows: $a/(a + c)$ and $d/(d + b_+ + b_-)$, respectively. Discrepant-analysis-based estimates of sensitivity and specificity are calculated as follows: $(a + b_+)/(a + b_+ + c)$ and $d/(d + b_-)$, respectively.

^b b_+ and b_- denote MOMP test-positive and MOMP test-negative results, respectively.

ysis classification of persons as infected or uninfected with that in estimates based on the culture classification. Comparisons were made over realistic ranges of values for culture sensitivity and specificity, NAA-test sensitivity and specificity, and the prevalence of *C. trachomatis* infection in the study population.

In this article, LCR is used as an example of an NAA test. All such references to LCR, however, can be considered to apply to any of the NAA tests for *C. trachomatis*.

MATERIALS AND METHODS

We considered an experiment in which each specimen would be tested by both culture and LCR and in which culture-negative, LCR-positive specimens would be subjected to a single confirmation test, a MOMP test. Table 1 depicts the data that would be collected from such an experiment, along with two sets of estimates of LCR sensitivity and specificity. The culture-based estimates use the culture test to classify persons as infected or uninfected, while the discrepant-analysis-based estimates classify a person as infected when either the culture test is positive or both the LCR and MOMP tests are positive.

The bias in an estimate is the difference between its expected value (i.e., its value in the absence of sampling variability) and the actual value of the parameter being estimated. An estimate is biased upward or downward according to whether its expected value is greater than or less than the actual value of the parameter being estimated. Since the estimates under consideration are proportions, their expected values are conditional probabilities; for example, the expected value of the culture-based estimate of sensitivity is the probability of a positive LCR test result given a positive culture test result. For this analysis, we used elementary rules of probability to express each of these conditional probabilities in terms of the prevalence of infection and the sensitivities and specificities of the culture, LCR, and MOMP tests (see the Appendix for derivations). These expressions were then used to evaluate the bias in both the culture-based and discrepant-analysis-based estimates of LCR sensitivity and specificity when the values of test performance characteristics and prevalence of infection listed in Table 2 are assumed.

In selecting values for test performance characteristics and prevalence of infection, we attempted to reflect both what is included in the manufacturers' package inserts and what has been published in peer-reviewed articles by independent investigators. Values for the prevalence of infection were chosen to reflect typical values that have been reported in studies of both low- and high-risk populations and to cover generally accepted definitions of low and high prevalence (2, 3, 13, 18). Similarly, we selected two values for culture sensitivity, 70 and 85%, that represent the range of values observed in experienced laboratories (3, 15, 16). We also included a lower value of 60% to reflect what is likely to occur in less experienced laboratories. We used the generally accepted culture specificity value of 100% but also allowed for a slight degradation of this value. Although the appearance of fluorescence-stained inclusion bodies is held to be highly characteristic of *C. trachomatis* infection, cross contamination of specimens, misclassification due to the presence of cell artifacts that resemble inclusions, and clerical errors are all difficult to eliminate entirely. The amount of degradation in culture specificity was increased with increasing prevalence of infection, since the risk of splash-over contamination or of mistakenly assigning a positive test result to a specimen from an uninfected person might increase with the number of specimens from infected persons being handled. For actual LCR sensitivity and specificity, we used broad ranges of values that extend well below most published estimates. Since no data on the dependence of culture and LCR are available, we included the case in which LCR sensitivity is the same for culture-positive specimens as for culture-negative specimens as well as the case in which LCR sensitivity is moderately higher for culture-positive specimens than for culture-negative specimens. In both cases, we assumed that LCR specificity is not affected by whether the culture test result is positive or negative. Finally, since few performance data on MOMP tests have been reported to date, we set the MOMP-test sensitivity and specificity for culture-negative, LCR-positive

TABLE 2. Values of test performance characteristics and prevalence of infection for which bias is evaluated

Characteristic	Value(s)
Prevalence of infection.....	2, 5, and 10%
Culture sensitivity.....	60, 70, and 85%
Culture specificity.....	(i) 100% (ii) 99.92, 99.8, and 99.6% when prevalence of infection is 2, 5, and 10%, respectively
LCR sensitivity.....	75–95%, in increments of 1%
LCR specificity.....	90–100%, in increments of 0.5%
Culture-LCR dependence.....	(i) LCR sensitivity the same for culture-positive and culture-negative specimens (ii) LCR sensitivity 10% lower for culture-negative specimens than for culture-positive specimens
MOMP-test sensitivity ^a	85%
MOMP-test specificity ^a	95%

^a For culture-negative, LCR-positive specimens.

specimens to the midpoint of the ranges used for overall LCR sensitivity and specificity.

RESULTS

As indicated in Table 1, discrepant analysis removes the culture-negative, LCR-positive, MOMP-test-positive specimens from the denominator of the culture-based LCR specificity estimate and adds them to both the numerator and the denominator of the culture-based LCR sensitivity estimate. As a result, the discrepant-analysis-based estimates of both LCR sensitivity and LCR specificity are always greater than or equal to the culture-based estimates, with equality holding only if no persons are reclassified as a result of the confirmation testing.

The culture-based estimate of LCR specificity is biased downward throughout the indicated range. This bias increases as the prevalence of infection increases and as culture sensitivity decreases, i.e., as more specimens from infected persons are culture negative (Fig. 1). On the other hand, the discrepant-analysis-based estimate of LCR specificity may be biased upward or downward, but what bias exists is small (–0.63 to +0.43 percentage points) and is generally less than that of the culture-based estimate. Neither a slight degradation of culture specificity, the actual LCR sensitivity, nor a moderate dependence of this sensitivity on the culture test result has any noticeable effect on the bias in either the culture-based or the discrepant-analysis-based estimate (data not shown).

For estimates of LCR sensitivity, prevalence of infection and actual LCR specificity have relatively little effect on the bias in either the culture-based or the discrepant-analysis-based estimate (data not shown). In contrast, however, the bias in these estimates is greatly affected by small variations in culture specificity (Fig. 2).

If culture specificity is 100%, the expected value of the culture-based estimate of LCR sensitivity is equal to the actual LCR sensitivity for culture-positive specimens. Therefore, if LCR sensitivity is the same for culture-positive specimens and culture-negative specimens, the culture-based estimate is un-

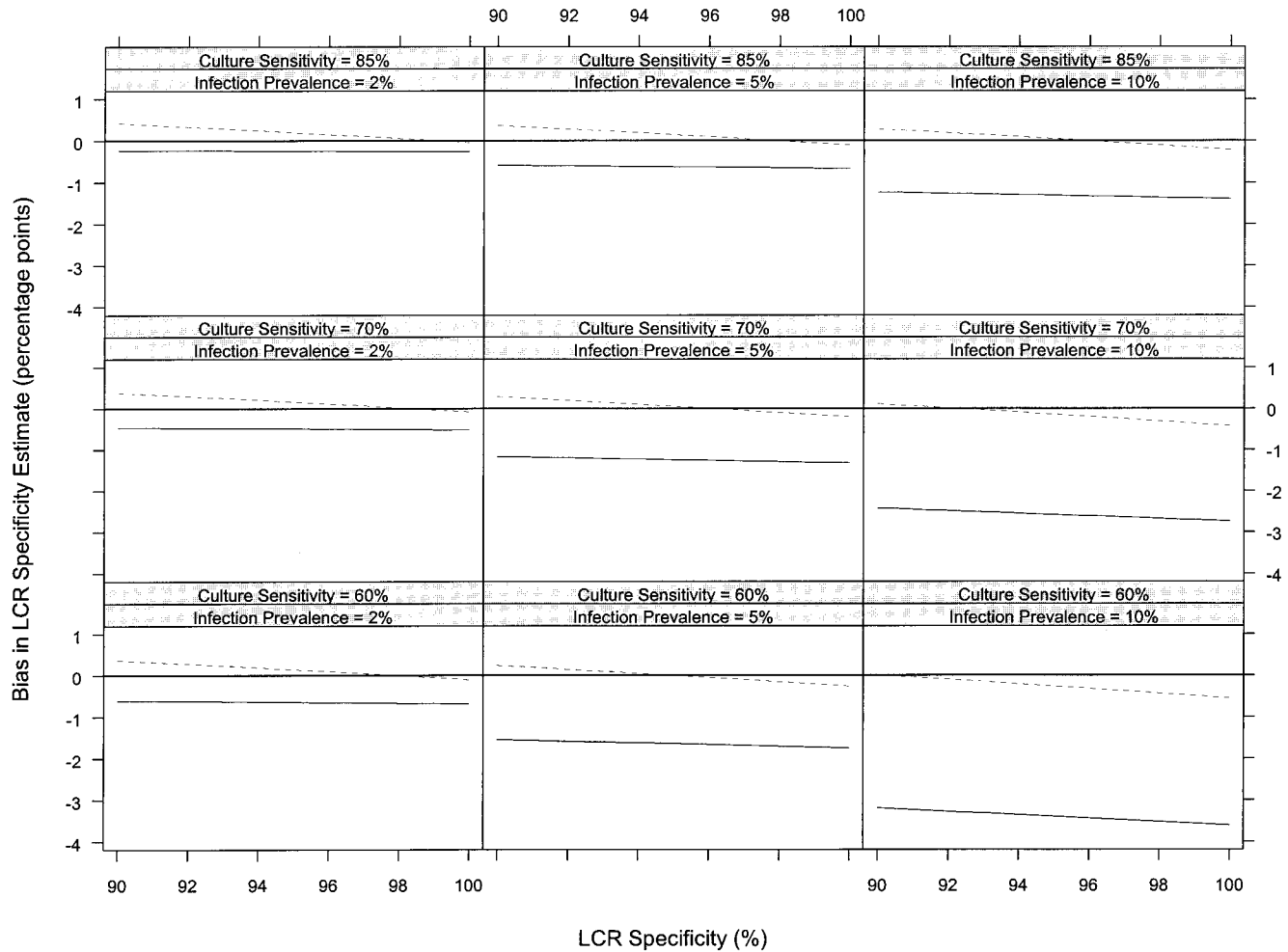


FIG. 1. Bias in culture-based (—) and discrepant-analysis-based (---) estimates of LCR specificity (culture specificity = 100%; LCR sensitivity = 85%; LCR sensitivity is the same for culture-positive and culture-negative specimens).

biased, and if it is higher for culture-positive specimens than for culture-negative specimens, the culture-based estimate is biased upward. Consequently, since the discrepant-analysis-based estimate is always greater than or equal to the culture-based estimate, it is biased upward in both cases and the bias is at least as large as any bias exhibited by the culture-based estimate. These statements, along with the magnitudes of the various biases, are illustrated in the left half of Fig. 2.

Conversely, if culture specificity is <100% (right half of Fig. 2), the expected values of the culture-based and discrepant-analysis-based estimates of LCR sensitivity are 2.6 to 6.1 percentage points and 2.3 to 4.1 percentage points, respectively, lower than the expected values of the corresponding estimates for 100% culture specificity. As a result, which of the estimates is less biased depends on the values assumed for other test performance characteristics and the prevalence of infection. For example, if the actual LCR sensitivity is moderately higher for culture-positive specimens than for culture-negative specimens (and culture specificity is <100%), then the culture-based estimate of LCR sensitivity is generally less biased if culture sensitivity is low but the discrepant-analysis-based estimate is generally less biased if culture sensitivity is high (Fig. 3). In each case, however, which estimate is less biased also depends to some extent on actual LCR sensitivity and speci-

ficity and the prevalence of infection, and it may be that the bias in neither estimate is acceptably small.

DISCUSSION

NAA tests such as LCR have great potential because their performance characteristics are inherently better than those of nonamplification tests (3, 4, 18). In addition, NAA tests can be used with specimens collected by noninvasive means, providing an important opportunity for screening asymptomatic persons and populations outside a clinical setting. For these tests to be widely accepted, however, samples with positive test results should not require confirmation testing since this adds to the already relatively high cost of an NAA test. However, removing the need for confirmation testing requires that the tests exhibit near-perfect specificity since the public health implications of even a very low rate of false positivity are enormous, particularly given the potential legal, medical, and social ramifications of the diagnosis of a sexually transmitted disease.

Estimates of the specificity of an LCR test for *C. trachomatis* infection, with culture used as the reference test, do not meet this standard (1, 4, 5, 9, 17). Because the sensitivity of culture is low, microbiologists have suspected that culture-based estimates of both LCR sensitivity and LCR specificity are biased

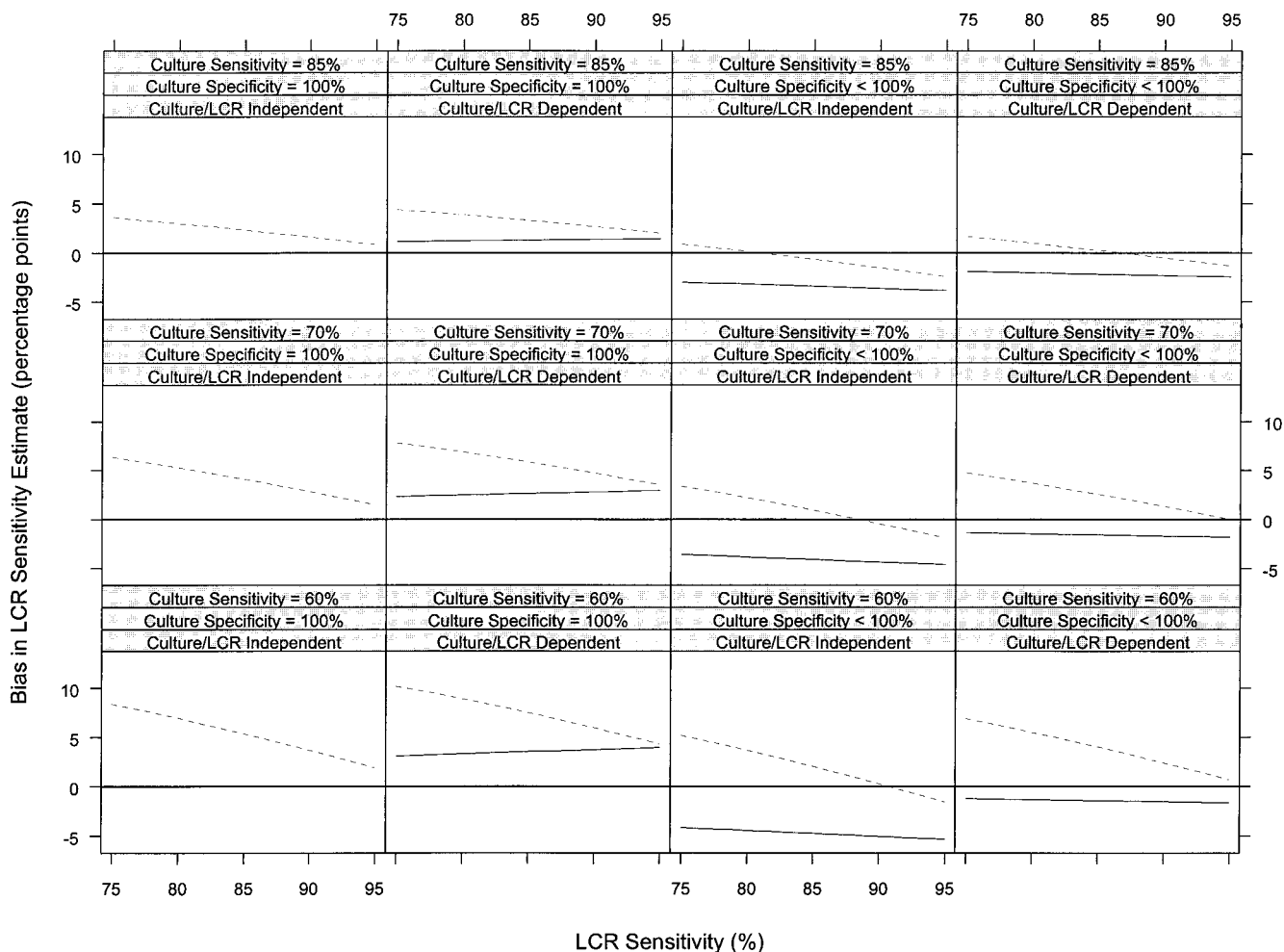


FIG. 2. Bias in culture-based (—) and discrepant-analysis-based (---) estimates of LCR sensitivity (prevalence of infection = 5%; LCR specificity = 95%). Culture/LCR independent refers to cases in which LCR sensitivity is the same for culture-positive and culture-negative specimens. Culture/LCR dependent refers to cases in which LCR sensitivity is moderately higher for culture-positive than for culture-negative specimens.

downward since a substantial number of LCR-positive specimens from infected persons are misclassified as uninfected. As a result of this suspicion, discrepant analysis has been used to improve the estimates. Discrepant analysis, as defined in this article, permits persons to be reclassified from uninfected to infected on the basis of a confirmation test applied to culture-negative, LCR-positive specimens; i.e., it attempts to properly classify what is presumed to be the most numerous group of misclassified specimens while disregarding misclassified specimens having other test-result combinations. For estimates of LCR specificity, this approach is sound. When realistic values of test performance characteristics and prevalence of infection are assumed, the bias in the discrepant-analysis-based estimate of LCR specificity is acceptably small and is generally less than that of the culture-based estimate. This is because removing LCR-positive specimens from the denominator of the LCR specificity estimate, even when an imperfect confirmation test is used, largely eliminates the underestimation of LCR specificity caused by culture-negative specimens from infected persons. Furthermore, other biases, particularly the overestimation caused by not removing similarly misclassified LCR-negative specimens from both the numerator and the denominator of the estimate, are negligible.

The effect of discrepant analysis on estimates of LCR sensitivity is more complicated. The ideal estimate of LCR sensitivity would be based exclusively on specimens from infected persons and would include all such specimens; such an estimate would be unbiased. If culture specificity is 100%, the culture-based estimate of LCR sensitivity is based exclusively on specimens from infected persons but only includes specimens that are culture positive. If LCR is equally sensitive for culture-positive and culture-negative specimens, the inclusion of only the culture-positive specimens does not introduce any bias; therefore, the culture-based estimate of LCR sensitivity remains unbiased. If, instead, LCR is more sensitive for culture-positive than for culture-negative specimens, the inclusion of only the culture-positive specimens causes the culture-based estimate to be biased upward. Most laboratory investigators and clinicians would expect this latter scenario to be true, particularly since culture positivity has been shown to correlate with higher numbers of organisms in the patient's specimen (10, 12). Since discrepant analysis adds culture-negative, LCR-positive specimens to both the numerator and the denominator of the culture-based estimate, it increases the estimate. As a result, discrepant analysis either creates or increases upward bias. This may seem counterintuitive to microbiologists who,

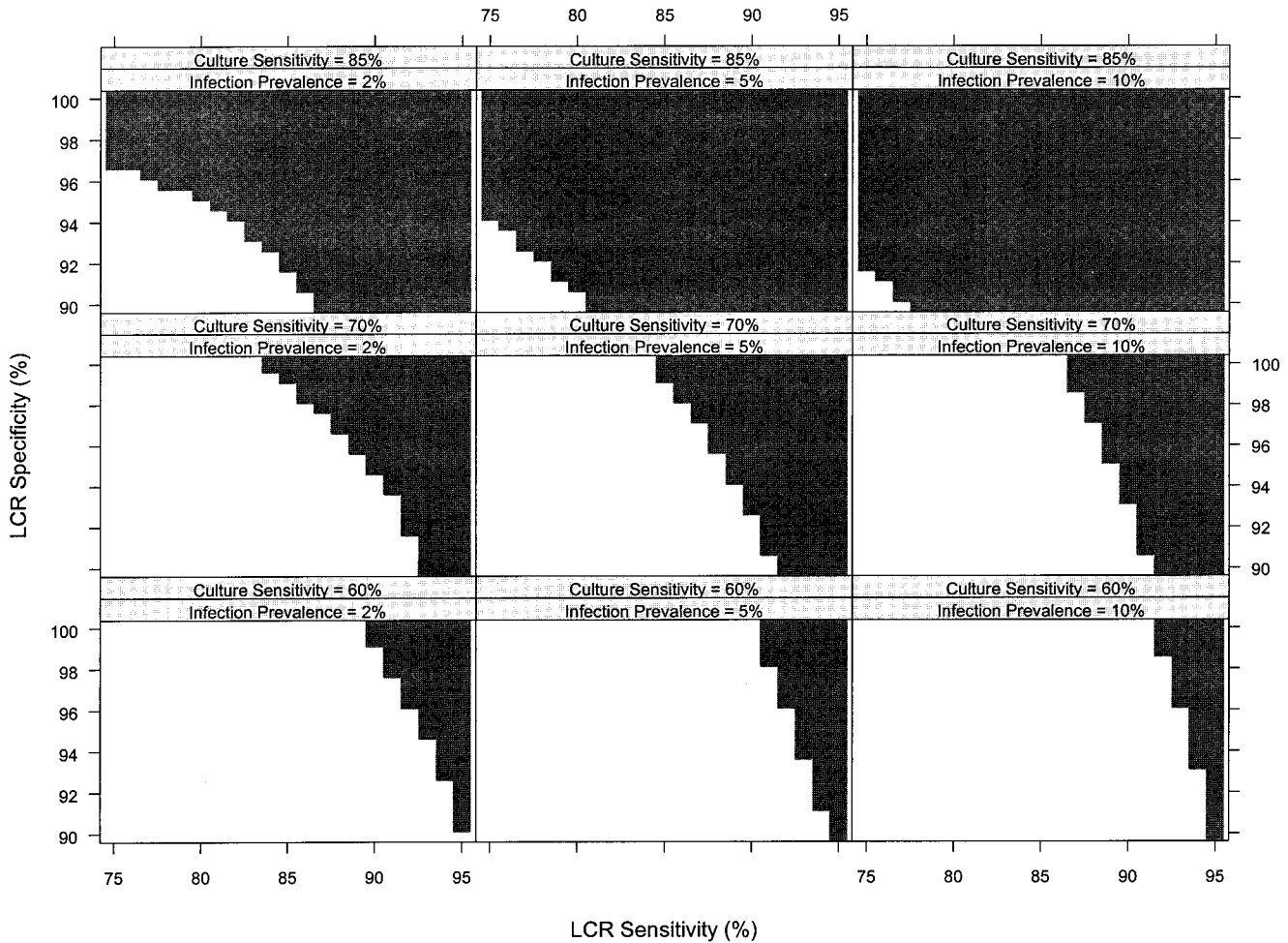


FIG. 3. Comparison of bias in culture-based and discrepant-analysis-based estimates of LCR sensitivity (culture specificity <100%; LCR sensitivity is moderately higher for culture-positive than for culture-negative specimens). The shaded areas indicate combinations of test performance characteristics and prevalence of infection for which the discrepant-analysis-based estimate is less biased than the culture-based estimate.

bearing in mind the biologically inherent improvement in the limit of detection between NAA technology and culture, believe a confirmation test that uses amplification technology would always alleviate, at least to some degree, the misclassification problem created by the insensitivity of culture. However, if LCR sensitivity for culture-positive specimens is either equal to or greater than that for culture-negative specimens, there exists a number (albeit a small number) of culture-negative, LCR-negative specimens from infected persons. Failing to add these specimens to the denominator of the culture-based estimate of LCR sensitivity is detrimental to the accuracy of the discrepant-analysis-based estimate.

Conversely, the direction of any bias in LCR sensitivity estimates is less predictable if culture is even slightly less than 100% specific. The presence of as few as 1 to 4 culture-positive test results per 1,000 specimens from uninfected persons, rates that most microbiologists would believe to be realistic in any laboratory, however skilled, introduces a substantial downward bias in the culture-based estimate of LCR sensitivity. This bias is downward because most of these culture-positive specimens will be LCR negative and thus included only in the denominator of the culture-based estimate; it is substantial to the extent that applying even a very low false-positive culture rate

to the large number of specimens from uninfected persons in low-prevalence settings produces a substantial number of culture-positive specimens compared to the much smaller number of specimens from infected persons. In this case, discrepant analysis may improve the culture-based estimate of LCR sensitivity by introducing an upward bias that offsets the downward bias caused by culture-positive specimens from uninfected persons. The estimate may thus be reasonably accurate, but only to the extent that competing biases not fully taken into account by discrepant analysis cancel each other out. This seems a poor justification for using discrepant analysis to estimate LCR sensitivity.

A possible limitation of this study stems from the use of a single, somewhat arbitrary set of performance characteristics for the confirmation test. The general conclusions of our analysis did not change, however, when the MOMP-test sensitivity and specificity values were varied over the broad ranges used for the initial LCR test. In particular, the upward bias of discrepant-analysis-based estimates of LCR specificity exceeded one-half a percentage point only when the actual LCR specificity was less than 95%. Therefore, there is little danger that the use of discrepant analysis will cause a test with unac-

ceptably low specificity to be judged as having near-perfect specificity.

The purpose of this article is to provide guidance for assessing the many published studies that have used discrepant analysis to evaluate NAA tests, with culture used as the reference test. Studies that use discrepant analysis contribute important and accurate information on NAA-test specificity. However, the accuracy of information on NAA-test sensitivity provided by these studies depends critically on whether culture specificity is equal to or is slightly less than 100%, and it is affected by competing biases that are not fully taken into account by discrepant analysis.

To increase readers' confidence in the accuracy of discrepant-analysis-based estimates, some investigators have performed a third test with a sample of the typically much more numerous culture-negative, NAA-test-negative specimens (5, 9). While this may provide some assurance that few infected persons remain misclassified, it is not clear whether a patient's classification should be changed when the result of a test used primarily as a tiebreaker contradicts the results of both the reference test and the test under evaluation. However, study designs in which three or more tests are performed with all specimens allow the application of more sophisticated statistical techniques that do not require a strict classification of persons as infected or uninfected (14, 19, 20). Such designs offer a better, albeit expensive, solution to the problem of imperfect reference tests. Ultimately, the ideal method for evaluating new tests may involve both a more accurate reference test and a comparative evaluation of multiple tests all performed with the same specimens.

APPENDIX

Derivation of bias formulas. Let C+, L+, and M+ denote a positive result by culture, LCR, and MOMP testing, respectively, and C-, L-, and M- denote the corresponding negative test results. Let S+ and S- denote whether a person is infected or uninfected, respectively, and let D+ and D- denote whether a person is classified as infected or uninfected by discrepant analysis, respectively. Let P(A|B) denote the conditional probability of the occurrence of A given the occurrence of B. Using elementary rules of probability, we derived the following expressions for the expected values of the culture-based and discrepant-analysis-based estimates of LCR sensitivity and specificity.

The culture-based estimate of LCR sensitivity was calculated as follows:

$$\begin{aligned} P(L+|C+) &= P(C+,L+)/P(C+) \\ &= [P(C+,L+|S+)P(S+) + P(C+,L+|S-)P(S-)] \\ &\quad \div [P(C+|S+)P(S+) + P(C+|S-)P(S-)] \\ &= [P(L+|S+,C+)P(C+|S+)P(S+) \\ &\quad + P(L+|S-,C+)P(C+|S-)P(S-)] \\ &\quad \div [P(C+|S+)P(S+) + P(C+|S-)P(S-)] \end{aligned}$$

The discrepant-analysis-based estimate of LCR sensitivity was calculated as follows:

$$\begin{aligned} P(L+|D+) &= P(L+,D+)/P(D+) \\ &= [P(C+,L+) + P(C-,L+,M+)] \\ &\quad [P(C+) + P(C-,L+,M+)] \\ &= \{[P(C+,L+|S+) \\ &\quad + P(C-,L+,M+|S+)]P(S+) \\ &\quad + [P(C+,L+|S-) \\ &\quad + P(C-,L+,M+|S-)]P(S-)\} \end{aligned}$$

$$\begin{aligned} &\div \{[P(C+|S+) + P(C-,L+,M+|S+)]P(S+) \\ &\quad + [P(C+|S-) + P(C-,L+,M+|S-)]P(S-)\} \\ &= \{[P(L+|S+,C+)P(C+|S+) \\ &\quad + P(M+|S+,C-,L+)P(L+|S+,C-)P(C-|S+)]P(S+) \\ &\quad + [P(L+|S-,C+)P(C+|S-) \\ &\quad + P(M+|S-,C-,L+)P(L+|S-,C-)P(C-|S-)]P(S-)\} \\ &\quad \div \{[P(C+|S+) + P(M+|S+,C-,L+) \\ &\quad \times P(L+|S+,C-)P(C-|S+)]P(S+) \\ &\quad + [P(C+|S-) + P(M+|S-,C-,L+) \\ &\quad \times P(L+|S-,C-)P(C-|S-)]P(S-)\} \end{aligned}$$

The culture-based estimate of LCR specificity was calculated as follows:

$$\begin{aligned} P(L-|C-) &= P(C-,L-)/P(C-) \\ &= [P(C-,L-|S-)P(S-) + P(C-,L-|S+)P(S+)] \\ &\quad \div [P(C-|S-)P(S-) + P(C-|S+)P(S+)] \\ &= [P(L-|S-,C-)P(C-|S-)P(S-) \\ &\quad + P(L-|S+,C-)P(C-|S+)P(S+)] \\ &\quad \div [P(C-|S-)P(S-) + P(C-|S+)P(S+)] \end{aligned}$$

The discrepant-analysis-based estimate of LCR specificity was calculated as follows:

$$\begin{aligned} P(L-|D-) &= P(L-,D-)/P(D-) \\ &= P(C-,L-)/[P(C-,L-) + P(C-,L+,M-)] \\ &= [P(C-,L-|S-)P(S-) + P(C-,L-|S+)P(S+)] \\ &\quad \div \{[P(C-,L-|S-) + P(C-,L+,M-|S-)]P(S-) \\ &\quad + [P(C-,L-|S+) + P(C-,L+,M-|S+)]P(S+)\} \\ &= [P(L-|S-,C-)P(C-|S-)P(S-) \\ &\quad + P(L-|S+,C-)P(C-|S+)P(S+)] \\ &\quad \div \{[P(L-|S-,C-) + P(M-|S-,C-,L+) \\ &\quad \times P(L+|S-,C-)]P(C-|S-)P(S-) \\ &\quad + [P(L-|S+,C-) + P(M-|S+,C-,L+) \\ &\quad \times P(L+|S+,C-)]P(C-|S+)P(S+)\} \end{aligned}$$

For each bias evaluation, the actual value of the test performance characteristic being estimated was subtracted from the expected value computed by using the appropriate expression.

REFERENCES

1. Bassiri, M., H. Y. Hu, M. A. Domeika, J. Burczak, L.-O. Svensson, H. H. Lee, and P.-A. Mårdh. 1995. Detection of *Chlamydia trachomatis* in urine specimens from women by ligase chain reaction. *J. Clin. Microbiol.* **33**:898-900.
2. Bauwens, J. E., A. M. Clark, and W. E. Stamm. 1993. Diagnosis of *Chlamydia trachomatis* endocervical infections by a commercial polymerase chain reaction assay. *J. Clin. Microbiol.* **31**:3023-3027.
3. Black, C. M. 1997. Current methods of laboratory diagnosis of *Chlamydia trachomatis* infections. *Clin. Microbiol. Rev.* **10**:160-184.
4. 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. Clin. Microbiol.* **32**:2682-2685.
5. 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.
6. **Dille, B. J., C. C. Butzen, and L. G. Birkenmeyer.** 1993. Amplification of *Chlamydia trachomatis* DNA by ligase chain reaction. *J. Clin. Microbiol.* **31**:729–731.
 7. **Hadgu, A.** 1996. The discrepancy in discrepant analysis. *Lancet* **348**:592–593.
 8. **Hadgu, A.** 1997. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. *Stat. Med.* **16**:1391–1399.
 9. **Lee, H. H., M. A. Chernesky, J. Schachter, J. D. Burczak, W. W. Andrews, S. Muldoon, G. Leckie, and W. E. Stamm.** 1995. Diagnosis of *Chlamydia trachomatis* genitourinary infection in women by ligase chain reaction assay of urine. *Lancet* **345**:213–216.
 10. **Lin, J.-S. L., W. E. Jones, L. Yan, K. A. Wirthwein, E. E. Flaherty, R. M. Haivanis, and P. A. Rice.** 1992. Underdiagnosis of *Chlamydia trachomatis* infection. Diagnostic limitations in patients with low-level infection. *Sex. Transm. Dis.* **19**:259–265.
 11. **Loeffelholz, M. J., C. A. Lewinski, S. R. Silver, A. P. Purohit, S. A. Herman, D. A. Buonagurio, and E. A. Dragon.** 1992. Detection of *Chlamydia trachomatis* in endocervical specimens by polymerase chain reaction. *J. Clin. Microbiol.* **30**:2847–2851.
 12. **Magder, L. S., K. C. Klontz, L. H. Bush, and R. C. Barnes.** 1990. Effect of patient characteristics on performance of an enzyme immunoassay for detecting cervical *Chlamydia trachomatis* infection. *J. Clin. Microbiol.* **28**:781–784.
 13. **Mahony, J. B., K. E. Luinstra, J. W. Sellors, L. Pickard, S. Chong, D. Jang, and M. A. Chernesky.** 1994. Role of confirmatory PCRs in determining performance of Chlamydia Amplicor PCR with endocervical specimens from women with a low prevalence of infection. *J. Clin. Microbiol.* **32**:2490–2493.
 14. **Qu, Y., M. Tan, and M. H. Kutner.** 1996. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* **52**:797–810.
 15. **Schachter, J.** 1984. Biology of *Chlamydia trachomatis*, p. 243–257. In K. K. Holmes, P.-A. Mårdh, P. F. Sparling, and P. J. Wiesner (ed.), *Sexually transmitted diseases*. McGraw-Hill Book Co., New York, N.Y.
 16. **Schachter, J.** 1986. Rapid diagnosis of sexually transmitted diseases—speed has a price. *Diagn. Microbiol. Infect. Dis.* **4**:185–189.
 17. **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.
 18. **Skulnick, 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.
 19. **Walter, S. D., and L. M. Irwig.** 1988. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol.* **41**:923–937.
 20. **Yang, L., and M. P. Becker.** 1997. Latent variable modeling of diagnostic accuracy. *Biometrics* **53**:948–958.