A pooled analysis to compare the clinical characteristics of human papillomavirus-positive and -negative cervical precancers

Running Title: HPV-positive and -negative cervical precancers

Philip E. Castle[1*], Amanda J. Pierz [1], Rachael Adcock[2], Shagufta Aslam[3], Partha S. Basu[4], Jerome L. Belinson[5], Jack Cuzick[2], Mariam El-Zein[6], Catterina Ferreccio[7,8], Cynthia Firnhaber[9], Eduardo L. Franco[6], Patti E. Gravitt[10], Sandra D. Isidean[6], John Lin[11], Salaheddin M. Mahmud[12], Joseph Monsonego[13], Richard Muwonge[4], Samuel Ratnam[6], Mahboobeh Safaeian[3], Mark Schiffman[14], Jennifer S. Smith[15], Avril Swarts[16], Tom Wright[17], Vanessa Van De Wyngard[7,8], Long Fu Xi[11]

[1]Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Bronx, New York, USA

[2]Queen Mary University of London, Centre for Cancer Prevention, London, United Kingdom

[3]Roche Molecular Diagnostics, Pleasanton, CA, USA

[4]International Agency for Research on Cancer, Screening Group, Lyon, France

[5]Preventive Oncology International and the Cleveland Clinic, Cleveland, Ohio, USA

[6]McGill University, Division of Cancer Epidemiology, Montréal, Quebec, Canada.

[7]Advanced Center for Chronic Diseases, ACCDiS, Santiago, Chile

[8]Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

[9]University of Colorado, School of Medicine, Aurora, Colorado, USA

[10]Johns Hopkins University, Department of Epidemiology, Baltimore, Maryland, USA

[11]University of Washington, HPV Research Group, Seattle, Washington, USA

[12]University of Manitoba, Max Rady College of Medicine, Rady Faculty of Health Sciences, Department of Community Health, Winnipeg, Manitoba, Canada

[13]Institute of the Cervix, Federation Mutualiste Parisienne, Paris, France

[14]National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology & Genetics, Bethesda, Maryland, USA

[15]University of North Carolina at Chapel Hill, Department of Epidemiology, Chapel Hill, North Carolina, USA

[16]University of the Witwatersrand, Department of Medicine, Johannesburg, South Africa

[17]Columbia University, Department of Pathology and Cell Biology, New York, New York, USA

*Corresponding Author: National Cancer Institute/National Institutes of Health, 9609 Medical Center Dr., Room 5E410, Rockville, MD 20850, USA; phone: +1 (703) 772-0611; email: castle.philip@gmail.com

Conflict of Interest Statements: Dr. Castle has received HPV tests and assays for research at reduced or no cost from Roche, Becton Dickinson, Cepheid, and Arbor Vita Corporation. Dr. Belinson received grant support from Hologic and in-kind support from Qiagen and BGI Shenzhen for the studies cited. NCI (Dr. Schiffman) has received cervical screening results and supplies for independent research at reduced or no cost from Roche, Becton Dickinson, Qiagen, MobileODT, and Arbor Vita Corporation. Dr. Ratnam received research grants, honoraria and

travel assistance from Roche Diagnostics and Abbott Molecular. Dr. Franco served as an occasional advisor for companies involved with HPV vaccines (Merck, GSK) and HPV diagnostics (Roche Diagnostics) and as a Steering Committee Member for a publicly funded study in Finland that received support from GSK. Dr. Franco and Dr. El-Zein hold a patent related to the discovery "DNA methylation markers for early detection of cervical cancer", registered at the Office of Innovation and Partnerships, McGill University, Montreal, Quebec, Canada (October, 2018). A provisional utility patent application before the United States Patent & Trademark Office was also filed (November, 2018). Dr. Mahmud has received unrestricted research grants from Merck, GlaxoSmithKline, Sanofi Pasteur, Pfizer and Roche-Assurex for unrelated studies. SMM has received fees as an advisory board member for Sanofi Pasteur. Dr. Ferreccio received screening supplies for independent research at reduced cost from Qiagen. Dr. Monsonego has served as an advisor for Merck, Sanofi Pasteur MSD, Gen-Probe, Roche Diagnostics, Qiagen, and Becton Dickinson. Dr. Wright is a consultant and speaker for Roche and BD Diagnostics and a consultant to Inovio. Dr. Smith has received consultancies, donations-in-kind, and contracts from Hologic and Becton Dickenson, and donations-in-kind from Arbor Vita. Dr. Cuzick reports grants and personal fees from Qiagen, grants from Hologic, grants and personal fees from Becton Dickinson (BD), grants and personal fees from Genera Biosystems (GB), grants from Gene First, and grants from Trovagene, outside the submitted work.

Acknowledgements: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of

the International Agency for Research on Cancer / World Health Organization. Dr. Mahmud's work is supported, in part, by funding from the Canada Research Chairs Program.

Funding: None

Abstract (n=223)

Given that high-risk human papillomavirus (HPV) is the necessary cause of virtually all cervical cancer, the clinical meaning of HPV-negative cervical precancer is unknown. We therefore conducted a literature search in Ovid MEDLINE, PubMed Central®, and Google Scholar to identify English-language studies in which 1) HPV-negative and positive, histologically confirmed cervical intraepithelial neoplasia grade 2 or more severe diagnoses (CIN2+) were detected and 2) summarized statistics or de-identified individual data were available to summarize proportions of biomarkers indicating risk of cancer. Nineteen studies including 3,089 (91.0%) HPV-positive and 307 (9.0%) HPV-negative CIN2+ were analyzed. HPV-positive CIN2+ (vs. HPV-negative CIN2+) was more likely to test positive for biomarkers linked to cancer risk: a study diagnosis of CIN3+ (vs. CIN2) (18 studies, 0.56 vs. 0.24, $p<0.001$) preceding HSIL+ cytology (15 studies, 0.54 vs. 0.10, $p<0.001$); and high-grade colposcopic impression (13 studies, 0.30 vs. 0.18, $p=0.03$). HPV-negative CIN2+ was more likely to test positive for low-risk HPV genotypes than HPV-positive CIN2+ ($p<0.001$). HPV-negative CIN2+ appears to have lower cancer risk than HPV-positive CIN2+. Clinical studies of human high-risk HPV testing for screening to prevent cervical cancer may refer samples of HPV test-negative women for disease ascertainment to correct verification bias in the estimates of clinical performance. However, verification-bias adjustment of the clinical performance of HPV testing may over-correct/underestimate its clinical performance to detect truly precancerous abnormalities.

Introduction

The discovery that specific, high-risk human papillomavirus (HPV) genotypes cause virtually all cervical cancer, as well as most anal, vaginal, vulvar, and penile cancers, and a significant proportion of oropharyngeal cancers, has led to changes in how we prevent these cancers, including prophylactic HPV vaccination for primary prevention and HPV detection screening for secondary prevention of cervical cancer. Of the latter, there are now several HPV tests that have FDA approval for cervical screening either alone ("primary HPV testing") or in combination with concurrent cytologic/Pap testing ("co-testing"). The primary advantage of including HPV detection in routine cervical screening is that a negative HPV test or co-test provides better reassurance against cervical cancer and its immediate precursor abnormalities, cervical intraepithelial neoplasia (CIN) grade 3 (CIN3), CIN grade 2 (CIN2), and adenocarcinoma *in situ* (AIS)(1-6). Thus, using an HPV test in routine cervical screening safely allows for a lengthening of screening intervals between negative results.

In theory, calculation of diagnostic performance indices, such as sensitivity and specificity, must take into account the possibility of verification bias, which results from unequal verification of the presence of disease between test positive and test negative subjects (7-10). Clinical trials to evaluate the accuracy of HPV and other screening tests often have included verification-bias adjustment (VBA) in the study design. VBA is an imputation method intended to correct for the inability of the investigator to verify the presence of disease among those who tested negative on screening. It relies on randomly sampling those who were in the latter category, and reweighting before calculating test performance, to simulate complete disease ascertainment.

This is accomplished by sending subjects/patients with negative screening test results for further evaluation by colposcopy and biopsy. Using the sampling fraction, one can then estimate via

extrapolation the number of true negative (TN) and false negative (FN) cases, thus enabling the reconstitution of the unobserved underlying 2x2 table that would ideally be used to measure the clinical performance of the test under evaluation.

Although VBA is a standard statistical method used in screening studies, one caveat is that it assumes that the cases with FN and true–positive (TP) results have the same clinical/biological importance. In the case of cancer prevention, this would mean that FN and TP cases have similar invasive or oncogenic potential. Specifically, for the evaluation of an HPV test, this would imply that HPV-negative (FN-test) CIN2 or more severe (CIN2+) cervical abnormalities have similar risk of becoming invasive cervical cancer as HPV-positive (TP-test) CIN2+. As an example of the impact on VBA, the crude versus VBA-adjusted sensitivity of high-risk HPV by the cobas test (Roche Molecular Systems, Pleasanton, CA, USA) for CIN2+ was 92.0% and 75.1%, respectively (11).

However, given that high-risk HPV causes virtually all cervical cancer, the clinical meaning of HPV-negative CIN2/3 is uncertain, particularly given the subjective nature of both the colposcopic impression that guides sampling for diagnostic biopsy (12-16) and the histopathologic interpretation of the tissue sample (17-21). Therefore, it is unclear whether these assumptions for VBA are valid in the context of evaluating the performance HPV tests. To address this question, we conducted a literature search to identify studies that diagnosed both HPV-positive and –negative CIN2+ for a pooled analysis to examine whether in this context the basic tenet that FN and TP cases are biologically equivalent and therefore can be used to correct via VBA the clinical performance of HPV tests for screening to prevent cervical cancer.

Methods

An extensive literature search, using the following search string "(HPV OR (human AND papillomavirus)) AND ((verification AND bias) OR (Cytology OR PAP OR VIA or visual inspection) AND (CIN2 OR "CIN 2" OR CIN-2 OR CIN3 OR "CIN 3" OR CIN-3 OR precancer OR pre-cancer))" was conducted in Ovid MEDLINE, PubMed Central®, and Google Scholar for all relevant studies in the English language. The goal of the search was to identify studies evaluating HPV tests in which the study design included colposcopic referral of HPV-positive and –negative CIN2+, with the latter as the result of direct referral or the result of another assay testing positive (while the primary study HPV test was negative). Studies were included if histologic endpoints were available and either summary statistics or de-identified individual data were provided for the CIN2+ cases diagnosed in the study.

Proportions of diagnoses (CIN2 vs. CIN3 for primary or secondary endpoint diagnoses), other tests (e.g., a second HPV test or visual inspection after acetic acid (VIA)), cytology (i.e., high-grade cytology vs. not), and colposcopic impression (i.e., high-grade vs. not) were compared between HPV-positive and –negative-CIN2+ using Metaprop (22), a STATA command for pooling binomial data. We classified HPV genotyping results hierarchically according to cancer risk (23, 24): HPV16 positive, else HPV16 negative but positive for other high-risk HPV types, else negative for all high-risk HPV types but positive for low-risk HPV types, else negative for all HPV genotypes tested (HPV16 > other high-risk HPV > low-risk HPV > negative). However, we tested proportion of each HPV genotype category for HPV-negative and HPV-positive CIN2+ by running separate binomial models for each i.e., each category was independent. Thus the sum of the HPV categories is not constrained to equal one (unity).

158 STATA (version 15.1; College Station, TX, USA) was used for all analyses. A $p<0.05$ was
159 considered statistically significant.

Results

Nineteen studies included 3,396 cases of CIN2+, 3,089 (91.0%) that tested positive and 307 (9.0%) that tested negative by the primary study HPV test, were included in this analysis (**Table 1**) (25-43). We included one study conducted in human immunodeficiency virus (HIV)-infected women, in which women with abnormal cytology, a positive VIA, and a 25% random sample of the cytology- and VIA-negatives were referred to colposcopy and HPV testing was done on all women but was not the basis of referral to colposcopy (40). Exclusion of this study did not appreciably change our findings. We also included colposcopy referral arm of one RCT for the management of minor cytological abnormalities (26). Exclusion of this study did not appreciably change our findings.

Table 2 summarizes the main results. Women diagnosed with HPV-positive CIN2+ were less likely to be aged 40 years and older compared with HPV-negative CIN2+ (0.30 vs. 0.34, respectively, 19 studies, p=0.03). There was no difference in the proportion of women who smoked (9 studies, p=0.62) between those diagnosed with HPV-positive and –negative CIN2+.

Figure 1-4 shows forest plots (blobbograms) for some of the main comparisons, of the analysis. Women with a HPV-positive CIN2+ were two-fold more likely than those with HPV-negative CIN2+ to have a diagnosis of CIN3+ (vs. CIN2) (18 studies, 0.56 vs. 0.24, respectively, p<0.001) (**Figure 1**). HPV-positive CIN2+ was three-fold more likely than HPV-negative CIN2+ to have an antecedent (referral) HSIL+ cytology (15 studies, 0.34 vs. 0.10, respectively p<0.001) (**Figure 2**). HPV-positive CIN2+ was less likely than HPV-negative CIN2+ to be VIA positive (6 studies, 0.57 vs. 0.83, respectively p<0.001) (**Figure 3**); exclusion of the Chile study (38), which referred HPV-negative women to colposcopy based on a stratified sampling of VIA results (all VIA positives [n=117], VIA indeterminate [n=110], and VIA negative with cervical

cancer risk factors [n=68]) did not appreciably change these results (5 studies, 0.63 vs. 0.81, respectively, p<0.001). HPV-positive CIN2+ was more likely than HPV-negative CIN2+ to have a high-grade colposcopic impression (13 studies, 0.30 vs. 0.18, respectively, p=0.03) (**Figure 4**). Exclusion of the two studies (31, 32) in which the colposcopists were not masked to the HPV results did not appreciably change the relationship of HPV status and the appearance of high-grade colposcopic impression.

In 6 studies, a second clinical HPV test was done on the same cervical specimen and was more likely to test positive for HPV-positive CIN2+ compared to the HPV-negative CIN2+ (0.97 vs. 0.36, respectively, p<0.001) (**Table 2 and Supplemental Figure 1**). In 4 studies, HPV testing of self-collected cervicovaginal specimens were also more likely to be positive for HPV-positive CIN2+ compared to the HPV-negative CIN2+ although the differences were surprisingly small (0.95 vs. 0.84, respectively, p<0.001) (**Table 2 and Supplemental Figure 1**).

A secondary analysis was conducted on the HPV genotyping results that were available from 9 studies (26, 30, 33-35, 37, 39, 42, 43), of which one (39) conducted HPV genotyping on biopsied tissue that led to the diagnosis (**Table 2 and Supplemental Figure 2**). HPV genotyping results in the individual studies were grouped hierarchically into broad categories of cancer risk (HPV16>other high-risk HPV>low-risk HPV>negative) (n.b., because we ran separate models for each HPV genotyping category, the sum of categories does not equal unity and therefore the results do not represent attributable fractions of each category. Only the proportion within HPV genotyping category can be compared between HPV-negative and HPV-positive CIN2+.). HPV-positive CIN2+ was more likely than HPV-negative CIN2+ to test positive for HPV16 (0.46 vs. 0.09, respectively, p<0.001) and other high-risk HPV genotypes (0.49 vs 0.32, respectively, p<0.001). HPV-positive CIN2+ was less likely than HPV-negative CIN2+ to test positive for

low-risk HPV genotypes (0.00 vs. 0.13, respectively, p<0.001). HPV-positive CIN2+ was less likely than HPV-negative CIN2+ to test negative for any HPV genotype (0.02 vs. 0.33, respectively, p<0.001). Exclusion of the one study (30) in which the HPV test that was under evaluation also provided the HPV genotyping data for these correlative analyses did not appreciably change the results.

Several studies included marker/biomarker testing results that, while not sufficiently commonly done for data pooling, are summarized in **Table 3**. Data on p16 immunohistochemistry of the CIN2+ was available from two studies. One study found that HPV-positive CIN2+ was significantly more likely to test positive by p16 immunohistochemistry than HPV-negative CIN2+ (37) while the other found similar but non-significant difference (34). There was no difference in detection of HPV16/18/45 E6 in HPV-positive CIN2+ as compared to HPV-negative CIN2+ based on one study (39).

Discussion

We combined data from several epidemiologic studies and clinical trials of HPV testing to compare the distribution of other biomarkers of cervical cancer risk among HPV-positive and HPV-negative CIN2+. The latter was detected because the source studies included in their design a protocol to perform a verification-bias adjustment (e.g., a sample of HPV-negative women were referred to colposcopy) or because referral to colposcopy was based on the positive result of another (non-HPV) test. We found evidence that the HPV-negative CIN2+ is a distinct biological and clinical entity compared with HPV-positive CIN2+ and thus would likely carry lower invasive cervical cancer risk. While the validity-seeking exercise of the VBA is meant to bring sensitivity and specificity estimates closer to the theoretical parameters that represent the performance of the test in detecting the full spectrum of disease, it introduces a departure from the real-world conditions of screening practice. The distortion comes from ignoring the heterogeneity of current histopathologic standards of cervical precancer, specifically leading to the unintended clinical consequence of considering TP-test and FN-test cases equivalent. That is, the VBA correction, when applied to HPV testing, makes FN-test cases become an overrepresentation in the totality of discoverable true disease, leading to an underestimation of the test performance relative to benchmarks of clinical utility. Strategies for better estimation of test performance are discussed below.

Our analysis shows that HPV-negative CIN2+ results were a mixture of CIN2+ diagnoses that 1) tested falsely negative (FN) for high-risk HPV, 2) were caused by HPV types not considered of high carcinogenic risk, and 3) epithelial changes that have the appearance of CIN2/3 but are benign mimics, such as immature squamous metaplasia, atrophy, reparative epithelial changes, and/or tangential sectioning on routine staining (44-46). FN-test CIN2+ are those that the HPV

test should have detected and as such, should count against its clinical performance. These FN-test CIN2+ include some CIN2/3 that are likely to be low-volume, low-area that as a consequence are poorly sampled and test falsely HPV negative i.e., they are truly HPV-positive CIN2/3 but test HPV negative due to poor representation in the cervical exfoliative sample (45, 47). This subset of FN-test CIN2+ are true test failures and cannot be discounted.

We observed that only about one-third of the HPV-negative CIN2+ tested high-risk HPV positive by a second, clinical test whereas virtually all of the HPV-positive CIN2+ tested positive. Thee data confirm that HPV-negative CIN2+ cases included some FN-test CIN2+ and TN-test CIN2+. Interestingly, in the small sub-group of studies, the proportion of HPV-negative CIN2+ that tested high-risk HPV positive on a self-collected cervicovaginal was high albeit still less than the for HPV-positive CIN2+. We speculate that the high proportion of high-risk HPV positivity for the self-collected specimen among the HPV-negative CIN2+ is due in part to the detection of vaginal HPV infections unrelated to the CIN2+ (48, 49).

Because those CIN2+ diagnoses caused by other types not classified as high-risk HPV are very unlikely to cause invasive cancer (50, 51), these cases should not be counted strictly as test failures. The goal of screening is ultimately to prevent cancer, via the detection of cervical pre-cancer that have significant propensity to progress to cancer, thus enabling treatment to arrest their development. Thus, these HPV types should not be included in clinical HPV tests as they can have the potential to classify more women as HPV positive and result in diagnosis of CIN2 with low progression potential. As a consequence, these diagnoses potentially would result in additional (unnecessary) treatments, which is a risk factor for pre-term delivery (52, 53), without the compensatory benefit of cancer prevention. A case-in-point is the inclusion of HPV66 in the current clinical high-risk HPV tests, which was momentarily believed to be another high-risk

HPV type (54). However, it is now recognized that HPV66 very rarely causes cancer but does increase the test positivity (50, 51, 54).

Some HPV-negative CIN2+ could be the result of abnormalities of smaller volume and/or lower viral shedding. We would expect that these, like truly HPV-negative CIN2+, would be of lower invasive potential. Larger CIN3, for example, found in older women have been hypothesized to have much greater invasive potential than early, small, incipient CIN3 diagnosed in young women (55, 56).

Finally, some epithelial changes appear to be visual and morphological look-alikes of CIN2/3 but may be less likely to represent true precursors to cancer (44, 45). Like those CIN2/3 caused by HPV types not classified as high risk, these "mimics" are best left undiagnosed, given that they will not cause cancer, and thereby avoiding any unnecessary treatment. These changes may be due in part to a common cause, atrophic cervical epithelial changes in peri- and post-menopausal women. These epithelial changes appear as acetowhitened cervical tissue, which are called positive by VIA and giving colposcopic impression of cervical abnormality, albeit not high-grade, leading to a biopsy and diagnosis.

Another implication of our findings is that the endpoint of CIN2/3 is very heterogeneous in its clinical importance i.e., invasive cervical cancer risk, likely even more so than the heterogeneity of CIN3, a more certain diagnosis of pre-cancer (55, 56). Thus, simple accounting of the detected and missed CIN2+ does not reflect accurately the true sensitivity of a cervical screening test *to prevent cervical cancer*. Rather than use crude sensitivity, a measure of clinically-relevant sensitivity, one in which all endpoints of CIN2/3 (regardless of the result of the HPV test being evaluated) are weighted according to their invasive cervical cancer risk or potential, should be used:

$$\frac{\left(\sum_{i=1}^{y} x_i\, n_i\right)_{pos}}{\left(\sum_{i=1}^{y} x_i n_i\right)_{All}}$$

in which x_i is the specific weighting factor and n_i is the number of abnormalities with the characteristics i.

However, we know very little about the factors that determine invasive potential of CIN2/3 and therefore their corresponding weighting. Almost certainly, how long the CIN2/3 has been present (a surrogate for which is the woman's age), its size (56), the causal HPV type, whether it is CIN2 or CIN3, if the woman has human immunodeficiency virus co-infection and her immune competency (57), etc. likely influence the risk of invasion. Unfortunately, we do not yet know how to weight the invasive risk of CIN2/3 for many of these factors.

However, one possible way to improve the estimated performance for the prevention of cervical cancer is to weight individual CIN2/3 diagnoses based on the HPV type present. Doing so, the above equation then would be written as:

$$\frac{\left(\sum_{i=1}^{y} X_{HPVi}\, n_{HPVi}\right)_{pos}}{\left(\sum_{i=1}^{y} X_{HPVi} n_{HPVi}\right)_{All}}$$

in which X_{HPVi} is the weighting factor for HPV type i.

Intuitively, it is clear that HPV16 is the most important type for causing cancer, i.e., more carcinogenic than all other types (23, 24). Likewise, prophylactic vaccines against HPV16 and HPV18 are projected to prevent approximately 70% of cervical cancers, not based on the number of percentage of HPV infections that are HPV16 and/or HPV18 but the fraction of cervical cancers they cause. Therefore, more importance should be placed on the detection or missed

detection of HPV16-related cervical abnormalities than comparable abnormalities caused by other HPV types.

To this point, we note that the HPV-negative CIN2+ cases from this analysis were unlikely to test HPV16 positive whereas approximately half of the HPV-positive CIN2+ tested HPV16 positive, which is generally the expected proportion of HPV16 positives (51). Each individual study included in this analysis found a greater proportion of HPV16 among the HPV-positive CIN2+ than HPV-negative CIN2+. This was true even for the study (39) in which the tissues were HPV genotyped, which found 17% of the HPV-negative CIN2+ and 62% of the HPV-positive CIN2+ tested HPV16 positive, which cannot be explained by differences in viral shedding. Again, these data support the inference that these HPV-negative CIN2+ have lower invasive cancer risk on the whole and a significant proportion carry virtually zero risk. Although we do not have prospective data to calculate type-specific transition probabilities from CIN2 or CIN3 to cancer for each HPV type, HPV type-specific weights might be derived from cross-sectional studies of HPV types and grade of disease as crude approximations. For example, using data from a large meta-analysis by Guan *et al*. (51), the ratio of invasive cervical cancer to CIN3 is 1.08 for HPV16 and 0.31 for HPV68. Therefore, a HPV68-related CIN3 should be discounted (weighted) by 0.29 compared to a HPV16-related CIN3; the weighting and that HPV16-related CIN3 is more common than HPV68-related CIN3 might reasonably approximate the relative importance of the two HPV type-specific CIN3. Using these same or similar data, the abnormality could be discounted further (weighted less) if it is CIN2 rather than a CIN3 diagnosis. Although these weighting factors are likely to be rough approximations of the relative importance of abnormalities by HPV type and diagnosis, using them would provide a more accurate representation of the potential value in detecting a given abnormality and initiating

treatment than treating all CIN3 equally or worse yet, all CIN2/3 equally. Future clinical evaluations of individual HPV tests might consider such an approach.

As corollary of our observations with cervical cancer screening, we propose that the same principle could be applied to other screening interventions focused on detection of precursors of other cancers, with the weighting informed by either empirical (natural history) data or modeling on the invasive cancer risk. For example, ductal carcinoma *in situ* of the breast is known to be heterogeneous in terms of cancer risk and in certain lower-risk groups, analogous to surveillance rather than immediate treatment of CIN2 (58), treatment is being "de-escalated" (59). Likewise, colorectal cancer risk following diagnosis of polyp(s) varies by size, type (adenomatous vs. non-adenomatous), number, histological type, and location of the polyp(s) as well as patients' age and sex (60-62).

The most important limitation of this analysis is that the actual invasive potential of each CIN2/3 diagnosed is unknown i.e., correlative measures were used to judge the relative clinical importance of each CIN2+ diagnosis. It is unethical to observe the development of cervical cancer from high-grade cervical abnormalities as was previously done (55). Although CIN2 diagnosed in women under the age of 30 years can be followed according to certain guidelines (58) because it is highly regressive (63), CIN3 is typically treated by excision therapies to avert invasive cancer. Even CIN3, the most severe pre-malignant cervical diagnosis, is not pre-cancer *per se* i.e., it is not synonymous with having or developing invasive cervical cancer as only approximately one-third of CIN3 diagnosed in older women will become invasive cervical cancer if untreated (55). We therefore used biomarkers that are associated with higher risk of cancer rather than directly observing which CIN2+ would develop into frank cancer.

There were other important limitations, notably heterogeneity between studies e.g., differences in HPV tests used and the percentage of sampling and selection of cases among the HPV-negative CIN2/3. There were not enough studies to conduct separate stratified analyses for each HPV test; however, many of the HPV tests used in studies included in this analysis have been shown to have good agreement with one another because they were developed in accordance with international standards for clinical HPV tests (64). A study using a less sensitive HPV test would have resulted in more FN-CIN2+ that are clinically relevant.

In addition, some data were not available to us and therefore we cannot call this a true systematic review. Although there was significant heterogeneity in the relative and absolute effects between studies, patterns/trends were generally consistent across studies. We therefore think that it is unlikely that the few missing studies biased our findings, although we cannot rule out this possibility entirely either.

In conclusion, we pooled data across many studies to demonstrate that HPV-negative CIN2+ systematically were less like to test positive for any biomarkers associated with invasive cancer risk compared to HPV-positive CIN2+. Thus, the use of HPV-negative CIN2+ in VBA calculations may result in systematic over-correction of sensitivity for CIN2+ due to presumptive FN HPV results. Although VBA is an established approach to derive numerically correct estimates of test performance in screening studies, the corrective effect comes at the expense of over-estimating the clinical relevance of insipient, non-progressive CIN2/3. *The true limitation of VBA for HPV testing validation is not VBA itself, which is mathematically correct, but rather our limited ability to differentiate by any means the clinically relevant from the irrelevant CIN2/3*. Therefore, its use must be cautiously interpreted because the VBA correction distorts the

371 disease detection reality by giving excessive value to low-grade disease i.e., false-positive
372 diagnoses.

373 Biomarkers applied to the diagnostic tissue may help. However, p16 immunohistochemistry,
374 which is recommended for differentiating high-risk and low-risk CIN2 (44), does so imperfectly
375 (65). Hence, using VBA, with or without p16 immunohistochemistry, still will result in some
376 underestimation of clinical performance of a cervical cancer screening test, such as an HPV
377 assay. A more general framework of mathematical adjustment of the clinical performance for the
378 detection of pre-invasive disease, using weighting based on the risk of invasive disease derived
379 from the natural history of the cancer, may provide a better estimate of the true effectiveness for
380 an intervention to prevent cancer.

Reference List

(1) Sherman ME, Lorincz AT, Scott DR, et al. Baseline cytology, human papillomavirus testing, and risk for cervical neoplasia: a 10-year cohort analysis. J Natl Cancer Inst 2003;95(1):46-52.

(2) Dillner J, Rebolj M, Birembaut P, et al. Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study. BMJ 2008;337:a1754. doi: 10.1136/bmj.a1754.:a1754.

(3) Sankaranarayanan R, Nene BM, Shastri SS, et al. HPV screening for cervical cancer in rural India. N Engl J Med 2009;360(14):1385-94.

(4) Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383(9916):524-32.

(5) Gage JC, Schiffman M, Katki HA, et al. Reassurance against future risk of precancer and cancer conferred by a negative human papillomavirus test. J Natl Cancer Inst 2014;106(8):dju153.

(6) Dijkstra MG, van ZM, Rozendaal L, et al. Safety of extending screening intervals beyond five years in cervical screening programmes with testing for high risk human papillomavirus: 14 year follow-up of population based randomised cohort in the Netherlands. BMJ 2016;355:i4924. doi: 10.1136/bmj.i4924.:i4924.

(7) Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983;39(1):207-15.

(8) Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. Stat Methods Med Res 1998;7(4):337-53.

(9) O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. BMJ Evid Based Med 2018;23(2):54-5.

(10) Franco EL. Statistical issues in human papillomavirus testing and screening. Clin Lab Med 2000;20(2):345-67.

(11) Castle PE, Stoler MH, Wright TC, Jr., et al. Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: a subanalysis of the ATHENA study. Lancet Oncol 2011;12(9):880-90.

(12) Liu AH, Gold MA, Schiffman M, et al. Comparison of Colposcopic Impression Based on Live Colposcopy and Evaluation of Static Digital Images. J Low Genit Tract Dis 2016;20(2):154-61.

(13) Jeronimo J, Massad LS, Castle PE, Wacholder S, Schiffman M. Interobserver agreement in the evaluation of digitized cervical images. Obstet Gynecol 2007;110(4):833-40.

(14) Hopman EH, Voorhorst FJ, Kenemans P, Meyer CJ, Helmerhorst TJ. Observer agreement on interpreting colposcopic images of CIN. Gynecol Oncol 1995;58(2):206-9.

(15) Garutti P, Cristiani P, Fantin GP, et al. Interpretation of colposcopy in population-based cervical screening services in north-eastern Italy: an online interregional agreement study. Eur J Obstet Gynecol Reprod Biol 2016;206:64-9.

(16) Cristiani P, Costa S, Schincaglia P, et al. An online quality assurance program for colposcopy in a population-based cervical screening setting in Italy: results on colposcopic impression. J Low Genit Tract Dis 2014;18(4):309-13.

(17) Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. JAMA 2001;285(11):1500-5.

(18) Stoler MH, Ronnett BM, Joste NE, et al. The interpretive variability of cervical biopsies and its relationship to HPV status. Am J Surg Pathol 2015;39(6):729-36.

(19) Cai B, Ronnett BM, Stoler M, et al. Longitudinal evaluation of interobserver and intraobserver agreement of cervical intraepithelial neoplasia diagnosis among an experienced panel of gynecologic pathologists. Am J Surg Pathol 2007;31(12):1854-60.

(20) Klaes R, Benner A, Friedrich T, et al. p16INK4a immunohistochemistry improves interobserver agreement in the diagnosis of cervical intraepithelial neoplasia. Am J Surg Pathol 2002;26(11):1389-99.

(21) Stoler MH, Wright TC, Jr., Ferenczy A, et al. Routine use of adjunctive p16 immunohistochemistry improves diagnostic agreement of cervical biopsy interpretation: Results from the CERTAIN study. Am J Surg Pathol 2018;42(8):1001-9.

(22) Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. Arch Public Health 2014;72(1):39-72.

(23) Munoz N, Bosch FX, de Sanjose S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 2003;348(6):518-27.

(24) de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 2010;11(11):1048-56.

(25) Cuzick J, Beverley E, Ho L, et al. HPV testing in primary screening of older women. Br J Cancer 1999;81(3):554-8.

(26) Schiffman M, Adrianza ME. ASCUS-LSIL Triage Study. Design, methods and characteristics of trial participants. Acta Cytol 2000;44(5):726-42.

(27) Ratnam S, Franco EL, Ferenczy A. Human papillomavirus testing for primary screening of cervical cancer precursors. Cancer Epidemiol Biomarkers Prev 2000;9(9):945-51.

(28) Schneider A, Hoyer H, Lotz B, et al. Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. Int J Cancer 2000;89(6):529-34.

(29) Belinson J, Qiao YL, Pretorius R, et al. Shanxi Province Cervical Cancer Screening Study: a cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. Gynecol Oncol 2001;83(2):439-44.

(30) Kulasingam SL, Hughes JP, Kiviat NB, et al. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. JAMA 2002;288(14):1749-57.

(31) Belinson JL, Qiao YL, Pretorius RG, et al. Shanxi Province cervical cancer screening study II: self-sampling for high-risk human papillomavirus compared to direct sampling for human papillomavirus and liquid based cervical cytology. Int J Gynecol Cancer 2003;13(6):819-26.

(32) Cuzick J, Szarewski A, Cubie H, et al. Management of women who test positive for high-risk types of human papillomavirus: the HART study. Lancet 2003;362(9399):1871-6.

(33) Gravitt PE, Paul P, Katki HA, et al. Effectiveness of VIA, Pap, and HPV DNA testing in a cervical cancer screening program in a peri-urban community in Andhra Pradesh, India. PLoS One 2010;5(10):e13711.

(34) Belinson JL, Wu R, Belinson SE, et al. A population-based clinical trial comparing endocervical high-risk HPV testing using hybrid capture 2 and Cervista from the SHENCCAST II Study. Am J Clin Pathol 2011;135(5):790-5.

(35) Monsonego J, Hudgens MG, Zerat L, et al. Evaluation of oncogenic human papillomavirus RNA and DNA tests with liquid-based cytology in primary cervical cancer screening: the FASE study. Int J Cancer 2011;129(3):691-701.

(36) Mahmud SM, Sangwa-Lugoma G, Nasr SH, et al. Comparison of human papillomavirus testing and cytology for cervical cancer screening in a primary health care setting in the Democratic Republic of the Congo. Gynecol Oncol 2012;124(2):286-91.

(37) Wright TC, Jr., Stoler MH, Behrens CM, et al. The ATHENA human papillomavirus study: design, methods, and baseline results. Am J Obstet Gynecol 2012;206(1):46.

(38) Ferreccio C, Barriga MI, Lagos M, et al. Screening trial of human papillomavirus for early detection of cervical cancer in Santiago, Chile. Int J Cancer 2013;132(4):916-23.

(39) Zhao FH, Jeronimo J, Qiao YL, et al. An evaluation of novel, lower-cost molecular screening tests for human papillomavirus in rural China. Cancer Prev Res (Phila) 2013;6(9):938-48.

(40) Firnhaber C, Mayisela N, Mao L, et al. Validation of cervical cancer screening methods in HIV positive women from Johannesburg South Africa. PLoS One 2013;8(1):e53494.

(41) Basu P, Mittal S, Banerjee D, et al. Diagnostic accuracy of VIA and HPV detection as primary and sequential screening tests in a cervical cancer screening demonstration project in India. Int J Cancer 2015;137(4):859-67.

(42) Isidean SD, Mayrand MH, Ramanakumar AV, et al. Human papillomavirus testing versus cytology in primary cervical cancer screening: End-of-study and extended follow-up results from the Canadian cervical cancer screening trial. Int J Cancer 2016;139(11):2456-66.

(43) Stoler MH, Wright TC, Jr., Parvu V, et al. The Onclarity Human Papillomavirus Trial: Design, methods, and baseline results. Gynecol Oncol 2018;149(3):498-505.

(44) Darragh TM, Colgan TJ, Cox JT, et al. The Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. J Low Genit Tract Dis 2012;16(3):205-42.

(45) Khieu M, Butler SL. High Grade Squamous Intraepithelial Lesion (HSIL).Treasure Island, FL, USA: StatPearls Publishing LLC; 2020.

(46) 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):847-56.

(47) Lonky NM, Felix J, Tsadik GW, Lonky S. False-negative hybrid capture II results related to altered adhesion molecule distribution in women with atypical squamous cells pap smear results and tissue-based human papillomavirus-positive high-grade cervical intraepithelial neoplasia. J Low Genit Tract Dis 2004;8(4):285-91.

(48) Castle PE, Schiffman M, Glass AG, et al. Human papillomavirus prevalence in women who have and have not undergone hysterectomies. J Infect Dis 2006;194(12):1702-5.

(49) Castle P. Cervical cancer screening among women without a cervix. JAMA 2004;292(13):1550-1.

(50) Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. Infect Agent Cancer 2009;4:8.

(51) Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. Int J Cancer 2012;131(10):2349-59.

(52) Kyrgiou M, Athanasiou A, Paraskevaidi M, et al. Adverse obstetric outcomes after local treatment for cervical preinvasive and early invasive disease according to cone depth: systematic review and meta-analysis. BMJ 2016;354:i3633. doi: 10.1136/bmj.i3633.:i3633.

(53) Sasieni P, Castanon A, Landy R, et al. Risk of preterm birth following surgical treatment for cervical disease: executive summary of a recent symposium. BJOG 2016;123(9):1426-9.

(54) IARC. Human Papillomaviruses. In: IARC/WHO, editor. 2007. p. 1-636.

(55) 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):425-34.

(56) Schiffman M, Rodriguez AC. Heterogeneity in CIN3 diagnosis. Lancet Oncol 2008;9(5):404-6.

(57) Kelly HA, Sawadogo B, Chikandiwa A, et al. Epidemiology of high-risk human papillomavirus and cervical lesions in African women living with HIV/AIDS: effect of anti-retroviral therapy. AIDS 2017;31(2):273-85.

(58) Massad LS, Einstein MH, Huh WK, et al. 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. J Low Genit Tract Dis 2013;17(5 Suppl 1):S1-S27.

(59) Punglia RS, Bifolck K, Golshan M, et al. Epidemiology, Biology, Treatment, and Prevention of Ductal Carcinoma In Situ (DCIS). JNCI Cancer Spectr 2018;2(4):ky063.

(60) Nusko G, Mansmann U, Partzsch U, et al. Invasive carcinoma in colorectal adenomas: multivariate analysis of patient and adenoma characteristics. Endoscopy 1997;29(7):626-31.

(61) Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366(8):687-96.

(62) Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993;329(27):1977-81.

(63) Tainio K, Athanasiou A, Tikkinen KAO, et al. Clinical course of untreated cervical intraepithelial neoplasia grade 2 under active surveillance: systematic review and meta-analysis. BMJ 2018;360:k499.:k499.

(64) 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):516-20.

(65) Castle PE, Adcock R, Cuzick J, et al. Relationships of p16 Immunohistochemistry and Other Biomarkers With Diagnoses of Cervical Abnormalities: Implications for LAST Terminology. Arch Pathol Lab Med 2020;144(6):725-34.

(66) Petry KU, Cox JT, Johnson K, et al. Evaluating HPV-negative CIN2+ in the ATHENA trial. Int J Cancer 2016;138(12):2932-9.

Table 1. List of studies included in these analyses. Abbreviations: HC, Hybrid Capture; HC2, Hybrid Capture 2; HR-HPV, high-risk human papillomavirus; Cyto+, borderline or atypical squamous cells of undetermined significance or more severe cytology; ≥ASC-US, ASC-US or more severe; RS, random sample.

Study Location[reference]	Study Population Type	Enrollment Period	Enrolled Population (Age Criteria)	Main HPV Test	Primary Histological Endpoint Diagnosis*	Cytology	VIA	Other (Clinical) Tests	Colposcopy Referral Criteria	HPV-positive CIN2+		HPV-negative CIN2+		p[∞]
										N	Age (mean; median) (Years)	N	Age (mean; median) (Years)	
England (25)	Screening	1994-1997	2,988 (≥35 years)	HC	Consensus Review	Yes	No	Yes; PCR and HC MY09/11 PCR	Cyto+ and/or HPV+ (PCR)	23	44.4;43	3	41.0;42	0.6
USA (26)	Referral for ASCUS or LSIL Pap**	1996-1998	1,836** (≥18 years)	HC2	Consensus Review[‡]	Yes	No	No	All**	222	24.6; 23	16	25.6; 23.5	0.5
Canada (27)	Screening	1996-1998	2,098 (18-69 years)	HC//HC2	Community Diagnosis	Yes	No	No	Cyto+ and/or HPV+	26	28.1; 27	5	31.2; 30	0.5
Germany (28)	Screening	1996-1998	4,761 (≥35 years)	GP 5+/6+ PCR	Consensus Review	Yes	No	PAPNET	All	108	30.8; 31	6	30.0; 32.5	0.8
China (29)	Screening	1999	1,997 (35-45 years)	HC2	Study Pathologist	Yes	Yes	Self-Collection	All	82	39.5; 40	2	38.5; 38.5	0.6
USA (30)	Screening	1997-2000	4,075 (18-50 years)	MY09/11 PCR	Study Pathologist	Yes	No	No	Cyto+ and/or HPV+[£]; ~10% RS of Screen-	190	25.2; 23.5	26	24.0; 24.5	1.0
China (31)	Screening	2001-2002	8,497 (27-56 years)	HC2	Study Pathologist	Yes	No	Self-Collection	All	306	42.0; 42	10	40.8; 40	0.5
England (32)	Screening	1998-2001	11,085 (30-60 years);	HC2	Consensus Review	Yes	No	No	Cyto+ and/or HPV+; ~5% (n=414) screen-negatives	87	36.2;34	49	47;49	0.02
India (33)	Screening	2005-2007	2,331 (≥25 years)	HC2	Two independent reads; worst histology	Yes	Yes	Linear Array	Cyto+, HC2+, and/or VIA+; 20% RS of Screen-	14	41.1; 37.5	4	37.5; 37.5	0.7
China (34)	Screening	2009-2010	8.556 (25-59 years)	HC2	Consensus Review	Yes	No	Cervista; MALDI-TOF; Self-Collection	Cyto+ and/or HPV+ by Cervista and/or MALDi-TOF on self-collection and/or provider collection and/or HC2 on provider collection	225	39.5;39	11	37.4; 36	0.4
France (35)	Screening	2008-2009	4,950 (20-65 years)	HC2	Consensus Review[¥]	Yes	No	Aptima	Cyto+ and/or HPV+ by HC2 and/or Aptima	96	34.9; 33	4	44.3; 44	0.04
Democratic Republic of the Congo (36)	Screening	2003-2004	1,699 (≥30 years)	HC2	Study Pathologist	Yes	Yes	No	All	21	59.7; 45	3	62.8; 44	0.9
USA (37)	Screening	2008-2009	41,955	cobas	Consensus	Yes	No	Linear Array;	Cyto+ and/or HPV+	578	32.9; 31	63	37.6; 36	0.003

			(≥25 years)		Review			AMPLICOR	by AMPLICOR and/or Linear Array positive; ~2.5% RS of Screen-					
Chile (38)	Screening	2009-2010	8,309 (25-64 years)	HC2	Community Pathology	Yes	Yes	No	Cyto+, HC2+, and/or VIA+; 68 high-risk, screen–negative women	91	36.8; 25	5	36.0; 35	1.0
China (39)	Screening	2010-2011	7,541 (25-65 years)	HC2	Study Pathologist	No	Yes	Self-collected specimens tested by HC2 and careHPV; provider-collected specimens tested by careHPV; HPV16/18/45 E6 Test; HPV genotyping of the biopsy by SPF10/LiPA	Women who tested positive for any of the 6 screening tests conducted (VIA, HPV E6, and HC2 and careHPV on clinician-collected and self-collected specimens); ~10% RS of Screen-	138	46.5; 46.5	6	47.7; 45.5	0.8
South Africa (40)	Screening	2009-2011	1,202[†] (18-65 years)	HC2	Community Pathology	Yes	Yes	No	Cyto+ and/or VIA+; ~25% RS of Screen-	291	37.0; 36	21	39.6; 40	0.1
India (41)	Screening	2010-2014	39,740 (30-60 years)	HC2	Consensus Review	No	Yes	No	HC2+ and/or VIA+	202	36.0; 34	74	42.2; 40	<0.001
Canada (42)	Screening	2002-2005	10,154 (30-69 years)	HC2	Community Pathology[‡]	Yes	No	Linear Array	≥ASC-US cytology, HPV+; ~10% (St. John's) and ~20% (Montreal) RS of Screen-	71	36; 37.0	11	50; 47.2	0.002
USA (43)	Screening	2013-2015	33,858 (≥21 years)	Onclarity	Consensus Review	Yes	No	HC2; Onclarity on ThinPrep Specimen; bidirectional Sequencing for HPV genotyping	≥ASC-US cytology, HPV+ by Onclarity on SurePath and/or ThinPrep) and/or HC2 on ThinPrep; ~5% RS of Screen-	316	32;34.0	36	25; 36.9	0.05

*Single=biopsy diagnosis based on a review by single pathologist as part of routine care or by a single study pathologist; Consensus=panel review of the biopsy diagnosis by two or more pathologists and some method of adjudication of discordant results

**Only those in the immediate colposcopy arm were considered in this analysis

[†]Women living with HIV

[£]HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and/or HPV68 positive

[‡]Worst diagnosis on biopsy or excised tissue

[∞]Kruskal-Wallis test for differences in the median age

[¥]Diagnoses of CIN2+ were confirmed by p16 immunohistochemistry

Table 2. Summary of results of the pooled analyses to compare human papillomavirus (HPV)–positive and HPV-negative cervical intraepithelial neoplasia grade 2 or more severe diagnoses (CIN2+).

	Number of Studies	HPV-positive CIN2+		HPV-negative CIN2+		p
		N	Fraction (95%CI)	N	Fraction (95%CI)	
Demographics						
Aged ≥40 years at Diagnosis	19	3,089	0.30 (0.29-0.32)	307	0.34 (0.28-0.41)	0.03
Ever Smoked	9	1,873	0.40 (0.38-0.42)	185	0.30 (0.30-0.46)	0.6
Clinical Correlates						
CIN3+ (Primary Endpoint)	18	3,062	0.56 (0.54-0.58)	304	0.24 (0.19-0.30)	<0.001
Antecedent HSIL+ Cytology	15	2,552	0.34 (0.32-0.36)	217	0.10 (0.05-0.15)	<0.001
VIA positive	6	739	0.57 (0.53-0.60)	113	0.83 (0.73-0.90)	<0.001
High-Grade Colposcopic Impression	13	1,800	0.30 (0.28-0.33)	174	0.18 (0.12-0.30)	0.03
Other Pathology Review	7	1,471	0.51 (0.48-0.53)	135	0.42 (0.33-0.52)	0.09
Other Clinical HPV*	6	1,449	0.97 (0.96-0.98)	134	0.39 (0.31-0.48)	<0.001
Self-Collection and HPV Testing**	4	721	0.95 (0.93-0.96)	58	0.84 (0.71-0.94)	<0.001
HPV Genotyping Category†‡						
HPV16	9	1,556	0.46 (0.43-0.48)	261	0.09 (0.05-0.14)	<0.001
Other High-Risk HPV			0.49 (0.47-0.52)		0.32 (0.26-0.38)	<0.001
Low-Risk HPV			0.00 (0.00-0.01)		0.13 (0.09-0.18)	<0.001
HPV Negative			0.02 (0.01-0.02)		0.33 (0.27-0.39)	<0.001

*Results from provider-collected cervical specimens tested by AMPLICOR (26, 37), MALDI-TOF (34), Aptima (35), careHPV (39), and HC2 (43)

**Self-collected cervicovaginal specimens tested by HC2 (29, 31, 39) and MALDI-TOF (34)

†HPV genotype results were categorized hierarchically according to cancer risk:

‡The totals do not add to 100%, even though individual studies add up to 100%. A separate binomial model was run for each HPV genotyping category and therefore independent. Thus the sum of the HPV categories is not constrained to equal one (unity).

571 **Table 3**. Comparison of results of biomarker testing from individual studies between human
572 papillomavirus (HPV)–positive and HPV-negative cervical intraepithelial neoplasia grade 2 or
573 more severe diagnoses (CIN2+).
574

	HPV-positive CIN2+		HPV-negative CIN2+		p
	N	%Positive	N	%Positive	
p16 immunohistochemistry*(37)	52	84.6%	62	61.3%	0.007
p16 immunohistochemistry (34)	161	96.3%	6	83.3%	0.2
HPV16/18/45 E6 (39)	138	42.8%	6	33.3%	1.0

575 *Only a stratified sample of cases were tested (66).
576

Figure Legends:

Figure 1. Forest plots for the proportion of cervical intraepithelial neoplasia (CIN) grade 2 or more severe (CIN2+) diagnoses that had a CIN grade 3 or more severe (CIN3+) diagnosis, stratified on the human papillomavirus (HPV) test result.

Figure 2. Forest plots for the proportion of cervical intraepithelial neoplasia (CIN) grade 2 or more severe (CIN2+) diagnoses that had an antecedent high-grade squamous intraepithelial lesion (HSIL) or more severe cytologic interpretation (HSIL+), stratified on human the papillomavirus (HPV) test result.

Figure 3. Forest plots for the proportion of cervical intraepithelial neoplasia (CIN) grade 2 or more severe (CIN2+) diagnoses that was positive by visual inspection by acetic acid (VIA), stratified on the human papillomavirus (HPV) test result.

Figure 4. Forest plots for the proportion of cervical intraepithelial neoplasia (CIN) grade 2 or more severe (CIN2+) diagnoses that had a high-grade colposcopic impression, stratified on the human papillomavirus (HPV) testing result.

Figure 1

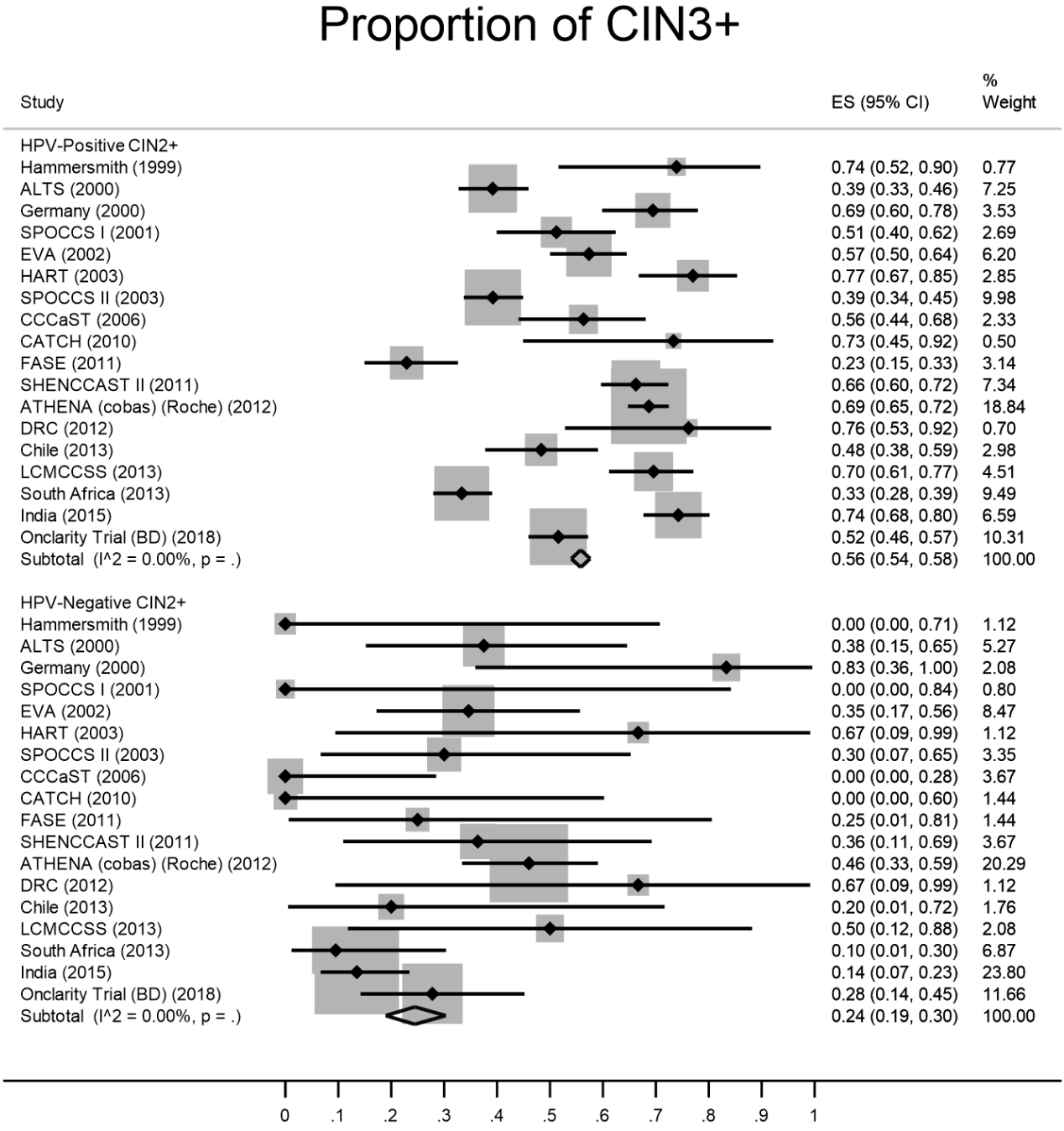
Proportion of CIN3+
Study
ES (95% CI)
% Weight
HPV-Positive CIN2+
Hammersmith (1999) 0.74 (0.52, 0.90) 0.77
ALTS (2000) 0.39 (0.33, 0.46) 7.25
Germany (2000) 0.69 (0.60, 0.78) 3.53
SPOCCS I (2001) 0.51 (0.40, 0.62) 2.69
EVA (2002) 0.57 (0.50, 0.64) 6.20
HART (2003) 0.77 (0.67, 0.85) 2.85
SPOCCS II (2003) 0.39 (0.34, 0.45) 9.98
CCCaST (2006) 0.56 (0.44, 0.68) 2.33
CATCH (2010) 0.73 (0.45, 0.92) 0.50
FASE (2011) 0.23 (0.15, 0.33) 3.14
SHENCCAST II (2011) 0.66 (0.60, 0.72) 7.34
ATHENA (cobas) (Roche) (2012) 0.69 (0.65, 0.72) 18.84
DRC (2012) 0.76 (0.53, 0.92) 0.70
Chile (2013) 0.48 (0.38, 0.59) 2.98
LCMCCSS (2013) 0.70 (0.61, 0.77) 4.51
South Africa (2013) 0.33 (0.28, 0.39) 9.49
India (2015) 0.74 (0.68, 0.80) 6.59
Onclarity Trial (BD) (2018) 0.52 (0.46, 0.57) 10.31
Subtotal (I^2 = 0.00%, p = .) 0.56 (0.54, 0.58) 100.00
HPV-Negative CIN2+
Hammersmith (1999) 0.00 (0.00, 0.71) 1.12
ALTS (2000) 0.38 (0.15, 0.65) 5.27
Germany (2000) 0.83 (0.36, 1.00) 2.08
SPOCCS I (2001) 0.00 (0.00, 0.84) 0.80
EVA (2002) 0.35 (0.17, 0.56) 8.47
HART (2003) 0.67 (0.09, 0.99) 1.12
SPOCCS II (2003) 0.30 (0.07, 0.65) 3.35
CCCaST (2006) 0.00 (0.00, 0.28) 3.67
CATCH (2010) 0.00 (0.00, 0.60) 1.44
FASE (2011) 0.25 (0.01, 0.81) 1.44
SHENCCAST II (2011) 0.36 (0.11, 0.69) 3.67
ATHENA (cobas) (Roche) (2012) 0.46 (0.33, 0.59) 20.29
DRC (2012) 0.67 (0.09, 0.99) 1.12
Chile (2013) 0.20 (0.01, 0.72) 1.76
LCMCCSS (2013) 0.50 (0.12, 0.88) 2.08
South Africa (2013) 0.10 (0.01, 0.30) 6.87
India (2015) 0.14 (0.07, 0.23) 23.80
Onclarity Trial (BD) (2018) 0.28 (0.14, 0.45) 11.66
Subtotal (I^2 = 0.00%, p = .) 0.24 (0.19, 0.30) 100.00
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1

Figure 2

Proportion of HSIL+ Cytology

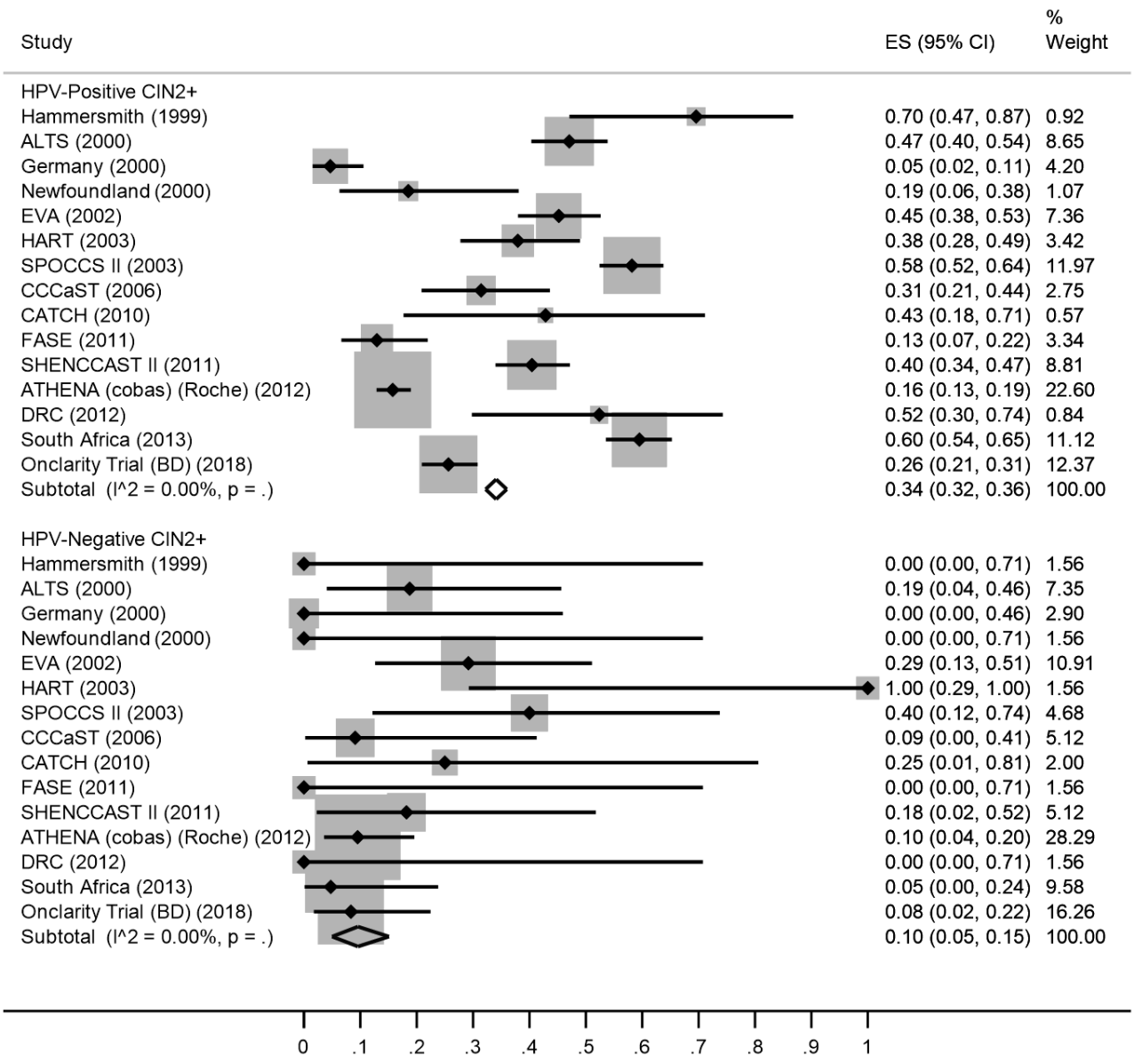

Figure 3

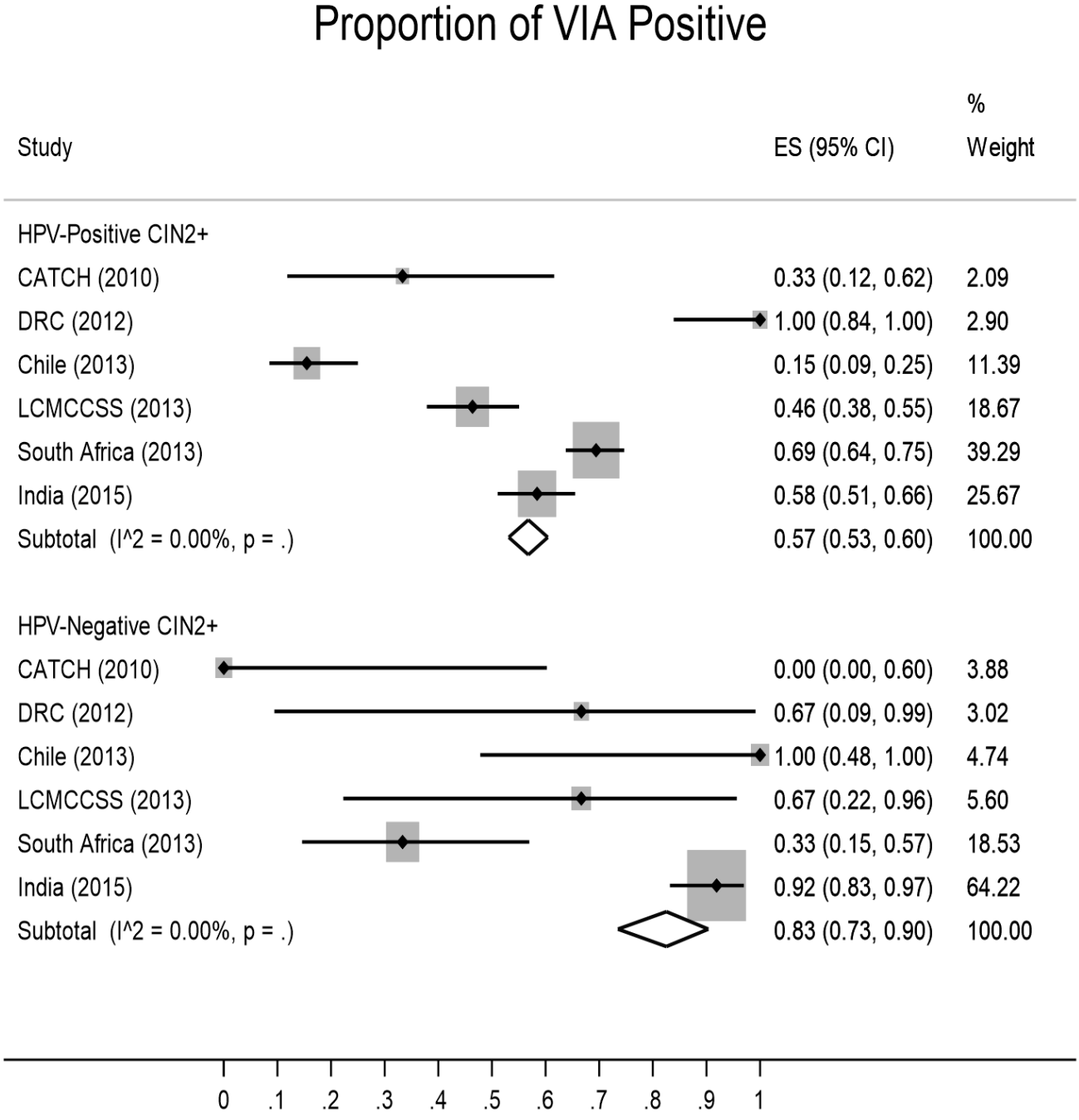
Proportion of VIA Positive
Study
ES (95% CI)
% Weight
HPV-Positive CIN2+
CATCH (2010) 0.33 (0.12, 0.62) 2.09
DRC (2012) 1.00 (0.84, 1.00) 2.90
Chile (2013) 0.15 (0.09, 0.25) 11.39
LCMCCSS (2013) 0.46 (0.38, 0.55) 18.67
South Africa (2013) 0.69 (0.64, 0.75) 39.29
India (2015) 0.58 (0.51, 0.66) 25.67
Subtotal (I^2 = 0.00%, p = .) 0.57 (0.53, 0.60) 100.00
HPV-Negative CIN2+
CATCH (2010) 0.00 (0.00, 0.60) 3.88
DRC (2012) 0.67 (0.09, 0.99) 3.02
Chile (2013) 1.00 (0.48, 1.00) 4.74
LCMCCSS (2013) 0.67 (0.22, 0.96) 5.60
South Africa (2013) 0.33 (0.15, 0.57) 18.53
India (2015) 0.92 (0.83, 0.97) 64.22
Subtotal (I^2 = 0.00%, p = .) 0.83 (0.73, 0.90) 100.00
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1

Figure 4

Proportion of High-Grade Colposcopy