JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 1995, p. 545–550 Vol. 33, No. 3
0095-1137/95/$04.00+0

Accuracy and Interlaboratory Reliability of Human Papillomavirus DNA Testing by Hybrid Capture

MARK H. SCHIFFMAN,[1]* NANCY B. KIVIAT,[2] ROBERT D. BURK,[3] KEERTI V. SHAH,[4] RICHARD W. DANIEL,[4] RENEE LEWIS,[3] JANE KUYPERS,[2] M. MICHELE MANOS,[4] DAVID R. SCOTT,[5] MARK E. SHERMAN,[4]† ROBERT J. KURMAN,[4] MARK H. STOLER,[6]‡ ANDREW G. GLASS,[5] BRENDA B. RUSH,[5] IWONA MIELZYNSKA,[7] AND ATTILA T. LORINCZ[7]

Epidemiology and Biostatistics Program, National Cancer Institute, Bethesda, Maryland[1]; University of Washington, Seattle, Washington[2]; Albert Einstein College of Medicine, Bronx, New York[3]; Johns Hopkins Medical Institutions, Baltimore, Maryland[4]; Portland Kaiser-Permanente, Portland, Oregon[5]; Cleveland Clinic, Cleveland, Ohio[6]; and Digene Diagnostics, Inc., Silver Spring, Maryland[7]

Received 13 July 1994/Returned for modification 29 September 1994/Accepted 6 December 1994

Epidemiologists and clinicians wishing to introduce human papillomavirus (HPV) testing into cervical cancer prevention programs need standardized, reliable, and accurate HPV DNA tests that can detect the full spectrum of pathogenic HPV types. The Hybrid Capture System assay from Digene (hybrid capture assay) is a nonradioactive kit designed to detect 14 HPV types in two groups: a mix of 9 high-risk types associated with anogenital cancer (HPV types 16, 18, 31, 33, 35, 45, 51, 52, and 56) and another group of 5 low-risk types associated with condyloma acuminatum (HPV types 6, 11, 42, 43, and 44). The assay yields quantitative data meant to reflect viral concentration. In a study of 199 cervical specimens from women with concurrent Pap smears, we assessed the reliability of the new assay by comparing the hybrid capture assay results from three laboratories. We assessed the accuracy of the hybrid capture assay in comparison with a reference standard of HPV DNA content (multiple testing by several methods in two reference laboratories). We also compared the hybrid capture assay results with the concurrent cytologic diagnoses on the basis of an independent review of each smear by five pathologists. Pairwise interlaboratory agreement rates on HPV positivity for either high-risk or low-risk types ranged from 87 to 94%, and kappa values ranged from 0.61 to 0.83. Among specimens positive for high-risk types (the most important clinical outcome), the interlaboratory correlations of the quantitative data ranged from 0.60 to 0.90. Test results from all three laboratories were strongly associated with those of the HPV DNA reference standard and with the concurrent cytopathologic diagnoses. The most common errors were sporadic, apparently false-positive results.

Human papillomavirus (HPV) infection is the major cause of most cervical cancer and cervical intraepithelial neoplasia (CIN) worldwide (3, 12, 15, 17, 23). Consequently, there is strong motivation to evaluate the use of HPV testing in cervical cancer screening. Presently, HPV testing depends on the detection of HPV DNA in exfoliated cervical cell specimens (16).

The application of HPV DNA testing to cervical cancer screening will require further improvements in and standardization of testing methods. HPV testing has generally been conducted by Southern blotting or PCR methods, with very few attempts at interlaboratory standardization of these research techniques (4, 16). Instead, some researchers have chosen to use a U.S. Food and Drug Administration-approved dot blot kit called ViraPap/ViraType (11, 14). However, the kit detects only seven HPV types (types 6, 11, 16, 18, 31, 33, and 35). Neither nonstandardized research assays nor the ViraPap/ViraType kit are adequate for routine clinical use. An assay for routine clinical use will require reliable and accurate detection of the broad range of pathogenic HPV types infecting the cervix.

Newly developed, the Hybrid Capture System HPV DNA detection test (hybrid capture assay) from Digene Diagnostics, Inc. (Silver Spring, Md.), is a nonradioactive, relatively rapid, liquid hybridization assay designed to detect 14 HPV types divided into high-risk and low-risk groups. High-risk HPV types are those associated with cervical cancer. The hybrid capture assay kit detects high-risk types 16, 18, 31, 33, 35, 45, 51, 52, and 56, but does not yet include probes for a few other high-risk types such as types 39 and 58 (3, 13). The low-risk group detects the types more commonly associated with condyloma acuminatum (venereal warts) of the external genitalia: HPV types 6, 11, 42, 43, and 44 (13). Both the high-risk and low-risk groups, along with other rarer and/or uncharacterized types of HPV, can be found in women with CIN or normal cervical cytologic diagnoses.

In addition to detecting a wide range of HPV types associated with the risk of CIN (5), early experiments suggest that the hybrid capture test is reliable when it was performed in masked experiments at Digene Diagnostics, Inc., the originating laboratory (9a). As a possible unique advantage compared with other available HPV test kits, the hybrid capture test is also designed to provide quantitative estimates of viral load, which may correlate with the grade and natural history of cervical pathology (5, 6). Given the apparent promise of the hybrid capture test, the present investigation was conducted to assess the assay's interlaboratory reliability and accuracy when

* Corresponding author. Mailing address: Epidemiology and Biostatistics Program, National Cancer Institute, Executive Plaza North Room 443, Bethesda, MD 20892. Phone: (301) 496-1691. Fax: (301) 402-0916.

† Present address: Department of Pathology, George Washington University Medical Center, Washington, DC 20032.

‡ Present address: Department of Pathology, University of Virginia Health Sciences Center, Charlottesville, VA 22903.

it was performed on realistic clinical specimens at independent testing sites.

MATERIALS AND METHODS

Digene Diagnostics, Inc., provided the equipment, reagents, and training for the interlaboratory comparison of the hybrid capture test. To ensure an independent evaluation of the assay, all testing and data collection steps were masked, and all statistical analyses were performed at the National Cancer Institute (NCI).

Study population and collection of specimens. The study population included 200 women participating in an NCI-sponsored cohort study of HPV infection and cervical neoplasia (17). Briefly, enrollment of more than 23,000 women receiving cervical cytologic screening was conducted between April 1989 and November 1990 at seven Kaiser-Permanente clinics in Portland, Oreg. The 200 subjects originally selected for the present investigation were randomly chosen from those study participants with enrollment cervical smears routinely diagnosed by Portland Kaiser-Permanente pathology staff as equivocal (mild or severe squamous atypia) (21). This study group was considered appropriate for the interlaboratory study because one possible, major use of HPV testing might be to clarify equivocal smears (5). The cervical smear for one woman could not be found, leaving smears for 199 subjects in the analytic data set.

In addition to a cervical smear obtained with an endocervical cytobrush and ectocervical Ayre spatula, a cervicovaginal lavage specimen was concurrently obtained from all subjects for HPV DNA testing. The lavage was performed by rinsing the external cervical os with sterile saline by using a 10-ml syringe equipped with an intracatheter extender. The pooled fluid was collected from the posterior vaginal fornix and was divided into aliquots for HPV DNA testing. A 1-ml unprocessed aliquot of each lavage specimen was frozen at −70°C. The remaining 9 ml of the lavage specimen was centrifuged and was divided to yield to two cell pellets. The processed specimens were transported on dry ice to the NCI repository and were stored at −70°C until they were sent for testing.

The unprocessed 1-ml aliquot of the lavage specimen was tested for HPV DNA in 1990 at Cetus by the L1 consensus primer PCR-based test of Manos and colleagues as described in previous reports (17, 18). This labor-intensive, research PCR method is capable of detecting and typing a wide range of HPV types and also detects many uncharacterized HPVs.

The first of the two specimen pellets was tested for HPV DNA at Digene Diagnostics (1990 to 1992) by three separate methods: low-stringency Southern blotting, which is capable of detecting a wide range of HPV types (18, 21), dot blotting for the 14 types in consideration, and the prototype of the 14-type hybrid capture HPV DNA test (the predecessor of the format submitted for clearance by the U.S. Food and Drug Administration). These Digene HPV DNA results were combined with the Cetus PCR results to compose the HPV DNA reference standard for the present study (see below).

When the present investigation began in 1992, the remaining specimen pellet was resuspended and mixed vigorously in 1 ml of Digene sample transport medium. Each of the laboratories participating in the study received 160 μl of the suspension diluted to 230 μl with an additional 70 μl of transport medium. The specimens were held at −20°C or colder until testing in 1992 or early 1993. One hundred microliters of the aliquot received was used for each hybrid capture test (high-risk or low-risk groups). The amount tested represented approximately 3.5% of the original 10 ml of the cervicovaginal lavage specimen.

HPV testing in the three laboratories. Before testing the study specimens, each laboratory completed a training period that included on-site instruction and test runs consisting of at least two quality assurance panels: a clean panel of 24 specimens containing 100 μg of sonicated herring sperm DNA per ml spiked with HPV plasmid DNA of various types and concentrations and a 24-specimen clinical panel of cervical samples with known HPV test results. The laboratories tested the study specimens only after successful results were obtained with the quality assurance panels. Specifically, one of the laboratories completed testing of two 24-specimen clinical panels before testing the study specimens.

Each laboratory tested the study specimens without duplicates by using the hybrid capture assay kit for the high-risk and low-risk HPV groups. The following uniform protocol, detailed more fully in the instructions accompanying the kit, was used. Specimens were treated with sodium hydroxide to hydrolyze specimen RNA and to denature the DNA. The liberated single-stranded DNA was hybridized in solution with an RNA probe mix consisting of the high-risk or low-risk HPV types. Each reaction mixture, containing any RNA-DNA hybrids that formed, was transferred to a capture tube coated with antibodies to the RNA-DNA hybrids, which immobilized the hybrids. Bound RNA-DNA hybrids were then reacted with an alkaline phosphatase-conjugated antibody directed against RNA-DNA hybrids. Unreacted material was removed by washing, and a dioxetane-based chemiluminescent compound, Lumi-Phos 530, was added as a substrate for the alkaline phosphatase. The light produced by the ensuing reaction was measured by a luminometer. Light measurements were expressed as relative light units (RLUs). As a negative control, sonicated herring sperm DNA in Digene transport medium (100 μg/ml) was used. Triplicate specimens of HPV type 16 or HPV type 11 DNAs at 10 pg/ml served as the positive controls for high-risk HPV probes and low-risk probes, respectively.

All RLU measurements for specimens were divided by the mean RLU of the three appropriate positive controls (PCs) to give a ratio of specimen RLU/PC. A specimen RLU/PC ratio of 1.0 or greater was regarded as positive for HPV DNA, and a ratio of less than 1.0 was regarded as negative. In ancillary analyses, alternative RLU/PC cutoff points of 0.9 and 1.1 were examined and were found not to alter the conclusions, causing only minor changes in the reliability and accuracy of the assay (data not shown).

Since the amount of light produced by the hybrid capture assay is theoretically proportional to the amount of target HPV DNA in each clinical specimen, the results could possibly be viewed as quantitative. Of course, the use of hybrid capture quantitation to estimate viral load will be affected somewhat by variation in sample adequacy. Nonetheless, we considered it worthwhile to examine the interlaboratory reliability of the RLU/PC ratios and the correlation of the RLU/PC ratios to the cytologic diagnoses.

Creating a reference standard of HPV DNA content. The assessment of accuracy in HPV DNA testing required the creation of a reference standard, although, admittedly, no perfect standard of HPV infection currently exists.

To create an HPV DNA reference standard, the data from the Cetus PCR testing in 1990 as well as results from the three rounds of Digene testing in 1990 to 1992 (Southern blot, dot blot, prototype of hybrid capture) were evaluated. A summary judgment was made for each specimen by two of the investigators (A.T.L., M.H.S.) without knowledge of either the HPV results from the three laboratories or the specimen-specific cytologic data. Specifically, all previous testing results for a specific specimen were compared, and when a consensus or majority HPV testing decision was evident, it was taken to be true. In the few cases of truly equivocal specimens, i.e., those with severely discrepant previous results, the investigators tended to favor the dot blot and Southern blot diagnoses over the PCR or prototype hybrid capture assay results. As one possible consequence, some specimens classified as negative might truly have been very weak positives, which were detected by PCR only.

For the statistical analysis, specimens were classified as either positive or negative for the high-risk HPV group (HPV types 16, 18, 31, 33, 35, 45, 51, 52, and 56) and separately for the low-risk group (HPV types 6, 11, 42, 43, and 44). Any specimen containing an HPV type not in the 14-type mix (e.g., types 39 or 58) was called negative for the present purpose of assessing test accuracy for the 14 types. In an additional analysis, the Cetus PCR results were used alone as an alternative DNA reference standard to address the possible concern of using reference tests from Digene in judging the accuracy of the Digene kit.

Cytopathology as an ancillary reference standard. Cytopathologic diagnoses can also serve as a kind of reference standard for HPV DNA testing because of the strong and consistent association of HPV detection and prevalent CIN (17). When assayed by research tests that detect a wide range of HPV types (2, 9, 13, 21), the large majority of specimens from patients with CIN contain HPV DNA, while HPV prevalences in cytologically normal women are consistently much lower. Inaccurate assays, in contrast, find much weaker associations between HPV DNA and CIN (8, 19). Thus, one measure of the performance of an HPV DNA test is the strength of its association with CIN, including the percentage of women with confirmed diagnoses of CIN found to be HPV DNA positive.

To create a cytopathology reference standard for the present investigation, masked review of each cytology smear by five expert cytopathologists permitted the classification of the smears along a cytologic certainty scale (although all smears had originally been classified as atypical). The full details of this cytology scale, which correlated very well with both the Cetus research PCR and Digene Southern blot HPV DNA testing in previous work, have been published elsewhere (21). Briefly, the scale ranged from 0.0 to 5.0 in increments of 0.5, with larger numbers reflecting an increasing probability that the smear represented underlying CIN. For most analyses in the present investigation, the cytology scale was condensed into three categories of the proper size for analyses: probably normal (0.0 to 1.0), possibly CIN (1.5 to 3.0), or probably CIN (3.5 to 5.0). The cytopathology reference standard was not available for 10 women in the study because slides judged to be inadequate by any one of the experts had to be excluded. Thus, for analyses including the cytopathology reference standard, only 189 subjects remained in the study.

Statistical methods. Data were returned directly from each laboratory to NCI. Reliability was assessed by pairwise laboratory comparisons by using three statistics: the crude percentage of agreement, the percentage of agreement on positivity, and the kappa statistic. The two HPV type groups were considered independently for the analyses. The crude percentage of agreement was presented because it is a familiar statistic and comparable to that used in much of the literature. However, the crude percentages of agreement tended to be exaggerated by the large number of specimens called negative by all laboratories. Thus, we chose to present also the percentage of agreement on positivity, in which those specimens called negative by both laboratories in a pair were excluded. Finally, the unweighted kappa statistic was calculated. This method is designed to adjust the estimate of agreement for chance agreement that was bound to occur. As a general guide, a kappa value greater than 0.75 represents excellent agreement beyond chance, values below 0.40 represent poor agreement beyond chance, and values between 0.40 and 0.75 suggest fair to good agreement beyond chance (7).

Pairwise reliability of the quantitative data was assessed by calculating Pearson and Spearman correlation coefficients (r values). The Pearson statistic emphasizes the absolute numbers, while the Spearman statistic stresses the ranking of values. The high-risk and low-risk groups were considered separately. Each

TABLE 1. Pairwise agreement between three laboratories for HPV positivity of 199 specimens tested by hybrid capture assay

Laboratory pair	Low-risk group agreement (±)			High-risk group agreement (±)		
	% Crude agreement	% Agreement on positivity	Kappa statistic	% Crude agreement	% Agreement on positivity	Kappa statistic
A and B	94	56	0.67	87	66	0.69
A and C	93	48	0.61	93	78	0.83
B and C	94	54	0.65	89	69	0.73

comparison for high- or low-risk types was restricted to specimens judged to be positive for that type group by at least one of the three laboratories (i.e., probable negative specimens were deleted to avoid confusing the issue of positivity and negativity with quantitation).

Accuracy was assessed by comparing the hybrid capture assay results with those of the reference standard of HPV DNA content by using standard contingency tables and chi-square tests when they were informative. Comparisons with the HPV reference standard were made first for the results of each individual laboratory and then for the joint results of all three laboratories considered together.

The hybrid capture assay results were also compared with those of the cytopathology reference standard by using contingency tables. Correlation of the quantitative data to the cytopathology certainty scale was assessed by using the Spearman correlation coefficient, which was chosen because the values of the certainty scale were ordinal and not intrinsically quantitative.

Statistical analyses were conducted by using one of two readily available commercial statistical software packages, SAS or InStat (GraphPad Software).

RESULTS

The percentages of the 199 specimens found to be positive were similar for the three laboratories. Laboratory A found 29.6% of the specimens to be positive for the high-risk HPV group and 10.6% to be positive for the low-risk group. The corresponding percentages were 31.2 and 9.0% for laboratory B and 26.6 and 9.5% for laboratory C. These differences between laboratories were not statistically significant for either the high-risk ($P = 0.60$) or the low-risk ($P = 0.87$) HPV group.

The interlaboratory reliability of the hybrid capture assay is given in Table 1. Crude percentages of agreement were generally higher for the low-risk HPV group (93 to 94%) than for the high-risk HPV group (87 to 93%) because of the relative rarity of low-risk HPV types, leading to a high number of negative-negative agreements. Once the negative-negative agreements were eliminated, agreement on positivity was shown to be higher for the high-risk group (66 to 78%) than for the low-risk group (48 to 56%). Likewise, the kappa statistics were higher (0.69 to 0.83) for the high-risk group than for the low-risk group (0.61 to 0.67).

The interlaboratory reliability of the quantitative RLU/PC data for positive specimens was calculated for each laboratory pair. The Pearson correlation coefficients for high-risk and low-risk types were 0.90 and 0.96, respectively, for laboratories A and B, 0.60 and 0.95, respectively, for laboratories A and C, and 0.67 and 0.92, respectively, for laboratories B and C. The Spearman correlation coefficients for high-risk and low-risk types were 0.66 and 0.52, respectively, for laboratories A and B, 0.81 and 0.42, respectively, for laboratories A and C, and 0.67 and 0.56, respectively, for laboratories B and C. All of these correlations were highly significant, but the correlations were strongest for the Pearson statistics, particularly for laboratory A versus laboratory B.

Table 2 addresses the accuracy of the hybrid capture assay results compared with those of the HPV DNA reference standard. Crude percentages of agreement were high for both high-risk (88 to 92%) and low-risk (93 to 94%) groups. The percentages of agreement for specimens called positive by the reference standard, which can be viewed as the sensitivities of the three laboratories' results compared with those of the reference standard, ranged from 83 to 88% for the high-risk group and from 62 to 67% for the low-risk group.

The analyses in Table 2 were repeated for the (more important) high-risk group only by using the Cetus PCR results as an alternative HPV DNA standard. The PCR standard was restricted to the high-risk HPV types in the hybrid capture kit (e.g., four specimens found to have HPV-39 by PCR were defined as negative for this analysis). The crude percentages of agreement of the hybrid capture assay with the PCR standard for high-risk types ranged from 79 to 81%. The sensitivity of the hybrid capture assay ranged from 65 to 70% for detection of the 54 specimens called PCR positive with types included in the hybrid capture kit.

As another approach to assessing the accuracy of the hybrid capture assay, we tabulated the joint results from the three laboratories compared with those of the original HPV DNA reference standard. These data address the question of how accurate the assay would be, in aggregate, if it was performed in a group of laboratories. The comparison of the joint test results with the HPV DNA reference standard results is provided in Table 3. The associations of each group of types (high or low risk) with the corresponding DNA reference standards were very strong ($P < 0.001$). The most frequent errors were apparent false-positive results, in which a single laboratory classified a specimen as HPV positive, even though it was called negative by the reference standard and the other two laboratories. Of note, for the high-risk types, 2 of the 18 false-positive specimens were classified as probable CIN by the cytopathology reference standard, demonstrating how difficult it is to define a clear reference standard for HPV infection. Similarly, with regard to questionable false-negative results, six (29%) of the specimens classified by the DNA reference standard as low-risk HPV positive were called negative by all three laboratories. But three of these demonstrated possible CIN and three demonstrated probable CIN according to the cytopathology reference standard.

The direct comparisons of the hybrid capture assay results for each laboratory with those of the cytopathology reference standard are given in Table 4. The hybrid capture assay results were highly associated with the likelihood of CIN for all three laboratories. Among subjects with probable CIN, 71 to 74%

TABLE 2. Agreement and sensitivity of hybrid capture results for each of three laboratories compared with those of the HPV DNA reference standard

Laboratory	% Agreement[a]		% Sensitivity	
	Low-risk HPV group	High-risk HPV group	Low-risk HPV group ($n = 21$)	High-risk HPV group ($n = 58$)
A	93	92	67	88
B	93	88	62	83
C	94	92	67	83

[a] A total of 199 specimens were used in the test for agreement.

TABLE 3. Comparison of hybrid capture assay results in three laboratories with those of HPV DNA reference standard for the high-risk and the low-risk HPV type groups

HPV type group and HPV reference standard result	No. of specimens	No. (%) of specimens			
		Negative in all three laboratories	Positive in one laboratory only[a]	Positive in two laboratories only[b]	Positive in all three laboratories
High-risk group					
Negative	141	120 (85)	18 (13)	0 (0)	3 (2)
Positive	58	3 (5)	6 (10)	6 (10)	43 (74)
Low-risk group					
Negative	178	163 (92)	13 (7)	2 (1)	0 (0)
Positive	21	6 (29)	1 (5)	2 (10)	12 (57)

[a] Percentage of specimens found to be positive by any one, but only one, of the laboratories.
[b] Percentage of specimens found to be positive by any two, but only two, of the laboratories.

were HPV DNA positive, compared with 8 to 16% of the subjects with probably normal smears. For comparison, the corresponding figures for the original HPV DNA reference standard were 84% HPV positivity for probable CIN and 8% positivity for probably normal diagnoses. The Cetus PCR standard, restricted to HPV types also found in the hybrid capture kit (types 6/11, 42, 16, 18, 31, 33, 35, 42, 45, 51, 52, and 56), found 61% of probable cases of CIN and 7% of probably cytologically normal women positive (although PCR positivity was much higher in women with CIN or cytologically normal diagnoses before the type restrictions were made).

On closer examination, there were strong associations in all laboratories between the high-risk HPV group and the likelihood of CIN ($P < 0.001$) whether or not low-risk types were also present. The associations between low-risk HPV types alone and the likelihood of CIN were also statistically significant but much weaker. A similar contrast between high- and low-risk types was seen with the HPV DNA reference standards.

The comparison of the three laboratories' joint results with those of the cytopathology reference standard are shown in Table 5. Joint results for the high-risk HPV group (Table 5) were highly associated with the results of the cytopathology reference standard ($P < 0.001$). Among specimens from women with probable CIN, 55% were called HPV positive for high-risk types in all three laboratories, whereas only 3% of specimens from women with probably normal smears were called HPV positive. A much weaker but still significant relationship was seen for low-risk types ($P = 0.01$), as shown in Table 5.

To assess the possible utility of quantitation of HPV DNA as a predictor of cytopathology, independent of yes or no positivity, we correlated the quantitative RLU/PC results from each laboratory with the cytopathology certainty scale (0.0 to 5.0 in increments of 0.5). Negative specimens were excluded as noted in Materials and Methods. For high-risk HPV types, the Spearman correlations with the certainty scale were 0.51 for laboratory A, 0.30 for laboratory B, and 0.43 for laboratory C. Weaker correlations were seen for the low-risk types (Spearman correlation coefficients of −0.03, 0.11, and 0.34, respectively).

DISCUSSION

The development of a reliable, accurate, and cost-effective HPV test method is needed in order to move HPV testing into routine clinical practice. The present interlaboratory comparison demonstrated that the hybrid capture test has good reliability and accuracy, although room for improvement remains.

As shown by the relatively narrow range in measures of agreement between the three laboratories, their reliability in classifying specimens as positive or negative was fairly uniform, with no obvious outlier among them. The interlaboratory reliability of HPV positivity defined by the hybrid capture assay appeared to be somewhat better for the high-risk group than for the low-risk group, perhaps because low-risk positivity was

TABLE 4. Association of hybrid capture assay results from three laboratories with cytopathology review

Laboratory	Low-risk HPV group result	High-risk HPV group result	No. (%) of specimens		
			Probably normal ($n = 88$)	Possible CIN ($n = 50$)	Probable CIN ($n = 51$)
A	Negative	Negative	78 (88.6)	32 (64.0)	13 (25.5)
	Positive	Negative	3 (3.4)	3 (6.0)	3 (5.9)
	Negative	Positive	4 (4.6)	14 (28.0)	27 (52.9)
	Positive	Positive	3 (3.4)	1 (2.0)	8 (15.7)
B	Negative	Negative	74 (84.1)	31 (62.0)	13 (25.5)
	Positive	Negative	4 (4.6)	2 (4.0)	4 (7.8)
	Negative	Positive	9 (10.2)	16 (32.0)	28 (54.9)
	Positive	Positive	1 (1.1)	1 (2.0)	6 (11.8)
C	Negative	Negative	81 (92.0)	33 (66.0)	15 (29.4)
	Positive	Negative	2 (2.3)	2 (4.0)	4 (7.8)
	Negative	Positive	4 (4.6)	13 (26.0)	24 (47.1)
	Positive	Positive	1 (1.1)	2 (4.0)	8 (15.7)

TABLE 5. Comparison of hybrid capture assay results in three laboratories to cytopathology review for high-risk and low-risk HPV type groups

HPV type group and cytology diagnosis	No. of specimens	No. (%) of specimens			
		Negative in all three laboratories	Positive in one laboratory only[a]	Positive in two laboratories only[b]	Positive in all three laboratories
High-risk group					
Probably normal	88	72 (82)	13 (15)	0 (0)	3 (3)
Possible CIN	50	31 (62)	5 (10)	0 (0)	14 (28)
Probable CIN	51	12 (24)	5 (10)	6 (12)	28 (55)
Low-risk group					
Probably normal	88	81 (92)	3 (3)	1 (1)	3 (3)
Possible CIN	50	43 (86)	5 (10)	0 (0)	2 (4)
Probable CIN	51	35 (69)	6 (12)	3 (6)	7 (14)

[a] Percentage of specimens found to be positive by any one, but only one, of the laboratories.
[b] Percentage of specimens found to be positive by any two, but only two, of the laboratories.

uncommon, thereby accentuating the sporadic errors that did occur. In any case, the high-risk group is far more important clinically. For applications requiring a high positive predictive value, it is unlikely that inclusion of the less common, low-risk types will prove cost-effective when performing hybrid capture testing (21).

To assess accuracy, we attempted to create a reference standard of HPV DNA by a masked review of all available research DNA testing that had been performed previously on the study specimens. Hybrid capture test results in the three laboratories were very strongly associated with those of the reference standard, suggesting reasonably good accuracy.

However, the reference standard result depended partly on tests performed previously at Digene, including the prototype of the hybrid capture assay itself. Although we masked the specimen-by-specimen adjudication of test results that produced the reference standard result, Digene-developed test methods might be expected to agree better with each other than with assays developed and performed elsewhere. This could have inadvertently biased the reference standard.

Thus, we also compared the hybrid capture results with Cetus PCR results. As expected, hybrid capture results appeared to be less sensitive compared with PCR results than compared with the original reference standard results. It is probable that hybrid capture testing failed to detect some low-level infections that were detected by PCR. It is also probable that some discrepancies reflected false-positive PCR results. In this investigation, however, we did not try to determine how many of the additional positive specimens detected by PCR represented true versus false positivity, nor did we attempt to determine the clinical relevance of low-level infections detected only by PCR.

With regard to the cytology reference standard, the hybrid capture assay results were strongly associated with a likelihood of CIN in concurrently obtained cervical smears. The strengths of the associations for each laboratory were nearly as strong as the association between the HPV DNA reference standard result and the likelihood of CIN or between PCR and the likelihood of CIN. Of course, some cases of disagreement between DNA testing and cytopathologic diagnosis might not be due to DNA testing error. Rather, the disagreements could be due to either residual error in the cytologic diagnoses or true DNA positivity with no apparent cytologic abnormality. For example, two of the three specimens which all three laboratories called high-risk HPV positive, in the absence of cytologic abnormality, were also HPV positive according to the high-risk HPV DNA reference standard. Thus, these cases could represent vaginal or cervical lesions that were inapparent, incipient (12), or missed during cytologic screening.

Nonetheless, some other disagreements between hybrid capture assay results and the cytopathology reference standard results were likely false-negative test results linked to the still incomplete type range of the hybrid capture assay. Specifically, the current hybrid capture kit does not yet include some high-risk HPV types (e.g., types 39 and 58) which are commonly found in CIN (10). For example, HPV type 39 was found by the Cetus PCR assay in two of the nine cases of probable CIN called negative for high-risk HPVs by all three laboratories.

The most common error in the three laboratories appeared to be occasional false-positive results not predictive of CIN. The RLU/PC values of these false-positive results were sometimes quite high. The cause of these false-positive results is not known. Because the false-positive results were uncommon and apparently sporadic rather than related to specific specimens, they could in theory be detected by repeated testing of specimens with positive results (i.e., the false-positive specimens would be negative when retested). Unpublished results from the ongoing NCI prospective study suggest that this approach is feasible.

The quantitative information provided by the hybrid capture test had good reliability, given the crude division of the lavage specimen as provided to the laboratories and the lack of an internal DNA standard. Moreover, higher viral loads correlated with a likelihood of CIN. Other investigators have shown that a high viral load predicts an increased probability of histological confirmation of CIN (5) and a higher risk of high-grade CIN (6). Thus, the quantitative aspect of the hybrid capture test merits further development and evaluation.

Some additional cautions on the interpretation of the study should be mentioned. First, the population was highly selected, to represent equivocal cytologic diagnoses. This special population is a likely target of the application of HPV testing, but nonetheless, assay performance may be different when specimens from women in the general population are tested. Also, this experiment was based on aliquots of a nonuniform pellet taken from long-term storage rather than optimal specimens collected directly into the hybrid capture collection kit. Despite the strong correlations observed between laboratories, specimen heterogeneity and dehydration may have reduced the reliability of the assay as performed in the present study. On the other hand, the performances of the laboratories participating in the study were probably enhanced by the on-site training and pretests as well as the prior HPV testing expertise of the participants.

The impetus to this methodologic comparison was the recognition that HPV testing is becoming increasingly important. There are at least two promising clinical uses of HPV testing as an adjunct to cervical cytologic screening: the clarification of equivocal and low-grade Pap smears and general screening in older women, among whom the positive and negative predictive values of a positive HPV test might be high enough to justify general use (5, 20, 21). Large natural history studies and clinical trials of both applications are under way, and on the basis of the available data derived from research assays, the use of HPV testing to clarify equivocal Pap smears appears to be especially promising. If either of these two major applications of HPV testing gains widespread acceptance, HPV DNA tests will eventually be performed on millions of women a year.

In this context, another interlaboratory trial of the hybrid capture assay should be performed once the collection and testing protocols are available in final, U.S. Food and Drug Administration-ready kit formats. Then, it will be important to evaluate any residual error in hybrid capture testing in the proper context. HPV testing is designed to clarify and supplement the Pap smear. The reliability and accuracy of a single Pap smear are quite imperfect; thus, the two tests together can outperform either one alone (21). For instance, the same hybrid capture kit evaluated in the present study performed very well when it was recently applied to atypical triage Pap smears in a colposcopy clinic (5).

A broad and important question raised by the present study, therefore, is how reliable and accurate a DNA diagnostic test must be before it is judged to be ready for widespread clinical use. The screening literature regarding DNA diagnostics is expanding (1, 22), but few performance standards have been promulgated in the DNA diagnostics community on the basis of commonly used statistics such as percentages of agreement, sensitivity, and specificity. In particular, interlaboratory reliability has rarely been formally investigated and presented. Given the increasing clinical importance of HPV testing and other DNA diagnostic assays, it would be helpful to hold a public discussion of realistic target ranges of reliability and accuracy for the DNA diagnostic assays now approaching the clinical arena.

ACKNOWLEDGMENT

Digene Diagnostics, Inc., supported this work by providing the equipment and supplies used to perform the interlaboratory comparison.

REFERENCES

1. **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.
2. **Bergeron, C., R. Barrasso, S. Beaudenon, P. Flamant, O. Croissant, and G. Orth.** 1992. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct distribution in low- and high-grade lesions. Am. J. Surg. Pathol. **16:**641–649.
3. **Bosch, F. X., M. M. Manos, N. Munoz, M. E. Sherman, A. Jansen, J. Peto, M. H. Schiffman, V. Moreno, K. V. Shah, and the IBSCC Study Group.** Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J. Natl. Cancer Inst., submitted for publication.
4. **Brandsma, J., R. D. Burk, W. D. Lancaster, H. Pfister, and M. H. Schiffman.** 1989. Inter-laboratory variation as an explanation for varying prevalence estimates of human papillomavirus infection. Int. J. Cancer **43:**260–262.
5. **Cox, J. T., A. T. Lorincz, M. H. Schiffman, M. E. Sherman, R. J. Kurman, M. Hussein, and A. Cullen.** The use of hybrid capture, a newly-developed HPV DNA test, to triage women with a cytologic diagnosis of ASCUS. Am. J. Obstet. Gynecol., in press.
6. **Cuzick, J., G. Terry, L. Ho, T. Hollingworth, and M. Anderson.** 1992. Human papillomavirus type 16 DNA in cervical smears as a predictor of high-grade cervical intraepithelial neoplasia. Lancet **339:**959–960.
7. **Fleiss, J. L.** 1981. Statistical methods for rates and proportions. John Wiley & Sons, Inc., New York.
8. **Franco, E. L.** 1991. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. Epidemiology **2:**98–106.
9. **Fuchs, P. G., F. Girardi, and H. Pfister.** 1988. Human papillomavirus DNA in normal, metaplastic, preneoplastic and neoplastic epithelia of the cervix uteri. Int. J. Cancer **41:**41–45.

9a. **Hildesheim, A., et al.** Unpublished data.

10. **Hildesheim, A., P. Gravitt, M. H. Schiffman, R. J. Kurman, W. Barnes, S. Jones, J. G. Tchabo, L. A. Brinton, C. Copeland, J. Epp, and M. M. Manos.** 1993. Determinants of genital human papillomavirus infection in low-income women in Washington, D.C. Sex. Transm. Dis. **20:**279–285.
11. **Kiviat, N. B., L. A. Koutsky, C. W. Critchlow, D. A. Galloway, D. A. Vernon, M. L. Peterson, P. E. McElhose, S. J. Pendras, C. E. Stevens, and K. K. Holmes.** 1990. Comparison of Southern transfer hybridization and dot filter hybridization for detection of cervical human papillomavirus infection with types 6, 11, 16, 18, 31, 33, and 35. Am. J. Clin. Pathol. **94:**561–565.
12. **Koutsky, L. A., K. K. Holmes, C. W. Critchlow, C. E. Stevens, J. Paavonen, A. M. Beckmann, T. A. DeRouen, D. A. Galloway, D. Vernon, and N. B. Kiviat.** 1992. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. N. Engl. J. Med. **327:** 1272–1278.
13. **Lorincz, A. T., R. Reid, B. Jenson, M. D. Greenberg, W. Lancaster, and R. J. Kurman.** 1992. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. Obstet. Gynecol. **79:**328–337.
14. **Moscicki, A. B., J. Palefsky, G. Smith, S. Siboshski, and G. Schoolnik.** 1993. Variability of human papillomavirus DNA testing in a longitudinal cohort of young women. Obstet. Gynecol. **82:**578–585.
15. **Munoz, N., F. X. Bosch, S. deSanjose, L. Tafur, I. Izarzugaza, M. Gili, P. Viladiu, C. Navarro, C. Martos, N. Ascunce, L. C. Gonzalez, J. M. Kaldor, E. Guerrero, A. Lorincz, M. Santamaria, P. A. deRuiz, N. Aristizabal, and K. Shah.** 1992. The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Columbia and Spain. Int. J. Cancer **52:**743–749.
16. **Schiffman, M. H.** 1992. Validation of HPV hybridization assays: correlation of filter in situ, dot blot and PCR with Southern blot, p. 169–179. *In* N. Munoz, F. X. Bosch, K. V. Shah, and A. Meheus (ed.), The epidemiology of human papillomavirus and cervical cancer. IARC Scientific Publications, Lyon, France.
17. **Schiffman, M. H., H. M. Bauer, R. N. Hoover, A. G. Glass, D. M. Cadell, B. B. Rush, D. R. Scott, M. E. Sherman, R. J. Kurman, S. Wacholder, C. K. Stanton, and M. M. Manos.** 1993. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J. Natl. Cancer Inst. **85:**958–964.
18. **Schiffman, M. H., H. M. Bauer, A. T. Lorincz, M. M. Manos, J. C. Byrne, A. G. Glass, D. M. Cadell, and P. M. Howley.** 1991. Comparison of Southern blot hybridization and polymerase chain reaction methods for the detection of human papillomavirus DNA. J. Clin. Microbiol. **29:**573–577.
19. **Schiffman, M. H., and A. Schatzkin.** 1994. The critical importance of measurement error in molecular epidemiologic studies: lessons from studies of HPV and cervical neoplasia. Cancer Res. **54:**1944s–1947s.
20. **Schiffman, M. H., and M. E. Sherman.** 1994. HPV testing can be used to improve cervical cancer screening, p. 265–277. *In* S. Srivastava, S. M. Lippman, W. K. Hong, and J. L. Mulshine (ed.), Molecular markers of early detection of cancer. Futura Publishing Co., Armonk, N.Y.
21. **Sherman, M. E., M. H. Schiffman, A. T. Lorincz, M. M. Manos, D. R. Scott, R. J. Kurman, N. B. Kiviat, M. Stoler, A. G. Glass, and B. B. Rush.** 1994. Towards objective quality assurance in cervical cytopathology: correlation of cytopathologic diagnoses with detection of high risk HPV types. Am. J. Clin. Pathol. **101:**182–187.
22. **Vlaspolder, F., J. A. Mutsaers, F. Blog, and A. Notowicz.** 1993. Value of a DNA probe assay (Gen-Probe) compared with that of culture for diagnosis of gonococcal infection. J. Clin. Microbiol. **31:**107–110.
23. **zur Hausen, H.** 1994. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. Human pathogenic papillomaviruses. Curr. Top. Microbiol. Immunol. **186:**131–156.