American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2009.

Vol. 171, No. 2
DOI: 10.1093/aje/kwp390
Advance Access publication:
December 10, 2009

Practice of Epidemiology

Impact of Improved Classification on the Association of Human Papillomavirus With Cervical Precancer

Philip E. Castle*, Mark Schiffman, Cosette M. Wheeler, Nicolas Wentzensen, and Patti E. Gravitt

* Correspondence to Dr. Philip E. Castle, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Room 5004, MSC 7234, Bethesda, MD 20892-7234 (e-mail: castlep@mail.nih.gov).

Initially submitted June 1, 2009; accepted for publication August 5, 2009.

Misclassification of exposure and surrogate endpoints of disease can obscure causal relations. Using data from the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS, 1997–2001), the authors explored the impact of exposure (human papillomavirus (HPV) detection) and endpoint (histologic cervical precancer) classification on their mutual association. Women referred into this study with an atypical squamous cells of undetermined significance Papanicolaou test with satisfactory results for all 4 HPV tests were included in this analysis ($n = 3{,}215$; 92.2%). HPV testing results were related to different definitions of cervical precancer, based on paired, worst 2-year histologic diagnoses, by calculating clinical sensitivity, specificity, and odds ratios. The authors found that HPV test sensitivity increased and specificity decreased with increasing certainty of cervical precancer, with HPV testing having the highest sensitivity (92%–98%) and lowest specificity (46%–54%) for consensus cervical intraepithelial neoplasia grade 3 (CIN 3). The overall accuracy of each HPV test, as measured by odds ratios, was greatest for consensus CIN-3 diagnoses, from 2- to 4-fold greater than for a less stringent precancer definition of any diagnosis of CIN 2 or more severe. In summary, there was convergence of greater certainty of carcinogenic HPV with greater certainty of a precancerous diagnosis, such that all 4 HPV tests almost always tested positive in women most likely to have cervical precancer. Finding increasingly strong associations when both test and diagnostic misclassification are reduced is a useful sign of ''true association'' in molecular epidemiology.

cervical intraepithelial neoplasia; misclassification; papillomavirus infections; uterine cervical neoplasms

Abbreviations: ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study; ASC-US, atypical squamous cells of undetermined significance; CIN 2 and CIN 3, cervical intraepithelial neoplasia grades 2 and 3, respectively; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; Pap, Papanicolaou.

Editor's note: *An invited commentary on this article appears on page 164.*

Persistent cervical infections by approximately 15 carcinogenic human papillomavirus (HPV) genotypes are now well established as the essential cause of cervical cancer (1, 2). The natural history of HPV infection leading to invasive cancer is summarized by 4 reliably measured stages: 1) HPV acquisition, 2) HPV persistence (vs. clearance), 3) progression of a persistent infection to cervical precancer, and 4) invasion (3). On the basis of this causal relation and well-defined steps in cervical carcinogenesis, 2 HPV-based technologies have emerged: HPV vaccines to prevent HPV acquisition and HPV testing for sensitive detection of precancerous lesions to facilitate timely treatment.

Cervical precancer is the key endpoint to evaluate new prevention strategies for cervical cancer because screening has already reduced the incidence of cervical cancer in Western countries significantly such that cervical cancer is fairly uncommon (4), and it would be unethical to follow women with precancerous lesions to cancer although unfortunately this has occurred once (5, 6). However, the definition of a precancerous lesion is somewhat subjective.

Cervical intraepithelial neoplasia grade 3 (CIN 3) and carcinoma in situ, based on the aforementioned study (5, 6), are inarguably precancerous diagnoses. Nonetheless, CIN 3 is an imperfect diagnosis of precancer and an intermediate surrogate for cancer because not all CIN 3 invades (5, 6), such lesions are occasionally caused by noncarcinogenic HPV genotypes that are unlikely to invade (7), CIN 3 can be a false positive diagnosis (7), and the classification is not perfectly reproducible (8). CIN 2 is included in the definition of precancer for safety and is typically treated in the United States (9), yet CIN 2 is often caused by noncarcinogenic HPV genotypes (10), it is regressive, especially when it is caused by HPV genotypes other than HPV-16 (11, 12), and the diagnostic agreement among pathologists for CIN 2 is poor (8, 10, 13).

Likewise, detection of HPV viral DNA is not error free. No test is perfectly reproducible, and no 2 tests agree completely on the presence of HPV DNA. Random misclassification is known to reduce the magnitude of true associations. Real-life examples of how associations strengthen as misclassification is reduced are uncommon; the gradual recognition in epidemiologic studies of HPV as the central cause of cervical cancer and precancer has provided some of the best examples (14–17). The early articles focused mainly on the impact of improvement in HPV testing on the strength of associations. The impact of misclassification on both exposure and disease assessment, as well as how reduced misclassification can improve our ability to observe true causal associations, has not been illustrated as thoroughly with real data. To expand the literature on this important point, we analyzed the relation of carcinogenic HPV detection with histologically confirmed cervical precancer in the context of multiple measures of each.

In the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS), we have 4 measures of carcinogenic HPV using validated assays and 2 evaluations of histopathology among women referred into ALTS with an atypical squamous cells of undetermined significance (ASC-US) Papanicolaou (Pap) test result. This provided an excellent opportunity to explore how the classification of both the exposure (HPV) and the intermediate endpoint (histologically confirmed cervical precancer) influences the relation between exposure and disease.

MATERIALS AND METHODS

Study design and population

ALTS (1997–2001) was a multisite, randomized trial comparing 3 management strategies for women referred for ASC-US ($n = 3{,}488$) or low-grade squamous intraepithelial lesion ($n = 1{,}572$) conventional cytology (20–24). (Under the 1991 Bethesda System for Reporting Cervical Cytology (18), ALTS was slightly more inclusive, particularly of probable reactive changes and ASC-H (atypical squamous cells, cannot rule out high-grade intraepithelial lesion), than the ASC-US category of the 2001 Bethesda system (19).) Women were randomized to 1 of 3 study arms at enrollment: 1) immediate colposcopy arm (referral to colposcopy regardless of enrollment test results); 2) HPV triage (HPV arm) (referral to colposcopy if the enrollment HPV result was positive by Hybrid Capture 2 (hc2; Qiagen Corporation, Gaithersburg, Maryland) or missing, or if the enrollment cytology was high-grade squamous intraepithelial lesion (HSIL)); or 3) conservative management arm (referral to colposcopy only if the enrollment cytology was HSIL). The National Cancer Institute and local institutional review boards approved the study, and all participants provided written, informed consent.

At enrollment and follow-up visits over the 2-year duration, all women underwent a pelvic examination with collection of 2 cervical specimens: the first specimen in PreservCyt for ThinPrep cytology (Cytyc Corporation, Marlborough, Massachusetts; now Hologic) and the second in specimen transport medium (STM; Digene Corporation, Gaithersburg, Maryland; now Qiagen). Women in all 3 arms of the study were reevaluated by cytology every 6 months during the 2 years and sent to colposcopy if cytology was high-grade squamous intraepithelial lesion. An exit examination with colposcopy was scheduled for all women. We refer readers to other references for details on randomization, examination procedures, patient management, and laboratory and pathology methods (20). This analysis was restricted to women referred into ALTS for an ASC-US Pap smear.

HPV testing

Residual PreservCyt specimens, after being used for liquid-based cytology, were tested by hc2 (24), a pooled-probe, signal-amplification DNA test that targets a group of 13 carcinogenic HPV genotypes (HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59, and -68).

HPV genotyping by line blot assay (Roche Molecular Systems, Inc., Pleasanton, California) was performed on the specimen transport medium specimen as previously described (25). Amplicons were subjected to reverse-line blot hybridization for detection of 27 individual HPV genotypes (HPV-6, -11, -16, -18, -26, -31, -33, -35, -39, –40, -42, -45, -51 to -59, -66, -68, -73, -82 to -84) (26, 27). We also tested for an additional 11 noncarcinogenic HPV genotypes (HPV-61, -62, -64, -67, -69 to -72, -81, -82 variant (82v or IS39), and -89 (CP6108)) in 76% of the enrollment specimens from women referred into the study because of an ASC-US Pap test.

Aliquots of the archived, enrollment STM specimens were retested by using linear array, a next generation version of line blot assay that tests for 37 of 38 HPV genotypes detected by line blot assay (excluding HPV-57) as previously described (28), and AMPLICOR (Roche Molecular Systems, Inc.), a pooled test for the same 13 carcinogenic HPV genotypes targeted by hc2 (29). We used a positive cutpoint of 1.5 for AMPLICOR as we previously found that this cutoff was the most accurate for detection of precancerous lesions (29).

For this analysis, women were considered positive for carcinogenic HPV if they tested positive by hc2 or AMPLICOR or positive for HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59, and -68 by line blot assay or linear array (30, 31). We excluded HPV-66 from our definition of

carcinogenic HPV because AMPLICOR does not target it and because HPV-66 is not considered carcinogenic in the latest *IARC Monograph*, Volume 100b, regarding carcinogenicity of HPV types (32). However, we acknowledge that hc2 is well known to cross-react with untargeted HPV-66 and other HPV types that very rarely might be carcinogenic (33).

Pathology and treatment

Clinical management was based primarily on the clinical center pathologists' cytologic interpretations and histologic diagnoses as previously described (20–24).

Statistical analyses

Among women referred for an ASC-US Pap test, 3,326 (95.4%) had hc2 results, 3,436 (98.5%) had AMPLICOR results, 3,445 (98.8%) had linear array results, and 3,363 (96.4%) had line blot assay results. This analysis was restricted to the 3,215 women (92.2%) who had results for all 4 HPV tests. We calculated the pairwise percent of agreement and kappa values among the 4 HPV tests.

In addition to the certainty of the HPV test, we were interested in the other dimension of the standard 2 × 2 table, namely, how the definition and certainty of precancerous disease influenced the association with carcinogenic HPV. We used paired histopathologic diagnoses from clinical center and ALTS quality control pathology in our definition of endpoints to reflect the certainty of a precancerous diagnosis. Both reviews were used because neither pathology review is without misclassification (8, 10). Diagnoses from each pathology review were categorized as <CIN 2, CIN 2, or CIN 3. Thus, there were 9 combinations of histopathologic diagnoses from the 2 pathology reviews; the crude tabulation of the 4 test results and paired diagnoses are shown in Web Table 1 posted on the *Journal*'s website (http://aje.oxfordjournals.org). We defined different endpoints of precancer, from less certain to more certain, as follows (Web Table 1): any CIN-2 diagnosis or more severe (CIN 2+) by either clinical center or ALTS quality control pathology review (endpoint 1); a CIN 2+ on both pathology reviews (consensus CIN 2+) (endpoint 2); any diagnosis of CIN 3 by either clinical center or ALTS quality control pathology review (endpoint 3); a diagnosis of CIN 3 on one pathology review and CIN 2 or more severe on the other pathology review (endpoint 4) (excluding from endpoint 3 any cases of CIN 3 by one pathology group and <CIN 2 by the other pathology group); and CIN 3 diagnosed on both pathology reviews (consensus CIN 3) (endpoint 5).

We used the 2-year worst diagnoses for all participating women as our endpoints to account for the inaccuracies of a single colposcopic evaluation (34–36). Among women referred for an ASC-US Pap test, the 2-year endpoints (described above) included 2 cancers diagnosed during follow-up (37). We note that, because the conservative management arm sent only women with enrollment HSIL cytology to colposcopy, some CIN 2 probably regressed (11). We observed the same patterns and relations when the analysis was limited to the enrollment diagnoses for the women in the immediate colposcopy arm, all of whom underwent colposcopy (data not shown).

Table 1. The Kappa Values and the Percent Overall Agreement Among the 4 Tests for Detection of Carcinogenic HPV DNA, ALTS, 1997–2001

	hc2		AMP		LA	
	Kappa	Percent Overall Agreement	Kappa	Percent Overall Agreement	Kappa	Percent Overall Agreement
AMP	0.60	81				
LA	0.67	84	0.77	89		
LBA	0.70	85	0.71	85	0.77	88

Abbreviations: ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study; AMP, AMPLICOR (Roche Molecular Systems, Inc., Pleasanton, California); hc2, Hybrid Capture 2 (Qiagen Corporation, Gaithersburg, Maryland); HPV, human papillomavirus; LA, linear array; LBA, line blot assay.

We compared the number of positive HPV tests with different endpoints as defined above. The odds ratios with 95% confidence intervals were calculated as a measure of association between the number of positive HPV tests and the different endpoints, with 0 positive tests as the reference and women with less than CIN 2 by both pathology reviews serving as the "control."

In ancillary analyses, we explored associations of other measures beyond carcinogenic HPV testing in relation to improving the certainty of precancer diagnosis. We examined the detection of the most carcinogenic HPV genotype, HPV-16 (by either line blot assay or linear array) (3,215 of 3,215 included in the analysis had results, 100%), cytologic interpretation of HSIL by clinical center pathology (n = 3,206, 99.7%) and by ALTS quality control pathology (n = 3,176, 98.8%), and visual evidence of high-grade cervical neoplasia as determined by cervigram review of a digitized image of the cervix taken at colposcopy (n = 3,114, 96.9%), with paired histologic diagnoses. We tested the trend (38) for detection of HPV-16, HSIL cytology, or high-grade cervigram impression with increasing severity of histopathologic diagnosis by one pathology group when diagnosis by the other pathology group was CIN 3.

Finally, we examined the relation of clinical sensitivity and specificity of each HPV test with the different definitions of precancer as described above. For reference, we included the cytologic interpretation of the clinical center pathologists as another biomarker of HPV. As the metric of accuracy, we calculated the odds ratio and 95% confidence interval of each test with the different endpoint definitions.

RESULTS

Table 1 shows the agreement among the 4 HPV tests. The best agreement was between AMPLICOR and linear array, with a percent overall agreement of 89% and kappa of 0.77. The worst agreement was between hc2 and AMPLICOR, 2 commercial tests that purport to detect the same pool of carcinogenic HPV types, with a percent overall agreement of 81% and kappa of 0.60.

Table 2. The Relation of the Number of Positive Carcinogenic Human Papillomavirus Tests With Histologic Diagnoses of Precancer Over the 2-Year Duration of ALTS, 1997–2001

No. of Positive Tests	<CIN 2		Endpoint[a]										Total No.[b]
			1		2		3		4		5		
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	
0	995		29		18		13		9		3		1,024
Row		97.2		2.8		1.8		1.3		0.9		0.3	
Column		37.3		5.3		4.2		4.2		3.1		1.7	
1	423		24		17		9		7		3		447
Row		94.6		5.4		3.8		2.0		1.6		0.7	
Column		15.9		4.4		4.0		2.9		2.4		1.7	
2	135		14		8		5		4		0		149
Row		90.6		9.4		5.4		3.4		2.7		0.0	
Column		5.1		2.5		1.9		1.6		1.4		0.0	
3	200		51		37		20		20		9		251
Row		79.7		20.3		14.7		8.0		8.0		3.6	
Column		7.5		9.3		8.7		6.4		6.8		5.0	
4	872		432		345		266		252		164		1,304
Row		66.9		33.1		26.5		20.4		19.3		12.6	
Column		32.7		78.5		81.2		85.0		86.3		91.6	
Total	2,665		550		425		313		292		179		3,215
Odds ratio (0 vs. 1 positive test)[c]			1.9*		2.2*		1.6		1.8		2.3		
Odds ratio (0 vs. 2 positive tests)[c]			3.6*		3.2*		2.7		3.1		0.0		
Odds ratio (0 vs. 3 positive tests)[c]			8.7*		9.7*		6.7*		9.8*		13*		
Odds ratio (0 vs. 4 positive tests)[c]			17*		20*		20*		27*		49*		

Abbreviations: ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study; CIN 2 and CIN 3, cervical intraepithelial neoplasia grades 2 and 3, respectively.

* $P < 0.05$.

[a] Endpoint 1, any CIN-2 diagnosis or more severe (CIN 2+) by either clinical center or ALTS quality control pathology review; endpoint 2, a diagnosis of CIN 2 or more severe on both pathology reviews (consensus CIN 2+); endpoint 3, any diagnosis of CIN 3 by either clinical center or ALTS quality control pathology review; endpoint 4, a diagnosis of CIN 3 on one pathology review and CIN 2 or more severe on the other pathology review; and endpoint 5, CIN 3 diagnosed on both pathology reviews (consensus CIN 3).

[b] The total in each row is the sum of women with <CIN 2 and endpoint 1 (CIN 2+).

[c] Odds ratios were calculated as a measure of the association of the number of positive human papillomavirus tests and the definitions of precancer.

One-thousand twenty-four women (31.9%) were HPV negative by all 4 tests, 447 (13.9%) were positive by 1 HPV test, 149 (4.6%) were positive by 2 HPV tests, 251 (7.8%) were positive by 3 HPV tests, and 1,304 (40.6%) were positive by all 4 tests. The number of tests positive for HPV was strongly associated with lifetime number of sexual partners ($P < 0.001$) (data not shown).

Comparing the HPV testing results with the different definitions of precancer, we found that the percentage in which all 4 HPV tests were positive increased significantly with increasing certainty of cervical precancer. For 2-year worst diagnoses (Table 2), the percentage in which all 4 tests were positive for carcinogenic HPV increased from 32.7% among women with a diagnosis of <CIN 2 by both pathology groups to 91.6% among women with consensus CIN 3. Further, the percent total agreement for HPV testing increased from 70.1% among women with a diagnosis of <CIN 2 by both pathology groups to 93.3% among women with consensus CIN 3.

There was a general trend of increasing strength of association from 1 positive HPV test to 4 positive HPV tests for any endpoint definition; for example, the odds ratios for 1, 2, 3, and 4 positive HPV tests with endpoint 5 (consensus CIN 3) using the 2-year worst histologic diagnoses were 2.3, 0.0, 13, and 49, respectively. The association of 4 positive HPV tests results (vs. no positive tests) strengthened with increasing certainty of precancer, with an odds ratio of 17 for endpoint 1 (any CIN 2) and an odds ratio of 49 for endpoint 5 (consensus CIN 3) using the 2-year worst histologic diagnosis.

Women with a 2-year worst histologic diagnosis of consensus CIN 3 were the most likely at enrollment to be

Table 3. Association of Precancerous Diagnoses With Indicators of Cancer Risk, ALTS, 1997–2001

	Indicators of Cancer Risk[a]											
	HPV-16			HSIL Cytology (Quality Control)			HSIL Cytology (Clinical Center)			Cervigram[b]		
	No.[c]	No. Positive	%	No.[c]	No. Positive	%	No.[c]	No. Positive	%	No.[c]	No. Positive	%
<CIN 2	2,658	346	13.0	2,624	1,229	46.8	2,651	1,407	53.1	2,570	22	0.9
Endpoint 1[d]	557	243	43.6	552	450	81.5	555	460	82.9	544	36	6.6
Endpoint 2[d]	429	214	49.9	427	357	83.6	429	362	84.4	419	34	8.1
Endpoint 3[d]	315	175	55.6	314	261	83.1	315	265	84.1	307	32	10.4
Endpoint 4[d]	294	170	57.8	293	246	84.0	294	250	85.0	287	32	11.1
Endpoint 5[d]	179	116	64.8	178	154	86.5	179	162	90.5	174	26	14.9

Abbreviations: ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study; CIN 2 and CIN 3, cervical intraepithelial neoplasia grades 2 and 3, respectively; HPV-16, human papillomavirus 16; HSIL, high-grade squamous intraepithelial lesion.

[a] Indicators of cancer risk: 1) detection of HPV-16; 2) HSIL cytology rendered by the quality-control pathology review; 3) HSIL cytology rendered by the clinical center pathologists; and 4) high-grade impression on cervigram review.

[b] Cervigram impression of P2 (high grade) or P3 (cancer).

[c] Total number of women with that paired diagnosis and having the measurement.

[d] Endpoint 1, any CIN-2 diagnosis or more severe (CIN 2+) by either clinical center or ALTS quality control pathology review; endpoint 2, a diagnosis of CIN 2 or more severe on both pathology reviews (consensus CIN 2+); endpoint 3, any diagnosis of CIN 3 by either clinical center or ALTS quality control pathology review; endpoint 4, a diagnosis of CIN 3 on one pathology review and CIN 2 or more severe on the other pathology review; and endpoint 5, CIN 3 diagnosed on both pathology reviews (consensus CIN 3).

HPV-16 positive (64.8%), to have cytology results interpreted as HSIL by clinical center pathology (90.5%) and by ALTS quality control pathology (86.5%), and to have their cervigram called high grade or more severe (14.9%) than women without consensus CIN 3 (Table 3). In ancillary analyses, the increasing certainty of disease diagnosis led to increasing strength of association with tests other than grouped HPV detection. Among women diagnosed with CIN 3 by ALTS quality control pathology, there was an increasing proportion of women with enrollment HPV-16 ($P_{trend} = 0.001$), HSIL cytology by clinical center pathology ($P_{trend} < 0.001$) and by ALTS quality control pathology ($P_{trend} < 0.001$), and cervigrams called high grade ($P_{trend} = 0.004$) with increasing severity of clinical center pathology diagnosis (data not shown). Among women diagnosed with CIN 3 by clinical center pathology, there was an increasing proportion of women with enrollment HPV-16 ($P_{trend} = 0.001$) with increasing severity of ALTS quality control pathology diagnosis (data not shown). Taken together, consensus CIN 3 had the strongest association with biomarkers related to cervical cancer risk.

We also examined the impact of cervical precancer definition on clinical sensitivity and specificity of HPV detection, including cytology. For the 2-year worst diagnosis (Figure 1), clinical sensitivity increased and clinical specificity decreased with increasing certainty of precancer. However, as shown in Table 4, the overall accuracy of a test, as measured by odds ratios, increased with increasing certainty of precancer. The strength of association increased from 2- to 4-fold from endpoint 1 (any CIN 2+) to endpoint 5 (consensus CIN 3) for all tests. Of note, repeat cervical cytology findings at any threshold of positivity (we show ASC-US) were less accurate than HPV tests in the context of finding CIN 3 among women referred for a previous ASC-US cytology evaluation.

Finally, we compared the few cases of consensus 2-year worst diagnosis of CIN 3 in which not all HPV tests tested positive ($n = 15$) versus those in which all 4 tests tested positive ($n = 164$). Although the statistical power to detect true differences was limited, those few cases with at least 1 negative test were more likely than those cases in which all 4 tests were positive to have 1) the enrollment cervigram interpreted as negative (36% vs. 21%; $P = 0.4$), 2) a negative colposcopic impression at enrollment (18% vs. 5%; $P = 0.2$), 3) the enrollment cytology interpreted as negative by the ALTS quality control pathology group (40% vs. 11%; $P = 0.007$), 4) the enrollment cytology interpreted as negative by the clinical center (20% vs. 9%; $P = 0.2$), and 5) the diagnosis made during follow-up (53% vs. 38%; $P = 0.2$) versus enrollment (data not shown).

DISCUSSION

We used the data collected in ALTS, including dual pathology readings and 4 HPV tests, to examine the impact of the accuracy of exposure and outcome classification on the relation of cervical precancer definition with detection of carcinogenic HPV. There was convergence of greater certainty of carcinogenic HPV with greater certainty of a precancerous diagnosis, such that all 4 tests, including the less sensitive predecessor to linear array, line blot assay (28), almost always tested positive on enrollment specimens from women with consensus CIN 3 diagnosed during the 2 years of ALTS. The few cases of consensus CIN 3 in which at least one baseline HPV test was negative characteristically had fewer indications of abnormality, suggesting that these cases were the most likely to be incident or very small CIN 3 at enrollment with a low viral burden and/or poorly sampled. Among the hc2-positive consensus CIN 3, the hc2 signal strength, a semiquantitative measure of viral load (39), for those in

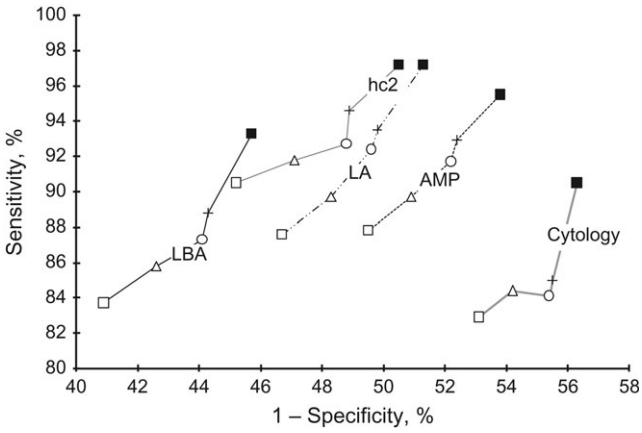

Figure 1. Receiver-operator curves for the 4 human papillomavirus tests and the clinical center pathology-read cytology for detection of cervical precancer over the 2-year duration of ALTS, 1997–2001. Four definitions were used for precancer: □, a CIN-2 diagnosis or more severe on either pathology review (endpoint 1); Δ, a CIN-2 diagnosis or more severe on both pathology reviews (consensus CIN 2+) (endpoint 2); ○, a CIN-3 diagnosis or more severe on either pathology review (endpoint 3); +, a diagnosis of CIN 3 on one pathology review and CIN 2 or more severe on the other pathology review (endpoint 4); and ■, a CIN-3 diagnosis or more severe on both pathology reviews (endpoint 5) (consensus CIN 3). ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study; AMP, AMPLICOR (Roche Molecular Systems, Inc., Pleasanton, California); CIN 2 and CIN 3, cervical intraepithelial neoplasia grades 2 and 3, respectively; Cytology, clinical center pathology interpretation of enrollment ThinPrep cytology (Cytyc Corporation, Marlborough, Massachusetts (now Hologic)), with atypical squamous cells of undetermined significance or worse considered as positive; hc2, Hybrid Capture 2 (Qiagen Corporation, Gaithersburg, Maryland); LA, linear array; LBA, line blot assay.

which all 3 other tests were positive ($n = 164$) was nonsignificantly higher than in those in which at least one of the other tests was negative (201 vs. 64 relative light units/positive control, respectively; $P = 0.16$). Similarly, the sensitivity of enrollment cytology findings also improved as the diagnosis of precancer became more certain.

Misclassification of both the exposure (40) and the endpoint (41) can obscure even strong, causal associations. With regard to HPV research, early associations between HPV DNA positivity and cervical cancer in case-control studies (42) were 2 orders of magnitude lower than current estimates that rely on state-of-the-art testing. The earliest studies failed to demonstrate the strong relation between sexual behavior and HPV infection, which is sexually transmitted. Franco (14, 15) and Franco et al. (16) reported a series of important analyses showing how misclassification weakened the association of HPV with sexual behavior and with cervical cancer and precancer. A similar point was made by Schiffman et al. (17), who clarified that accurate classification of a strong risk factor is important when assessing its role as a confounder or intermediate endpoint of other exposure-disease associations; for example, the association between sexual behavior and cervical precancer is explainable by HPV positivity if testing measures infection history accurately.

Here, we show that the opposite is also true: Improved classification of both the exposure (HPV) and the endpoint (cervical precancer) leads to stronger epidemiologic evidence of causal associations (43). The detection of HPV by multiple assays was more strongly associated with endpoints 4 and 5 than HPV detected by only 1 or 2 HPV tests. Detection of HPV by any single assay was more strongly associated with consensus CIN 3 than with less severe paired diagnoses still considered ''precancerous.'' Previous studies that explored the impact of misclassification on the association of HPV and cervical cancer used HPV assays that were less accurate. By comparison, we used validated HPV assays, some of which are used or are being considered for use in cervical cancer screening. In general, the increase in sensitivity offset the decrease in specificity such that the overall accuracy, as measured by odds ratios, improved for all tests with an increasingly more rigorous definition of the precancer endpoint.

Table 4. Association of Carcinogenic Human Papillomavirus Test Detection by hc2, AMPLICOR, Linear Array, and Line Blot Assay, as Measured by Odds Ratios and 95% Confidence Intervals, With Different Disease Endpoints, ALTS, 1997–2001

	hc2		AMP		LA		LBA		Cytology	
	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval
Endpoint 1[a]	12	8.6, 15	7.3	5.6, 9.6	8.1	6.2, 10	7.4	5.8, 9.4	4.3	3.4, 5.4
Endpoint 2[a]	13	8.9, 18	8.4	6.1, 12	9.4	6.8, 13	8.1	6.1, 10	4.6	3.5, 6.0
Endpoint 3[a]	13	8.7, 21	10	6.8, 15	12	8.1, 19	8.7	6.2, 12	4.3	3.1, 5.8
Endpoint 4[a]	18	11, 30	12	7.5, 19	15	9.1, 23	10	6.9, 14	4.6	3.3, 6.3
Endpoint 5[a]	34	14, 83	18	9.0, 37	33	14, 81	17	9.2, 30	7.4	4.5, 12

Abbreviations: ALTS, Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study; AMP, AMPLICOR (Roche Molecular Systems, Inc., Pleasanton, California); CIN 2 and CIN 3, cervical intraepithelial neoplasia grades 2 and 3, respectively; hc2, Hybrid Capture 2 (Qiagen Corporation, Gaithersburg, Maryland); LA, linear array; LBA, line blot assay.

[a] Endpoint 1, any CIN-2 diagnosis or more severe (CIN 2+) by either clinical center or ALTS quality control pathology review; endpoint 2, a diagnosis of CIN 2 or more severe on both pathology reviews (consensus CIN 2+); endpoint 3, any diagnosis of CIN 3 by either clinical center or ALTS quality control pathology review; endpoint 4, a diagnosis of CIN 3 on one pathology review and CIN 2 or more severe on the other pathology review; and endpoint 5, CIN 3 diagnosed on both pathology reviews (consensus CIN 3). (Women with <CIN 2 on both pathology reviews are the reference group.)

One consideration in interpreting this analysis was that this study population was referred into ALTS because of an ASC-US Pap test and was mostly HPV positive (2,191 of 3,215 (68.2%) women tested positive for carcinogenic HPV by at least 1 test at baseline). There are 2 implications. First, there was evidence of some cytologic abnormalities in this population, which generally correlates with higher HPV viral loads. Therefore, the agreement between HPV tests might have been better than what might be expected if the evaluation were to be done in a general screening population, especially among women without precancerous lesions. Second, the control population, women with <CIN 2, was still highly exposed to carcinogenic HPV (62.7% were positive by at least 1 test, and 32.7% were positive by all 4 tests for carcinogenic HPV). As a consequence, the degree to which HPV detection was associated with consensus CIN 3 was most likely muted in this population compared with what might have been observed in a general screening population.

There are several implications of this analysis. The first and most obvious is that the choice of endpoints influences estimates of test performance. Diagnoses of CIN 2 and CIN 3 are the surrogate endpoints for cervical cancer used in clinical trials evaluating new screening tests. However, there is an increasing recognition that CIN 2 is an equivocal diagnosis of precancer, an admixture of HPV infections by both carcinogenic and noncarcinogenic HPV and misclassified CIN 3 (10). Approximately 25%–50% of CIN 2 will regress within a year or 2 years (11, 12). Although CIN 3 is a better surrogate for invasive potential, it is increasingly clear that it too is heterogeneous (44). Only a third of CIN 3/carcinoma in situ diagnosed in women with a median age in the late 30s invaded over 30 years (5). In this study, CIN 3 was diagnosed in much younger women (median and mean enrollment ages of 23.0 and 25.5 years) and was very small (45), presumably less mature, and therefore of less invasive potential than those cases of CIN 3/carcinoma in situ that unfortunately were observed without treatment in New Zealand (5).

Although there is no way to predict the invasive potential of the different categorizations of cervical lesions in this study, it is reasonable to assume that there is a spectrum of invasive potential with consensus CIN 3 perhaps having the greatest risk. To that point, consensus CIN-3 lesions were the most likely to be HPV-16 positive (the most carcinogenic of all HPV genotypes (46)) and the most likely to be accompanied by visual and microscopic evidence of precancer. Inclusion of CIN 2 (the recommended clinical threshold for treatment in the United States) in the definition of an endpoint in clinical trials likely results in a high proportion of diagnoses with little or no invasive potential. Yet, there was evidence of heterogeneity even among the consensus CIN-3 diagnoses, with some consensus CIN 3 not testing carcinogenic HPV positive by all tests, which correspondingly were associated with negative cytology and colposcopy results.

Thus, we must consider misclassification in trials and guidelines (47, 48). Current screening and vaccine trials are underpowered because of misclassification to an extent that is rarely appreciated, if the goal is to assess the clinical performance for detection of cervical precancerous lesions as proxy for *invasive* cervical cancer. Ironically, as biomarkers become more specific for cervical carcinogenesis, assays for their detection will appear to underperform compared with carcinogenic HPV DNA detection because there will be an increasing number of seemingly (''false'') negative results associated with CIN 2 and perhaps even with CIN 3, since not all CIN 3 is truly precancerous (i.e., having the potential to invade). We need such specific biomarkers to screen HPV–16/18 vaccinated populations, because the predictive value of positive cytology or HPV test results is expected to be diminished with the elimination of the most carcinogenic HPV genotypes (49, 50). The challenge for validating the next generation of biomarkers will be distinguishing between false and true negative test results in women with diagnoses of cervical precancerous lesions.

It will be especially important to avoid nonrandom misclassification. Such correlated errors (biases) can be severe and have been observed to affect clinical performance estimates of visual inspection after acetic acid when disease is assessed by colposcopy (51), both of which rely on the same measurement (visualization of the cervix).

A biorepository or ''biobank'' of Pap specimens, if properly constructed with the correct buffer and storage of specimens and thorough disease ascertainment, might be a better choice by which to assess the clinical performance for prevention of cervical cancer because more cases of rigorously defined precancerous lesions identified over time can be accrued. Such an approach, if validated and also accepted by regulatory entities, might achieve 2 admirable goals: more rigorous evaluations against the most important preinvasive disease and reducing the costs of validation, which will encourage promising screening tests to be developed and validated.

The importance of the example of HPV and cervical cancer extends to molecular epidemiology in general. Remembering how weak the initial association between HPV and cervical precancer seemed, we can wonder what other strong and important epidemiologic relations we might be missing because of type II errors resulting from misclassification of our tests and surrogate endpoints of disease (40, 41).

ACKNOWLEDGMENTS

Author affiliations: Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland (Philip E. Castle, Mark Schiffman, Nicolas Wentzensen); Departments of Molecular Genetics and Microbiology and of Obstetrics and Gynecology, School of Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico (Cosette M. Wheeler); and Departments of Epidemiology and of Molecular Microbiology and Immunology, Johns Hopkins University, Baltimore, Maryland (Patti E. Gravitt).

ALTS was supported by contracts CN-55153, CN-55154, CN-55155, CN-55156, CN-55157, CN-55158, CN-55159, and CN-55105 from the National Cancer Institute, National Institutes of Health, Department of Health and Human

Services. This research was also supported (in part) by the Intramural Research Program of the National Cancer Institute, National Institutes of Health. Some of the equipment and supplies used in these studies were donated or provided at reduced cost by Digene Corporation, Gaithersburg, Maryland; Cytyc Corporation, Marlborough, Massachusetts; National Testing Laboratories, Fenton, Missouri; DenVu, Tucson, Arizona; TriPath Imaging, Inc., Burlington, North Carolina; and Roche Molecular Systems, Inc., Alameda, California.

The authors thank the ALTS Group Investigators for their help in planning and conducting the trial.

Conflict of interest: none declared.

REFERENCES

1. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. 1999;189(1):12–19.
2. Muñoz N, Bosch FX, de Sanjosé S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003;348(6):518–527.
3. Wright TC Jr, Schiffman M. Adding a test for human papillomavirus DNA to cervical-cancer screening. *N Engl J Med*. 2003;348(6):489–490.
4. Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics, 2002. *CA Cancer J Clin*. 2005;55(2):74–108.
5. McCredie MR, Sharples KJ, Paul C, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncol*. 2008;9(5):425–434.
6. Green GH. The progression of pre-invasive lesions of the cervix to invasion. *N Z Med J*. 1974;80(525):279–287.
7. Castle PE, Cox JT, Jeronimo J, et al. An analysis of high-risk human papillomavirus DNA-negative cervical precancers in the ASCUS-LSIL Triage Study (ALTS). *Obstet Gynecol*. 2008;111(4):847–856.
8. Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*. 2001;285(11):1500–1505.
9. Wright TC Jr, Massad LS, Dunton CJ, et al. 2006 consensus guidelines for the management of women with abnormal cervical screening tests. *J Low Genit Tract Dis*. 2007;11(4):201–222.
10. Castle PE, Stoler MH, Solomon D, et al. The relationship of community biopsy-diagnosed cervical intraepithelial neoplasia grade 2 to the quality control pathology-reviewed diagnoses: an ALTS report. *Am J Clin Pathol*. 2007;127(5):805–815.
11. Castle PE, Schiffman M, Wheeler CM, et al. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol*. 2009;113(1):18–25.
12. Trimble CL, Piantadosi S, Gravitt P, et al. Spontaneous regression of high-grade cervical dysplasia: effects of human papillomavirus type and HLA phenotype. *Clin Cancer Res*. 2005;11(13):4717–4723.
13. Carreon JD, Sherman ME, Guillén D, et al. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *Int J Gynecol Pathol*. 2007;26(4):441–446.
14. Franco EL. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology*. 1991;2(2):98–106.
15. Franco EL. Measurement errors in epidemiological studies of human papillomavirus and cervical cancer. *IARC Sci Publ*. 1992;119:181–197.
16. Franco EL, Rohan TE, Villa LL. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. *J Natl Cancer Inst*. 1999;91(6):506–511.
17. Schiffman MH, Bauer HM, Hoover RN, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst*. 1993;85(12):958–964.
18. Kurman RJ, Malkasian GD Jr, Sedlis A, et al. From Papanicolaou to Bethesda: the rationale for a new cervical cytologic classification. *Obstet Gynecol*. 1991;77(5):779–782.
19. Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA*. 2002;287(16):2114–2119.
20. Schiffman M, Adrianza ME. ASCUS-LSIL Triage Study. Design, methods and characteristics of trial participants. *Acta Cytol*. 2000;44(5):726–742.
21. Human papillomavirus testing for triage of women with cytologic evidence of low-grade squamous intraepithelial lesions: baseline data from a randomized trial. The Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study (ALTS) Group. *J Natl Cancer Inst*. 2000;92(5):397–402.
22. ASCUS-LSIL Triage Study (ALTS) Group. A randomized trial on the management of low-grade squamous intraepithelial lesion cytology interpretations. *Am J Obstet Gynecol*. 2003;188(6):1393–1400.
23. ASCUS-LSIL Triage Study (ALTS) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. *Am J Obstet Gynecol*. 2003;188(6):1383–1392.
24. Solomon D, Schiffman M, Tarone R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst*. 2001;93(4):293–299.
25. Schiffman M, Wheeler CM, Dasgupta A, et al. A comparison of a prototype PCR assay and hybrid capture 2 for detection of carcinogenic human papillomavirus DNA in women with equivocal or mildly abnormal Papanicolaou smears. *Am J Clin Pathol*. 2005;124(5):722–732.
26. Gravitt PE, Peyton CL, Alessi TQ, et al. Improved amplification of genital human papillomaviruses. *J Clin Microbiol*. 2000;38(1):357–361.
27. Peyton CL, Gravitt PE, Hunt WC, et al. Determinants of genital human papillomavirus detection in a US population. *J Infect Dis*. 2001;183(11):1554–1564.
28. Castle PE, Gravitt PE, Solomon D, et al. Comparison of linear array and line blot assay for detection of human papillomavirus and diagnosis of cervical precancer and cancer in the Atypical Squamous Cell of Undetermined Significance and Low-Grade Squamous Intraepithelial Lesion Triage Study. *J Clin Microbiol*. 2008;46(1):109–117.
29. Wentzensen N, Gravitt PE, Solomon D, et al. A study of Amplicor human papillomavirus DNA detection in the Atypical Squamous Cells of Undetermined Significance-Low-Grade Squamous Intraepithelial Lesion Triage Study. *Cancer Epidemiol Biomarkers Prev*. 2009;18(5):1341–1349.
30. Bosch FX, Manos MM, Muñoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International Biological Study on Cervical Cancer (IBSCC) Study Group. *J Natl Cancer Inst*. 1995;87(11):796–802.

31. Cogliano V, Baan R, Straif K, et al. Carcinogenicity of human papillomaviruses. *Lancet Oncol*. 2005;6(4):204.
32. Bouvard V, Baan R, Straif K, et al. A review of human carcinogens—Part B: biological agents. *Lancet Oncol*. 2009; 10(4):321–322.
33. Castle PE, Solomon D, Wheeler CM, et al. Human papillomavirus genotype specificity of hybrid capture 2. *J Clin Microbiol.* 2008;46(8):2595–2604.
34. Pretorius RG, Zhang WH, Belinson JL, et al. Colposcopically directed biopsy, random cervical biopsy, and endocervical curettage in the diagnosis of cervical intraepithelial neoplasia II or worse. *Am J Obstet Gynecol*. 2004;191(2):430–434.
35. Gage JC, Hanson VW, Abbey K, et al. Number of cervical biopsies and sensitivity of colposcopy. *Obstet Gynecol*. 2006; 108(2):264–272.
36. Guido R, Schiffman M, Solomon D, et al. Postcolposcopy management strategies for women referred with low-grade squamous intraepithelial lesions or human papillomavirus DNA-positive atypical squamous cells of undetermined significance: a two-year prospective study. *Am J Obstet Gynecol*. 2003;188(6):1401–1405.
37. Atkins KA, Jeronimo J, Stoler MH. Description of patients with squamous cell carcinoma in the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study. *Cancer.* 2006;108(4):212–221.
38. Cuzick J. A Wilcoxon-type test for trend. *Stat Med*. 1985;4(1): 87–90.
39. Gravitt PE, Burk RD, Lorincz A, et al. A comparison between real-time polymerase chain reaction and hybrid capture 2 for human papillomavirus DNA quantitation. *Cancer Epidemiol Biomarkers Prev.* 2003;12(6):477–484.
40. Schiffman MH, Schatzkin A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res.* 1994;54(7 suppl):1944s–1947s.
41. Schatzkin A, Freedman LS, Dorgan J, et al. Surrogate end points in cancer research: a critique. *Cancer Epidemiol Biomarkers Prev.* 1996;5(12):947–953.
42. Reeves WC, Brinton LA, García M, et al. Human papillomavirus infection and cervical cancer in Latin America. *N Engl J Med*. 1989;320(22):1437–1441.
43. Bosch FX, Lorincz A, Muñoz N, et al. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol*. 2002;55(4):244–265.
44. Schiffman M, Rodriguez AC. Heterogeneity in CIN3 diagnosis. *Lancet Oncol*. 2008;9(5):404–406.
45. Sherman ME, Wang SS, Tarone R, et al. Histopathologic extent of cervical intraepithelial neoplasia 3 lesions in the Atypical Squamous Cells of Undetermined Significance Low-Grade Squamous Intraepithelial Lesion Triage Study: implications for subject safety and lead-time bias. *Cancer Epidemiol Biomarkers Prev.* 2003;12(4):372–379.
46. Smith JS, Lindsay L, Hoots B, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer.* 2007; 121(3):621–632.
47. Stoler MH, Castle PE, Solomon D, et al. The expanded use of HPV testing in gynecologic practice per ASCCP-guided management requires the use of well-validated assays. *Am J Clin Pathol*. 2007;127(3):335–337.
48. Meijer CJ, Berkhof J, Castle PE, et al. Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. *Int J Cancer.* 2009;124(3):516–520.
49. Khan MJ, Castle PE, Lorincz AT, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. *J Natl Cancer Inst*. 2005;97(14):1072–1079.
50. Castle PE, Solomon D, Saslow D, et al. Predicting the effect of successful human papillomavirus vaccination on existing cervical cancer prevention programs in the United States. *Cancer.* 2008;113(10 suppl):3031–3035.
51. Pretorius RG, Kim RJ, Belinson JL, et al. Inflation of sensitivity of cervical cancer screening tests secondary to correlated error in colposcopy. *J Low Genit Tract Dis*. 2006;10(1):5–9.