



Published in final edited form as:

*Int J Cancer*. 2017 February 01; 140(3): 718–725. doi:10.1002/ijc.30456.

## Proof-of-principle study of a novel cervical screening and triage strategy: computer-analyzed cytology to decide which HPV-positive women are likely to have CIN2

Mark Schiffman<sup>1,\*</sup>, Kai Yu<sup>1,\*</sup>, Rosemary Zuna<sup>2</sup>, S. Terence Dunn<sup>3</sup>, Han Zhang<sup>1</sup>, Joan Walker<sup>3</sup>, Michael Gold<sup>4</sup>, Noorie Hyun<sup>1</sup>, Greg Rydzak<sup>5</sup>, Hormuzd A. Katki<sup>1</sup>, and Nicolas Wentzensen<sup>1</sup>

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

<sup>2</sup>Oklahoma University Health Sciences Center, Oklahoma City, OK

<sup>3</sup>Oklahoma State Department of Health, Oklahoma City, OK

<sup>4</sup>Tulsa Cancer Institute, Tulsa, OK

<sup>5</sup>Information Management Services, Calverton, MD

### Abstract

A challenge in implementation of sensitive HPV-based screening is limiting unnecessary referrals to colposcopic biopsy. We combined 2 commonly recommended triage methods: partial HPV typing and “reflex” cytology, evaluating the possibility of automated cytology. This investigation was based on 1,178 exfoliated cervical specimens collected during the enrollment phase of The Study to Understand Cervical Cancer Early Endpoints and Determinants (SUCCEED, Oklahoma City, Oklahoma). We chose a colposcopy clinic population to maximize number of outcomes, for this proof-of-principle cross-sectional study. Residual aliquots of PreservCyt were HPV-typed using Linear Array (LA, Roche Molecular Systems, Pleasanton, CA). High-risk HPV typing data and cytologic results (conventional and automated) were used jointly to predict risk of histologically defined CIN2. We developed a novel computer algorithm that uses the same optical scanning features that are generated by the FocalPoint Slide Profiler (BD, Burlington, NC). We used the LASSO (Least Absolute Shrinkage and Selection Operator) method to build the prediction model based on a training dataset (n=600). In the validation set (n=578), for triage of all HPV-positive women, a cytologic threshold of ASC-US had a sensitivity of 0.94, and specificity of 0.30, in this colposcopy clinic setting. When we chose a threshold for the severity score (generated by the computer algorithm) that had an equal specificity of 0.30, the sensitivity was 0.91. Automated cytology also matched ASC-US when partial HPV typing was added to the

Correspondence to either: Mark Schiffman, MD, MPH, Clinical Genetics Branch, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, schiffmm@mail.nih.gov or Kai Yu, PhD, Biostatistics Branch, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, yuka@mail.nih.gov.

\*Contributed equally

Conflict of Interest Statement: Drs. Schiffman and Wentzensen have received HPV testing, typing, biomarker testing, and automated cytology data at reduced or no cost for independent evaluations, from a variety of commercial groups, including Roche, BD, Hologic, Cepheid, and Qiagen.

triage strategy, and when we re-defined cases as CIN3. If this strategy works in a prospective screening setting, a totally automated screening and triage technology might be possible.

## Keywords

cervical; cancer; screening; cytology; HPV; automation

---

## Introduction

Cervical cancer is caused by infection with a group of high-risk human papillomavirus (HPV) infections; 12 types of HPV (types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) are considered proven cervical carcinogens while HPV68 is considered a probable carcinogen (1,2). Vaccination to prevent cervical HPV acquisition is the ultimate preventive strategy (3). However, vaccine implementation has been slow worldwide, and a major impact on cancer rates from vaccination of adolescents will take decades to achieve. For the foreseeable future, cervical cancer screening will remain essential. Increasingly, DNA or RNA testing of cervicovaginal cells for carcinogenic HPV is being introduced as an adjunct or replacement of cytology (Pap) screening.

The goal of cervical screening is to detect and treat precancers and early cancers in order to reduce cancer mortality and morbidity, while minimizing intervention in women that do not need it (4). Because of the sensitivity of HPV tests, women screening HPV-negative using any of the well validated DNA or RNA assays can be reassured of very low risk of cervical cancer for several years until the next screen (5). Given the low risk predicted by a negative HPV test result alone, adding a second test decreases risk only marginally more (6); therefore, most countries, other than the U.S., that are considering HPV testing are introducing it as a stand-alone primary screening test.

For most of the world, the critical challenge in implementation of HPV-based screening lies in the management of women that test positive. Approximately 90% of HPV infections are controlled immunologically or cleared by other unknown mechanisms within a few years (7). The issue is distinguishing the small proportion of high-risk HPV-infected women who have, or will soon develop, cervical precancer (or rarely cancer) that requires treatment.

Colposcopic biopsy remains the clinical reference standard to define which HPV-positive women (detected by HPV tests, or related cytologic and/or visual abnormalities) have underlying cervical precancer/cancer requiring ablational or excisional treatment (8). Colposcopy involves magnified examination of the cervix, which is stained with acetic acid and sometimes Lugol's iodine, and targeted biopsy of visible lesions. As a visual method, its performance is challenged by the introduction of sensitive molecular HPV-based screening, which detects HPV-associated precancer early in its development (9). To achieve good program performance in a colposcopy referral population where lesions are increasingly early and small (and possibly more likely to regress without intervention), substantial expertise and access to very good histopathology services are required.

Given the prevalence of typically benign infections with “high-risk HPV types”, the low probability of CIN2 following infections, and the growing awareness that most cases of CIN2 would not cause harm if left untreated, we wish to prevent cancer but limit the number of women screening HPV-positive that are referred to colposcopic biopsy and treatment. An intermediate step called “triage” is introduced when affordable, meant to divide HPV-positive women according to their chance of having treatable precancer, if colposcopy is performed. The highest-risk women are referred to colposcopy directly while lower-risk women are retested, often after approximately 6–12 months, in the hope of clearance of infection (4). A 1-year delay permits viral clearance and return to routine screening intervals in approximately half the women (10); persistent infection (monitored either by repeat HPV testing or, less directly, by cytology) indicates an elevated risk and mandates re-consideration of colposcopy referral (11).

Several candidate triage strategies are being proposed. Perhaps the 2 most commonly recommended methods are partial HPV typing and “reflex” cytology (12). High-risk HPV types differ substantially in carcinogenic potential; immediate colposcopic referral may be recommended if the types predicting the highest risk for invasive cancer are found (i.e., HPV16 or HPV18) (13,14). Cytology can be used to triage either all HPV-positive women or only the fraction not already referred to colposcopy because highest-risk types have been found. Most commonly, the recommended referral cytology threshold when used for triage is (HPV-positive) ASC-US or worse (ASC-US).

The accuracy of cytology varies widely between laboratories, regions, and countries (15,16). For reliable triage of HPV infections, we hypothesized that automated cytology might help standardization if computer-generated classification could be made sensitive and specific enough. The approach would require the collection of an exfoliated cervical specimen, like current cervical cytology. But, a single integrated piece of test equipment, combining HPV partial typing and computer-interpreted cytology, could be useful in well-resourced settings that nonetheless do not have strong conventional cytotechnology and cytopathology workforces. The present proof-of-principle investigation explores the promise of this approach.

## Material and Methods

### Study Population and Procedures in SUCCEED

This investigation was based on exfoliated cervical specimens collected during the cross-sectional enrollment phase of a study of 3,000 women at Oklahoma University Health Sciences Center (OUHSC) (17,18). The Study to Understand Cervical Cancer Early Endpoints and Determinants (SUCCEED) was designed to search for biomarkers influencing or reflecting the transitions from HPV infection to cervical precancer, and from precancer to cancer. OUHSC serves as the tertiary referral center for the eastern part of the state of Oklahoma. SUCCEED was conducted among women either referred to colposcopy for cervical cytologic screening abnormalities (with HPV testing of ASC-US commonly used in the community to triage these common equivocal results) or for treatment of already diagnosed precancer/cancer. The enrollment period for the project, which was conducted in two different discrete phases, stretched from November 2003 to the summer of 2011.

At the time of the enrollment clinical evaluation or treatment visit, a broom device was used to collect cervical cells into PreservCyt (Hologic, Marlborough, MA). ThinPrep cytology slides (Hologic) were prepared, interpreted in a standard manner without knowledge of HPV test results using the Bethesda System classification (19), and stored. Residual aliquots of PreservCyt were HPV-typed using Linear Array (LA, Roche Molecular Systems, Pleasanton, CA), as previously described (17). We considered high-risk HPV types to include HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68, with a LA positivity signal intensity threshold of “extremely weak”. We included HPV66 as well, although it rarely causes cancer, because its misclassification (14) as an important carcinogenic type led to its inclusion in most HPV tests including cobas™ (Roche Molecular Systems, Pleasanton CA), Aptima™ (Hologic), and Onclarity™ (BD Diagnostics, Sparks MD).

For the present study, the HPV typing data and cytologic results were used jointly to predict risk of histologically defined CIN2 among women found to be HPV-positive at enrollment; risk of CIN3 was evaluated as well, although the number of cases was smaller. The performance of automated cytology for detection of prevalent CIN2 (and CIN3) among HPV-positive women was compared to conventionally interpreted cytology.

To evaluate the potential of automated cytology as an alternative to conventional interpretation for triage of HPV-positive women, we developed a novel computer algorithm that uses the same optical scanning data that are generated by the FocalPoint Slide Profiler (BD, Burlington, NC). The FDA-approved algorithm in-use today ranks slides for severity, typically within batches of about 150 stained slides. The initial indication was to define women whose cytology indicated such low risk that cytotechnologist review was not needed (20). However, instead we developed a new algorithm that defines the most severely abnormal slides among HPV-positive women.

The study population is pictured in Figure 1. We opted to study a population greatly enriched for disease endpoints. The SUCCEED population represents a group of women presenting, on the basis of abnormal cytologic or histologic results, for colposcopy or for treatment to the OUHSC gynecologic referral service. Thus, this population contains many more cases of invasive cancer and CIN2/3 than the general population, and all women even if currently cytologically and histologically normal had a preceding cervical abnormality. We aimed to explore the use of automated cytology as it might be used for triage of women testing positive using HPV DNA tests like those approved (or currently seeking approval) by the U.S. Food and Drug Administration (FDA). Because HPV primary screening is recommended (at least in the U.S.) starting at age 25 (12), we restricted the main analysis to 1,178 slides from high-risk HPV-positive women aged 25 or older. A separate ancillary analysis of performance in younger women was also performed.

### **FocalPoint Scan**

The archived ThinPrep slides were scanned at BD (North Carolina) in a few large batches or “run sets”. In anticipation of high-throughput, computerized slide production with low variability between run sets, we ignored batch differences in the analysis.

FocalPoint performs a high-speed, magnified scan, and outputs 160 binary or quantitative “features” of the cytology slide, such as presence of different cell types, nuclear size, chromatin density, and nuclear contour (20). The preset algorithm is designed for general screening, i.e., review of batches of mainly negative slides from mainly HPV-negative women. Because cytologic results of HSIL are uncommon in the general population compared with minor changes, almost all abnormal results in any batch of slides represent equivocal (ASC-US), or at most, minor (LSIL) HPV-associated changes (data not shown). Consequently, the severity rank using the preset method is not specific for precancer/cancer. To identify precancer/cancer effectively in a group of HPV-positive women, a new algorithm that uses the same optical features to target severe pathologic changes was required.

Of archived ThinPrep slides scanned by FocalPoint, the percentage judged to be suboptimal (failing “process review”) was 28.1%, mainly because of pale nuclear staining, probably related to slide age. We opted to include the suboptimal slides in the main analysis and conducted a sensitivity analysis that excluded them.

To analyze the FocalPoint scan data and construct a new algorithm that targeted precancer/cancer, we followed standard practice for prediction modeling, i.e., we divided the test slides into a training set and a validation set. The training set arbitrarily included random draws of 300 cases of CIN2, and 300 controls that were HPV-positive but were <CIN2. This draw left an available validation set of 379 CIN2 (including 162 CIN2, 156 CIN3, 6 AIS, 6 adenocarcinoma, and 49 squamous cell cancers) and 169 <CIN2 (63 negative, 14 atypical metaplasia, 91 CIN1, and 1 “uncertain histology”).

### **Building the Prediction Algorithm Using the Training Set**

We used the LASSO (Least Absolute Shrinkage and Selection Operator) method (21) to build the prediction model based on the training dataset, without consideration of HPV typing. (HPV type group was added as a risk stratifier in later analyses). The LASSO procedure fits an additive model (in the logit scale in the case of binary outcome) with the purposes of both variable selection and coefficient estimation. We used the R package GLMNET to fit the LASSO model, and 5-fold cross-validation to identify the optimal prediction model with performance measured by area under the receiver operating characteristic (ROC) curve (AUC).

### **Statistical Analysis in the Validation Dataset**

The analyses reported here represent the use of the algorithm created in the training set, when applied in the validation dataset. The main goal of the study was to compare the performance of automated cytology with the performance of conventionally interpreted cytology, divided at the ASC-US vs. NILM threshold, for the identification of HPV-infected women who needed immediate colposcopy because of elevated risk of precancer/cancer (CIN2) versus <CIN2.

We first considered the use of FocalPoint among all HPV-positive women, ignoring HPV type, and compared the distribution of severity scores by increasingly severe histologic outcomes. The trend was tested using linear regression. While recognizing that differences exist between CIN2, CIN3/AIS, and cancer, for the clinical question addressed here, we

subsequently dichotomized the histologic outcomes for the main analysis into CIN2 versus <CIN2. In ancillary analyses, we evaluated the same algorithm for prediction of CIN3 (including uncommon AIS) versus <CIN2, excluding CIN2.

Next, the degree of association between the 2 triage alternatives, severity score and conventionally interpreted cytology, was examined. We explored also whether severity score varied by 4 HPV type groups representing increasing risk of precancer/cancer, before conducting the major analyses regarding the use of FocalPoint to triage HPV-positive women stratified by these type groups. In other words, we examined the distribution of the severity scores as a triage method for HPV-positive women either ignoring type or stratifying the high-risk types by partial typing reported in 4 type-groups that vary in risk for CIN3, (HPV16, HPV18/45, HPV31/33/52/58, or HPV35/39/51/56/59/66/68) (Schiffman et al., IJC, e-published).

In the main analysis, we compared the two alternative cytologic triage methods for HPV-positive women, FocalPoint severity scores and conventionally-interpreted cytologic results from OUHSC cytopathologists, as follows: Receiver operating characteristic (ROC) curves, and areas under the curves (AUC) were used to show the sensitivity/specificity trade-offs for the continuous severity scores. Categorical analyses required that we set cut-points of the continuous FocalPoint score to compare with the conventionally interpreted cytology results. We considered ASC-US versus NILM to be the threshold of particular importance among HPV-positive women; therefore, we compared the sensitivity for CIN2 of automated vs. conventionally interpreted cytology at the level of specificity achieved by a ASC-US referral threshold. Because HSIL, although uncommonly found, elicits special concern, we also considered whether automated cytology identified a subset similar to the high-risk group signaled by HSIL.

## Results

We applied the LASSO model on the training set with 160 features. The final model consisted of 73 features, which were used to create the severity score for the prediction of risk of CIN2. The 73 selected features and their corresponding coefficients are given in Supplementary Table 1.

As shown in Figure 2, the severity score differed significantly and strongly between the histologic outcomes of <CIN2 and CIN2 ( $p < 2e-16$ ). Within the CIN2 category, the score tended to increase with finer divisions of increasing severity (CIN2, CIN3/AIS, cancer). However, further division of <CIN2 into finer categories (CIN1, Atypical Metaplasia, Negative, No Biopsy Taken) did not reveal differences in severity scores.

The severity score was strongly associated with the standard triage method, conventionally interpreted cytology (Table 1). Severity scores tended to increase monotonically with increasing abnormal cytology result, classified in the order of NILM, ASCUS, LSIL, ASC/AGC, HSIL, and Cancer, in both <CIN2 and CIN2 groups. If we treated the cytology outcome as a continuous variable, with 0 to 4 representing the cytology outcome in the order of NILM, ASCUS, LSIL, ASC/AGC, HSIL+, we detected significant trend effects

comparing the severity score to the cytology result ( $p=8.6e-7$  in <CIN2,  $p=2.0e-12$  in CIN2).

HPV-negative women overall tended to have lower severity scores than HPV-positive women ( $p=0.002$  for cases, and  $p=0.01$  for controls, Table 2). However, among HPV-positive women, severity scores were not associated with HPV type groups, when stratified by <CIN2/ CIN2 outcome.

Figures 3a–e address the central question, i.e., the direct comparison between severity score and cytology for triage of HPV-positive women; the panels of ROC curves represent comparisons of the sensitivity/specificity trade-off for diagnosis of CIN2 for increasing values of the severity score. When considering all HPV-positive subjects in the validation set, the AUC of the severity score was 0.73 (95% CI 0.69–0.78); values for each type-group varied somewhat around this value. When cases were defined as CIN3 (excluding cases of CIN2), the AUC value for all HPV-positive women was 0.76, and values were slightly greater for each type-group than for CIN2.

In each panel of Figure 3, the cytologic threshold of ASC-US is shown. High levels of sensitivity for detection of CIN2 were achieved only in combination with poor specificity, for both conventionally-interpreted and automated cytologic methods, as would be anticipated in this colposcopy referral population. Overall, a cytologic threshold of ASC-US had a sensitivity for detection of CIN2 of 0.94, and specificity of 0.30. When we chose a threshold for the severity score that had an equal specificity of 0.30, the sensitivity was 0.91, which was only slightly less than the value of 0.94 generated by ASC-US ( $p=0.06$  based on the McNemar test).

We conducted the same analysis for each of the 4 risk-based HPV type groups: picking a threshold for the severity score that resulted in the same specificity as cytology ASC-US in the subgroup, the sensitivity of the severity score did not differ significantly from the sensitivity for detection of CIN2 afforded by ASC-US ( $p=1.00$  within HPV16,  $p=1.00$  within HPV18/45,  $p=0.34$  within HPV31/33/52/58, and  $p=1.00$  within HPV35/39/51/56/59/66/68). The comparisons of sensitivity of severity score to ASC-US were similarly close when cases were re-defined as CIN3.

Severity scores generated by automated cytology did *not* identify a subgroup of HPV-positive women analogous to those identified by a conventionally interpreted result of HSIL (marked on each panel of Figure 3). In this referral population, with a large number of cancers among the CIN2 group, conventionally interpreted HSIL was not rare as is usually the case, and HSIL yielded moderate sensitivity for histopathology of CIN2 or CIN3; the ROC curves of severity scores failed by a large margin to match the sensitivity of HSIL when specificity was held constant.

We were concerned that including the 28.1% of slides judged to be less than optimal (i.e., that failed FocalPoint process review) would affect the conclusions. However, repeating the main analysis excluding the lower-quality slides did not change the results (data not shown).

Finally, confounding by age was a theoretical possibility. The mean age of the women in the validation set was 33.5 years (range 25–81). However, severity score increased only weakly with increasing age in the group of women with CIN2-3 (Spearman correlation of 0.11,  $p=0.04$ ) and increased non-significantly among women with <CIN2 (correlation of 0.14,  $p=0.07$ ) or cancer (0.26,  $p=0.05$ ). This indicated that age-adjustment was not needed for the dataset. Finally, when we applied the same algorithm to scanned optical features from the women <25 years of age ( $n > 1100$ ), the performance was reduced compared to older women (e.g., AUC for all types of 0.66 compared with 0.73, and reduced sensitivity compared to ASC-US).

## Discussion

Our data suggest that automated cervical cytology, such as FocalPoint with a re-designed algorithm, might serve to triage HPV-positive women aged 25 and older. The method, when applied to a population of women referred for colposcopy or treatment, identified those women at currently elevated risk of precancer/cancer (defined broadly as histologic CIN2 or more stringently as CIN3), with a sensitivity and a specificity similar to that of conventionally-interpreted ASC-US cytology. The automated algorithm we used performed very slightly worse than conventionally interpreted cytology for HPV-positive women overall, but matched cytology within strata when defined by HPV type groups. We doubt that automated cytology would prove cost-effective among HPV-negative women as a form of cotesting; the target of treatable precancers is so uncommon among women negative for high-risk HPV types that use of a second method for all women as opposed to the 5–10% that are HPV-positive is likely to be very expensive per additional case found (6).

The introduction of primary HPV screening raises the possibility of over-treatment. Even if screening begins at 25 or 30 years old, which is considered past the peak of HPV acquisition that occurs in the first decade following initiation of sexual intercourse, a small if declining percentage of screened women will be found to have high-risk HPV at each screening round (22). Incidence declines with age but over a lifetime of repeated screening, a large proportion of women could test positive for new or re-appearing HPV infections (23), most of which would be ultimately benign. Colposcopy referral rates could rise if HPV-positive women are referred straight to colposcopy.

Moreover, once referred to the colposcopy clinic, the histologic diagnosis of precancer is made deliberately broad, in recognition of our current inability to predict which lesions pose a true cancer risk. Thus, if HPV-positive women are referred to colposcopy, the likelihood of treatment is increased. Specifically, the most commonly chosen treatment threshold is cervical intraepithelial neoplasia grade 2 or worse (CIN2) (13). This definition of precancer implicitly accepts a considerable amount of effort and over-treatment in order to reduce the population risk of cervical cancer (24). Only about 15% of HPV-positive women referred to colposcopy have CIN2 diagnosed within 3 years (25). Moreover, most of those lesions would not cause invasive cancer, if left untreated (24).

Thus, a triage step between screening and colposcopic biopsy is essential. Based on the successful results of this proof-of-principle evaluation of automated cytology, the next step



in establishing whether it could be an effective triage technique for screening HPV-positive women will be to evaluate the strategy in a large screening setting. Our initial algorithm development and validation was performed in a colposcopy and treatment clinic population, usefully maximizing the number of CIN2 outcomes and including numerous women referred with invasive cancer. This had the effect of skewing the results toward higher sensitivity (due to more severe cases) but very low specificity. In a screening population, the histologic category of CIN2 would contain many fewer cancers as a proportion of the total case group. Also, those women referred to colposcopy with HPV-positive NILM might include many with lesions too early and small to be seen and biopsied. Ideally, the algorithm should be designed in a population similar to the one in which it will be used (i.e., women in a screening population aged 25 and older, found to be high-risk HPV-positive), using a firmer definition of precancer, i.e., CIN3/AIS or worse and prospective follow-up. We are conducting such a study of absolute risks ascertained prospectively in a screening population in collaboration with Kaiser Permanente Northern California.

It is noteworthy that the FocalPoint severity score substantially stratified risk of precancer among all HPV-positive women referred to colposcopy, regardless of HPV type group. Because it stratified risk even among women that might be referred immediately to colposcopy because of HPV16 or HPV18/45 typing results, knowing the severity score, possibly combined into meaningful categories, might usefully alert the colposcopist to a particularly high-risk patient. However, arguing somewhat against this point is the failure of the FocalPoint to predict risk of precancer as well as a conventional interpretation of HSIL. Because the algorithm stratified risk among women positive for types not targeted by current prophylactic HPV vaccines, it might continue to be useful in a partially vaccinated population. The eventual role of screening in well-vaccinated populations, once those birth cohorts constitute the majority of women, is a future question beyond the scope of this study.

The FocalPoint optical scanner appears to work reasonably well with ThinPrep slides, although it was developed for conventional Pap smears or SurePath liquid-based preparations. A non-negligible fraction of this aging archived collection could not be analyzed and, among those analyzed, a sizable proportion of the stored slides were judged to be suboptimal (although excluding them from the validation set did not noticeably affect algorithm performance). Nonetheless, ThinPrep slides differ enough from SurePath slides, for which FocalPoint is optimized, to merit a thorough investigation of whether the kind of liquid-based cytology is critical to FocalPoint performance. We anticipate improved performance from FocalPoint using SurePath slides in our on-going work, and are testing this expectation directly at this time.

In terms of long-range goals, we can imagine a totally automated screening and triage technology, for use in places that cannot maintain first-rate, conventional cervical cytology screening programs that require a large and skilled workforce of cytotechnologists and cytopathologists. A hybrid molecular processor/optical scanning machine could yield partial HPV typing and automated cytology results without the need for conventional interpretation, except in unusual circumstances. Such an instrument could be maintained in a regional diagnostics laboratory, eliminating the need for large-scale cervical cytology screening by cytotechnologists and interpretation by cytopathologists. The combination of molecular and

imaging technologies could thereby extend high-quality screening to more women, if the equipment and reagents can be offered affordably.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Funding: The study was supported by the Intramural Research Program of the National Cancer Institute, NCI.

## References

1. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Cogliano V. WHO International Agency for Research on Cancer Monograph Working Group. A review of human carcinogens--Part B: biological agents. *Lancet Oncol.* 2009 Apr; 10(4):321–2. [PubMed: 19350698]
2. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet.* 2007 Sep 8; 370(9590):890–907. Review. [PubMed: 17826171]
3. Bosch FX, Broker TR, Forman D, Moscicki AB, Gillison ML, Doorbar J, Stern PL, Stanley M, Arbyn M, Poljak M, Cuzick J, Castle PE, Schiller JT, Markowitz LE, Fisher WA, Canfell K, Denny LA, Franco EL, Steben M, Kane MA, Schiffman M, Meijer CJ, Sankaranarayanan R, Castellsague X, Kim JJ, Brotons M, Alemany L, Albero G, Diaz M, de Sanjose S. authors of ICO Monograph Comprehensive Control of HPV Infections and Related Diseases Vaccine Volume 30, Supplement 5, 2012. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine.* 2013 Dec 31; 31(Suppl 7):H1–31. DOI: 10.1016/j.vaccine.2013.10.003 [PubMed: 24332295]
4. Wentzensen N, Schiffman M, Palmer T, Arbyn M. Triage of HPV positive women in cervical cancer screening. *J Clin Virol.* 2016 Mar; 76( Suppl 1):S49–55. Epub 2015 Nov 28. Review. DOI: 10.1016/j.jcv.2015.11.015 [PubMed: 26643050]
5. Ronco G, Dillner J, Elfstrom KM, Tunesi S, Snijders PJ, Arbyn M, Kitchener H, Segnan N, Gilham C, Giorgi-Rossi P, Berkhof J, Peto J, Meijer CJ. International HPV screening working group. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet.* 2014 Feb 8; 383(9916):524–32. Epub 2013 Nov 3. Erratum in: *Lancet.* 2015 Oct 10;386(10002):1446. DOI: 10.1016/S0140-6736(13)62218-7 [PubMed: 24192252]
6. Gage JC, Schiffman M, Katki HA, Castle PE, Fetterman B, Wentzensen N, Poitras NE, Lorey T, Cheung LC, Kinney WK. Reassurance against future risk of precancer and cancer conferred by a negative human papillomavirus test. *J Natl Cancer Inst.* 2014 Jul 18.106(8) pii: dju153. Print 2014 Aug. doi: 10.1093/jnci/dju153
7. Rodriguez AC, Schiffman M, Herrero R, Wacholder S, Hildesheim A, Castle PE, Solomon D, Burk R. Proyecto Epidemiologico Guanacaste Group. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst.* 2008 Apr 2; 100(7):513–7. Epub 2008 Mar 25. DOI: 10.1093/jnci/djn044 [PubMed: 18364507]
8. Jeronimo J, Schiffman M. Colposcopy at a crossroads. *Am J Obstet Gynecol.* 2006 Aug; 195(2): 349–53. Epub 2006 May 3. [PubMed: 16677597]
9. Petry KU, Luyten A, Scherbring S. Accuracy of colposcopy management to detect CIN3 and invasive cancer in women with abnormal screening tests: results from a primary HPV screening project from 2006 to 2011 in Wolfsburg, Germany. *Gynecol Oncol.* 2013 Feb; 128(2):282–7. Epub 2012 Oct 23. DOI: 10.1016/j.ygyno.2012.10.017 [PubMed: 23088926]
10. Bulkman NW, Berkhof J, Rozendaal L, van Kemenade FJ, Boeke AJ, Bulk S, Voorhorst FJ, Verheijen RH, van Groningen K, Boon ME, Ruitinga W, van Ballegooijen M, Snijders PJ, Meijer CJ. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. *Lancet.* 2007 Nov 24; 370(9601):1764–72. Epub 2007 Oct 4. [PubMed: 17919718]

11. Kjaer SK, Frederiksen K, Munk C, Iftner T. Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: role of persistence. *J Natl Cancer Inst.* 2010 Oct 6; 102(19):1478–88. Epub 2010 Sep 14. DOI: 10.1093/jnci/djq356 [PubMed: 20841605]
12. Huh WK, Ault KA, Chelmow D, Davey DD, Goulart RA, Garcia FA, Kinney WK, Massad LS, Mayeaux EJ, Saslow D, Schiffman M, Wentzensen N, Lawson HW, Einstein MH. Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. *Obstet Gynecol.* 2015 Feb; 125(2):330–7. DOI: 10.1097/AOG.0000000000000669 [PubMed: 25569009]
13. Massad LS, Einstein MH, Huh WK, Katki HA, Kinney WK, Schiffman M, Solomon D, Wentzensen N, Lawson HW. 2012 ASCCP Consensus Guidelines Conference. 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. *Obstet Gynecol.* 2013 Apr; 121(4):829–46. DOI: 10.1097/AOG.0b013e3182883a34 [PubMed: 23635684]
14. Schiffman M, Burk RD, Boyle S, Raine-Bennett T, Katki HA, Gage JC, Wentzensen N, Kornegay JR, Aldrich C, Tam T, Erlich H, Apple R, Befano B, Castle PE. A study of genotyping for management of human papillomavirus-positive, cytology-negative cervical screening results. *J Clin Microbiol.* 2015 Jan; 53(1):52–9. Epub 2014 Oct 22. DOI: 10.1128/JCM.02116-14 [PubMed: 25339396]
15. Stoler MH, Schiffman M. Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA.* 2001 Mar 21; 285(11):1500–5. [PubMed: 11255427]
16. Scott DR, Hagmar B, Maddox P, Hjerpe A, Dillner J, Cuzick J, Sherman ME, Stoler MH, Kurman RJ, Kiviat NB, Manos MM, Schiffman M. Use of human papillomavirus DNA testing to compare equivocal cervical cytologic interpretations in the United States, Scandinavia, and the United Kingdom. *Cancer.* 2002 Feb 25; 96(1):14–20. [PubMed: 11836698]
17. Wentzensen N, Schiffman M, Dunn T, Zuna RE, Gold MA, Allen RA, Zhang R, Sherman ME, Wacholder S, Walker J, Wang SS. Multiple human papillomavirus genotype infections in cervical cancer progression in the study to understand cervical cancer early endpoints and determinants. *Int J Cancer.* 2009 Nov 1; 125(9):2151–8. DOI: 10.1002/ijc.24528 [PubMed: 19585494]
18. Wentzensen N, Walker JL, Gold MA, Smith KM, Zuna RE, Mathews C, Dunn ST, Zhang R, Moxley K, Bishop E, Tenney M, Nugent E, Graubard BI, Wacholder S, Schiffman M. Multiple biopsies and detection of cervical cancer precursors at colposcopy. *J Clin Oncol.* 2015 Jan 1; 33(1):83–9. Epub 2014 Nov 24. DOI: 10.1200/JCO.2014.55.9948 [PubMed: 25422481]
19. Nayar R, Solomon D. Second edition of ‘The Bethesda System for reporting cervical cytology’ - atlas, website, and Bethesda interobserver reproducibility project. *Cytojournal.* 2004 Oct 21.1(1):4. [PubMed: 15504231]
20. Patten SF Jr, Lee JS, Wilbur DC, Bonfiglio TA, Colgan TJ, Richart RM, Cramer H, Moinuddin S. The AutoPap 300 QC System multicenter clinical trials for use in quality control rescreening of cervical smears: I. A prospective intended use study. *Cancer.* 1997 Dec 25; 81(6):337–42. [PubMed: 9438458]
21. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997 Feb 28; 16(4):385–95. [PubMed: 9044528]
22. Clifford GM, Gallus S, Herrero R, Munoz N, Snijders PJ, Vaccarella S, Anh PT, Ferreccio C, Hieu NT, Matos E, Molano M, Rajkumar R, Ronco G, de Sanjose S, Shin HR, Sukvirach S, Thomas JO, Tunsakul S, Meijer CJ, Franceschi S. IARC HPV Prevalence Surveys Study Group. Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. *Lancet.* 2005 Sep 17–23; 366(9490):991–8. [PubMed: 16168781]
23. Gage JC, Katki HA, Schiffman M, Fetterman B, Poitras NE, Lorey T, Cheung LC, Castle PE, Kinney WK. Age-stratified 5-year risks of cervical precancer among women with enrollment and newly detected HPV infection. *Int J Cancer.* 2015 Apr 1; 136(7):1665–71. Epub 2014 Sep 2. DOI: 10.1002/ijc.29143 [PubMed: 25136967]

24. Schiffman M, Rodriguez AC. Heterogeneity in CIN3 diagnosis. *Lancet Oncol*. 2008 May; 9(5): 404–6. DOI: 10.1016/S1470-2045(08)70110-4 [PubMed: 18452848]
25. Wright TC, Stoler MH, Behrens CM, Sharma A, Zhang G, Wright TL. Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test. *Gynecol Oncol*. 2015 Feb; 136(2):189–97. Epub 2015 Jan 8. DOI: 10.1016/j.ygyno.2014.11.076 [PubMed: 25579108]

Author Manuscript

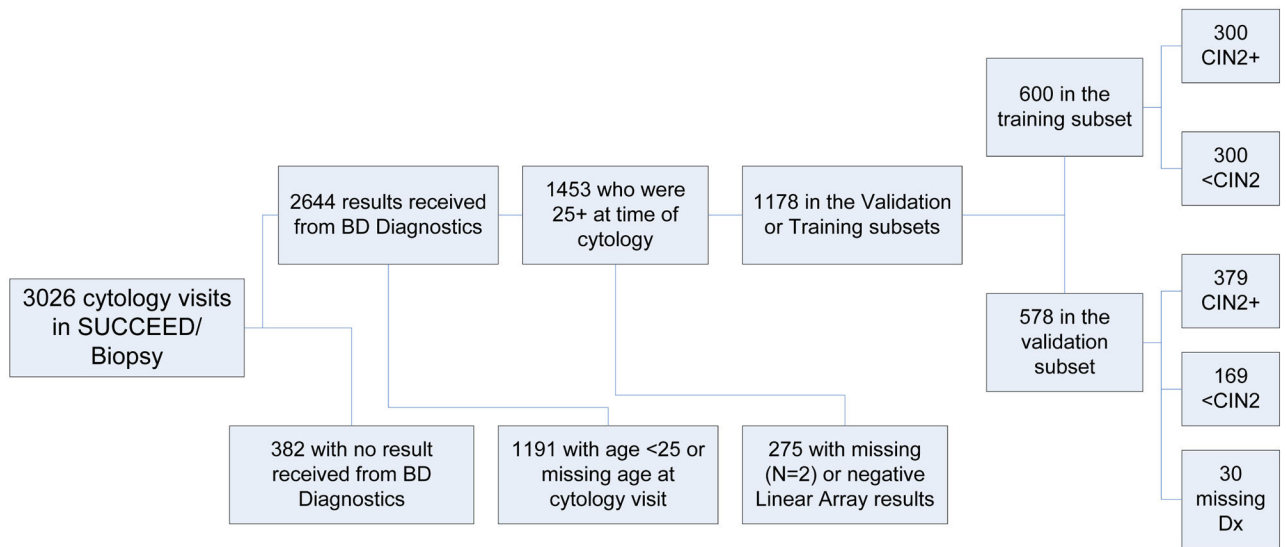
Author Manuscript

Author Manuscript

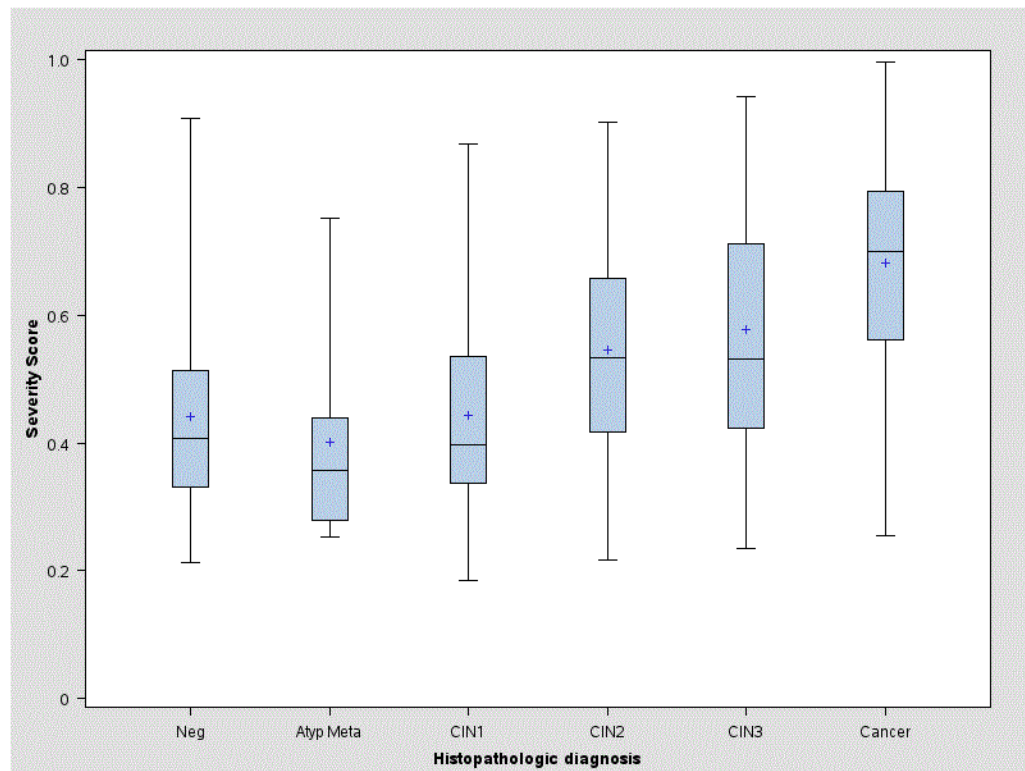
Author Manuscript

### Novelty and Impact Statement

Risk-stratification (“triage”) of HPV-positive women using cytology (ASC-US threshold) has been recommended to reduce excessive colposcopy referral and over-treatment that would result from HPV screening alone. In a proof-of-principle study within a colposcopy referral population, we show that automated cytology (a re-programmed BD FocalPoint™ algorithm) performs as well as conventionally interpreted cytology for triage of HPV-positive women, whether or not partial HPV typing is included. This raises the possibility of completely automated cervical cancer screening.

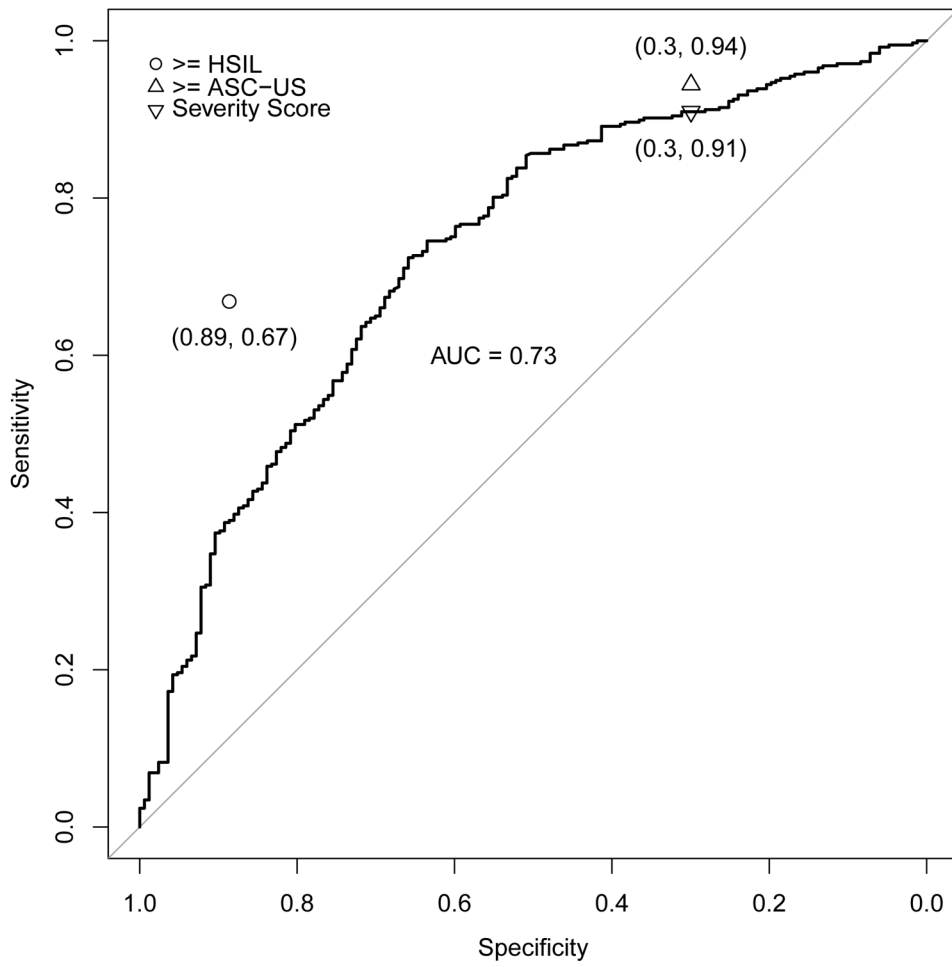


**Figure 1.** Study Population. Creation of a training set and a validation set from slides collected as part of the Study to Understand Cervical Cancer Early Endpoints and Determinants (SUCCEED), which was conducted within a colposcopy clinic population in Oklahoma City, Oklahoma.

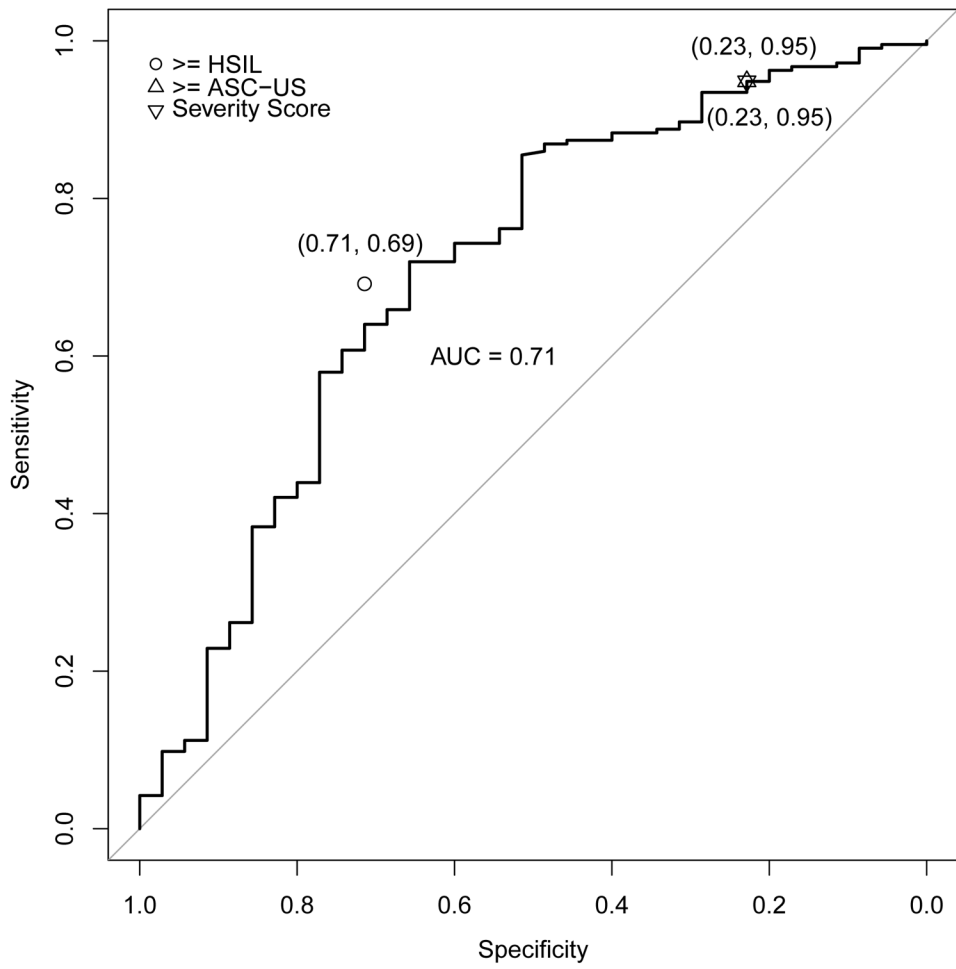


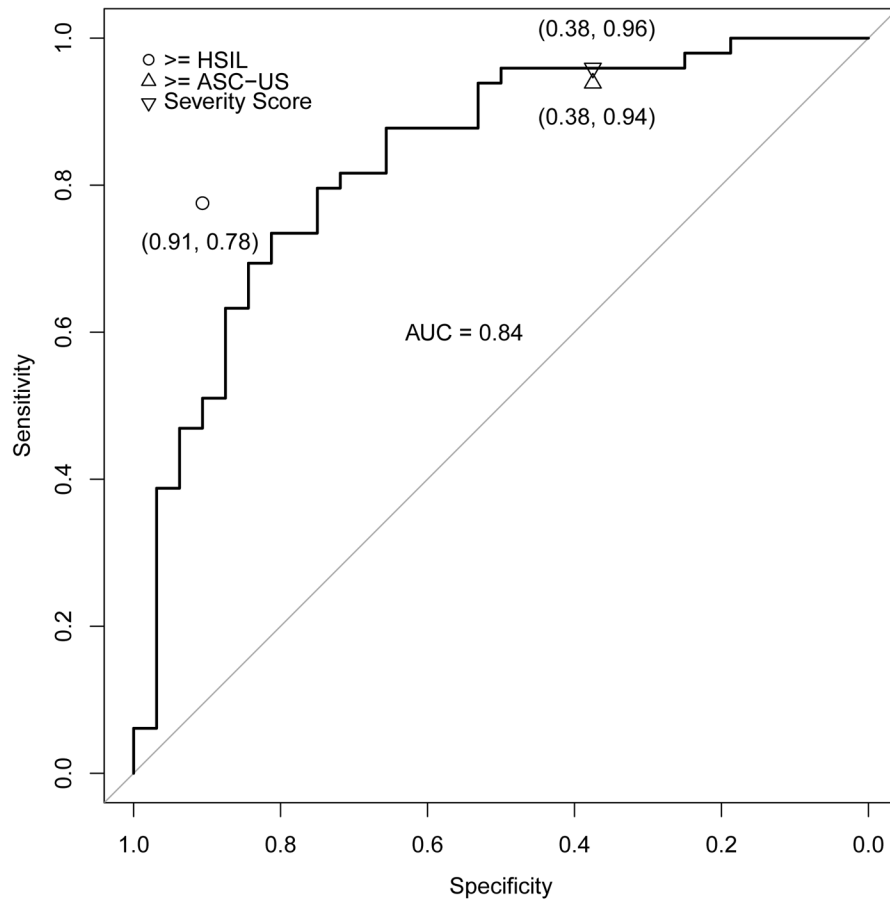
**Figure 2.**

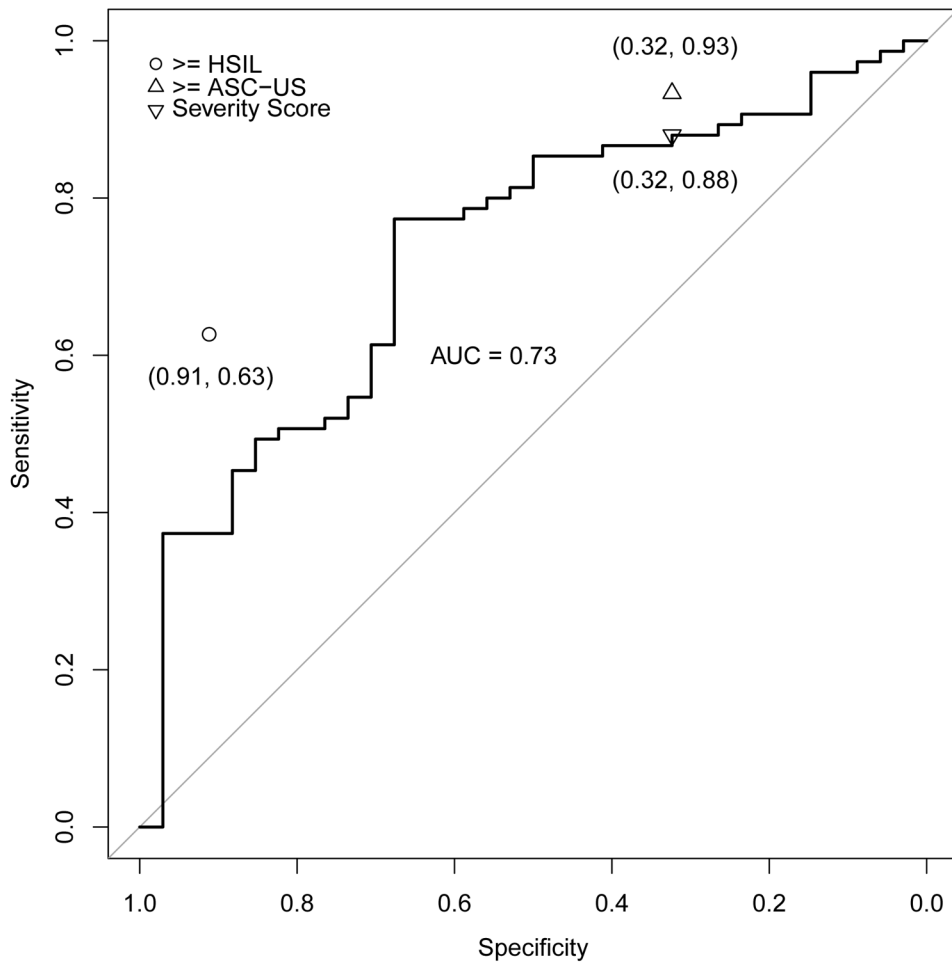
Average severity score for each histopathologic diagnosis in the validation set. The histopathologic diagnosis was that assigned routinely during clinical care by the OUHSC pathologists. For each diagnosis, we plotted the distribution of severity scores produced by the LASSO-generated algorithm (varying from 0 to 1). The average score tended to increase with increasingly severe histopathologic diagnosis within cases (>CIN2) but did not markedly increase within controls (<CIN2). Of note, in this referral population, the control group was not representative of the general population, and all women in the study were currently positive for high-risk types of HPV.

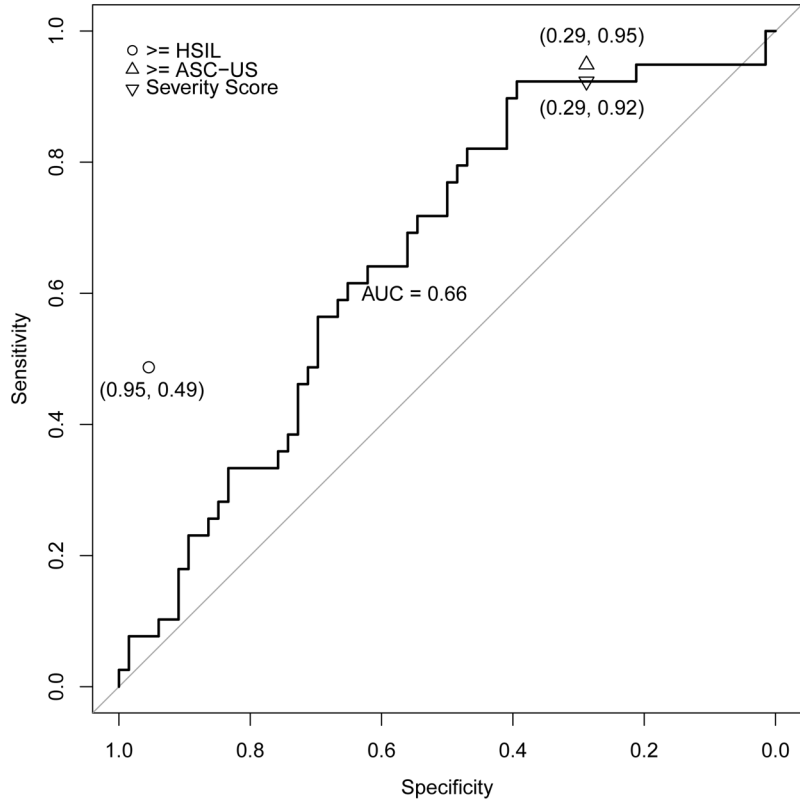












**Figure 3.**

ROC curve for severity score, with area under the curve (AUC with 95% confidence interval), and comparison of severity score cut-point to conventionally-read cytology (ASC-US, HSIL) thresholds. Figure 3a. All high-risk HPV-positive, AUC = 0.73 (0.69, 0.78); Figure 3b. HPV16, AUC = 0.71 (0.61, 0.81); Figure 3c. HPV18/45, AUC = 0.84 (0.74, 0.93); Figure 3d. HPV31/33/52/58, AUC = 0.73 (0.63, 0.83); Figure 3e. HPV35/39/51/56/59/66/68, AUC = 0.66 (0.55, 0.77). The results indicate that in this referral population, the computer-generated severity score yielded discrimination of cases of CIN2 versus non-cases with a combination of sensitivity and specificity similar to conventional cytologic results of ASC-US+ versus NILM. The inverted triangle is to denote the hypothetical choice of a cutpoint to simulate a standard reading of the ThinPrep cytology slide at an ASC-US threshold.

**Table 1**

Mean and standard deviation of severity score, by cytology result and case/control status in the validation set

<b>Cytology Result</b>	<b>Case-Ctl</b>	<b>Obs. (N)</b>	<b>Mean</b>	<b>Std Dev</b>
NILM	<CIN2	50	0.39	0.13
	CIN2+	21	0.44	0.15
ASC-US/LSIL	<CIN2	85	0.42	0.13
	CIN2+	60	0.50	0.14
AGUS/ASC-H	<CIN2	13	0.53	0.18
	CIN2+	44	0.49	0.15
HSIL	<CIN2	18	0.57	0.19
	CIN2+	222	0.61	0.17
Cancer	<CIN2	1	0.68	.
	CIN2+	30	0.71	0.13

Increasingly severe cytologic results were tended to have higher average severity scores ( $p=8.6e-7$  in <CIN2,  $p=2.0e-12$  in CIN2).

**Table 2**

Mean and standard deviation of severity score, by HPV type group and case/control status in the validation set

HPV Type Group	Case-Ctl	Obs (N)	Mean	Std Dev
Negative	<CIN2	233	0.40	0.13
	CIN2+	36	0.51	0.19
35/39/51/56/59/66/68	<CIN2	68	0.45	0.14
	CIN2+	39	0.53	0.15
31/33/52/58	<CIN2	34	0.43	0.14
	CIN2+	75	0.57	0.17
18/45	<CIN2	32	0.40	0.14
	CIN2+	49	0.61	0.17
16	<CIN2	35	0.46	0.18
	CIN2+	216	0.58	0.18

HPV-negative women overall tended to have lower severity scores than HPV-positive women ( $p = 0.002$  among cases, and  $p = 0.01$  among controls). Among HPV-positive women, severity scores were not associated with HPV type groups, when stratified by <CIN2/ CIN2 outcome.