JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2004, p. 4749–4758 Vol. 42, No. 10
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.10.4749–4758.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Use of Multiple Nucleic Acid Amplification Tests To Define the Infected-Patient "Gold Standard" in Clinical Trials of New Diagnostic Tests for *Chlamydia trachomatis* Infections

David H. Martin,[1]* Malanda Nsuami,[1] Julius Schachter,[2] Edward W. Hook III,[3]
Dennis Ferrero,[4] Thomas C. Quinn,[5,6] and Charlotte Gaydos[5]

Health Sciences Center, Louisiana State University, New Orleans, Louisiana[1]; University of San Francisco, San Francisco,[2] and San Joaquin County Public Health Department, San Joaquin,[4] California; University of Alabama at Birmingham, Birmingham, Alabama[3]; and Johns Hopkins University, Baltimore,[5] and National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda,[6] Maryland

Received 11 October 2003/Returned for modification 7 January 2004/Accepted 5 July 2004

Nucleic acid amplification tests (NAATs) can be used to define the infected-patient "gold standard" for the purpose of designing studies of the performance of *Chlamydia trachomatis* diagnostic tests. It is unclear how many test results run by different NAATs and what combinations of specimens comprise the best infected-patient gold standard. We approached this question with data from a large study of the performance of a new NAAT. Data were available from three endocervical swabs and a urine specimen collected from each of 1,412 women and tested by three different NAATs. Results from all three assays were used equally in a rotating fashion to define the infected-patient gold standard. Multiple different infected-patient gold standards for estimating swab and urine specimen sensitivity and specificity for one NAAT method were created by varying the number and combinations of swab and urine comparator results with two different NAATs, The effect of changing the infected-patient gold standard definition was determined by constructing receiver-operator-like curves with calculated sensitivities and specificities for each test. The one-positive-of-two-results or two-positive-of-two-results (same or two different assays) infected-patient gold standard definitions produced low sensitivity and low specificity estimates, respectively. If four comparator NAAT results were used, the any-three-positive-of-four-results definition or the at-least-one-specimen-positive-by-each-of-two-comparator-assays definition appeared to provide better combinations of sensitivity and specificity estimates. The any-two-positive-out-of-three-results definition resulted in estimates that were as good as produced with the former two definitions. This analytic approach provides a means of clearly visualizing the effects of changing NAAT-based infected-patient gold standards and should be helpful in designing future studies of new *C. trachomatis* diagnostic tests.

Determination of the infected patient "gold standard" for measuring the performance of new nucleic acid amplification tests (NAATs) for the detection of *Chlamydia trachomatis* in genitourinary specimens has proven difficult. Historically, culture was the gold standard for determining the performance characteristics of new diagnostic tests for this organism. Based on the analytical sensitivity of NAATs in addition to the fact that it was known that culture was not optimally sensitive, it was reasonably clear initially that these assays would be more sensitive than *C. trachomatis* culture. Indeed, studies showed that the use of culture alone as a reference standard resulted in significant underestimates of the specificity for NAATs as many infected patients were considered falsely negative by culture (7, 13).

To address this problem, investigators applied an alternative target amplification assay to the putative false-positive results (discrepancy analysis). There was no alternative at the time, as tests with similar sensitivity were not available for use in a composite infected-patient definition along with culture. This approach later was shown to be biased towards overestimating both sensitivity and specificity (6). Though the extent of bias appeared to be minimal, the ensuing controversy resulted in a lack of acceptance of this approach for determining the performance characteristics of new NAATs (5, 10, 11, 12).

Now that multiple NAATs are available and cleared for clinical use by the Food and Drug Administration, it is possible to design protocols to assess the performance of newer NAATs for the detection of *C. trachomatis* and *Neisseria gonorrhoeae* without the use of culture. Johnson et al. have shown that a single NAAT substituted for culture significantly improved performance estimates of another NAAT (8). In this study the combined results of two NAATs were used to estimate the performance of a third NAAT. It is now recognized that the problem with this approach is that variation in the sensitivity and specificity of the comparator NAATs could significantly influence the performance estimates of the other test.

Recently a multicenter trial was carried out to determine the performance of the APTIMA Combo 2 (Combo 2) transcription-mediated amplification assay (Gen-Probe Incorporated, San Diego, Calif.) for detection of *C. trachomatis* and *N. gonorrhoeae* in endocervical swabs, male urethral swabs, and urine from both men and women (3). Both the Abbott LCx ligase chain reaction (Abbott Laboratories Inc., Abbott Park, Ill.)

* Corresponding author. Mailing address: LSU Health Sciences Center, Department of Medicine, Section of Infectious Diseases, 1542 Tulane Ave., New Orleans LA 70112. Phone: (504) 568-5031. Fax: (504) 568-6752. E-mail: dhmartin@lsuhsc.edu.

TABLE 1. Definitions

Type	Definition
Definitions using four comparator assay results	a: All four comparator results had to be positive.
	b: At least three of the four results tested had to be positive.
	c: At least one result (swab or urine) from each comparator assay had to be positive.
	d: At least two of the four comparator results had to be positive.
	e: At least one of the four comparator results had to be positive.
Definitions using three comparator assay results	f: All three comparator results had to be positive.
	g: At least one result from by each comparator assay had to be positive.
	h: At least two of the three comparator results had to be positive.[a]
	i: At least one of the three comparator results had to be positive.
Definitions using two comparator results, from a different assay	j: Both comparator results had to be positive.
	k: At least one of the two results had to be positive.
Definitions using two comparator results, both from the same assay	l: Both comparators results had to be positive.
	m: At least one of the two results had to be positive.

[a] As is demonstrated in figures 8 and 9, optimal estimates were obtained by using two urine and one swab comparator for the evaluation of urine specimens and two swabs and one urine comparator for swab specimens when three test results were used to comprise the comparator. The optimal comparator combinations were used for definition h in Fig. 1 to 7.

and the Roche Amplicor PCR (Roche Diagnostic Systems, Indianapolis, Ind.) assays were used to devise a comparator standard for *C. trachomatis* that did not include culture. Women were defined as infected if any two of four comparator test results (endocervical swab or urine by PCR or ligase chain reaction) were positive. For men, a PCR urethral swab was not obtained, so the definition for infected status was if any two of the three comparator test results were positive. While these definitions seemed rational, they were not evidence based.

Since these data were derived from a large trial in which

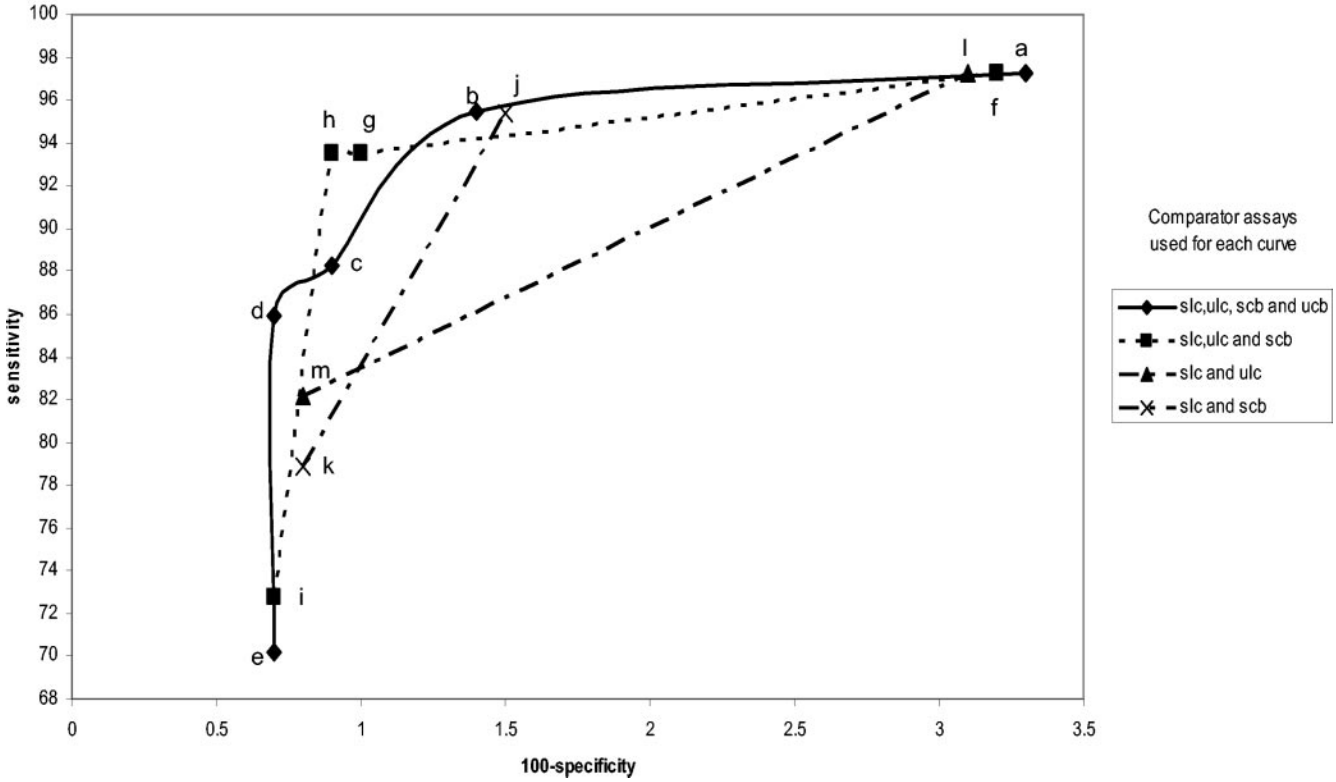


FIG. 1. Amplicor female swab specimen performance curves. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

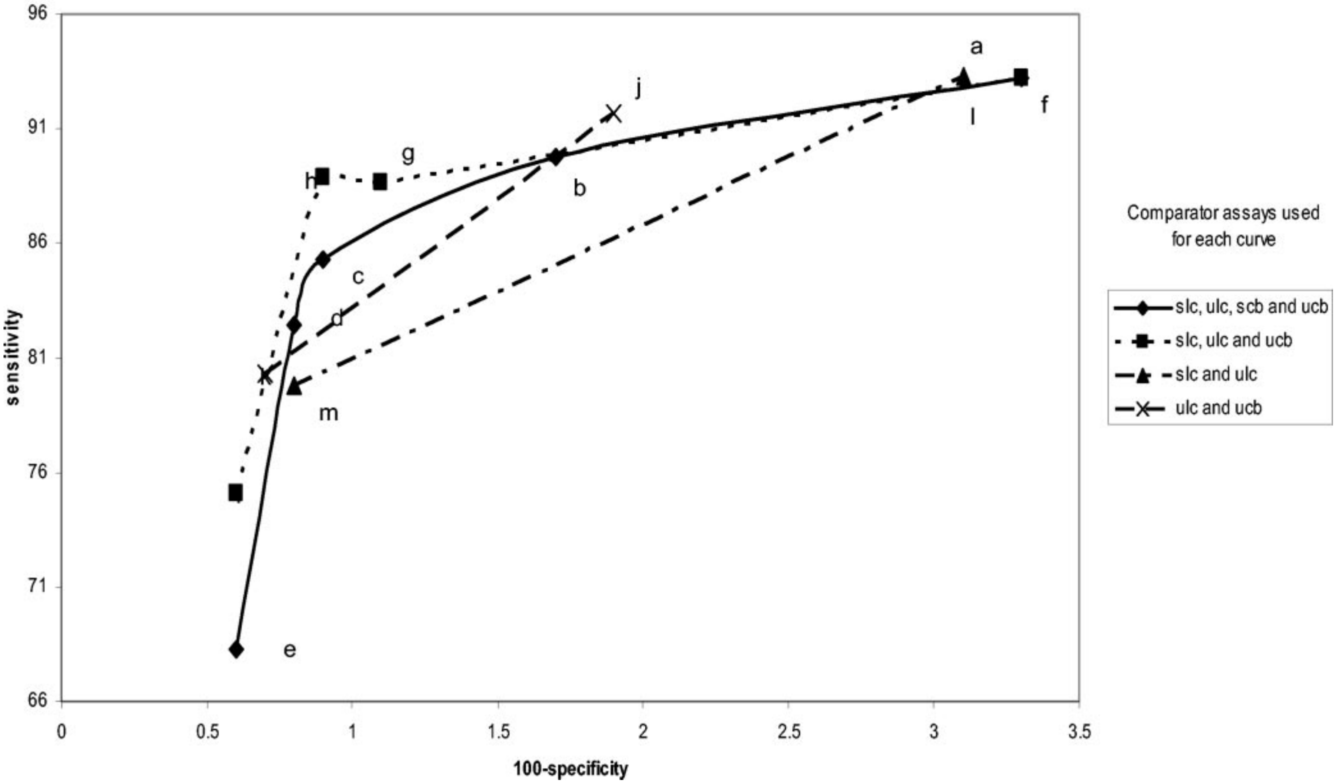


FIG. 2. Amplicor female urine specimen performance curves. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

patients were tested with three different NAATs and two different specimens were tested in most cases, they provided a unique opportunity to look at the effect of varying the infected-patient definition on the performance estimates of a third NAAT. Therefore, we performed an analysis to better understand the use of NAATs as the infected-patient gold standard for measuring the performance of new *C. trachomatis* diagnostic assays. Additionally, the data provided an opportunity for a head-to-head comparison of the performance of Combo 2 with both Amplicor and LCx.

TABLE 2. Comparison of sensitivity and specificity of the Combo 2 and LCx assays for female urine and swab specimens[a]

Specimen and assay	Sensitivity (%)	Specificity (%)
Female endocervical swab		
Combo2	86.7	97.3
LCx	81.9	99.1
Female urine		
Combo2	88.1	98.6
LCx	80.5	99.1

[a] Sensitivity and specificity were calculated with Amplicor endocervical swab and urine results. The infected-patient standard was defined as at least one of these two tests positive.

MATERIALS AND METHODS

Urine and urogenital swab specimens were collected from male and female patients at seven clinical sites in the United States, including sexually transmitted disease clinics and family planning clinics (3). Specimens were excluded if the patient had urinated within 1 h before providing the specimen or had taken antibiotics within the previous 21 days or if collection, storage, or transport requirements were not met.

Male and female patients provided 25 ml of a first-catch urine. Three urethral

TABLE 3. Comparison of sensitivity and specificity of the Combo 2 and Amplicor assays for female and male urine specimens and female swab specimens

Specimen and assay	Sensitivity (%)	Specificity (%)
Female endocervical swab		
Combo2	86.9	97.8
Amplicor	82.2	99.2
Female urine		
Combo2	87.3	98.7
Amplicor	79.8	99.2
Male urine		
Combo2	94.8	98.2
Amplicor	87.6	99.5

[a] Sensitivity and specificity were calculated with LCx endocervical swab and urine results. The infected-patient standard was defined as at least one of these two tests positive.

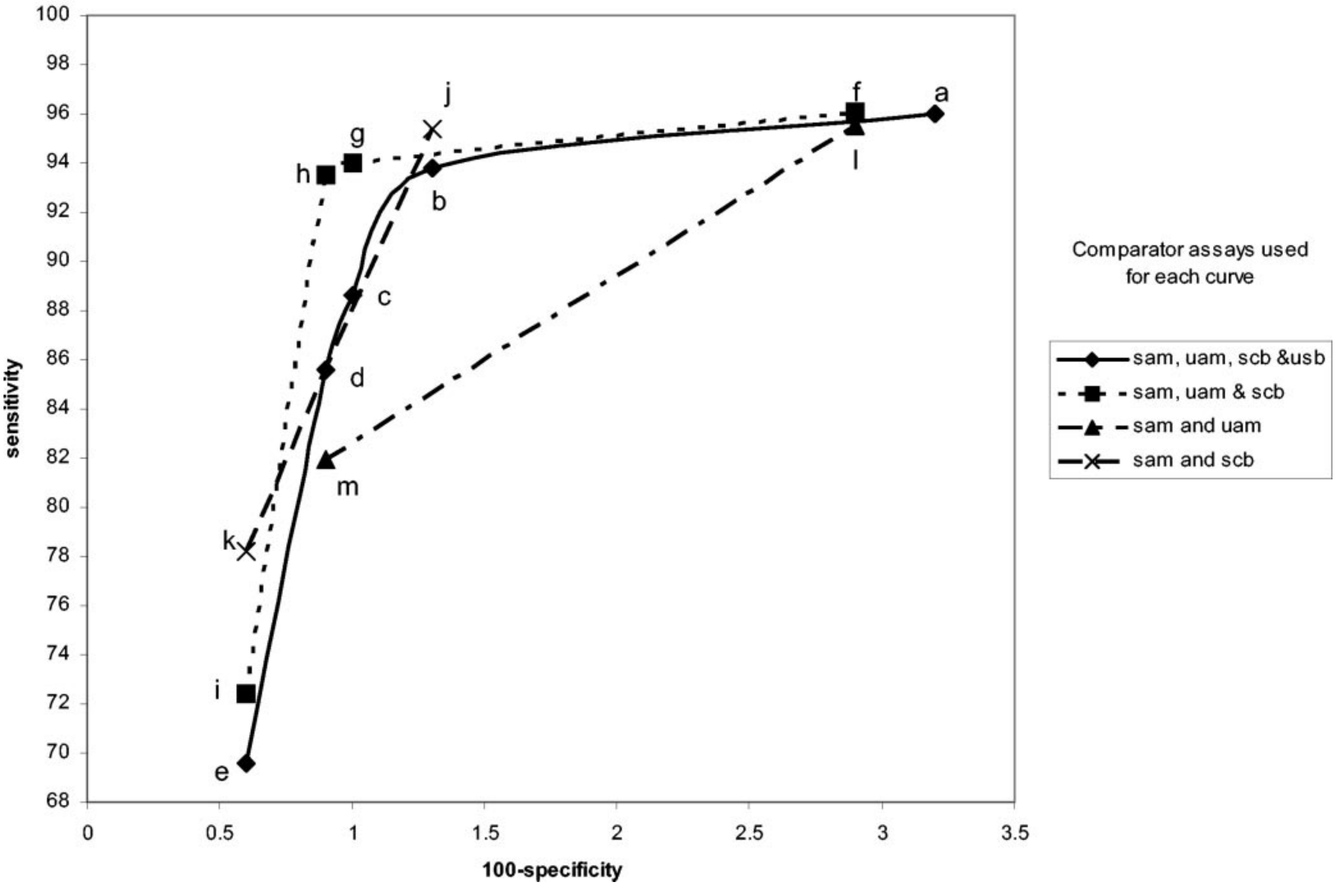


FIG. 3. LCx female swab specimen performance curves. Each figure represents the evaluation of a single specimen by LCx, Amplicor, or Combo 2. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

swab specimens were obtained from males for the following assays: *N. gonorrhoeae* culture, Combo 2, and LCx. Swab specimens were collected before the first-catch urine specimen. Females provided four endocervical swab specimens for one *N. gonorrhoeae* culture and all three NAAT assays (Combo 2, LCx, and Amplicor). For women, the first-catch urine specimen was collected before the swab specimens. For men and women, the *N. gonorrhoeae* culture swab was collected first, and the collection order of the subsequent swabs was randomized.

Collection, storage, and transport of the GC culture swab followed site-specific protocols. All other specimens were collected, stored, and transported to the laboratory according to each assay manufacturer's instructions. Male swabs were not collected for Amplicor testing. Only the chlamydia data are analyzed here.

Main analysis. The effect of reducing the available NAATs used to define the infected patient from four tests to three tests to two tests was explored. The details of the definitions used are provided in Table 1. With these definitions, curves were constructed by plotting sensitivity on the *y* axis against 1 − specificity on the *x* axis. It should be noted that the resultant curves resemble receiver-operator curves but they are distinct. Receiver-operator curves are based on a single gold standard test and display the effect of changing the definition of positive for an evaluated test. What we have done in this study is different. Here, for each family of curves, we are using a single "evaluated" test which has a predetermined definition of positive to assess multiple different gold standard definitions. In a sense, our analysis is the opposite of a receiver-operator curve analysis. The importance of this distinction is the fact that the points on our curves (each representing a different infected-patient gold standard) are not equally accurate and that one of the points is likely to be more accurate than the others. However, it should be noted that it may not be possible to determine from our analysis which definition truly is the best.

Swab and urine specimens tested by two different assays were used to assess the performance of the swab and urine specimens tested by a third assay. This was done in a rotating fashion in order to generate a family of curves for each swab and urine specimen. For women, the analysis resulted in six families of curves, four for each specimen type, as shown in Fig. 1 to 6. A complete family of curves could be generated from the male urine data only, since an Amplicor swab was not obtained (Fig. 7). Multiple families of curves were generated to ensure consistency of the observations.

Comparison of transcription-mediated amplification assay (Combo 2) to PCR (Amplicor) and ligase chain reaction (LCx). Combo 2 was compared to Amplicor with the LCx swab and urine results. This comparison could be done for female urine and cervical swabs and male urine. Combo 2 was also compared to LCx with Amplicor cervical swab and urine results for females.

RESULTS

A total of 2,932 patients were enrolled. Of those enrolled, 2,457 (84%) subjects had a complete set of *C. trachomatis* tests. Of these 1,412 were women and 1,045 were men. The results from these cases were the subject of these analyses.

Figures 1, 2, 3, 4, 5, and 6 show the families of curves generated by calculating sensitivity and specificity for each assay with a decreasing number of comparator assay results and/or different combinations of specimens to define the infected female patient. Each point on each curve represents a different definition of the infected patient (Table 1). There

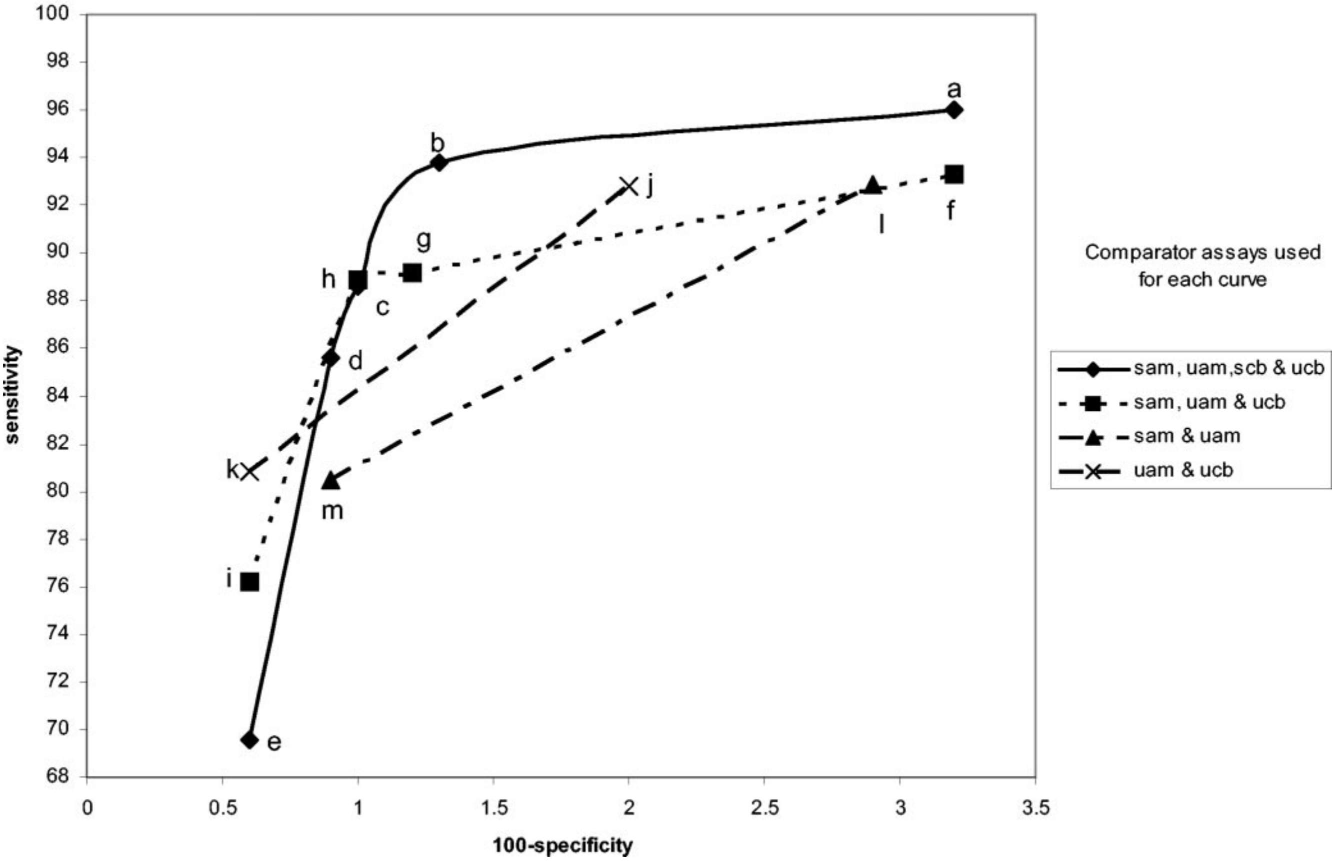


FIG. 4. LCx female urine specimen performance curves. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

appear to be only four points on the four comparator curves for Combo 2 because the sensitivity and specificity for definitions c and d are exactly the same. Only one family of curves could be generated for males (Amplicor urine evaluation), since only two urethral swabs were obtained (Fig. 7). The results of this analysis closely matched those derived from the female data.

Comparisons of these curves revealed several facts. Curves based on three comparators closely paralleled curves based on four comparators. Requiring that all comparators be positive regardless of the number of comparators used in the definition maximized sensitivity estimates but resulted in lower specificity estimates. Requiring that three of four results or two of three results be positive provided higher specificity estimates without lowering the sensitivity estimates significantly. The one-positive-of-four-results, the one-positive-of-three-results, and the one-positive-of-two-results definitions (definitions e, i, k, and m, respectively) maximized specificity estimates but resulted in low sensitivity estimates. The two-positives-of-four-results definition and the requirement that the two positive results each come from two different NAATs (definition c) or the any-three-positives-of-four-results definition (definition b) provided higher combined estimates of sensitivity and specificity, as the points for these definitions appeared to be closest to the ideal of 100% sensitivity and 100% specificity.

Interestingly, the any-two-positive-of-three-results definition (definition h) appears to perform as well as definitions b and c, which require four comparator assay results. However, there are two different ways of creating the infected-patient definitions with three comparators; one swab and two urine comparators or two swabs and one urine comparator could have been used. As can be seen in Fig. 8 and 9, this choice does have an effect on the sensitivity and specificity estimates. The highest combined estimates of swab sensitivity and specificity are derived by using two swab specimens and one urine specimen as comparators. Similarly, the highest combined estimates of urine performance are provided by two urine specimens and one swab specimen as comparators.

Inspection of the curves suggested that Combo 2 may be more sensitive for the detection of chlamydiae and less specific than either Amplicor or LCx, while the latter two assays appear to perform very similarly. Since these curves were generated with all of the available data to create infected-patient definitions, such conclusions may not be valid, as all the assays were being used as components of the infected-patient definitions for each other. Table 2 shows the results of a head-to-head

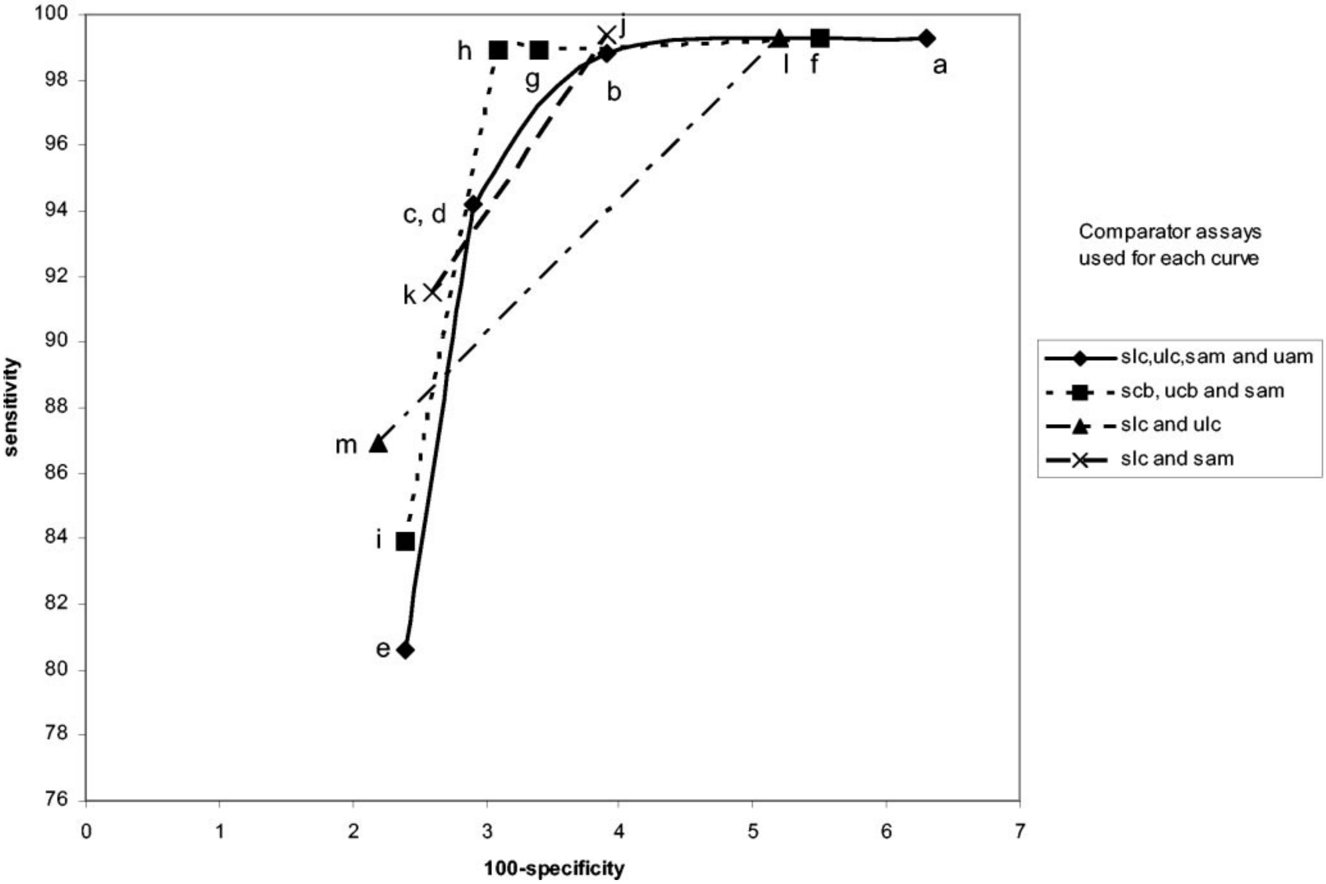


FIG. 5. Combo 2 female swab specimen performance curves. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

comparison of Combo 2 to LCx with the Amplicor swab and urine results and infected-patient gold standard definition m as the comparator. Table 3 shows the results of a head-to-head comparison of Combo 2 to Amplicor with the LCx swab and urine results and infected-patient gold standard definition m as the comparator. These head-to-head comparisons confirmed that Combo 2 has greater sensitivity than both LCx and the Amplicor assays but lower specificity.

DISCUSSION

Initially NAAT performance characteristics were determined with chlamydial cultures and direct fluorescent microscopy of urine and swab transport medium sediments to resolve discrepancies (1, 4). The first studies of both the PCR and ligase chain reaction assays utilized alternative target amplification assays to resolve discrepant results (1). From a statistical point of view this approach was flawed, though the biases introduced were very small (5). More recently, assessment of newer NAATs used previously cleared NAAT assays to define the infected-patient gold standard (2, 8, 14). Now that several of these assays have been cleared by the Food and Drug Administration, it is possible to eliminate culture as a part of the infected-patient gold standard altogether (3). However, it is not clear how many different assays should be used and which combination of specimen types should constitute the infected patient definition.

It is known that in some infected women, *C. trachomatis* can be found only in the endocervix, while in others it can be detected only in the urine specimen (9). Therefore, exclusive use of multiple swab specimens or multiple urine specimens could significantly bias performance estimates of a new test. The dilemma of what does constitute the best definition of a NAAT-based infected-patient gold standard then arises. If it is necessary to use both a swab and urine specimen to define the infected-patient gold standard, is a single Food and Drug Administration-cleared NAAT adequate for these tests? If so, should it be required that both tests be positive, or is only one positive of the two adequate? If two specimens tested by only one NAAT is inadequate and more than one assay is to be used to define the infected-patient gold standard, is it necessary to test both urine and swab specimens by both assays? For males, the more urethral swabs required by a clinical protocol, the more difficult it is to recruit study subjects. Could an adequate male infected-patient gold standard be created by testing a urine sample by two different Food and Drug Administration-

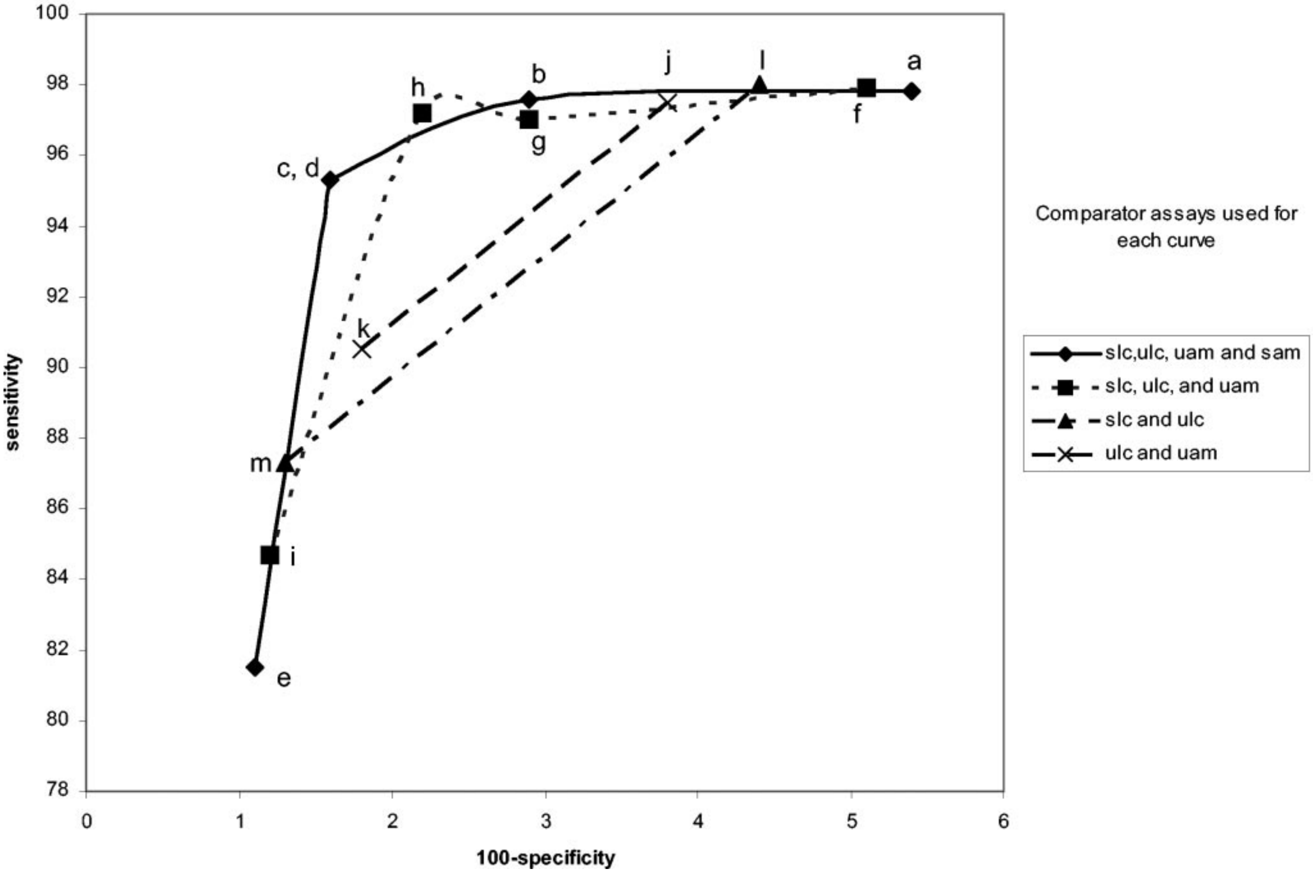


FIG. 6. Combo 2 female urine specimen performance curves. Each figure represents the evaluation of a single specimen by LCx, Amplicor or Combo 2. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.

cleared NAATs plus a single urethral swab specimen tested by only one of the methods?

In this study we attempted to answer these questions by examining the effect of varying the number of comparator assays and specimen types used to define the infected patient. These results are summarized in Fig. 1 through 7. Theoretically, the more points available to construct such curves, the more reliable the results. Based on this consideration, the curves generated with four available comparator results would be considered the standard for comparison with curves that are constructed with fewer comparators. As can be seen from the figures, curves with only three comparator results closely approximated the curves with four comparators results. On the other hand, using only two comparators results to define the infected patient does not appear to be adequate. Requiring that two of two assays be positive (definitions j and i) biases results towards low specificity, while requiring that only one of two be positive (definitions k and m) has the opposite effect.

If three or four comparator results are used to formulate the infected-patient gold standard definition, the effect of requiring that all comparator results be positive (definitions a and f) biases performance estimates towards high sensitivity and low specificity. Requiring only one test to be positive (definitions e and i) has the opposite effect. An ideal infected-patient gold standard would result in estimates of 100% sensitivity and 100% specificity for a perfect *C. trachomatis* test. It follows that infected-patient gold standard definitions resulting in estimates that are nearest to the ideal might be the most accurate. With four comparator results, the points on the curves defined by infected-patient gold standard definitions b and c appear to be closest to meeting this criterion. Using three comparator results and defining the infected patient as any two positives of the three possible results (definition h) appear to provide estimates for both the sensitivity and specificity between those of definitions b and c. A three-component infected-patient gold standard would be less costly than a four-component infected-patient gold standard.

If three comparator amplification assay results are adequate for defining the infected patient, is there a difference if two urine results and one swab result are used as opposed to one urine and two swab results? The curves in Fig. 8 and 9 suggest that there is. If swab specimens are being evaluated, using two swabs and one urine specimen as the comparators will result in higher combined sensitivity and specificity estimates than one swab and two urine comparators. Similarly, for evaluating urine specimens, two urine specimens and one swab specimen

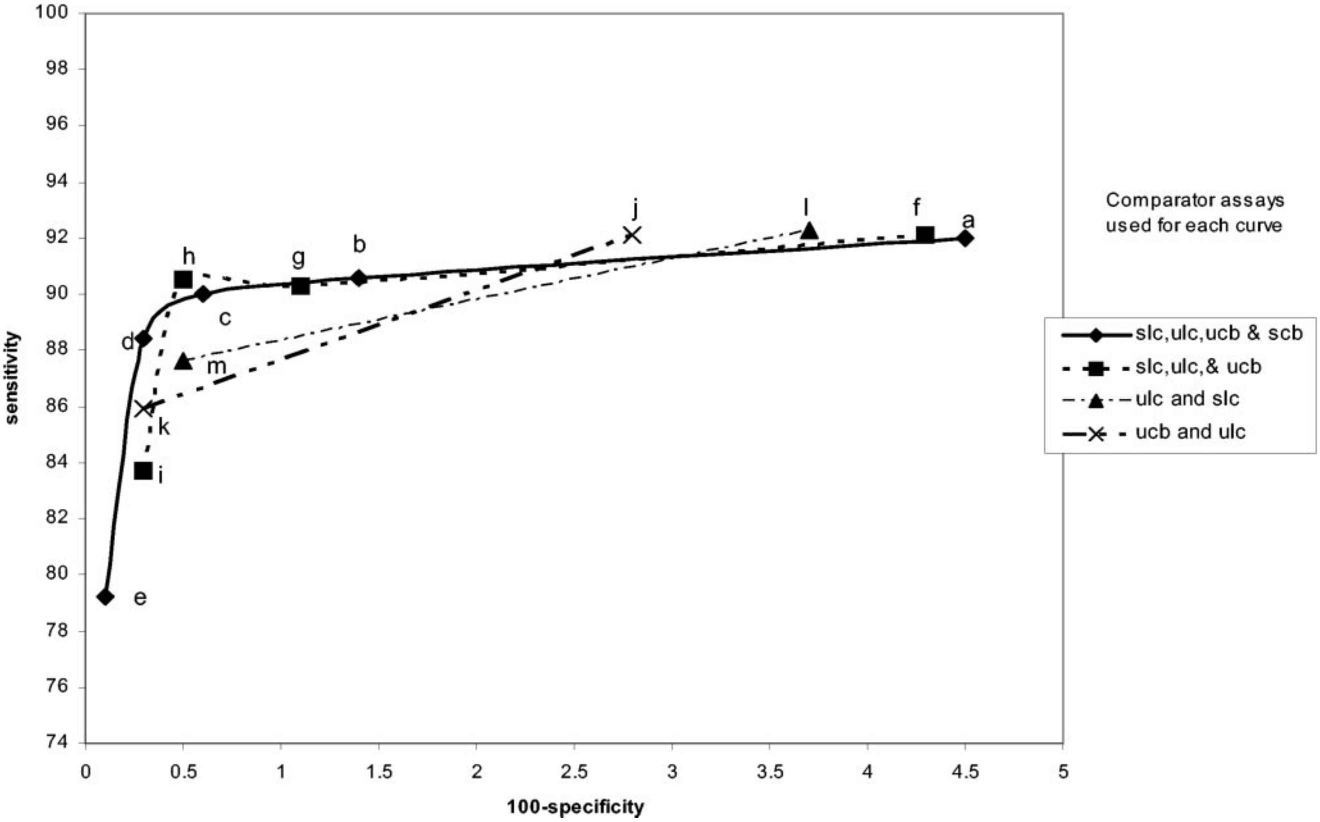
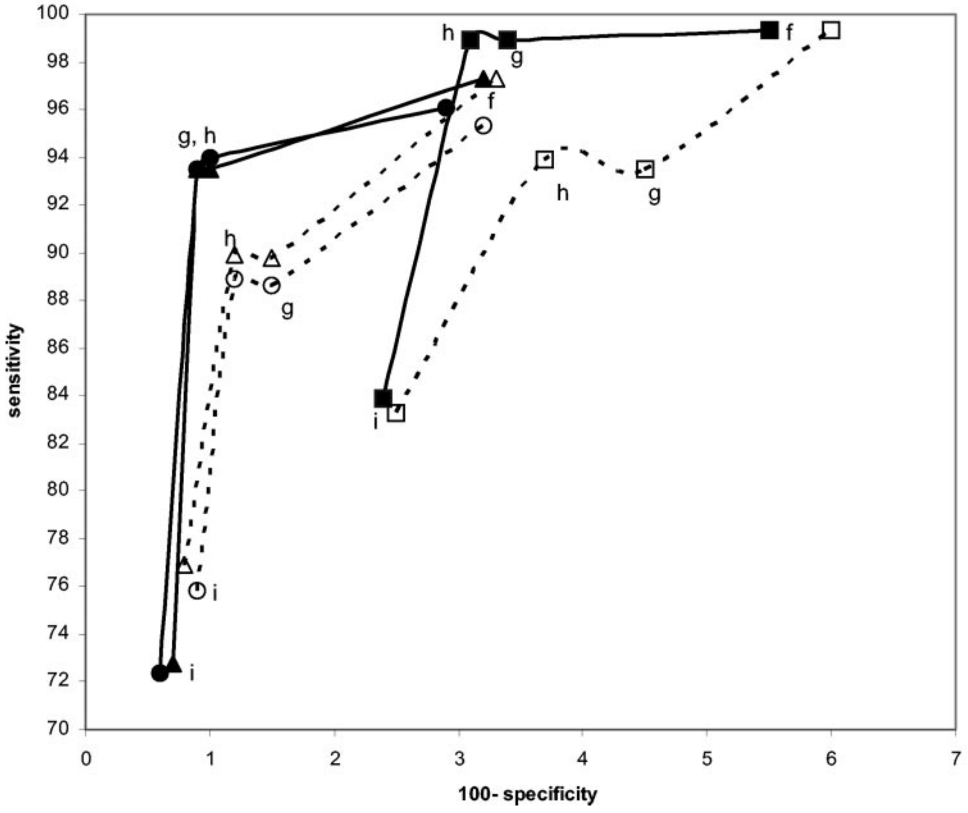


FIG. 7. Amplicor male urine specimen performance curves. The legend to the right shows which specimens and assays were used as comparators for each curve. sam, swab by Amplicor; uam, urine by Amplicor; slc, swab by LCx; ulc, urine by LCx; scb, swab by Combo 2; ucb, urine by Combo 2. The individual points in each curve were determined as described in Table 1. The letters refer to the specific infected-patient definitions used to calculate each point on the curves as detailed in Table 1.



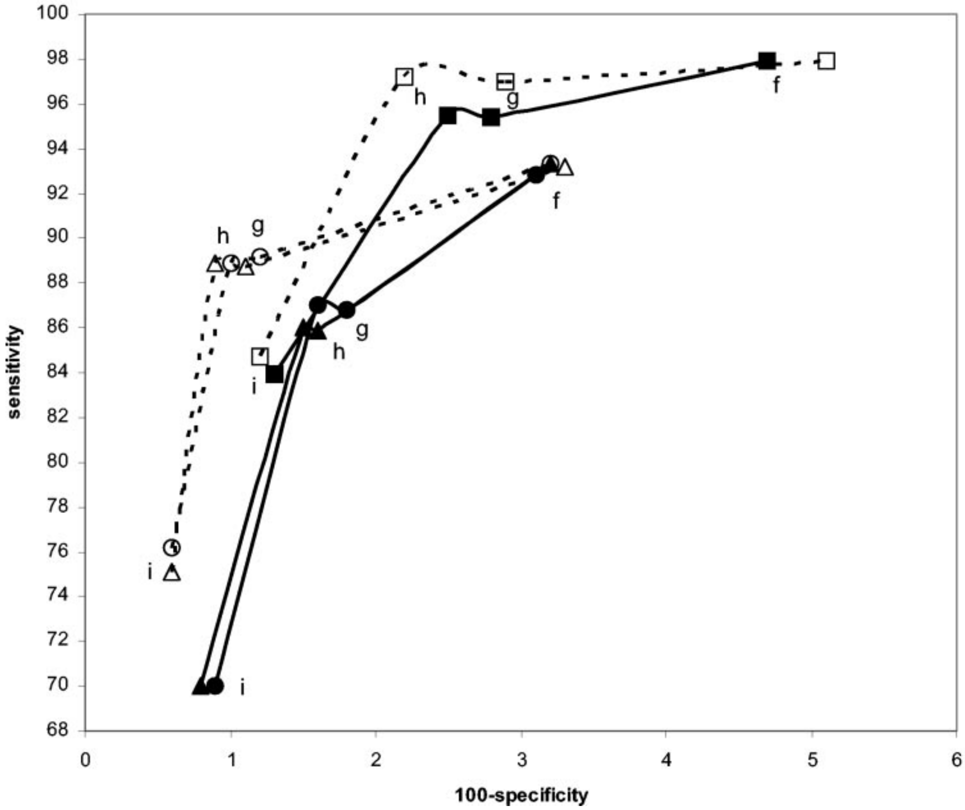


FIG. 9. Effect on performance estimates for urine specimens of varying the definition of the infected patient by changing the mix of urine and swab specimens used as comparators. All curves were constructed with the three-comparator definitions f, g, h, and i (see Table 1). Solid lines represent data based on two swabs and one urine comparator. Dashed lines represent data based on one swab and two urine comparators. Solid square, swab by Combo 2 versus swab by LCx, urine by LCx, and swab by Amplicor; open square, swab by Combo 2 versus swab by LCx, urine by LCx, and urine by Amplicor; solid circle, swab by LCx versus swab by Amplicor, urine by Amplicor, and swab by Combo 2; open circle, swab by LCx versus swab by Amplicor, urine by Amplicor, and urine by Combo 2; solid triangle, swab by Amplicor versus swab by LCx, urine by LCx, and swab by Combo 2; open triangle, swab by Amplicor versus swab by LCx, urine by LCx, and urine by Combo 2.

as the comparators result in higher combined sensitivity and specificity results than the converse.

Based on these observations, we recommend the following approach to evaluation of new diagnostic tests for *C. trachomatis* in women. Since multiple endocervical swabs are not difficult to obtain, a swab specimen result for the evaluated assay in women can be compared to one urine and two swab results with two different Food and Drug Administration-cleared NAATs. The evaluated test's urine result would be compared to one swab and two urine results. Any two positive results out of the possible three comparator results would define the infected-patient gold standard (definition h). Based on our analysis, this algorithm appears to provide estimates for a new diagnostic test's performance with both female swab and urine specimens that are as good as or better than those of any other combination of assays and specimens.

For male studies the infected-patient gold standard could also be defined by three comparators, including a swab and urine run by one Food and Drug Administration-cleared NAAT and urine run by another. This strategy would result in optimal performance estimates for new urine tests. However, our data indicate that the swab performance estimates with this approach will be slightly lower than they would be if one urine and two swab specimens were used to formulate the infected male patient definition. This is a reasonable trade-off given the fact that it is difficult to obtain more than two urethral swab specimens from men for the purposes of a clinical trial.

Of course these recommendations are not based on a rigor-

FIG. 8. Effect on performance estimates for swab specimens of varying the definition of the infected patient by changing the mix of urine and swab specimens used as comparators. All curves were constructed with the three-comparator definitions f, g, h, and i (see Table 1). Solid lines represent data based on two swabs and one urine comparator. Dashed lines represent data based on one swab and two urine comparators. Solid square, swab by Combo 2 versus swab by LCx, urine by LCx, and swab by Amplicor; open square, swab by Combo 2 versus swab by LCx, urine by LCx, and urine by Amplicor; solid circle, swab by LCx versus swab by Amplicor, urine by Amplicor, and swab by Combo 2; open circle, swab by LCx versus swab by Amplicor, urine by Amplicor, and urine by Combo 2; solid triangle, swab by Amplicor versus swab by LCx, urine by LCx, and swab by Combo 2; open triangle, swab by Amplicor versus swab by LCx, urine by LCx, and urine by Combo 2.

ous statistical analysis of the data. Given the novelty of our analysis, there do not appear to be well-established mathematical approaches to the data. It has been suggested that the latent class model approach could be applied, but this is relatively new and it is not clear that how it would be applied to our data or that the end result would lead to conclusions that would be any more acceptable than the opinions offered above. Our data are available to anyone with an interest in developing such analytic approaches. In the meantime it is our hope that the graphic presentation of the data shown here will enable anyone with an interest in developing new diagnostic tests for *C. trachomatis*, other sexually transmitted diseases, and possibly infectious diseases in general to gain a sense of how differences in NAAT-based infected-patient gold standard definitions affect sensitivity and specificity estimates.

Comparison of the performance curves for the Combo 2 assay to those for the Amplicor and LCx assays in Fig. 8 and 9 suggested that Combo 2 might be more sensitive but less specific than these other two tests. Direct comparisons of Combo 2 and LCx were done with the two Amplicor results. Similarly, Combo 2 was compared to the Amplicor assay with the LCx results. While the estimates derived from these comparisons suffer from bias towards lower sensitivity and higher specificity, as discussed above, relative performance comparisons between any two assays with a third assay remain valid. On this basis, the APTIMA Combo 2 does appear to be a more sensitive test. The lower specificity may reflect the greater sensitivity of the assay (the infected-patient definition is missing some truly infected cases) or could reflect more false-positive results. Testing of specimens that appeared false positive by Combo 2 in a transcription-mediated amplification assay that targets alternative nucleic acid sequences suggests that the former is the case (3). This suggests that the true specificity of APTIMA Combo 2 is higher than that reflected by the analyses shown here.

REFERENCES

1. **Black, C. M.** 1997. Current methods of laboratory diagnosis of *Chlamydia trachomatis* infections. Clin. Microbiol. Rev. **10:**160–184.
2. **Black, C. M., J. Marrazzo, R. E. Johnson, E. W. Hook III, R. B. Jones, T. A. Green, J. Schachter, W. E. Stamm, G. Bolan, M. E. St Louis, and D. H. Martin.** 2002. Head-to-head multicenter comparison of DNA probe and nucleic acid amplification tests for *Chlamydia trachomatis* infection in women performed with an improved reference standard. J. Clin. Microbiol. **40:**3757–3763.
3. **Gaydos, C. A., T. C. Quinn, D. Willis, A. Weissfeld, E. W. Hook, D. H. Martin, D. V. Ferrero, and J. Schachter.** 2003. Performance of the APTIMA Combo 2 assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in female urine and endocervical swab specimens. J. Clin. Microbiol. **41:**304–309.
4. **Gaydos, C. A., C. A. Reichart, J. M. Long, L. E. Welsh, T. M. Neumann, E. W. Hook III, and T. C. Quinn.** 1990. Evaluation of Syva enzyme immunoassay for detection of *Chlamydia trachomatis* in genital specimens. J. Clin. Microbiol. **28:**1541–1544.
5. **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.
6. **Hadgu, A.** 1997. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. Stat. Med. **16:**1391–1399.
7. **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.
8. **Johnson, R. E., 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.
9. **Jones, R. B., B. P. Katz, B. van der Pol B, V. A. Caine, B. E. Batteiger, and W. J. Newhall.** 1986. Effect of blind passage and multiple sampling on recovery of *Chlamydia trachomatis* from urogenital specimens. J. Clin. Microbiol. **24:**1029–1033.
10. **McAdam, A. J.** 2000. Discrepant analysis: how can we test a test? J. Clin. Microbiol. **38:**2027–2029.
11. **Miller, W. C.** 1998. Can we do better than discrepant analysis for new diagnostic tests evaluation? Clin. Infect. Dis. **27:**1186–1193.
12. **Schachter, J.** 1998. Two different worlds we live in. Clin. Infect. Dis. **27:** 1181–1185.
13. **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.
14. **Van Der Pol, B., D. V. Ferrero, L. Buck-Barrington, E. Hook III, C. Lenderman, T. Quinn, C. A. Gaydos, J. Lovchik, J. Schachter, J. Moncada, G. Hall, M. J. Tuohy, and R. B. Jones.** 2001. Multicenter evaluation of the BDProbeTec ET System for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in urine specimens, female endocervical swabs, and male urethral swabs. J. Clin. Microbiol. **39:**1008–1016.