REDEFINING THE GOLD STANDARD FOR DIAGNOSIS AND RISK FACTOR ANALYSIS IN CERVICAL DYSPLASIA

By

Emily E. King

A THESIS

Presented to the Department of Public Health & Preventive Medicine
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Master of Public Health

March 2010

Department of Public Health and Preventive Medicine

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's thesis of

Emily E. King

has been approved

Michelle Berlin, ~~MD~~, MPH
Thesis Chair

Terry Morgan, MD, PhD
Thesis Advisor

Tomi Mori, PhD
Thesis Advisor

TABLE OF CONTENTS

Figures, Tables & Appendices

Abbreviations & Glossary

Cervical intraepithelial neoplasia (CIN) – A term used to categorize degree (grade 1, 2 or 3) of dysplasia observable microscopically in the epithelium of the exocervix.

"CIN 2+" – A diagnostic designation that indicates that the specimen has histologic features that warrant a diagnosis of at least CIN 2 but where the pathologist does not specifically differentiate between CIN 2 and CIN 3. CIN 2+, CIN 2, and CIN 3 are equivalent to high grade squamous intraepithelial lesion (HSIL).

Human papillomavirus (HPV) – A double-stranded DNA virus (Papillomaviridae) that infects cells of human epidermis and mucosal membranes and thought to be a necessary cause of human cervical cancer.

Dysplasia – A term used to describe abnormal cellular morphologic features observable with a microscope such as nuclear and cellular pleomorphism, decreased nuclear to cytoplasm ratio, increased mitotic activity, and loss of other normal cellular features.

Hematoxylin & eosin (H&E) – Two chemical stains applied to tissue sections on microscopic slides that allow conventional evaluation and diagnosis of the tissue.

Immunohistochemistry (IHC) – A laboratory technique with allows visualization and localization of cellular antigens by staining tissue with chromagen-tagged antibodies.

p16 – A tumor suppressor protein that is overexpressed in cervical keratinocytes that have undergone malignant transformation. IHC can be used to detect overexpression of p16.

Ki-67 – A nuclear protein involved in cellular proliferation. Dysplastic, inflammatory, and reparative processes in cervical keratinocytes will stain for Ki-67 by IHC

Intraobserver reproducibility – A term used to describe the degree to which a single observer (eg, a single pathologist) will be able to reproducibly categorize or diagnose a single specimen on multiple reviews of the specimen.

Interobserver reproducibility – A term used to describe the degree to which multiple observers (eg, several pathologists) will be able to reproducibly categorize or diagnose a single specimen on single or multiple reviews of the specimen.

ASCUS/LSIL Traige Study for Cervical Cancer (ALTS) – A multicenter randomized trial comparing the efficacy and cost-effectiveness of different management strategies for low-grade lesions to guide appropriate treatment and early detection of high-grade lesions

Measurement Error – The difference between the true quantity being assessed and the quantity that can be measured using available measurement methods.

Cohen's kappa statistic – A statistic that can be used to mathematically quantify the degree of reproducibility in measurement or diagnostic assessment.

Misclassification bias – A term used to describe bias in the estimation of risk that is due to improper categorization of either exposure or outcome status of patients. Depending on degree of misclassification, this bias may significantly alter the magnitude of the risk estimate and affect interpretation of the results of the investigation.

Misclassification structure – A diagrammatic representation of outcome misclassification detailing the movement of cases between diagnostic categories due to use of different diagnostic methods.

Overall kappa estimates – In this investigation, this term refers to assessment of reproducibility that can be achieved when considering all grades of CIN (1-3) simultaneously.

Grade-specific kappa estimates – In this investigation, this term refers to assessment of reproducibility that can be achieved when considering a single grade of CIN (1, 2 or 3) individually.

Kaiser Permanente-Northwest (KP-NW) – A health maintenance organization in the Pacific Northwest with >450,000 enrolled members in Oregon and Washington.

Kaiser Permanente Center for Health Research (KPCHR) – A division of KP-NW that manages and coordinates clinical information being used for investigations involving KP-NW patients.

Oregon Health & Science University (OHSU)

Oregon Clinical and Translational Research Institute (OCTRI)

Acknowledgments

My thesis committee, Michelle Berlin, Terry Morgan, Motomi Mori, who have given me invaluable epidemiologic, pathologic, biostatisical, and professional advice and mentorship over the course of this thesis.

Rob Krum, gynecologic pathologist and our collaborator at Kaiser-Permanente, who has contributed significantly to data collection and recommended modifications for methods of data collection.

Denise Schwarzkopf & Dan Sapp, project managers at KPCHR and Parker Pettus, research analyst at KPCHR, who were essential for acquisition of clinical data and instrumental in teaching me strategies for operationalizing variables

Cindy Morris, Mary Samuels & OCTRI, who through the Oregon Students Learn & Experience Research (OSLER) TL1 program, provided invaluable professional, logistic, and administrative support and mentorship

Ethan Siefert, who provided essential programming and database maintenance support for RedCAP through OCTRI

The many members of the Christopher Corless laboratory and Terry Morgan laboratory who provided assistance preparing and staining microscopic slides for this investigation.

The faculty of the OHSU Public Health & Preventive Medicine Department including, John Stull, Bill Lambert, Katie Riley, and Tom Becker, who have all significantly contributed to my professional development and to my education in epidemiology and biostatistics

My friends and family, who have counseled, supported, and loved me unconditionally through the process of my thesis, medical school, and before; especially my parents, Anna, Sarah, Jodge, Jessica, Bjorn, and Jack – thank you.

Abstract

Introduction: Cervical intraepithelial neoplasia (CIN) diagnostic methods are suboptimal due to limitations in reproducibility and accuracy, which affects clinical management and understanding of risk factors for CIN. These limitations are especially relevant for the histologic diagnosis of moderate dysplasia (CIN 2). We hypothesize that molecular markers, such as p16 and Ki-67, may improve diagnostic reproducibility and accuracy and also lead to a better understanding of cervical cancer epidemiology. **Methods:** A randomly selected retrospective cohort of 300 women with cervical dysplasia diagnosed by colposcopic biopsy and at least five years of clinical follow-up at Kaiser-Permanente Northwest in Portland, Oregon was obtained. Two experienced gynecologic pathologists (A & B) independently reviewed histologic sections of the colposcopic biopsy material while blinded to each other's assessments and long-term clinical outcome. Diagnoses made using routine H&E stained histologic slides were compared to diagnoses made using 1) H&E plus p16 and 2) H&E plus p16 and Ki-67. Cohen's kappa statistic was used to quantify diagnostic reproducibility. Measures of test accuracy were determined by comparing colposcopic biopsy diagnoses to a consensus outcome from 5 years of clinical follow-up. Multivariate logistic regression was used to model the relationship between risk factors and reviewer's CIN diagnoses. **Results:** The reproducibility of CIN 2 diagnosis was significantly improved by using H&E plus p16 (κ=0.4783) compared to H&E only (κ=0.4041, $p<0.05$). The reproducibility of CIN 2 diagnosis using H&E plus p16 and Ki-67 was significantly improved compared to H&E only (κ=0.5204, $p<0.05$), but not compared to H&E plus p16 (κ=0.4783, $p>0.05$). H&E plus p16 significantly improved sensitivity and negative predictive value compared to H&E only for both reviewers. Unadjusted odds ratio estimates for CIN 2+ if exposed to high (>$45,000 annual) family income was not significant (OR 0.892, $p>0.05$) using H&E only as diagnostic method but was significant when using H&E plus p16 (OR 0.585, $p<0.05$). Multivariate logistic regression showed a similar trend in magnitude of ORs in this pilot study. **Conclusions:** This pilot study suggests that misclassification in diagnosis of CIN may have significant ramifications for epidemiologic research. For example, p16 contributed to significant changes in the estimated risk of CIN 2+ for women with low family income, which would have been missed by conventional diagnostic methods. It demonstrated that improvements in diagnostic precision and accuracy gained with markers like p16 may provide a "new gold standard" with which to evaluate strength of associations.

Introduction

I. Overview

Cervical cancer screening programs which have relied on conventional cervical cytologic and colposcopically-directed biopsy diagnoses have successfully reduced the incidence and mortality of invasive cervical cancer worldwide. (1) This reduction has consequently shifted the emphasis of cervical cancer prevention toward a better understanding of natural history and biology of cervical precancerous lesions (eg, cervical dysplasia). Despite the success of screening programs, however, current screening programs rely on diagnostic methods with low sensitivity and poor diagnostic reproducibility.

The American Society for Colposcopy and Cervical Pathology (ASCCP) published revised recommendations in 2007 for the management of women with cervical dysplasia (cervical intraepithelial neoplasia, CIN). (2) Prior to these guidelines, CIN 2 and CIN 3 lesions were grouped together histopathologically as "CIN 2+," because of poor interobserver reproducibility in differentiating CIN 2 from CIN 3, and at the time, CIN 2 and CIN 3 received the same clinical management (surgical excision). The revised recommendations now substantially change clinical management strategies for CIN 2 and CIN 3: clinical observation with serial Pap smears or colposcopy for CIN 2 and definitive surgical excision for CIN 3. (2)

The accurate classification of CIN 2 and CIN 3 is also important in epidemiologic studies to evaluate the risk of specific exposures associated with the pathogenesis of cervical cancer. The validity of these epidemiologic studies relies on precise and accurate CIN diagnoses by the pathologist to minimize sources of misclassification. (3) Despite generally accepted morphologic criteria, the distinction between CIN 1, 2, and 3 lesions remains challenging for pathologists. (4) This leads to persistent poor reproducibility, particularly in the diagnosis of CIN 2, which introduces misclassification that may bias the conclusions of epidemiologic studies. (3, 5-13) There is a clear need to improve the reproducibility and accuracy of histopathologic CIN diagnoses to better guide patient care and to minimize misclassification bias in cervical cancer risk estimations.

For example, a recently published study identified stronger epidemiologic associations between HPV status and CIN when using exposure and outcome measures that have greater certainty of correct classification. (14) This study demonstrated the well-established epidemiologic principle that minimizing misclassification, both for a given exposure and given outcome, will result in stronger measures of association between that exposure and outcome. It

suggests that the degree of misclassification currently present in epidemiologic studies of cervical dysplasia may be sufficient to significantly alter estimates of relative risk and bias current understanding of cervical cancer epidemiology. (15, 16, 16)

One solution to this problem may be molecular-based methods, which appear to improve cervical biopsy diagnostic reproducibility and accuracy. (12, 13, 17-23) The purpose of this thesis is to explore the implications of these molecular-based methods to diagnose CIN on the clinical management of individual patients and the conduct of epidemiologic studies. This thesis will add to the existing body of literature by:

1. Evaluating the effect of a novel combination of two molecular diagnostic markers (p16 and Ki-67) added to conventional diagnostic methods on interobserver reproducibility, sensitivity, specificity, and positive and negative predictive values when doing histologic assessment of CIN lesions,

2. Determining the amount and kind of diagnostic reclassification of CIN lesions that occurs when using conventional diagnostic methods only (H&E only) compared to a combination of conventional and molecular and conventional diagnostic methods: 1) H&E plus p16 and 2) H&E plus p16 and Ki-67,

3. Determining whether using a diagnostic method (either H&E plus p16 or H&E plus p16 and Ki-67) that demonstrates significant improvements in interobserver reproducibility, sensitivity, specificity, and positive and negative predictive values in the assessment of CIN significantly changes the observed strength of association between risk factors and CIN. We anticipate that stronger associations will be observed when CIN is diagnosed using molecular and conventional diagnostic methods compared to when CIN is diagnosed using the conventional diagnostic method alone.

II. Human papillomavirus (HPV) and cervical cancer

HPV is a prevalent pathogen in the United States and worldwide (11-27% and 44%, respectively). (24-26) Moreover, HPV prevalence estimates are as high as 90% among certain high risk patient populations. (27) The incidence of HPV infection is estimated to be greater than 5.5 million cases annually in the United States. (28) Of the millions of new HPV infections each year in the United States, only infections with fifteen genotypes of HPV (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82) are considered oncogenic ("high-risk"). (29-31) High-risk HPV infection is estimated to be prevalent in 90-99% of invasive cervical cancer, with HPV 16 and 18 being the most prevalent subtypes. (29, 31-35)

HPV is a double-stranded DNA virus with >100 genetic subtypes. (36) Thirty to 40 of the HPV subtypes preferentially infect mucosa of the anogenital region, including the cervix.(26, 36) When HPV infects a basal cell in the cervical epithelium, the viral genome localizes to the nucleus, but remains as an extrachromosomal element. In a productive infection, viral DNA is amplified and host cellular machinery is used to produce viral progeny. Ultimately, mature cells containing infectious virus particles are shed from the surface of the cervix. (37) In order to induce cervical carcinogenesis, it is hypothesized that the viral genome must first integrate with host DNA, leading to expression of viral genes *E6* and *E7*. (37, 38) The cellular effects of E6 and E7 are complex. E6 is known to bind and degrade host tumor-suppressor p53. E7 is known to functionally inactivate host tumor-suppressor retinoblastoma (pRb), leading to subsequent overexpression of p16 protein (Figure 1). (12, 37) Through incompletely characterized pathways, the normal tumor suppressor effect of p16 is overcome, allowing unchecked progression through the cell cycle. Using animal models to study the isolated effect of E6 and E7, it appears that the E6 and E7 pathways overlap and potentiate the processes of immortalization and malignant transformation. (39-45) Unfortunately, there are no reliable genetic assay methods to distinguish between cells that have undergone malignant transformation by high risk HPV and those in which HPV remains an extrachromosomal element in non-dysplastic cells (eg, clinical HPV testing of cervical Pap smears is more sensitive, but not as specific for dysplasia as microscopic evaluation of the specimen (46)).

HPV is transmitted via skin-to-skin contact and therefore may be transmitted via non-intercourse foreplay, contact with skin not covered by barrier contraceptives, and sex toys. (47) Most women acquire their first HPV infection shortly after coitarche, usually as adolescents, and they may acquire infections of different HPV genotypes or become reinfected with the same

genotype with subsequent sexual exposure. The adolescent cervix is uniquely susceptible to HPV infection, because during puberty the cervix undergoes change from predominately glandular to predominately squamous mucosa. The process of squamous metaplasia involves cellular proliferation that both supports HPV replication and the likelihood of virus-induced genetic alterations. It is hypothesized that virus-induced genetic alterations in the setting of a persistent infection, may ultimately lead to high-grade CIN lesions and invasive carcinoma. (47-49)

Serial HPV testing has shown that >80% of HPV infections become undetectable within 2 years of the incident infection. (26, 50) Two potential explanations for the apparent resolution of HPV infection exist. The first potential explanation is that the virus is controlled and cleared by the host immune system. (50-53) It is unclear which host immunity factors contribute to successful clearance of virus infection in some women, while other women experience persistently detectable high-risk HPV infection. (54) Alternatively, it has been hypothesized, based on observations in immunosuppressed patients, that in immunocompetent hosts, the virus enters a latent state and maintains a viral load undetectable by current testing methods.(55) Despite this confusion, various sources have estimated that among all ages of women 57% of CIN 1 lesions, 43-58% of CIN 2 lesions, and 32-47% of CIN3 lesions may be cleared and resolve. (56-59) The remaining infections are considered persistent and may progress to a higher grade CIN or invasive carcinoma. Of CIN 2 lesions, specifically, it is estimated that approximately 35% will persist unchanged and 22% will progress to CIN3.(57)

The highest risk of progression to high-grade CIN is during the first year after incident infection with a high-risk HPV subtype. (60) Infection with low-risk HPV genotypes such as HPV 6 and HPV 11 is commonly associated with genital warts (condylomata acuminata), occasionally associated with CIN 1, and rarely associated with CIN 2, but not associated with CIN 3 or invasive cervical cancer. (33, 61, 62) High-risk HPV genotypes have been identified in both CIN 1 (27%) and CIN 2+ (47%) lesions. (61)

It is important to emphasize that adolescent patients appear to represent a subgroup with unique biology, requiring different screening and management strategies. Among adolescents, the mean duration of incident HPV infection is estimated to be less than 2 years and more than 90% of HPV infections are cleared spontaneously by the immune system without causing cytologic or histologic abnormalities on screening examinations. (47, 61) In contrast to adult populations, various sources have estimated that greater than 92-94% of CIN 1 lesions in adolescents will convert to normal. (47) Fuchs et al. (2007) estimated the regression rate of CIN

2 among adolescents to be as high as 92%.(63) The regression rate of CIN 3 in adolescents is not known because the perceived risk of CIN 3 progressing to invasive cancer has prevented conservative management in this group.

In summary, persistent infection with high-risk oncogenic subtypes of HPV is a necessary cause of invasive cervical cancer. The biologic pathways and risk factors allowing for viral persistence are not well understood. Moreover, diagnostic criteria currently used to assess morphologic changes in cervical cells may not be adequate for predicting the biologic behavior of CIN lesions, especially in adolescents.

III. Conventional cervical intraepithelial neoplasia (CIN) diagnosis & management

The goal of cervical cancer prevention programs is to identify women with precancerous lesions. These so-called dysplastic lesions of human cervical squamous epithelial cells are currently classified by consensus cytologic and histomorphological criteria.(4, 48) Two schemes exist (CIN and SIL), each reflecting a different understanding of the nature of the pathophysiology that leads to the observable cellular changes. In the first scheme, precancerous lesions are classified into three grades of cervical intraepithelial neoplasia: low grade (CIN 1), moderate (CIN 2), and severe dysplasia (CIN 3), which is equivalent to carcinoma *in situ*(Figure 2). (4) This scheme was developed to reflect the belief that each grade represents a progressive stage on a single biological continuum and was based on observations that 1) any grade of CIN apparently could precede development of cervical carcinoma, 2) many women appeared to have progressively higher grade of CIN on biopsies preceding a diagnosis of cervical carcinoma, and 3) CIN and cervical carcinoma often had similar molecular abnormalities. (64) In the SIL scheme, precancerous lesions are classified into two grades of squamous intraepithelial lesion (SIL). Lesions that have similar histologic features as CIN 1 are called low-grade SIL, and lesions with similar histologic features as CIN 2 and CIN 3 are called high-grade SIL.(48)

The 2006 revised management guidelines acknowledge that "…histologic distinction between CIN 2 and CIN 3 lesions is poorly reproducible… Therefore, CIN 2 is utilized as the threshold for treatment in the United States to provide an added measure of safety." (2) However, these same guidelines offer different management strategies for CIN 2 and CIN 3 lesions in certain patient populations (eg, adolescents). In order to make this histologic distinction accurately and consistently, improvements in the currently available diagnostic methods are required.

Conventionally, pathologists use hematoxylin & eosin (H&E) stained microscope slides of cervical biopsy tissue to diagnose CIN lesions. Generally accepted histologic features of CIN 1 (LSIL) include condyloma morphology (flat or exophytic) and koilocytotic atypia (nuclear hyperchromasia, karyomegaly, binucleation, irregular nuclear contour, and perinuclear clearing) in the superficial epithelial layer. These features are believed to represent the effects of HPV replication within the cell cytoplasm and not neoplastic transformation. (48, 64) CIN 2 and CIN 3 lesions, in contrast, may show increased mitotic activity in the upper half of the epithelium, abnormal mitotic figures, loss of cell polarity, and occasional atypical, bizarre cells. (4, 65) The distinction between CIN 2 and CIN 3 relies on the relative thickness of high grade dysplasia. CIN 2 lesions maintain some degree of superficial epithelial maturation, while CIN 3 lesions show full thickness high grade dysplasia from the basal layer to the surface (Figure 2).

Using the conventional diagnostic method (H&E only), overall interobserver reproducibility, meaning agreement between two or more pathologists on any single CIN diagnosis for a single biopsy, ranges between 0.46 and 0.88, indicating moderate to excellent reproducibility (as measured by Cohen's kappa statistic). (5-11, 16, 16, 66) The range in overall diagnostic reproducibility can be partially explained by study design. For example, the kappa statistic will be higher with the use of fewer diagnostic categories, the use of fewer reviewers, the use of "expert" reviewers versus generalists, and the use of agreed upon consensus criteria for each type of diagnostic category before initiating review of the study cohort. Previously published investigations considering interobserver reproducibility have used diagnostic schemes with variable numbers of categories, between two and 22 observers, a mixture of expert gynecologic pathologists and general surgical pathologists, and variable diagnostic criteria, making comparisons between published results very challenging. (9, 67)

When considering grade-specific kappa estimates, CIN 2 is consistently the least reproducible diagnosis. (6-8) Among expert gynecologic pathologists, reproducibility of CIN 2 diagnosis is 0.38 (95% CI: 0.33-0.44). (6) Among all possible pairs from a pool of 15 pathologists, reproducibility of CIN 2 diagnosis ranged from <0.1 to 0.4. (12)

It warrants mention that calculation of diagnostic accuracy depends on knowledge of long-term clinical outcome (the patient's "true" disease state). If determination of outcome is only by colposcopic biopsy, the sensitivity of colposcopy is significant. Unfortunately, the sensitivity of colposcopic directed cervical biopsy is only moderate. The ASCUS/LSIL Triage Study (ALTS) demonstrated that among women with dysplasia detected by screening Pap smear,

nearly half (47%) of subsequent colposcopically-directed biopsies lacked pathologic lesions.(11, 68) Therefore, the most accurate "gold standard" outcome diagnosis used in studies of cervical dysplasia is often the highest grade of dysplasia obtained in long-term clinical follow-up.(68)

Until recently, to err on the side of safety, management guidelines and the standard of care categorized CIN 2 and CIN 3 lesions together under a single diagnostic category ("CIN 2+"). It was agreed that CIN 3 and CIN 2+ should be treated with definitive surgical excision due to the diagnostic limitations of Pap smear, colposcopy, and colposcopic-biopsy.(2) However, there are patient subgroups that may significantly benefit from more conservative therapy (clinical monitoring with serial Pap smears rather than surgery) if a diagnosis of CIN 2 could be reliably distinguished from CIN 3. First, adolescents (females $\leq$ 20 years old) represent a group that experiences a high rate of incident HPV infection but low rate of cervical cancer. Because of the morbidity associated with cervical surgery, such as increased risk of cervical incompetence and preterm delivery, management guidelines have been revised to support different clinical management strategies for CIN 2 and CIN 3 in adolescents. (2, 63, 69-71).

Conservative therapy may also benefit women with a CIN 2 diagnosis, who have a low clinical risk factor profile and who are likely to return for serial monitoring. In this group, a woman with CIN 2, in discussion with her gynecologist, may chose conservative clinical monitoring over a surgical excision procedure. The distinction between CIN 2 and CIN 3 is also relevant for women who have undergone surgical excision, but CIN 2+ is identified at the surgical margins. In these cases, it is prudent to excise additional tissue at the positive margin if CIN 3 is present, but CIN 2 may be managed with serial Pap smears and/or colposcopy. (72) If the CIN at the margin could be specified as CIN 2, which does not warrant re-excision, and differentiated from CIN 3, these patients would be spared the morbidity and cost associated with the re-excision procedure.

Unfortunately, the current H&E-based gold-standard diagnostic method is inadequate for differentiating between each grade of CIN. A new gold-standard diagnostic method, which would make diagnoses more accurate and more reproducible among pathologists would have significant impact on management of women with CIN. Ultimately, women who could be reliably identified as having $\leq$CIN 2 could be spared an excision procedure, while women with CIN 3 would be adequately treated.

When indicated, surgical treatment of CIN2+ is effective. A recent case-control study of 7,104 women treated for CIN and 35,437 individually matched controls concluded that there

was no difference in mortality due to cervical cancer after treatment of CIN compared to controls. (73) Although this observation suggests that current diagnostic methods and management of CIN are effective, these data are hard to reconcile with the wide variation in rates of progression for various grades of CIN and the substantial diagnostic discordance among pathologists. (5-13)

In total, this suggests that histomorphology as a sole diagnostic method may not provide adequate information to categorize lesions according to similar biologic behavior. Therefore, regardless of histomorphologic changes, women who have a high-risk HPV infection, especially if they 30 years or older, are considered to be at higher risk of progression to higher grade CIN or cervical cancer. In the absence of a diagnostic method that can reliably distinguish CIN 2 from severe dysplasia (CIN 3), which requires surgery, and low grade dysplasia (CIN 1), which can be followed clinically, management decisions will be flawed.

IV. Molecular-based diagnosis of CIN

Accumulating evidence suggests that diagnostic methods capitalizing on our current understanding of HPV pathophysiology will lead to substantial improvements in diagnostic reproducibility and accuracy. (12, 13, 18, 74) There are many candidate antigens involved in HPV-induced carcinogenesis that might be useful as diagnostic markers, including p16 and Ki-67, as well as topoisomerase II-α (TPIIα), minimicrosome maintenance protein 2 (MCM2), MCM7, retinoblastoma (Rb), cyclin E, and a cocktail stain of MCM2 and TPIIα called ProExC. (22, 23, 44, 75-77). These markers are usually assayed and visualized by the pathologist by a method called immunohistochemistry (IHC). IHC is a technique that relies on a labeled antibody binding its antigen. For example, a labeled p16 antibody will bind p16 protein in dysplastic cells. If this antibody is conjugated to a chromogenic reporter, binding of the p16 antibody to the p16 protein will create a brown precipitate, which can be seen in tissue sections. (78) It is possible to generate diagnostic information using IHC slides alone. However, with few exceptions, pathologists correlate IHC information with cellular morphology identified in H&E stained slides to generate a diagnosis.

p16 is becoming a widely employed tool to diagnose CIN 2+. Multiple studies have now demonstrated that diffuse strong staining for p16 of at least half the thickness of the epithelium correlates strongly with high-risk HPV infection and CIN 2+ in clinical follow-up. (21, 79-82) Therefore, diffuse, full-thickness epithelium positive p16 staining can be used to distinguish

CIN2+ lesions at high risk of becoming cervical cancer if left untreated. Scoring p16 IHC staining, which is a separate observation event from the H&E only slide, has also been demonstrated to have excellent interobserver reproducibility. (74) It has also been demonstrated that p16 staining is a sensitive marker for high-grade CIN lesions and a specific marker for HPV-infected cervical lesions. (17, 83) However, a major limitation of p16 is that the staining pattern for CIN 2 and CIN 3 lesions are similar in most cases such that p16 "cannot be used independently from morphologic interpretation to distinguish for example, between CIN 2 and CIN 3 lesions." (12) Therefore, it is necessary to consider p16 IHC information in conjunction with H&E slide morphologic information in order to improve diagnostic reproducibility in distinguishing CIN 2 from CIN 3. (12) The positive predictive value of p16 staining among CIN 1 and CIN 2 lesions is currently inadequate and must be used in combination with morphology to determine a diagnosis. (21, 84, 85) In adolescents, p16 has been shown to have similar discriminatory ability. (13) In this context, p16 may be most useful for its negative predictive value, in that a negative p16 IHC stain predicts that the lesion in question will not become CIN2+. (17, 86-89)

Most previous literature on the use of p16 to improve diagnostic reproducibility has compared diagnoses generated using H&E alone to diagnoses generated using p16 alone. (18, 74) In these investigations, as would be predicted, p16 generates the greatest improvement in kappa when the diagnostic categories are limited to $\leq$CIN 1 versus $\geq$CIN2. Three studies have investigated the use of p16 IHC in series with H&E slides to improve diagnostic reproducibility and accuracy of CIN. Sayed et al. (2007) considered the effect of p16 IHC on CIN diagnosis in adolescents. This group demonstrated improved reproducibility with the use of p16. This study did not distinguish between CIN 2 and CIN 3, but they estimated that 9-20% of cases scored CIN2+ with H&E only were downgraded to less than CIN 2 with p16. In addition, 16-43% of cases originally scored as less than CIN 2+ with H&E only were upgraded to CIN 2+ with p16. A limitation of their study was that they were not able to correlate the H&E only or H&E plus p16 diagnosis with any follow-up data to assess if either diagnostic method was more accurate. (13) The most recent study to investigate H&E plus p16 IHC in series demonstrated an improvement in overall kappa from 0.49 with H&E to 0.64 using H&E plus p16 IHC. (12) These data indicate improved precision with p16, but there are data questioning the accuracy of p16-based diagnoses. For example, Guedes et al. (2007) specifically considered the predictive value of CIN 2 diagnoses made using H&E plus p16. They concluded that p16 could be useful in making a diagnosis of CIN 2, but staining was not predictive of the eventual outcome. (84)

One reason why p16 may not accurately predict long term outcome may be its limitation to reliably distinguish CIN 2 from CIN 3. Immunostaining for a proliferation marker related to mitotic activity may provide a solution. This proliferation marker, Ki-67 (alternatively MIB-1), is an antigen expressed in the nuclei of proliferating cells. (22, 78, 90) If the cervical epithelium is divided into rough thirds (lower, middle, upper), CIN 1 lesions generally have Ki-67 staining restricted to the lower third, CIN 2 lesions have staining restricted to the lower two-thirds, and CIN 3 lesions show full thickness staining. (91-93) Alternatively, Ki-67 staining in the upper two-thirds of the cervical epithelium has been shown to be useful in distinguishing between negative mimics of CIN lesions and proliferating CIN lesions. (49, 83, 89) The limitation of Ki-67 staining is that it will stain proliferative metaplasia and hyperplasia in a similar fashion as CIN 2+ lesions.(74, 87) Indeed, cervical cells in smokers will often show increased staining for Ki-67 in the absence of high-risk HPV infection. (94) Ki-67 scoring has been demonstrated to have moderate to high interobserver reproducibility (κ=0.74), but this level of agreement may be difficult without computer digital quantification. (19, 91, 92)

In summary, both p16 and Ki-67 appear to improve CIN diagnostic precision and accuracy. However, both markers have certain limitations, namely that p16 IHC does not contribute to the distinction between CIN 2 and CIN 3, and that KI-67 IHC may require digital quantification to be useful. The combination of p16 and Ki-67 in conjunction with conventional H&E slides would be the preferred diagnostic approach to improve reproducibility and predictive value.

V. Risk factor modeling of CIN

After the establishment of HPV as the primary causal factor in cervical carcinogenesis, epidemiologic focus shifted toward the understanding of additional cofactors that contribute to development of CIN, including environmental, host, and viral factors. (95-97) Such studies have needed to distinguish between factors that affect the risk for acquisition of HPV and factors that affect the risk for persistent HPV infection. Risk factors for the acquisition of HPV include low socioeconomic status, younger age, >6 lifetime number of sexual partners, younger age at coitarche, and oral contraceptive use. (26, 98-101) Risk factors for persistent infection and progression to a high-grade CIN lesion include viral load $\geq$1 pg/mL among women with a monotype infection (OR 2.05, 95% CI 1.65-2.53), low social class (60% increased risk of dysplasia), less than high school education level (OR 2.4, 95% CI 1.5-3.7), history of injectable

contraceptive use (OR 1.5, 95% CI 1.0-2.3), history of >10 years oral contraceptive use (OR 4.03, 95% CI 2.09-8.02), being a former smoker (RR 3.3, 95% CI 1.6-6.7), and being a ≥1 pack per day current smoker (RR 4.3, 95% CI 2.0-9.3). (95, 99, 102-106) Notably, however, for several of these potential risk factors, the association is not present after adjusting for HPV status. (107, 108)

After adjusting for HPV status, there remains considerable inconsistency in the reported risk of CIN 2+ associated with certain exposures. A potential explanation for the observed inconsistency may be variables degrees of misclassification of CIN 2+ lesions due to inadequate diagnostic methods. Early reports of the association between parity and CIN after adjusting for HPV status reported a higher risk for women with history of 2 live births (OR 1.9, 95% CI: 1.1-3.3) and >4 live births (OR 2.8, 95% CI: 1.3-5.9). (107) In a pooled analysis of HPV-positive women, Munoz et al. reported an increased risk of cervical cancer among women with history of 7 or more live births (OR 3.8, 95% CI: 2.7-5.5). (109) However, in a review of the published literature, Castellsague et al. reported an association between ever being pregnant and CIN 3 (OR 4.6, 95% CI: 1.1-20) from a single study in Costa Rica, but there was no association in three other studies from Denmark, Manchester, UK, and Portland, OR, USA. (95) Others have also reported no association between parity and high-grade CIN after adjusting for HPV status. (105, 108) There are even more inconsistencies, for example, Belinson et al. observed a significant association between >3 pregnancies and CIN 3 (OR 1.76 95% CI: 1.36-2.16), but failed to observe a significant association between >3 pregnancies and CIN 2 (OR 1.06, 95% CI: 0.67-1.45). (110)

Similarly, for oral contraceptive use, some studies have reported a dose-dependent increased risk of cervical cancer associated with long term use [OR 2.82 for 5-9 years of use (95% CI: 1.46-5.42), OR 4.03 for >10 years of use (95% CI: 2.09-8.02)]. (95, 104) Castellsague et al. (2003) also concluded use of oral contraceptives was associated with CIN 3 when used for >5 years (OR 3.4, 95% CI: 2.1-5.5). These studies failed to find a consistent dose-response relationship when considering length of use, and other studies failed to find a significant association between oral contraceptive use and high-grade CIN, altogether. (95, 101, 103, 105, 108) However, there is an increased risk of CIN 3 associated with injectable hormonal contraceptive use (OR 1.5, 95% CI: 1.0-2.3) providing some additional evidence that exogenous hormones may yet play a role in development of CIN 3. (99)

Stronger relationships may survive the statistical noise generated by incorrect CIN 2+ diagnoses. The relationship between cigarette smoking and CIN2+ has been demonstrated by

several studies and has been demonstrated to be dose-dependent when controlling for high-risk HPV infection. (94, 95, 101, 103, 111-113) Among high-risk HPV infected women, current smoking is associated with CIN 3 in a dose-dependent manner (OR 1.7, 95% CI: 1.4-2.1). (113) Deacon et al. demonstrated a dose-response relationship between number of cigarettes smoked and CIN 3. (101) Interestingly, Kjellberg et al. demonstrated an association between smoking and CIN 2+ (OR 2.6, 95% CI: 1.7-4.0) and showed that this association remained regardless of method used to detect high-risk HPV. (108)

Importantly, it should be noted that Castle et al. (2002) investigated a cohort of women from the same health-management organization (Kaiser Northwest, Portland, OR) from which the patient sample for this thesis was selected. (103) However, they used surveys, instead of the medical records, to collect smoking information, and 45.6% of women in the study sample reported either being a current or former smoker. In this study, after adjusting for HPV status, the odds of CIN 3 among current $\geq$ 1 pack per day smokers was 4.3 (95% CI: 2.0-9.3). Since it has been reported that smoking is correlated with Ki-67 positivity in cervical squamous cells, it is necessary to be able to control for smoking status when modeling risk factors and using Ki-67 IHC as a component of the diagnostic method. (94) Furthermore, there is evidence that smoking status may also confound relationships between other risk factors and CIN, because of direct effects of cigarette smoking on cervical squamous cells. It is known that cigarette smoke contains carcinogens, and it has also been hypothesized that smokers have an impaired immune response to HPV-infection leading to longer high-risk infections. (95)

Of note, in parallel to the trend observed by Belinson et al. when considering risk of CIN associated with parity, McIntyre-Seltman et al. also observed a significant risk associated with smoking and CIN 3, but failed to observe a significant association between former smoking and CIN 2 (OR 1.5, 95% CI: 1.0-2.1). (110, 113) This suggests that the risk factor profiles for CIN 2 and CIN 3 lesions may be unique and gives further justification to the importance of developing diagnostic modalities that are capable of distinguishing between CIN 2 and CIN 3 lesions in epidemiologic studies of cervical dysplasia.

In summary, there are many factors which may contribute to HPV infection, persistence of high-risk HPV infection and progression to high grade dysplasia and invasive cancer. These factors may be environmental (eg, smoking, oral contraceptive use), host (eg, parity), or viral (eg, HPV genotype) in nature. For each of the risk factors considered above, there are studies to suggest that the risk factor is positively associated with development of high grade CIN, as well

as other studies that question the association, especially when considering CIN 2. This thesis will address for the first time whether a more precise and accurate diagnosis of CIN 2+ using p16 and Ki-67 leads to more valid modeling of cervical dysplasia risk factors and risk estimation. In turn, we anticipate that future epidemiologic studies using the new gold standard will provide better information for clinicians who advise patients about cervical cancer risks.

VI. Misclassification Bias

In epidemiologic studies, sources of bias in design or methods may obscure observation of the true nature of an association between an exposure and an outcome of interest. (3, 114, 115) Therefore, in order to determine the true nature of an association, epidemiologic studies should be conducted with the least amount of bias possible.

There are two general types of bias in epidemiology: selection bias and information bias. Selection bias may be introduced by methods by which patients available for the study are selected to or excluded from the analytic group. Information bias may be introduced by methods or criteria used to classify exposure and outcome status. (3) One important type of information bias is misclassification bias, which can affect the exposure assessment or the outcome assessment or both [eg, a patient is classified as outcome positive (CIN2+) when she is truly outcome negative (CIN 1)]. Pathologists are often unaware of the patient's exposure status when making a diagnosis (eg, outcome classification), which may contribute to non-differential misclassification bias, meaning that the misclassification is independent of the patient's exposure status. The potential for misclassification will be greater if the pathologist is using a diagnostic method that is either imprecise or inaccurate.

The net effect of non-differential misclassification is biasing of the relative risk estimate toward the null. (3, 116) In some cases, the bias is adequate to change the significance of the risk estimate (eg, with bias present the estimate is non-significant but when bias is minimized or absent the estimate is significantly different from the null). (115-117) The mathematical effect of outcome misclassification bias has been considered and statistical techniques have been proposed to adjust for this bias in observational studies, but a preferable alternative would simply be a reduction in misclassification. (15, 118, 119)

The failure of pathologists to reach agreement on a diagnosis, or the failure of an individual pathologist to assign the correct diagnosis represents avoidable opportunities for misclassification. There are few examples of epidemiologic studies of CIN that report diagnostic

interobserver reproducibility. (59) The frequency of reporting interobserver reproducibility statistics in epidemiologic studies has not been published. Due to study design and the use of large registries of patient information, calculation of reproducibility is often not possible.

Assuming large observational studies are subject to a similar degree of poor diagnostic reproducibility as studies using a small sample of CIN specimens, then a substantial source of misclassification bias is likely present in large epidemiologic studies of CIN. In fact, based on parameters that effect calculation of diagnostic reproducibility, one might expect significantly more imprecision in large epidemiologic studies that involve larger numbers of pathologists who are at various levels of expertise and may use slightly different diagnostic criteria.

Only one study has been published to date that specifically considers diagnostic reclassification in CIN lesions as a result of using p16 IHC in combination with conventional H&E slides. (120) The stated purpose of that paper was to determine if additional methods (p16 IHC, ProExC IHC, and deeper levels) could be used to resolve discordance between Pap smear diagnosis (HSIL) and the subsequent colposcopic biopsy diagnosis ($\leq$ CIN2+). With the use of the additional methods, only 2 of 57 discordant cases (3.5%) were resolved. It is important to note that this study investigated diagnostic discordance between two different types of samples – a Pap smear specimen and a colposcopic biopsy specimen. There are well described sources of discordance in this type of study design, such as regression of the lesion and sampling variation – meaning that the colposcopic biopsy did not sample the same lesion that the Pap smear sampled. The authors concluded that the additional methods (p16 IHC, ProExC IHC, and deeper levels) did not significantly reduce the discordance rate and that sampling variation was a legitimate explanation for the persistent discordance. This conclusion may only partially explain the observations and emphasizes the need for a study that further evaluates the amount of diagnostic reclassification achieved with p16 IHC when comparing the same types of specimens. Furthermore, given the subpopulations that may benefit from more conservative therapy of CIN 2, further evaluation of the amount of diagnostic reclassification achieved when IHC markers are used to differentiate between CIN 2 and CIN 3 is needed.

In a unique study design, Castle et al. (2010) recently published an elegant demonstration of the effect of misclassification bias on the strength of association between HPV status and CIN 3. (14) In this study, the exposure of interest was high-risk HPV infection. This group used four different laboratory modalities to measure the same exposure – line blot assay, linear array, hybrid capture II (HC2), and polymerase chain reaction (PCR – by AMPLICOR

analyzer). Patients were classified into exposure groups based on how many modalities by which they tested positive for high risk HPV infection (eg, 0, 1, 2, 3, or 4 modalities), assuming that a woman is more likely to truly have a high risk HPV infection if she tests positive by all four modalities than if she tests positive for only one of four modalities. For histopathologic outcome, two reviewers used conventional H&E stained slides, and cases were classified into 5 tiers of diagnostic certainty of high-grade CIN. A case was classified into the "least certain" tier when one reviewer diagnosed CIN 2 while the other reviewer diagnosed <CIN 2. The next most certain tier represented cases where both reviewers agreed on CIN 2. The "most certain" tier represented cases where both reviewers agreed on a CIN 3 diagnosis. The strength of associations between each number of modalities tested positive for high risk HPV infection and each tier of diagnostic certainty were calculated. By these methods, Castle et al. were able to demonstrate stronger associations between higher number of modalities tested positive for high risk HPV and greater certainty of high grade CIN diagnosis, indicating that when classification of exposure and outcome status is more certain, stronger, and more likely "true", associations will be calculated.

This thesis is a pilot investigation that will determine the amount and kind of diagnostic reclassification that results in CIN with the use of p16 and Ki-67 IHC. This information is valuable because it will highlight histomorphologic and IHC staining patterns that cause diagnostic disagreement and warrant additional investigation. We anticipate that p16 IHC will result in the reclassification of up to 25% of CIN cases, primarily influencing the distinction between Negative/CIN 1 and CIN 2+. Because the p16 staining pattern for CIN 2 and CIN 3 are often identical, we do not expect p16 IHC to significantly affect the distinction between CIN 2 and CIN 3. We anticipate that Ki-67 may result in reclassification of CIN – potentially at any diagnostic distinction point because the Ki-67 staining pattern for each grade of CIN is distinct. In a parallel fashion to Castle et al., the process of evaluating multiple measures of CIN (H&E only, H&E+p16, and H&E+p16 and Ki-67), will allow comparisons of reproducibility, sensitivity, specificity, positive predictive value, and negative predictive value achieved by each method. Comparison of the relative improvements in each of these measures will allow identification of the method with the best "diagnostic profile". We anticipate that bias will be minimized by the addition of molecular methods to conventional H&E slide review because the method with the best diagnostic profile will be expected to produce the least amount of outcome misclassification.

We anticipate that using H&E plus p16 and Ki-67 will be the method with the best diagnostic profile and should be considered the new "gold standard" method for diagnosis of CIN.

Furthermore, since this diagnostic method (H&E plus p16 and Ki-67) has not been used to determine outcome classification in any epidemiologic studies to date, it may be prudent to reconsider the strength of associations between pertinent social and biologic risk factors and CIN. In this thesis, as a secondary aim, we will be able to preliminarily compare the strength of association between commonly accepted risk factors (age, smoking history, education level, family income, etc.) and CIN 2+ as determined by each of the multiple measures of CIN. Finally, restricting the focus of this thesis to only outcome classification will add to the example detailed by Castle et al. and give additional understanding of the effect of outcome misclassification on the strength of association independent of exposure misclassification. Since this is a pilot study, evaluation of the effect of diagnostic reclassification on measures of association must be considered preliminary, but may warrant being repeated in a larger epidemiologic trial.

Specific Aims

The purpose of this thesis is to explore the implications of a new diagnostic modality for CIN on the management of individual patients and the conduct of epidemiologic studies. This thesis will add to the existing body of literature by:

1. Evaluating the effect of a novel combination of two molecular diagnostic markers (p16 and Ki-67) added to conventional diagnostic methods on interobserver reproducibility, sensitivity, specificity, and positive and negative predictive values when doing histologic assessment of CIN lesions,

2. Determining the amount and kind of diagnostic reclassification of CIN lesions that occurs when using conventional diagnostic methods only (H&E only) compared to a combination of conventional and molecular and conventional diagnostic methods: 1) H&E plus p16 and 2) H&E plus p16 and Ki-67,

3. Determining whether using the diagnostic method (either H&E plus p16 or H&E plus p16 and Ki-67) that demonstrates significant improvements in interobserver reproducibility, sensitivity, specificity, and positive and negative predictive values in the assessment of CIN significantly changes the observed strength of association between risk factors and CIN when CIN is diagnosed using molecular and conventional method compared to when CIN is diagnosed using the conventional diagnostic method alone.

Methods

I. Data Management

Data source: Kaiser-Permanent Northwest is a health management organization that serves approximately 100,000 women in the Pacific-Northwest, predominately in Portland, Oregon. Because of the organization of this health system, the majority of patients are believed to maintain long membership periods and to regularly receive standard preventive screenings, including Pap smears. The quality of the data is limited in that information is obtained from coded (ICD-9, CPT, NCD) clinical or pharmacy data. Therefore, non-coded data in a clinical note will not be obtained through chart extraction relying only on coded information. Because of patterns of physician practice and coding, certain variables are considered "weak" in these data because they are not consistently inquired about or coded, such as smoking status and HPV status. This patient population has been used in previous investigations where clinical information was obtained by survey. (100, 103, 107)

Subject selection & Genetic opt-out: The target population of this study is US adolescent and adult females, ages 15-80 years, who receive cervical cancer screening at Kaiser-Permanent Northwest (KP-NW). Using a protocol approved by both the Oregon Health & Science University (OHSU) (IRB# 2500) and KP-NW (IRB# NW-06TMorg-01) institutional review boards, all women with first time abnormal cervical cytology diagnoses with associated colposcopic biopsy at KP-NW between January 1, 1997 and December 31, 2003 were identified. Patients were excluded if there were fewer than two follow-up Pap smears in 5 years of total follow-up time after colposcopic biopsy in the absence of additional tissue specimen. Patients were also excluded if colposcopic biopsy tissue blocks were not locatable (Table 1). Based on limited reports of reclassification of CIN in the literature, it was estimated that 100 cases of each grade of CIN were necessary for 90% power (α=0.01). (13) From this pool of potential cases meeting inclusion criteria, random samples of 221 CIN 1 cases, 110 CIN 2 cases, and 108 CIN 3 cases (n=439) were retrieved from the archives of the Department of Pathology. For cases with more than 1 paraffin block, the diagnostically leading block containing the lesion was selected by one pathologist (RK). Per IRB approved protocol, case list was reviewed by Kaiser-Permanent Northwest to exclude patients who had completed a genetic opt-out form. No patients were excluded based on completion of a genetic-opt out form.

Chart Extraction: The patient list (n=439) was securely transferred to a research analyst as Kaiser-Permanent Center for Health Research (KPCHR). Demographic, clinical, and pathology variables were extracted from the Kaiser-Permanent Northwest medical, pharmacy, and pathology records. Extraction was based on operationalization of variables using appropriate ICD-9, CPT, and NDC codes. Independent variables were selected to characterize the study sample and to reconsider strength of association between pertinent risk factors and CIN. Risk factors were considered pertinent if there was a wide range in reported strengths of association in the literature or if there was a consistent null association in the literature, potentially indicating the effect of misclassification bias. For a complete list of extracted variables see Appendix 1. Demographic variables include age, race, and surrogate markers for family income and patient education level. Predictor variables include exposure to oral contraceptive pills, exposure to topical vaginal hormones, and history of multiple pregnancies. Adjustments for potential covariates will be performed and include: smoking status, history of sexually transmitted infection, and use of systemic steroids. Potential effect modifiers include: smoking status, parity, oral contraceptive use, hormone replacement therapy use, and topical vaginal hormones use. There are eight outcome variables of interest in this thesis. Temporally, the first outcome variable was obtained from KPNW Pathology archives and is the colposcopic biopsy diagnosis rendered by a pathologist at KPNW at the time of the biopsy that directed care for that woman. Six additional outcome variables include the diagnosis of grade of CIN by reviewer A and reviewer B using conventional methods (H&E only), diagnosis of grade CIN by reviewer A and reviewer B using molecular and conventional methods (H&E plus p16 and H&E plus p16 and Ki-67), and pattern of p16 and Ki-67 staining. The eight outcome variable (so-called "consensus diagnosis") will be the highest grade of CIN diagnosed by reviewer A or reviewer B during the 5 year follow-up period <u>after</u> the index colposcopic biopsy and will be used as the gold standard or "true measure of disease" for calculations of sensitivity, specificity, positive predictive value, and negative predictive value. This diagnosis could be made on tissue from a subsequent colposcopic biopsy or LEEP specimen depending on the course of management selected by each patient. If no additional tissue was obtained, the highest grade of CIN will be considered "negative" providing the patient had at least 2 negative Pap smears within the 5 year follow-up period (as per inclusion criteria). The majority of follow-up specimens were reviewed by both reviewer A and reviewer B and consensus diagnostic opinion was reached.

Data Storage: Chart extraction data, identified only by study identification (Morgan) number, were provided by KPCHR research analysts and imported into a web-based electronic data capture database (RedCAP). Each slide (H&E, p16, and Ki-67) prepared was labeled with the corresponding Morgan ID number. Slide review data was entered by reviewing pathologists at time of slide review into RedCAP under the appropriate Morgan ID number. Chart extraction data and slide review data were merged in RedCAP, producing a single export file (.xls, .dta) amenable for statistical analysis. [https://octri.ohsu.edu/redcap/]

II. Slide Preparation and Case Review

Conventional hematoxylin & eosin (H&E) stained slides: H&E slides and two unstained serial sections were prepared by standard methods from paraffin-embedded colposcopic biopsy specimens and follow-up specimens if available (eg. LEEP, conization, or additional colposcopic biopsy). The slides were numbered according to order in the cutting process. The first slide was stained with H&E by standard methods, the second slide was stained for $p16^{INK4a}$, and the third was stained for Ki-67. **Immunohistochemistry:** Tissue sections were deparaffinized and rehydrated through graded alcohols. Antigen retrieval was performed for 10 minutes in a high pressure cooker in 1x citrate buffer (pH 6.0); after which slides were allowed to cool for 20 minutes at room temperature and then washed with deionzed water. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide, rinsed, and then protein blocked (DAKO) before application of primary antibody. The $p16^{INK4a}$ primary antibody was obtained from Mtm Laboratories (Heidelberg, Germany). The Ki-67 primary antibody was obtained from DAKO Corporation (Carpenteria, CA). Biopsies were incubated with primary antibody for 30 minutes at room temperature, and then washed for 10 minutes in phosphate buffered saline (PBS) at room temperature. After washing, biopsies were incubated with secondary antibody (Dako EnVision+Dual Link System-HRP) for 30 minutes at room temperature, and then washed for 10 minutes in PBS at room temperature. Diaminobenzidine was used as a chromagen, and slides were counterstained with hematoxylin. The negative control was prepared identically, but using mouse IgG primary antibody. The positive control was prepared identically and was a CIN 3 biopsy obtained and confirmed at OHSU.

Case review: Routine hematoxylin & eosin (H&E) stained histologic sections were reviewed by reviewer A and reviewer B while blinded to the original anatomic pathologist's diagnosis (KP). Both pathologists classified each case (1 H&E slide) into one of four possible categories using conventional histomorphologic diagnostic criteria: negative, CIN 1, CIN 2, or CIN 3 (Figure 2).(4) After a wash-out period of at least 1 week, p16 immunohistochemical (IHC) slides were reviewed and scored by both pathologists according to a four tiered system: i) negative, ii) patchy basal, iii) diffuse basal, iv) diffuse full-thickness staining (Figure 3). Then the H&E slide and p16 IHC slide were considered together, and the case was classified into one of four possible diagnostic categories: negative, CIN 1, CIN 2, or CIN 3, based on the composite impression. After a second wash-out period of at least 3 weeks, Ki-67 IHC slides were reviewed and proliferation index was scored by estimating the highest zone of positive staining by a majority of nuclei, as described elsewhere.(19) Proliferation index scores were categorized according to a three-tiered system of nuclei staining: i) predominately lower third, ii) lower two-thirds, but involving the middle third, or iii) full thickness staining involving the upper third of the cervical squamous mucosa (Figure 4). A third diagnosis was generated using the combination of H&E impression, p16 score, and Ki-67 proliferation index. Each case was classified into one of four possible categories: negative, CIN 1, CIN 2, or CIN 3, based on composite impression. All slide review data were recorded in RedCAP (https://octri.ohsu.edu/redcap). Diagnostic slides from follow-up cervical surgeries (eg, LEEP, conization), and subsequent colposcopic biopsies were stained, coded, and reviewed in an identical manner to the primary colposcopic biopsy while blinded to patient diagnoses and outcomes. The "consensus diagnosis" was considered the highest grade of CIN diagnosed by reviewer A or reviewer B during the 5 year follow-up period. If a patient did not have a second tissue specimen collected, but had at least two negative Pap smears within 5 years of the index colposcopic biopsy, her final outcome ("consensus diagnosis") was considered negative for CIN. For calculations of sensitivity, specificity, and positive and negative predictive values, H&E only, H&E plus p16, and H&E plus p16 and Ki-67 diagnoses from reviewer A and reviewer B were used as the "test" diagnoses, and the "consensus diagnosis" was used as the "true" clinical outcome.

III. Statistical analysis

Reproducibility & Test Validation: From the sample of 439 cases, patients were excluded if any of the data necessary for outcome classification was missing (n=81 [52 CIN 1 cases, 15 CIN 2

cases, and 14 CIN 3 cases]). Cases were also excluded if reviewer A or reviewer B diagnosed a biopsy as negative for CIN lesion (n=57). At this point, there remained 121 CIN 1 cases, 84 CIN 2 cases, and 91 CIN 3 cases. Random samples of 84 CIN 1 and 84 CIN 3 cases were selected to produce a complete dataset (n=252 [n=84 for each grade of CIN]) to assess diagnostic reproducibility (Figure 5). The data were reviewed for additional missing or implausible values and none were identified. Reproducibility was measured by pairwise kappa statistic with >0.8 considered excellent; 0.6-0.8 substantial; 0.4-0.6 moderate; 0.2-0.4 poor; and <0.2 slight.(16) Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for molecular method of diagnosis using LEEP diagnosis or negative 5 year follow-up as outcomes. Data was analyzed using Stata 10.0 and 11.0 (StataCorp, College Station, TX).

Using the same subset of 252 cases, sensitivity, specificity, positive predictive value, and negative predictive values were calculated using H&E only, H&E plus p16, and H&E plus p16 and Ki-67 diagnoses from reviewer A and reviewer B as the "test" diagnoses and the "consensus diagnosis" as the "true" diagnosis.

Risk estimation: From the sample of 439 cases, patients were excluded if any of the data necessary for outcome classification was missing (n=81 [52 CIN 1 cases, 15 CIN 2 cases, and 14 CIN 3 cases]). This primarily included cases in which the tissue specimen was considered inadequate for diagnosis. The remaining 358 cases (169 CIN 1 cases, 95 CIN 2 cases, and 94 CIN 3 cases) were considered the complete dataset for multivariate analysis (Figure 5). The data were reviewed for additional missing or implausible values and none were identified. Appropriate descriptive statistics including frequency, range, mean, and standard deviation were obtained and used to evaluate the distribution of independent variables by reviewer A H&E only diagnosis (Table 2). The distribution of independent variables by reviewer A H&E plus p16, reviewer B H&E only, and reviewer B H&E plus p16 diagnoses were also considered. A subset of independent variables was selected based on biologic significance and previously published positive association with CIN. In order to construct 2x2 contingency tables, three variables were recategorized: age (<30 years old/$\geq$ 30 years old), gravidity (0 pregnancies/$\geq$ 1 pregnancy), and parity (0 deliveries/$\geq$1 delivery). All eight outcome classifications, including the "consensus diagnosis", were dichotomized as $\leq$CIN 1 or $\geq$CIN 2. Odds ratios were estimated from the resulting contingency tables.

Model building: Odds ratio estimates for selected predictor variables were calculated from 2x2 contingency tables. Outcome classification was dichotomized as CIN 1 and CIN 2+ for each of the three methods of diagnosis (H&E only, H&E plus p16, and H&E plus p16 and Ki-67) for both reviewer A and reviewer B. The same subset of selected predictor variables was then considered individually in simple logistic regression using the dichotomized reviewer A H&E only diagnosis to determine outcome. The associated between continuous independent variables and the dependent variable was considered in univariate analysis using ANOVA. Due to biological significance, *age* was recategorized as 2 dichotomous variables: age less than or greater than 21 and age less than or greater than 30. Due to small cell size and zero-cells, *gravidity* and *parity* were recategorized as binary variables (never pregnant/delivered or ever pregnant/delivered). If univariate p-value <0.25, the variable was included in the preliminary main effects model. If the variable maintained significance (Wald statistic <0.05) in the preliminary model, it was retained in the final model. All possible interaction terms and combinations of interaction terms were considered in the model and retained if $p<0.05$. This model was compared to the model generated using forward and backward stepwise procedures (pe=0.20 and pr=0.20). Once the final model was established using reviewer A's H&E only diagnosis to determine outcome classification, this same model was also run using reviewer A's H&E plus p16 diagnosis and reviewer A's H&E plus p16 and Ki-67 diagnosis to determine outcome classification. Additionally, the original Kaiser H&E only diagnosis, reviewer B's H&E only diagnosis, reviewer B's H&E plus p16 diagnosis, reviewer B's H&E plus p16 and Ki-67 diagnosis, and the "consensus diagnosis" were used to determine outcome classification. In total, 8 diagnostic methods were used to generate 8 different outcome classifications and 8 different logistic regression models that varied only by diagnostic method.

Results

Descriptive statistics:

Descriptive statistics are presented in Table 2. Cases not selected to the analytic group (158 random cases were not included in reproducibility and test accuracy statistics and 49 random cases were not included in regression analysis) were found to have no statistically significant differences in distributions of predictor variables compared to the cases used for analysis. Reviewer A and reviewer B produced non-significantly different distributions of predictor variables for all grades of CIN when using the same diagnostic method (data not shown). There were 62 (17%) adolescents (≤20 years old) in the analytic group used for regression analysis. The size of this group was considered inadequate for subgroup analysis.

Percentages of persistent dysplasia and percentages of dysplasia progressed to a higher grade or regressed to a lower grade were calculated for each grade of CIN for each reviewer using H&E only diagnosis compared to the "consensus diagnosis" (highest grade of dysplasia on subsequent tissue specimen or 5 years of clinical negative follow-up) (Table 3). Depending on the reviewer (A or B), CIN 1 progressed to a higher grade of dysplasia in 11-25% of cases, regressed to negative in <1% of cases, and was persistent in 74-89% of cases. CIN 2 progressed to a higher grade of dysplasia in 0-20% of cases, regressed to a lower grade of dysplasia in 5-15% of cases, and was persistent in 64-95% of cases. CIN 3 regressed to a lower grade of dysplasia in 0-42% of cases and was persistent in 58-100% of cases. There were no cases of invasive cervical carcinoma identified in this cohort. Therefore, no cases of CIN progressing to invasive carcinoma were identified.

Kappa statistic calculations (test reproducibility):

After excluding cases with missing values, a subset of the original cohort was used for kappa statistic calculations (n=252, 84 cases/grade CIN). Pairwise kappa statistics indicate that reviewers A and B have better agreement for each grade of CIN than either reviewer A versus Kaiser or reviewer B versus Kaiser (Table 4.1). Reproducibility between reviewers A and B was substantial for CIN 1 (κ=0.6406) and CIN3 (κ=0.6756) and moderate for CIN 2 (κ=0.4041). Reproducibility in the diagnosis of CIN 2+ was also substantial (κ=0.6677). Reproducibility in assessment of p16 stained slides was excellent (κ=0.8552) when negative staining pattern was grouped with patchy basal staining pattern and was substantial (κ=0.7255) for diffuse, full-thickness staining pattern. However, reproducibility was extremely poor for the pattern of

diffuse staining limited to just the basal layer (κ=0.1305) (Table 4.2). Due to lack of reproducibility in assessment of the basal p16 staining pattern, for subsequent analyses of p16 IHC staining patterns, cases having basal p16 staining were grouped with cases having negative/patchy staining. The diagnosis generated using composite H&E plus p16 IHC slides was significantly more reproducible (κ=0.4783) for only CIN 2 compared to diagnosis using H&E alone ($p<0.05$). Reproducibility in assessment of Ki-67 stained slides for every grade of CIN was moderate (κ=0.5481, 0.5305, and 0.5849 for CIN 1, 2 and 3, respectively) (Table 4.2). The diagnosis generated using composite H&E plus p16 and Ki-67 IHC slides was more reproducible for CIN 1 (κ=0.7453) and CIN 2 (κ=0.5204) compared to diagnosis using H&E alone ($p<0.05$). Diagnosis using H&E plus p16 and Ki-67 was, however, not significantly more reproducible than diagnosis using H&E plus p16 only.

Test accuracy:

Overall sensitivity and negative predictive value were significantly improved for both reviewers over H&E only when using p16 IHC alone, H&E plus p16, and H&E plus p16 and Ki-67. For reviewer A, p16 IHC alone, H&E plus p16, or H&E plus p16 and Ki-67 showed significant improvements in overall specificity and positive predictive value compared to H&E only (Table 5). CIN grade-specific sensitivity, specificity, positive predictive values, and negative predictive results are reported in Tables 6 and 7. The diagnostic method with the best test accuracy profile was reviewer A using H&E plus p16 IHC, which achieved overall sensitivity of 94.19%, specificity of 100%, positive predictive value of 100%, and negative predictive value of 88.89%. All of these measures were significantly improved compared to that achieved by reviewer A using H&E only. The diagnosis generated by reviewer A using H&E plus p16 and Ki-67 also showed significant improvements in measures of test accuracy compared to the diagnosis made using H&E only, but these improvements were not significantly better than that achieved with H&E plus p16. For reviewer B, the H&E plus p16 diagnostic method also appeared to have a better test profile compared to his own H&E only and H&E plus p16 and Ki-67 diagnoses. Reviewer B using H&E plus p16 IHC achieved overall sensitivity of 97.67% and negative predictive value of 93.22%, which were significantly improved compared to that achieved by reviewer B using H&E only. There were no demonstrated improvements in specificity or positive predictive value for reviewer B using H&E plus p16 compared to H&E only.

Reclassification structure:

As a result of using conventional H&E slides plus immunohistochemical staining pattern(s), the final diagnosis for each case could have remained the same as the original diagnosis rendered by Kaiser (using H&E slides alone), or the final diagnosis could have been reclassified as either a lower or higher grade CIN lesion. In this study, compared to the original H&E only diagnosis at Kaiser, 39% of CIN cases were reclassified by reviewer A using H&E only, 30% of cases were reclassified using H&E plus p16, and 27% of cases were reclassified using H&E plus p16 and Ki-67. For reviewer A, 26 CIN 1 cases (as determined by H&E only) were reclassified as CIN 2 when using H&E plus p16 due to full-thickness p16 staining. While 24 of these reclassified cases had middle third Ki-67 staining, 2 cases had lower-third Ki-67 staining and remained classified as CIN 2. For reviewer A, 10 CIN 2 cases (as determined by H&E only) were reclassified as CIN 3 on the basis of full-thickness p16 staining and upper third Ki-67 staining. Ten CIN 2 cases (as determined by H&E only) were reclassified as CIN 1 on the basis of negative/patchy or basal p16 staining and lower third Ki-67 staining. Four CIN 2 cases (as determined by H&E only) were reclassified as CIN 1 on the basis of basal p16 staining, but then reclassified again as CIN 2 on the basis of middle (1 case) or upper (3 cases) third Ki-67 staining. Additionally, 5 cases of CIN 1 with negative/patchy or basal p16 staining were reclassified as CIN 2 on the basis of middle (4 cases) and upper (1 case) third Ki-67 staining (Table 8, Figure 6).

Risk estimation:

For family income represented as a binary variable (<$45K/year, ≥$45K/year), the odds of CIN 2+ if the patient's family made ≥$45,000/year were 0.892 the odds of ≤CIN 1 as determined by conventional methods (H&E only) for reviewer A was not significantly different from 1. However, when calculated using H&E+p16 as the diagnostic method, the odds of CIN 2+ if the patient's family made ≥$45,000/year were 0.585 the odds of ≤CIN 1, which was significantly different from 1. And finally, when the same odds ratio was calculated using H&E+p16 and Ki-67 as the diagnostic method, the odds of CIN 2+ if the patient's family made ≥$45,000/year were 0.589 the odds of ≤CIN 1, which was significantly different from 1.0, although not significantly different form 0.585.

For education level represented as a binary variable (≤ HS graduate, >HS graduate), the odds of CIN 2+ if the patient's education level was > HS graduate were 0.571 the odds of ≤CIN 1 as determined by conventional methods (H&E only) for reviewer B, which was not significantly different from 1. However, for this reviewer, when using H&E plus p16, the odds of CIN 2+ if the

patient's education level was > HS graduate were 0.557 the odds of ≤CIN 1, which was significantly different from 1. And finally, when the same odds ratio was calculated using H&E plus p16 and Ki-67 as the diagnostic method, the odds of CIN 2+ if the patient's education level was > HS graduate were 0.600 the odds of ≤CIN 1, which was not significantly different from 1.0 or 0.557.

Exposure to vaginal hormones was significantly protective as determined by all 8 available outcome assessments. The remaining nine independent variables for which odds ratios were estimated showed neither meaningful trends nor values significantly different from 1 (Table 9), although this pilot study was limited by sample size.

Model building:

When using reviewer A H&E only diagnosis, only one independent variable met criteria for inclusion in the preliminary main effects model: *history of exposure to topical vaginal hormones* (Pearson χ^2_3 10.823, p=0.001). *Age* dichotomized at 30 years (Pearson χ^2_3 0.2844, p=0.594), *race* (Pearson χ^2_3 0.789, p=0.0714), *family income* (Pearson χ^2_3 0.2617, p=0.609), *education level* (Pearson χ^2_3 2.5344, p=0.282), *history of sexually transmitted infection* (Pearson χ^2_3 0.1197, p=0.729), *history of oral contraceptive use* (Pearson χ^2_3 0.5115, p=0.475), *parity* (Pearson χ^2_3 0.3791, p=0.538), and *smoking status* (Pearson χ^2_3 0.0843, p=0.772) did not meet criteria for consideration for inclusion.

Since a significant change was appreciated in the odds ratio estimate for *family income* in 2x2 contingency table calculations when using reviewer A's H&E plus p16 diagnosis for outcome classification, *age*, *education level*, *family income*, and *history of exposure to topical vaginal hormones* were included in the preliminary main effects model (LR χ^2_4 15.66, p=0.0035). By Wald statistic, only *history of exposure to topical vaginal hormones* would have been retained in the final model (p=0.002). Variable selection was confirmed using forward and backward stepwise procedures.

From this model, using reviewer A's H&E only diagnosis, the adjusted odds ratio for high family income (≥$45K/year) was not significant (OR 0.9987, p=0.996). When this model was run using reviewer A's H&E plus p16 diagnosis, in parallel to calculations from 2x2 contingency tables, the adjusted odds ratio for high family income (≥$45K/year) changed substantially but slightly missed reaching statistical significance (OR 0.6363, p=0.057). This change persisted when the model was run using reviewer A's H&E plus p16 and Ki-67 diagnosis (OR 0.6298, p=0.054). When the model was run using the original Kaiser H&E only diagnosis, as was seen in 2x2

contingency table calculations, the adjusted odds ratio for family income (≥$45K/year) was significant (OR 0.609, p=0.036). For reviewer B, although the odds ratio for family income (≥$45K/year) was significant using each diagnostic method in 2x2 contingency tables, the adjusted odds ratios in the logistic regression model failed to reach significance using H&E only (OR 0.681, p=0.109), H&E plus p16 (OR 0.686, p=0.131), and H&E plus p16 and Ki-67 (OR 0.689, p=0.137). By the "consensus diagnosis", the adjusted odds ratio for family income (≥$45K/year) was of similar magnitude but fell short of statistical significance (OR 0.6298, p=0.056) (Table 10).

Discussion

The field of pathology is situated at the leading edge of diagnostic method development and determination of criteria for histologic "gold standard" diagnoses. Pathologists are frequently involved in epidemiologic studies as contributors of histologic diagnoses that are used to determine outcome classification. They may also be involved in providing laboratory data (eg, HPV testing) used to determine exposure classification in epidemiologic studies. However, the process of evaluating new diagnostic methods and the subsequent effect on the strength of epidemiologic associations due to changing classification methods, especially in CIN, has not been thoroughly investigated. Without a precise and accurate gold standard diagnostic method to distinguish each grade of CIN, women with CIN may receive inappropriate treatment, and epidemiologic studies of risk factors for CIN will be biased preventing evaluation of the true strength of associations.

The purpose of this thesis was to provide an example of a method to improve diagnostic precision and accuracy in CIN. Our objective was to test whether molecular markers of cervical neoplastic transformation (p16 IHC) and cervical cellular proliferation (Ki-67 IHC) produced improvements in reproducibility, sensitivity, specificity, positive predictive value, and negative predictive value compared to the currently accepted diagnostic gold standard of conventional H&E slides alone. Our secondary objective was to illustrate how changing to a new "gold standard" diagnostic method would have appreciable and statistically significant effects on patient management and risk estimation.

Kappa statistic calculations (test reproducibility):

The results of this study indicate that the use of p16 IHC, in conjunction with H&E stained slides, leads to significant improvement in interobserver reproducibility of CIN 2, even between two experienced gynecologic pathologists. We also observed that certain p16 staining patterns are highly reproducible, consistent with the literature. (12, 17, 18, 74) The biologic significance of the basal p16 staining pattern is not well understood, and in this study, classification of this pattern demonstrated poor interobserver reproducibility, perhaps, in part, because differentiating this pattern from the negative/patchy p16 staining pattern is not routinely done in clinical practice. Fortunately, the basal p16 pattern was rare, representing 8.7% of the cases examined.

The use of Ki-67 IHC with H&E significantly improved reproducibility of CIN 1 and CIN 2 above the reproducibility of H&E alone, but it was not significantly different from the reproducibility of either grade of CIN obtained with H&E+p16. The use of p16 and Ki-67 in conjunction with H&E also did not significantly increase reproducibility in the diagnosis of CIN 2+. Furthermore, the reproducibility of distinguishing Ki-67 staining patterns was poor for all three patterns. Therefore, based on changes in reproducibility, the extra cost associated with obtaining a Ki-67 IHC slide does not seem warranted.

Test accuracy:

Similar trends were observed when assessing the sensitivity, specificity, and predictive values of H&E plus IHC. For reviewer A, for all grades of CIN, the use of H&E plus p16 produced significant improvements in sensitivity and specificity over H&E only. There were also significant improvements in both positive and negative predictive values (Tables 5 and 6). For reviewer A, the addition of Ki-67 produced significant improvements in each of these parameters compared to H&E only; however, the improvements were not significantly different from those obtained with H&E plus p16 alone. As with the changes observed in reproducibility above, the extra cost associated with obtaining a Ki-67 IHC slide does not seem warranted. Similar trends for significant improvements in sensitivity and negative predictive value with H&E plus p16 and Ki-67 compared to H&E only but not compared to H&E plus p16 were observed for reviewer B are reported in Tables 5 and 7.

Reclassification with the new “gold standard” diagnostic method:

The use of H&E plus p16 IHC was the most reproducible and cost-effective diagnostic method. With this method, reviewer A achieved improvements in all measures of test accuracy compared to H&E only. Because reviewer B achieved significant improvements only in sensitivity and negative predictive value, reviewer A’s H&E plus p16 diagnosis was designated the diagnostic method with the best test profile in this thesis and was used as the new “gold standard” method to determine the nature of diagnostic reclassification.

The degree of reclassification observed in this study is impressive. Up to 30% of cases were reclassified using reviewer A’s H&E plus p16 diagnosis. Additional reasons for diagnostic reclassification included the basal p16 staining pattern and discrepancies between the results of p16 IHC and Ki-67 IHC. Nine of 85 cases reclassified by reviewer A had basal p16 staining (41% of

all basal p16 staining cases in the sample). The biologic significance of this staining pattern is not known, however it is recognized that in clinical use of p16 IHC this pattern appears to contribute to diagnostic discordance. Further investigations are necessary to characterize the biologic and clinical behavior of CIN lesions with basal p16 staining.

Of the reclassified CIN cases, 21 CIN 3 lesions by morphology and with full-thickness p16 staining were reclassifed as CIN 2 due to the absence of upper third epithelial staining with Ki-67. This demonstrates the ideal use of this IHC panel. Because p16 IHC is inadequate for distinguishing between CIN 2 and CIN 3 lesions, Ki-67 identifies CIN 3 lesions with less cellular proliferation that should instead be categorized as CIN 2. But given the lack of significant improvements in reproducibility and test accuracy, this reclassification due to Ki-67 IHC may not be clinically significant. Additionally, with lower grade CIN, even if middle third Ki-67 staining was present, a negative/patchy p16 pattern often directed the diagnosis. We expected this among low grade CIN cases since prior to the study both reviewers were aware that p16 has been demonstrated to be specific with high negative predictive value. As such, in CIN 1 or CIN 2 cases with negative/patchy p16 staining, despite middle-third Ki-67 staining, both reviewers frequently made the diagnosis of CIN 1. However, reliance on p16 IHC was not perfectly used, since four CIN 1 cases that had negative/patchy, or basal p16 staining, were classified as CIN 2 because of middle-third Ki-67 staining. We interpret these outliers as evidence that patchy or basal staining yield discrepant results, especially when there is increased Ki-67 staining.

Implications of reclassification:

With the use of the new “gold standard” diagnostic method (reviewer A H&E plus p16), the diagnosis of CIN 2 as determined by H&E only at Kaiser had a 15% false positive rate. The diagnosis of CIN 1 as determined by Kaiser had 25% false negative rate. This suggests that 15% of women, who were diagnosed with CIN 2 by H&E only at Kaiser, in fact had CIN 1 by reviewer A H&E plus p16, and got unnecessary therapy. For CIN 1, 25% of women who were diagnosed with CIN 1 by H&E only at Kaiser, in fact had CIN 2+ as determined by reviewer A H&E plus p16 and did not get the appropriate therapy at the time of the original biopsy. The diagnosis of CIN 3 as determined by H&E only at Kaiser had 5% false positive rate, meaning that 5% of the women who were diagnosed with CIN 3, in fact were negative or CIN 1 by reviewer A H&E plus p16 and received unnecessary therapy at the time of the original biopsy.

Considering the trends for reviewer A: Reviewer A's H&E only diagnosis had a 25% false negative rate for CIN 1 and a 17% false positive rate for CIN 2 when compared to classification using H&E plus p16. There was no reclassification observed between CIN 2 and CIN 3 for reviewer A using H&E plus p16, and there were no significant changes in test profile when using H&E plus p16 and Ki-67. This suggests that the reclassification that occurred with the addition of Ki-67 (14% of cases) was not clinically relevant, and furthermore, when considering 5-year follow-up the addition of Ki-67 IHC to make CIN diagnoses was no more predictive of clinical outcome than H&E plus p16. Therefore, in this thesis, Ki-67 does not appear to provide additional valuable clinical information that should be used to guide patient management decisions.

Of the three subgroups of patients for whom the distinction between CIN 2 and CIN 3 is most important: adult women with low risk factor profile, women with excision margins positive for CIN2+, and adolescents, the implications for the results of this thesis for the first two groups is that H&E plus p16 is the best and most cost-effective diagnostic method and should be considered the new "gold standard" for diagnosis. The implications for adolescents are less clear since this thesis was unable to do a subgroup analysis because of the small number of adolescents in the sample. Given the differences in HPV natural history and CIN biology documented in adolescents, especially differences in smoking habits, Ki-67 may still play a useful role in diagnosis in this group. Evaluation of p16 and Ki-67 IHC in addition to H&E slides should be reconsidered in a larger population of adolescents.

Risk estimation:

Although underpowered for evaluation of any risk factors of CIN in logistic regression, this thesis has demonstrated that significant changes in strength of associations are possible with changes in diagnostic method. The binary variable family income demonstrates this trend well. (Tables 9 & 10) For this variable, the OR as calculated using reviewer A H&E only was not significantly different from the null value (OR = 0.892). However, when using the new "gold standard" diagnostic method (reviewer A H&E plus p16), the OR was determined to be significantly different from the null value (OR = 0.585). Although the OR calculated with H&E plus p16 was not significantly different from the OR calculated with H&E only, this trend demonstrated that without exploring the different diagnostic methods on calculation of strength of association, the conclusions about the risk of low family income from a study using

H&E slides only as the diagnostic method would have been significantly different from the conclusions about the same variable in a study using H&E plus p16 as the diagnostic method.

Furthermore, we observed that although the addition of Ki-67 resulted in improvements compared to H&E only, it did not result in significant improvements in reproducibility or accuracy over H&E plus p16 for both reviewers. This trend was preserved in that no significant change in the strength of association was calculated using H&E plus p16 and Ki-67; meaning that the odds ratio was still significantly different from the null, but not significantly different from the odds ratio calculated using H&E plus p16.

None of the predictor variables, except history of exposure to topical vaginal hormones, met criteria for inclusion in the preliminary main effects model, which was based on reviewer A's H&E only diagnosis. When the method of outcome classification was changed, the strength of the unadjusted association between family income and CIN changed significantly. After adjusting for key demographic variables and history of exposure to topical vaginal hormones, the strength of association between family income and CIN changed, but failed to reach statistical significance. Since many of the risk factors considered in this study have demonstrated significant association with CIN lesions in previous publications, this is likely due to inadequate sample size and misclassification, and this part of the thesis should be repeated in a much larger cohort adequately powered to consider these risk factors.

It is also important to note that improvements in interobserver reproducibility wouldn't necessarily be expected to manifest as changes in odds ratio estimates as determined by a single reviewer using multiple methods to diagnose CIN. Instead improvements in interobserver reproducibility will manifest as convergence on a similar point estimate by all reviewers in the study. As observed in this thesis, odds ratios for family income calculated using reviewer A and reviewer B H&E only diagnoses are very divergent (0.998 versus 0.681, respectively). However, when both reviewers used H&E plus p16 to make CIN diagnoses, the odds ratios for family income became more convergent (0.636 versus 0.686, respectively). Furthermore, the odds ratio calculated using the "consensus diagnosis" – which is the best measure of "true" disease outcome in this thesis was 0.6298, suggesting that both reviewers may be converging on a more accurate estimate of true risk associated with low family income in this sample.

Implications for understanding of risk factors:

Although underpowered, observing statistically significant changes in the strength of association between risk factors and histologic outcome due to alteration in diagnostic method is meaningful. This observation suggests that using only H&E slides to generate CIN diagnoses in epidemiologic studies contributes to outcome misclassification and is a source of significant bias in the study design. The implications of this observation for current understanding of risk for cervical precancerous lesions as well as conduct of future epidemiologic studies of CIN are substantial.

First, every major epidemiologic study of risk factors for CIN has used only H&E slides and various types of "consensus opinion" to determine histologic diagnosis. No study to date has considered the strength of association between risk factors and CIN when CIN is diagnosed using H&E slides and molecular methods. This thesis demonstrates that p16 IHC provides important information for individual patient management as well as classification of outcome in epidemiologic studies. Therefore, it would be prudent to reconsider many of the potential risk factors for CIN, especially those previously found to have no association with CIN, in epidemiologic studies that use H&E plus p16 to determine outcome classification. Reevaluation of these associations would yield less biased estimates of the true strength of association between risk factors and CIN. In some cases, previously null associations might be found to, in fact, have significant associations when a less biased outcome classification method is used.

In regards to conduct of epidemiologic studies of CIN, practically speaking, it may be logistically difficult and prohibitively expensive to locate and produce p16 IHC slides for the large number of patients in predominately retrospective cohorts used in epidemiologic studies of CIN. Therefore, new methods of data collection and data sharing will need to be developed. The web-based data collection and data storage system used in this thesis (https://octri.ohsu.edu/redcap) is an example of how these new methods may take shape. Moving forward, it would be ideal to prospectively collect digital images of H&E and p16 stained slides, which could then be collated with clinical information, and shared electronically with reviewers at multiple institutions. This would easily allow blinding of reviewers as well as circulation of slides multiple times, as has been done in previous investigations, and would also allow for quantification of intraobserver reproducibility. (6)

Limitations:

There are several limitations affecting this study. First, it is a pilot study designed to gather preliminary information, which will be used to justify and design an adequately powered protocol. This thesis was initially adequately powered only for calculations of reproducibility. However, due primarily to insufficient tissue material in the archived blocks and non-diagnostic tissue sections it was necessary to exclude cases that were missing diagnostic information (eg, if material on the p16 IHC slide was considered non-diagnostic then an H&E plus p16 diagnosis and an H&E plus p16 and Ki-67 diagnosis could not be made). This resulted in an analytic group of 252 cases (n=84 of each CIN grade). Although this was less than the 100 cases of each CIN grade determined necessary by sample size calculations, we still observed significant changes in kappa statistics compared to H&E only diagnosis with both H&E plus p16 and H&E plus p16 and Ki-67. Now that the reclassification achieved with these markers has been preliminarily quantified, it will be possible to refine sample size calculations for future studies.

A second methodological limitation is that the consensus diagnosis is not a true histologic consensus diagnosis. The highest grade of CIN diagnosed during the 5 year follow-up period on either excision procedure, subsequent colposcopy, or negative Pap smears was used as the "consensus diagnosis". Due to known limitations in diagnostic reproducibility and the desire to use this diagnosis as the best measure of "true disease", the intention was for all cases to be reviewed by both reviewer A and reviewer B and consensus opinion to be reached. To this end, all CIN 2 cases and CIN 1 cases that were not p16 negative were reviewed by both reviewer A and reviewer B and consensus opinion was reached. However, for the remaining CIN 1 cases and all CIN 3 cases, the "consensus diagnosis" is that determined by reviewer A only. It is difficult to estimate the effect of this on calculations of test accuracy. However, if we assume that reviewer A is more likely to agree with or be predictive of himself than with reviewer B, then estimates of sensitivity, specificity, positive predictive value, and negative predictive value for reviewer A are likely overestimated while those measures for reviewer B are likely underestimated, compared to what would have been calculated if a true "consensus" diagnosis had been used.

A third methodological limitation is that most of the prior research done by a single laboratory on Ki-67 with and without additional methods (eg, HPV testing) accomplished reproducible diagnoses with the assistance of digital imaging.(75, 91, 92) Digital imaging software was not used in this study. We might expect the reproducibility and accuracy of

diagnoses using Ki-67 to improve with the use of digital technology. However, this would be at potentially great cost to an institution which might not be warranted for the return of information provided by the stain.

The major limitation of the secondary aim of this study, analysis of risk factors for CIN, is sample size. As has been acknowledged, this pilot study was not powered to adequately consider risk factors in logistic regression. Despite this limitation, significant changes in strength of association were observed for single risk factors and emphasize the importance of repeating this study with a larger sample.

Again, it should be noted that Castle et al. (2002) investigated a cohort of women from the same health-management organization from which the patient sample for this thesis was selected. (103) Surveys, instead of the medical record, were used to collect smoking information, and 45.6% of the study sample reported being either current or former smokers. In the study by Castle et al., after adjusting for HPV status, the odds of CIN 3 among current $\geq$ 1 pack per day smokers was 4.3 (95% CI: 2.0-9.3) time the odds of CIN 3 among non-smokers. In this thesis, a comparable percentage of current and former smokers were identified (n=151, 42.4%); however, current or past smoking status was unknown for a large portion of the sample (n=103, 28.8%). Since history of smoking is the risk factor with the best established body of literature to support an association with CIN, the lack of association between smoking and CIN found in this thesis further supports the need for repetition of this aim with a larger sample size. In addition to inadequate power, we fully acknowledge that the quality of data on smoking available for the sample in this thesis was inadequate and should be improved upon in future investigations.

Finally, the rates of persistent, progressive, and regressive CIN in this thesis are different from what has been previously published, namely the rate of persistent dysplasia was higher among CIN 1 cases and the rate of regression for all grades of CIN was much lower than what has been previously published. These results are likely due to the short time window between biopsy and treatment for most women in this cohort. As such, the sensitivity, specificity, and predictive value results must be considered tentatively, since the “consensus diagnosis” despite the 5 year follow-up period and review by two pathologists, may not represent the true biologic potential of these lesions, especially those surgically excised within a short time window (eg, CIN 2). This is a limitation that will affect any study in which therapeutic intervention prohibits monitoring of the natural history of the disease process. Therefore, we acknowledge that the

estimates for persistence, progression, and regression of CIN in this thesis are potentially biased by the short time to treatment.

Public Health Implications:

The results of this thesis have substantial implications for the field of public health. First, these results, applied to current cervical cancer screening strategies, emphasize the need for molecular-based diagnostic methods. Histomorphology alone is likely no longer adequate for diagnosis of CIN lesions. Recommending that all CIN biopsies be evaluated with conjunction with p16 IHC slides would no doubt increase the cost of screeening for cervical cancer and its precursor lesions in the United States and around the world. However, as has been demonstrated here, significant diagnostic improvements are achieved when p16 IHC is used, which results in more appropriate management of women with CIN. We estimated that as many as 30% of women with CIN lesions may be reclassified as a different diagnosis with the use of molecular methods in addition to conventional H&E slides. This equates to a potential change in management for 450,000 of the 1.5 million women estimated to be diagnosed with new CIN in the United States each year. This estimate does not include evaluation of surgical margins, which would affect a smaller number of additional women each year. Since individual IHC slides cost significantly less than excision procedures, sparing women unnecessary procedures would also result in substantial potential for cost-savings. Additionally, the role of p16 and distribution of p16 staining patterns in HPV-mediated cervical cancers could be better characterized if p16 IHC were implemented on a national or international scale.

On a broader scale, the challenges of diagnosis of cervical dysplasia provide an excellent example that emphasizes the need for pathologists to critically examine their own diagnostic practices and asses their degree of diagnostic reproducibility and accuracy. With the ever-expanding body of molecular information, in the future, IHC and other molecular-based methods could conceivably be used to improve diagnostic accuracy and reproducibility for any human pathology (eg, breast "atypia").

This thesis also demonstrates empiric proof of the epidemiologic principle that less misclassification results in measuring stronger associations between risk factors and outcomes. Design of valid studies by reduction of bias is a major component of epidemiology. As applied to CIN, repeating analysis of risk factors with histologic outcomes determined by H&E plus p16

would allow for improved understanding of risk factors for CIN and would provide more valid information for clinicians to use to advise patients about modifiable risk factors. This should also encourage other researchers to explore diagnostic modalities which might, in the future, be anticipated to achieve even less misclassification than current modalities.

Summary & Conclusions

In summary, this study demonstrated that H&E plus p16 improves diagnostic precision and accuracy of cervical dysplasia. The degree of reclassification in this study was substantial (30%) and represents a source of misclassification that has significant effect on individual patient management and would result different treatment strategy for the reclassified patients. Use of p16 could contribute to reduction in the 25% observed false negative rate for CIN 1, 15% observed false positive rate for CIN 2, and 5% observed false positive rate for CIN 3. In effect, this would lead to more appropriate management of higher-grade lesions inappropriately classified as CIN 1, and potentially more conservative management of lower-grade lesions inappropriately classified as CIN 2 or CIN 3.

With the new "gold standard" diagnostic method, statistically significant changes in strength of associations between risk factors and CIN were observed. This thesis does not claim to provide additional insight into the true nature of the strength of associations between risk factors and CIN due to inadequate power, rather this is a successful demonstration of the effect misclassification bias has on calculations of strength of association and has ramifications for the validity of any epidemiologic study that uses pathologically determined outcome classifications based solely on H&E slides.

Tables & Figures

Table 1. Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Aged 15-80 years	Age <15 or >80 years
Initial abnormal pap smear (of any diagnosis) at Kaiser-Permanent Northwest between January 1997 and November 2003	History of prior cervical neoplasia of any type, cervical ablative or surgical excision procedures, or hysterectomy.
≥2 pap smears in 5 years of follow-up data when no excision procedure occurs.	Less than 2 pap smears in 5 years of follow-up data when no excision procedure occurs.
	Colposcopic biopsy blocks not locatable.

Table 2. Descriptive statistics.

Variable	Codes/Values	N=358	Cervical Intraepithelial Neoplasia Grade			
conv_diag_tm (categorical)	0=Negative, 1=CIN 1, 2=CIN 2, 3=CIN 3	Mean (SD) *Percent*	Negative [n=41] Mean (SD), *Percent*	CIN 1 [n=151] Mean (SD), *Percent*	CIN 2 [n=91] Mean (SD), *Percent*	CIN 3 [n=75] Mean (SD), *Percent*
age (continuous)	Years	31.3 (11.8)	38.1 (15.5)	30.9 (12.2)	29.6 (10.0)	30.6 (9.6)
race (categorical)	0=White, 1=Black, 2=Asian, 3=Unknwn, 4=Other	*70.4% White* *18.4% Other* *3.4% Asian*	*65.9% White* *19.5% Other* *7.3% Asian*	*70.9% White* *16.6% Other* *3.3% Black*	*70.3% White* *22% Other* *3.3% Black*	*72% White* *17.3% Other* *4% Asian*
gravity (continuous)	# of pregnancies	0.73 (1.5)	*80.5% G0* *7.3% G1* *4.9% G2* Max: G5	*72.2% G0* *5.3% G1* *10.6% G2* Max: G6	*75.8% G0* *8% G1* *7.7% G2* Max: G7	*72% G0* *8% G1* *2.7% G2* Max: G9
parity (continuous)	# of deliveries	0.31 (0.83)	*87.8% P0* *7.3% P1* *4.9% P2* Max: P2	*79.5% P0* *13.9% P1* *3.3% P2* Max: P3	*82.4% P0* *9.9% P1* *3.3% P2* Max: P6	*85.3% P0* *402% P1* *4% P2* Max: P7
faminc_bin (binary)	0= <$45K/yr, 1= ≥$45K/yr	*64.8% of families ≥$45K/yr*	*82.9% of families ≥$45K/yr*	*61.3% of families ≥$45K/yr*	*61.8% of families ≥$45K/yr*	*65.3% of families ≥$45K/yr*
edlev_bin (categorical)	0=HS grad, 1=some college, 2=college degree	*51.2% HS grad* *33.4% some college* *15.3% college degree*	*39% HS grad* *31.7% some college* *29.3% college degree*	*50.7% HS grad* *34.7% some college* *14.7% college degree*	*52.2% HS grad* *31.1% some college* *16.7% college degree*	*58.3% HS grad* *34.7% some college* *6.9% college degree*
smoke (categorical)	0=Never, 1=Yes (past or current), 2=Unknown	28.7% Never 42.5% Yes 28.8% Unknown	19.5% Never 41.5% Yes 39% Unknown	30.4% Never 42.4% Yes 27.2% Unknown	27.5% Never 41.8% Yes 30.8% Unknown	32% Never 44% Yes 24% Unknown
sti_any*	0=No, 1=Yes	*10.3% Yes*	*2.4% Yes*	*13.9% Yes*	*11% Yes*	*14.7% Yes*
candida_vag	0=No, 1=Yes	*19.6% Yes*	*19.5% Yes*	*20.5% Yes*	*22% Yes*	*18.7% Yes*
malig_any**	0=No, 1=Yes	*8.2% Yes*	*12.2% Yes*	*8.6% Yes*	*5.5% Yes*	*9.3% Yes*
any_ocp***	0=No, 1=Yes	*66.8% Yes*	*53.7% Yes*	*68.2% Yes*	*73.6% Yes*	*62.8% Yes*
depo_provera	0=No, 1=Yes	*21.8% Yes*	*19.5% Yes*	*23.2% Yes*	*19.8% Yes*	*22.7% Yes*
sys_estrogen	0=No, 1=Yes	*1.4% Yes*				
vag_horm	0=No, 1=Yes	*7.26% Yes*				
pap_count	# of pap smears in	7.3 (3.69)	8.1 (3.8); Max: 16	7.7 (3.5); Max: 15	7.18 (3.3); Max: 17	6.12 (4.3); Max:18

* Includes: HPV, nos (n=23), HPV (HR and LR) vaginal (n=0), anus (n=0), and cervix (n=0), HIV (n=3), primary syphilis (n=0), genital warts (n=29) chlamydia of anus (n=0) and lower GU tract (n=0), acute gonorrhea of lower GU (n=1) and upper GU (n=0) tract, and chronic gonorrhea of lower GU (n=0) and upper GU tract (n=0).

**Includes malignancy of any type (n=16), breast (n=10), esophagus (n=1), bladder, (n=1), lung (n=1), mouth (n=1). 19 patients documented to have prior history of malignant cervical neoplasia. Since inclusion criteria required de novo abnormality, will assume that this represents duplication in documentation.

*** Multiple patients used multiple types of oral contraceptive pills. Data are not available to assess dosage or sequence of medications.

Table 3. Rates of persistent, progressive, and regressive CIN as determined by H&E only compared to 5 year follow-up outcome.

		CIN 1	CIN 2	CIN 3
Kaiser H&E only	Persist	74%	64%	58%
	Progress	25%	20%	0%*
	Regress	1%	15%	42%
TM H&E only	Persist	84%	95%	100%
	Progress	15%	0%	0%*
	Regress	0%	5%	0%
RK H&E only	Persist	89%	87.5%	77%
	Progress	11%	5%	0%*
	Regress	0%	7.5%	23%

* No cases of invasive carcinoma were identified in the study.

Table 4.1 Kappa statistic (95% CI) by grade of CIN for each pair of reviewers.

	CIN 1	CIN 2	CIN 3	CIN 2+
	H&E only	H&E only	H&E only	H&E only
TM vs. RK	**0.6406 (0.0617)**	**0.4041 (0.0629)**	**0.6756 (0.0622)**	**0.6677 (0.0625)**
Kaiser vs. TM	0.5254 (0.0623)	0.2162 (0.0630)	0.4736 (0.0620)	0.5254 (0.0623)
Kaiser vs. RK	0.3086 (0.0666)	0.1882 (0.0630)	0.4852 (0.0630)	0.4909 (0.0629)
	H&E plus p16	H&E plus p16	H&E plus p16	H&E plus p16
TM vs. RK	0.6912 (0.0603)	**0.4783 (0.0626)***	0.6678 (0.0621)	0.6912 (0.0603)
	H&E plus p16 and Ki-67	H&E plus p16 and Ki-67	H&E plus p16 and Ki-67	H&E plus p16 and Ki-67
TM vs. RK	**0.7453 (0.0615)***	**0.5204 (0.0629)***	0.5997 (0.0605)	0.7453 (0.0615)

* Significant at $p<0.05$.

Table 4.2. Kappa statistic (95% CI) by staining pattern for p16 and Ki-67 immunohistochemical stained slides for reviewers A and B.

	p16 negative/patchy	p16 basal	p16 diffuse
A vs.B	0.8552 (0.0627)	0.1305 (0.0443)	0.7255 (0.0609)
	Ki-67 – lower third	Ki-67 – middle third	Ki-67 – upper third
A vs. B	0.5481 (0.0674)	0.5305 (0.0628)	0.5849 (0.0616

Table 5. Sensitivity, specificity, positive predictive value, and negative predictive value for each diagnostic method (including p16 IHC and Ki-67 IHC alone) using the "consensus diagnosis" determined over the 5 year follow-up period as the measure of "true disease".

	Kaiser		A H&E	A p16	A H&E+p16	A Ki-67	A H&E +p16/Ki-67		B H&E	B p16	B H&E+p16	B Ki-67	B H&E +p16/Ki-67
Sensitivity	87.79%		80.81%	**93.02%***	**94.19%***	31.98%	**97.09%***		88.37%	**95.93%***	**97.67%***	43.02%	**97.67%***
Specificity	78.75%		86.25%	**100%***	**100%***	**98.75%***	**95.00%***		72.50%	71.25%	68.75%	**96.25%***	68.75%
PPV	89.88%		92.67%	**100%***	**100%***	**98.21%***	**97.66%***		87.36%	87.77%	87.05%	**96.10%***	87.05%
NPV	75.00%		67.65%	**86.96%***	**88.89%***	40.31%	**93.83%***		74.36%	**89.06%***	**93.22%***	44.00%	**93.22%***

* Significant improvement over H&E only for that reviewer ($p<0.05$).

Table 6. For Reviewer A. "True" diagnosis determined by 5 year follow-up consensus diagnosis.

	CIN 1		CIN 2		CIN 3		CIN 2+	
	H&E only	H&E+p16	H&E only	H&E+p16	H&E only	H&E+p16	H&E only	H&E+p16
Sensitivity	52.38%	52.38%	76.06%	**100%***	92.50%	**100%***	80.81%	**94.19%***
Specificity	87.30%	**100%***	84.62%	**100%***	75.00%	**100%***	87.34%	**100%***
PPV	57.89%	**100%***	96.43%	100%	98.67%	100%	93.29%	100%
NPV	84.62%	86.30%	39.39%	**100%***	33.33%	**100%***	67.25%	**88.76%**

* Significant improvement over H&E only for reviewer A ($p<0.05$).

Table 7. For Reviewer B. "True" diagnosis determined by 5 year follow-up consensus diagnosis.

	CIN 1		CIN 2		CIN 3		CIN 2+	
	H&E only	H&E+p16	H&E only	H&E+p16	H&E only	H&E+p16	H&E only	H&E+p16
Sensitivity	71.43%	80.95%	84.51%	**100%***	96.25%	NC	88.37%	**97.67%***
Specificity	74.60%	82.54%	69.23%	**23.08%^**	50.00%	NC	73.42%	68.35%
PPV	48.39%	**60.71%***	93.75%	87.65%	97.47%	NC	87.86%	87.05%
NPV	88.68%	92.86%	45.00%	**100.00%**	40.00%	NC	74.36%	**93.10%***

* Significant improvement over H&E only for reviewer B ($p<0.05$). ^ Significantly worse. NC=not calculatable because reviewer B did not have any CIN3 cases which were not positive (full-thickness) p16 staining.

Table 8. Reclassification structure.

Kaiser H&E only diagnosis	Reviewer A H&E only diagnosis	N (% of 252)	Reviewer A H&E plus p16 diagnosis	N (% of 252)	Reviewer A H&E plus p16 and Ki-67 diagnosis	N (% of 252)
CIN 1	CIN 2	16 (6%)	CIN 2	8 (3%)	CIN 2	12 (5%)
	CIN 3	3 (2%)	CIN 3	3 (2%)	CIN 3	4 (2%)
CIN 2	CIN 1	28 (11%)	CIN 1	13 (5%)	CIN 1	10 (4%)
	CIN 3	17 (7%)	CIN 3	17 (7%)	CIN 3	4 (2%)
CIN 3	CIN 1	9 (4%)	CIN 1	4 (2%)	CIN 1	3 (2%)
	CIN 2	26 (10%)	CIN 2	31 (12%)	CIN 2	36 (14%)
Total reclassified		99 (39%)		76 (30%)		68 (27%)

Table 9. Odds ratio estimates (selected variables). All outcomes dichotomized as negative or CIN1 versus CIN2 or CIN 3.

Variable	Kaiser H&E		Reviewer A H&E	Reviewer A H&E+p16	A H&E+ p16/Ki-67		Reviewer B H&E only	Reviewer B H&E+p16	B H&E+ p16/Ki-67		Consensus diagnosis
Age (<30, ≥30)	0.909		0.892	0.84	0.748		0.784	0.758	0.741		0.738
Family income (<$45K/yr, ≥$45K/yr)	**0.539***		0.892	**0.585***	**0.589***		**0.61***	**0.607***	**0.618***		**0.569***
Education level (<HS grad, ≥HS grad)	**0.477***		0.650	0.630	0.70		0.571	**0.557***	0.600		0.583
Gravidity (0, ≥1)	0.787		0.993	0.928	0.82		0.947	0.888	0.874		0.987
H/o STI? (Never, ever)	1.28		1.12	1.25	1.19		1.05	1.14	1.13		1.53
Vaginal hormone use? (Never, ever)	**0.371***		**0.191***	**0.157***	**0.136***		**0.176***	**0.159***	**0.157***		**0.165***
OCP use? (Never, ever)	0.769		1.17	1.02	1.13		1.11	1.18	1.20		1.15
Race (White, all other)	0.949		0.94	0.811	0.877		0.804	0.842	0.828		0.703
Parity (0, ≥1)	0.84		0.842	0.924	0.784		0.824	0.980	0.966		1.05
H/o smoking? (Never, ever/unknown)	0.82		0.934	0.866	0.887		1.08	0.798	0.811		0.991
H/o malignancy? (Never, ever)	0.885		0.753	0.62	0.621		0.812	0.774	0.764		1.02
H/o depo provera use? (Never, ever)	0.666		0.926	0.645	0.638		0.954	0.638	0.628		0.778

* OR estimates significantly different from H_o: OR=1 ($p<0.05$).

Table 10. Model comparisons and adjusted odds ratio for family income (≥$45K/year) in comparison to unadjusted odds ratios from 2x2 contingency tables.

Diagnostic method	LR X^2_4	p-value of model	Adjusted OR *family income*	p-value of variable	Unadjusted OR *family income*
Kaiser H&E only	16.34	0.0026	**0.609**	**0.036**	**0.539***
Reviewer A H&E only	15.66	0.0035	0.998	0.996	0.892
Reviewer A H&E plus p16	23.03	0.0001	0.636	0.057	**0.585***
Reviewer A H&E plus p16 and Ki-67	25.39	0.0000	0.629	0.054	**0.589***
Reviewer B H&E only	22.26	0.0002	0.681	0.109	**0.61***
Reviewer B H&E plus p16	25.11	0.0001	0.686	0.131	**0.607***
Reviewer B H&E plus p16 and Ki-67	24.56	0.0001	0.689	0.137	**0.618***
"Consensus diagnosis"	24.51	0.0001	0.6298	0.056	**0.569***

* OR estimates significantly different from H_o: OR=1 ($p<0.05$).

Figure 1. Neoplastic transformation by HPV. After high-risk viral integration into the host genome, HPV E7 protein is upregulated, leading to inactivation of human pRb and subsequent overexpression of p16.

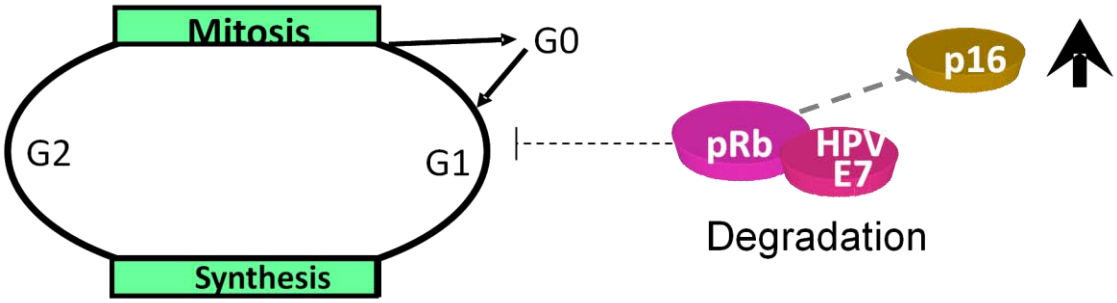


Figure 2. Photomicrographs of cervical epithepithelium showing morphology of CIN 1 (a), CIN 2 (b), and CIN 3 (c) on H&E stained slides.

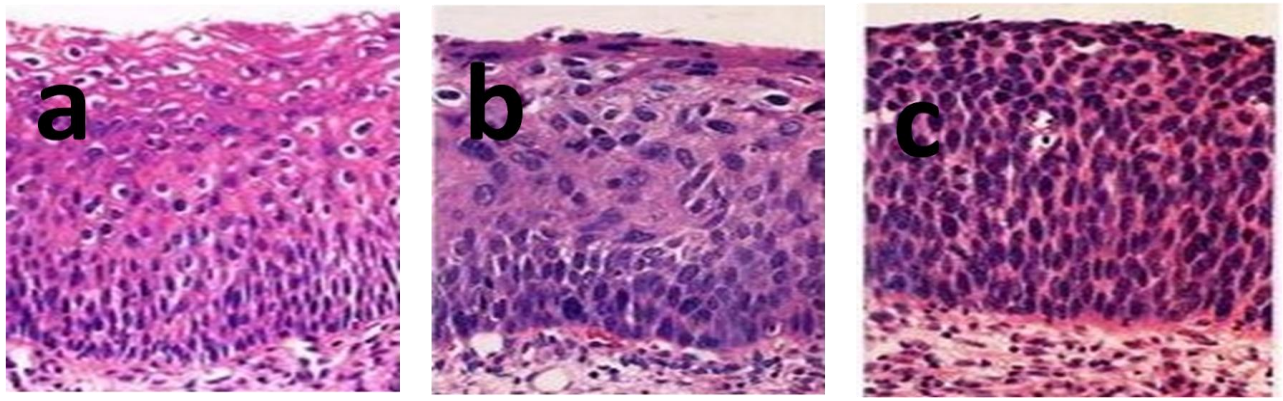


Figure 3. Photomicrographs of negative (a) p16 IHC staining pattern seen in CIN 1 and diffuse full-thickness p16 IHC staining pattern seen in CIN 2 (b) and CIN 3 (c).

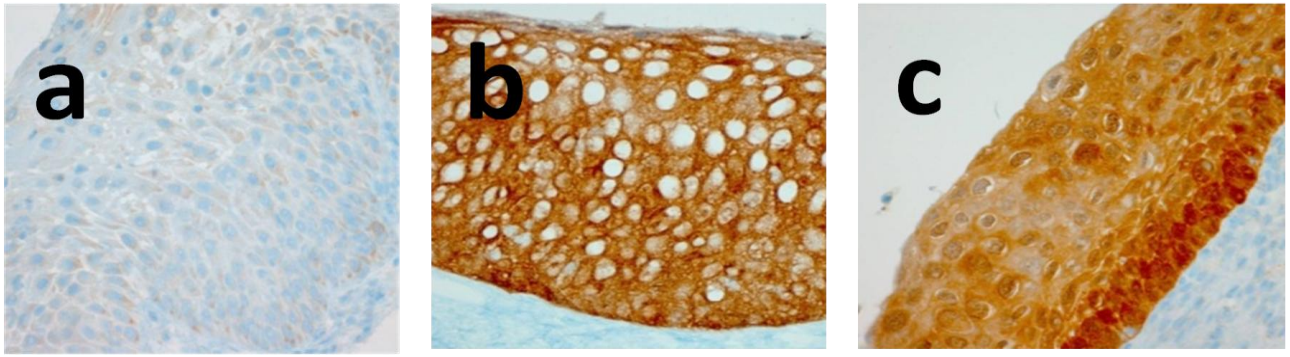


Figure 4. Photomicrographs of predominately lower-third Ki-67 IHC staining patterns (a), predominately lower two-thirds Ki-67 staining patterns (b), and full-thickness Ki-67 IHC staining pattern (c).

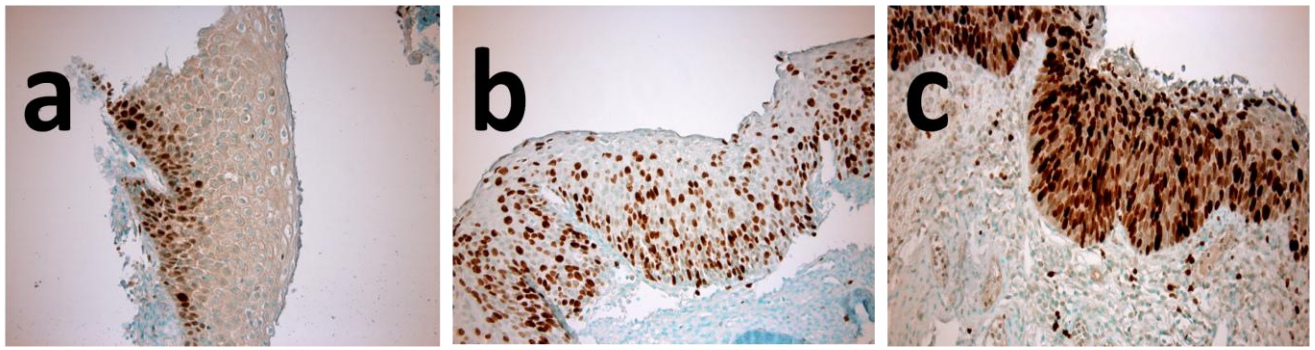


Figure 5. Analytic group diagram.

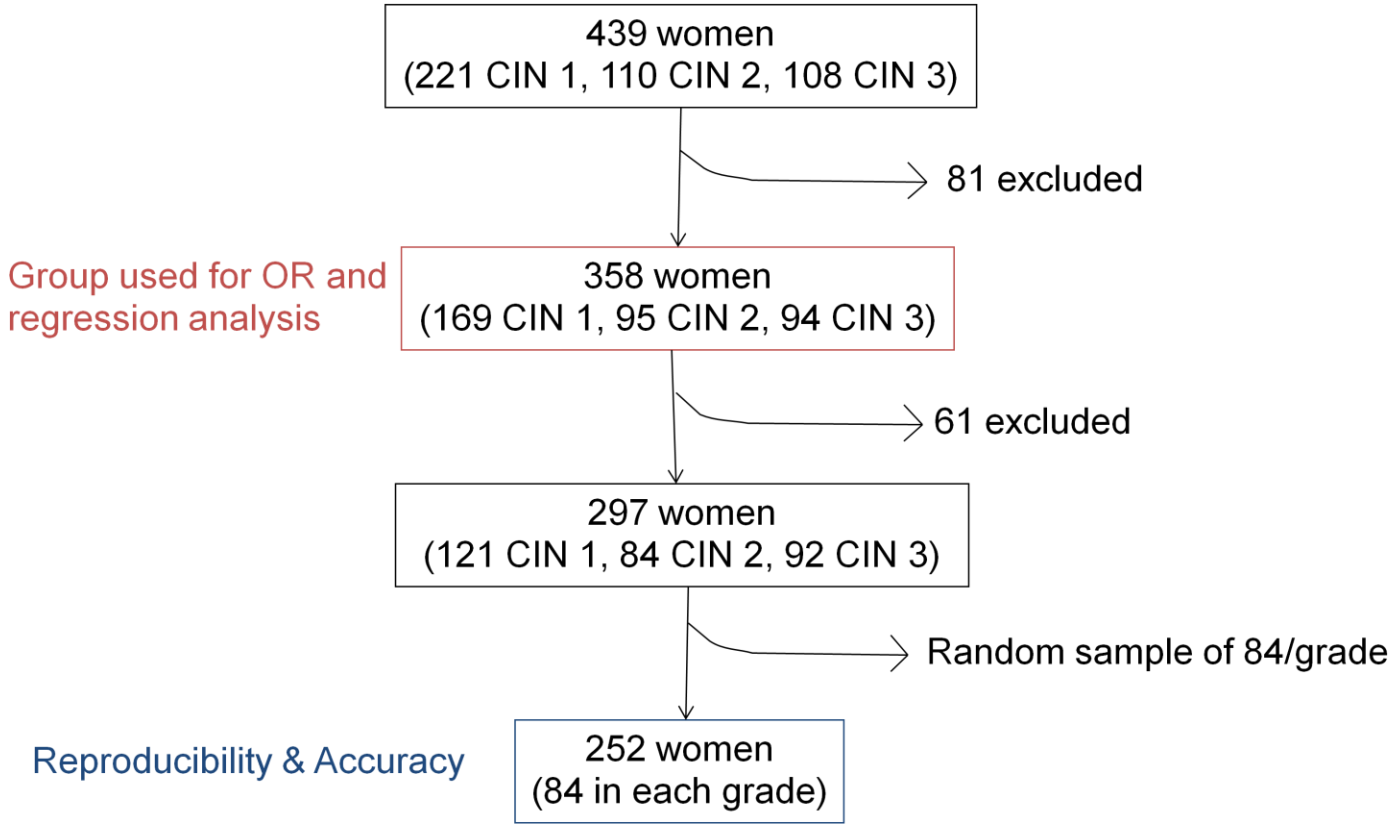
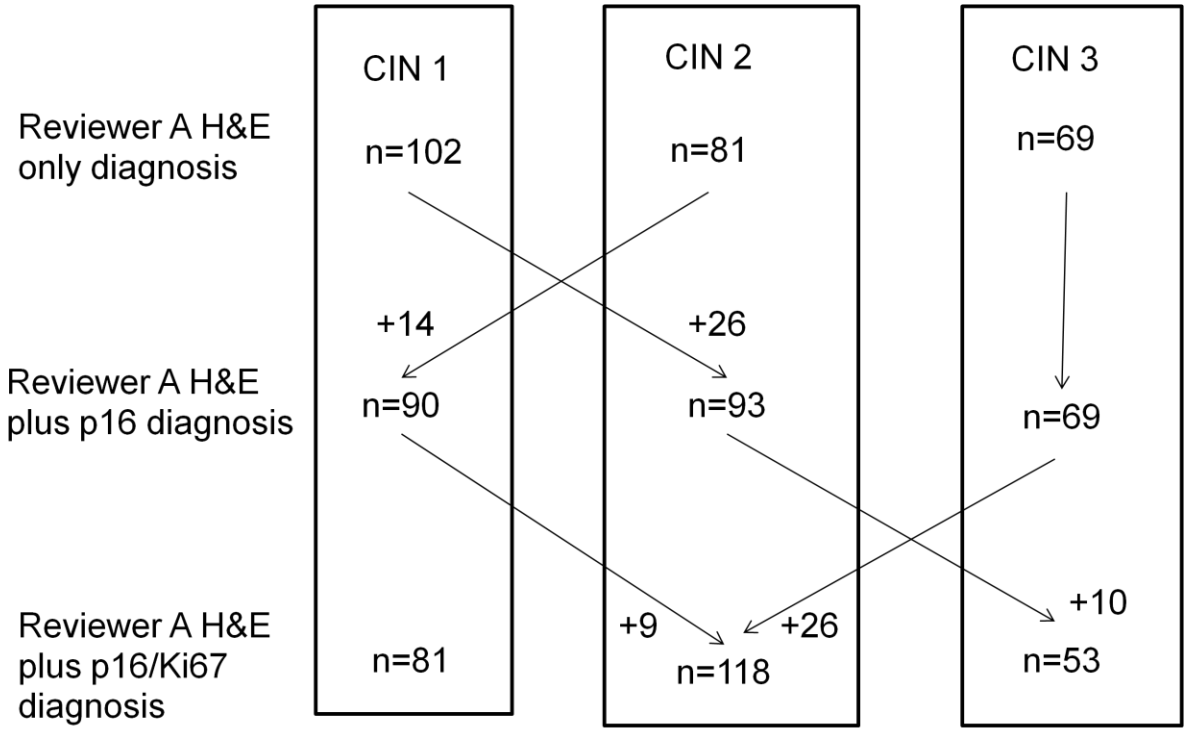


Figure 6. Reclassification structure for reviewer A. Of the original 252 cases (n=84/CIN grade), reviewer A reclassified the cases as 102 CIN 1, 81 CIN 2, and 69 CIN 3 using H&E only diagnosis. With the use of H&E plus p16, 26 CIN 1 cases were reclassified as CIN 2 and 14 CIN 2 cases were reclassified as CIN 1. With the use of H&E plus p16 and Ki-67, 9 CIN 1 cases were reclassified as CIN 2, 10 CIN 2 cases were reclassified as CIN 3, and 26 CIN 3 cases were reclassified as CIN 2.



Appendix 1. Table of variables. Relevance includes citations for selected papers.

Variable name	How queried?	Relevance	Label
age	Patient's age?	Standard demographic information	
race	Patient's race?	Standard demographic information	
faminc	Family income (surrogate)	Standard demographic information	
edlev	Education level (surrogate)	Standard demographic information	
gravidity	Gravidity	Standard demographic information	
parity	Parity	Standard demographic information, (53, 95, 105, 108-110)	Predictor, potential effect modifier (EM)
smoke	Smoking history	Standard demographic information, (94, 95, 108, 110, 111, 113)	Predictor, potential covariate
quadhpv	Quadrivalent HPV vaccine	Indirect medical history	
sti_*	History of sexually transmitted infection?	Indirect medical history, (96)	Potential covariate
hiv, hiv_asx	Positive HIV lab test?	Indirect medical history	Potential covariate
candida_vag	History of vaginal candidiasis?	Indirect medical history	Potential covariate
sind_^	Symptoms at time of index colposcopic biopsy?	Indirect medical history	
htn_aso_ihd	History of hypertension, atherosclerosis, or ischemic heart disease?	Indirect medical history	

asthma	History of asthma?	Indirect medical history	
cbronch	History of chronic bronchitis?	Indirect medical history	
cerebrovd	History of cerebrovascular disease?	Indirect medical history	
diabetes	History of diabetes mellitus?	Indirect medical history	
emphysema	History of emphysema?	Indirect medical history	
endomet	History of endometritis?	Indirect medical history	
do_immune	History of a disorder of the immune system?	Indirect medical history	
do_ovary	History of a disorder of ovarian function?	Indirect medical history	
comp_dtttrans	History of complications due to organ transplant?	Indirect medical history	
malig_any	History of any malignancy?	Indirect medical history	
malig_**	History of ** cancer?	Indirect medical history, History of cervical cancer	Criteria for exclusion
ho_cin1	History of CIN 1?	History of cervical dysplasia	Criteria for exclusion
ho_cin2	History of CIN 2?	History of cervical dysplasia	Criteria for exclusion
ho_cin3	History of CIN 3?	History of cervical dysplasia	Criteria for exclusion
ho_cin_nos	History of CIN, not otherwise specified?	History of cervical dysplasia	Criteria for exclusion
ho_exocerv_neo	History of neoplasm of the exocervix?	History of cervical dysplasia	Criteria for exclusion
pho_malig_cerv	Personal report of history of malignancy of the cervix?	History of cervical dysplasia	Criteria for exclusion
ho_malig_cervneo	Documented history of	History of cervical dysplasia	Criteria for exclusion

	malignancy of the cervix?		
antineo_^^	History of use of antineoplastic medication?	Indirect medical history, medication exposure history	
endocrine	History of use of any medications which effect any endocrine organ or system?	Indirect medical history, medication exposure history	
antiretro	History of use of antiretroviral medications?	Indirect medical history, medication exposure history	
cgmster	History of use of corticosteroids, glucocorticoids, or mineralicorticoids?	Medication exposure history	Potential covariate
vag_horm	History of use of topical vaginal hormones?	Medication exposure history	Primary predictor/EM
vag_anti	History of use of topical vaginal antimicrobials?	Medication exposure history	
vag_other	History of use of Vagisil?	Medication exposure history	
anti_etoh	History of use of medications to assist alcohol cessation?	Indirect medical history	
anti_cig	History of use of medications to assist smoking cessation?	Indirect medical history	
andro_oral, nonoral	History of use of andogens (oral or nonoral)?	Medication exposure history	Potential covariate
anabolic_oral, nonoral	History of use of anabolic steroids (oral or nonoral)?	Medication exposure history	Potential covariate
planb	History of use of Plan B?	Medication exposure history	
ortho_evra	History of use of Orthoevra?	Medication exposure history	
iud	History of IUD use?	Medication exposure history	

implanon	History of Implanon use?	Medication exposure history	
depo_provera	History of Depo Provera use?	Medication exposure history (99)	Primary predictor/EM
nuva_ring	History of Nuva Ring use?	Medication exposure history	
biphasic_oc	History of use of biphasic OCs?	Medication exposure history	Primary predictor/EM
combo_oc	History of use of combination OCs?	Medication exposure history (95, 99, 103-105, 108)	Primary predictor/EM
triphasic_oc	History of use of triphasic OCs?	Medication exposure history	Primary predictor/EM
progestin_oc	History of use of progestin OCs?	Medication exposure history	Primary predictor/EM
insulin	History of use of insulin?	Indirect medical history, medication exposure history	
anti_sugar	History of use of oral anti-hyperglycemics?	Medication exposure history	
sys_estrogen	History of use of systemic estrogens?	Medication exposure history	Primary predictor/EM
analg_antiinflam	History of use of anti-inflammatory analgesics?	Medication exposure history	
analg_nonnarc	History of use of non-narcotic analgesics?	Medication exposure history	
inhaled_ster	History of use of inhaled steroids?	Medication exposure history	
immune_serum	History of use of immune serums?	Medication exposure history	
cerv_cap	History of being dispensed a cervical cap?	Indirect measure of sexual behavior	
diaphragm	History of being dispensed a diaphragm?	Indirect measure of sexual behavior	

pharmacy_iud	History of being dispensed an IUD (per pharmacy records)?	Indirect measure of sexual behavior	
female_condom	History of being prescribed female condoms?	Indirect measure of sexual behavior	
male_condom	History of being prescribed male condoms?	Indirect measure of sexual behavior	
atyp_pap	History of having an atypical Pap smear?	History of cervical dysplasia	
pap_count	Number of Pap smears during study period?	Indirect measure of access to and regularity of medical care	
max_pap_diag	Highest grade of cervical abnormality diagnosed during study period?	History of cervical dysplasia	
colpo_diag	Grade of CIN diagnosed at initial colposcopic biopsy?	Method of diagnosis	Outcome classification 1
conv_diag_tm	Diagnosis assigned by reviewer A for H&E slide of initial colposcopic biopsy?	Method of diagnosis	Outcome classification 2
conv_diag_rk	Diagnosis assigned by reviewer B for H&E slide of initial colposcopic biopsy?	Method of diagnosis	Outcome classification 5
p16_colpo_tm	Diagnosis assigned by reviewer A for p16 slide?	p16 IHC staining pattern contributing to outcome classification 3 and 4	
p16_colpo_rk	Diagnosis assigned by reviewer B for p16 slide?	p16 IHC staining pattern contributing to outcome classification 6 and 7	
molec_c16_tm	Diagnosis assigned by reviewer A for H&E+p16 slides of initial colposcopic biopsy?	Method of diagnosis	Outcome classification 3
molec_c16_rk	Diagnosis assigned by reviewer B for H&E+p16 slides	Method of diagnosis	Outcome classification 6

	of initial colposcopic biopsy?		
Ki-67_colpo_tm	Diagnosis assigned by reviewer A for Ki-67 slide?	Ki-67 IHC staining pattern contributing to outcome classification 4	
Ki-67_colpo_rk	Diagnosis assigned by reviewer B for Ki-67 slide?	Ki-67 IHC staining pattern contributing to outcome classification 7	
molec_c67_tm	Diagnosis assigned by reviewer A for H&E+p16 and Ki-67 slides of initial colposcopic biopsy?	Method of diagnosis	Outcome classification 4
molec_c67_rk	Diagnosis assigned by reviewer B for H&E+p16 and Ki-67 slides of initial colposcopic biopsy?	Method of diagnosis	Outcome classification 7
highdx_tm	Highest grade of CIN diagnosed during entire study period?	Method of diagnosis	Outcome classification 8

* Includes: HPV (not otherwise specified), vaginal HPV (high risk [HR] and low risk [LR] types), anal HPV (HR and LR), and cervical HPV (HR and LR), HIV, primary syphilis, genital warts, chlamydia of anus, and chlamydida of the lower genitourinary (GU) tract, acute gonorrhea of the lower GU tract and upper GU tract, and chronic gonorrhea of lower GU and upper GU tract.

^Clinical symptoms at the time of index colposcopic biopsy including: cervical inflammation, vaginal discharge, a non-inflammatory cervical disorder, erosion or ectropion, uterine prolapse, abnormal bleeding, and genitourinary inflammation.

**Malignancies including: solid organ malignancy, anus, breast, esophagus, uterus, fallopian tube, ovary, vulva, vagina, liver, trachea/bronchus and lung, mouth, oropharyngeal, and skin.

^^Anti-neoplastic medications including hormonal (anti-androgen, anti-estrogen, anti-LHRH, and anti-progestin) and non-hormonal forms.

& All pharmacy data are dichotomous (ever exposed, never exposed).

Appendix 2. Selected statistical output.

Odds ratio calculations from 2x2 contingency tables for *family income*:

```
. tab faminc_bin colpo_diagbin
```

faminc_bin	colpo_diagbin 0	1	Total
0	47	77	124
1	121	107	228
Total	168	184	352

```
. csi 107 77 121 47, or
```

	Exposed	Unexposed	Total		
Cases	107	77	184		
Noncases	121	47	168		
Total	228	124	352		
Risk	.4692982	.6209677	.5227273		
	Point estimate		[95% Conf.	Interval]	
Risk difference	-.1516695		-.2588506	-.0444884	
Risk ratio	.755753		.6219594	.9183278	
Prev. frac. ex.	.244247		.0816722	.3780406	
Prev. frac. pop	.1582054				
Odds ratio	.539766		.3458643	.8424568	(Cornfield)

chi2(1) = 7.41 Pr>chi2 = 0.0065

```
. tab faminc_bin conv_diag_tmbin
```

faminc_bin	conv_diag_tmbin 0	1	Total
0	65	59	124
1	126	102	228
Total	191	161	352

```
. csi 102 59 126 65, or
```

	Exposed	Unexposed	Total		
Cases	102	59	161		
Noncases	126	65	191		
Total	228	124	352		
Risk	.4473684	.4758065	.4573864		
	Point estimate		[95% Conf.	Interval]	
Risk difference	-.028438		-.1374894	.0806133	
Risk ratio	.9402319		.7437669	1.188593	
Prev. frac. ex.	.0597681		-.1885929	.2562331	
Prev. frac. pop	.0387134				
Odds ratio	.8918483		.5758233	1.381257	(Cornfield)

chi2(1) = 0.26 Pr>chi2 = 0.6089

```
. tab faminc_bin molec_c67_tmbin
```

faminc_bin	molec_c67_tmbin 0	1	Total
0	47	77	124
1	116	112	228
Total	163	189	352

```
. csi 112 77 116 47, or
```

	Exposed	Unexposed	Total
Cases	112	77	189
Noncases	116	47	163
Total	228	124	352
Risk	.4912281	.6209677	.5369318

	Point estimate	[95% Conf.	Interval]	
Risk difference	-.1297397	-.2369888	-.0224905	
Risk ratio	.7910686	.6537351	.9572524	
Prev. frac. ex.	.2089314	.0427476	.3462649	
Prev. frac. pop	.1353306			
Odds ratio	.5893417	.3777361	.9195863	(Cornfield)

chi2(1) = 5.44 Pr>chi2 = 0.0197

```
. tab faminc_bin  molec_c16_tmbin
```

faminc_bin	molec_c16_tmbin 0	1	Total
0	51	73	124
1	124	104	228
Total	175	177	352

```
. csi 104 73 124 51, or
```

	Exposed	Unexposed	Total
Cases	104	73	177
Noncases	124	51	175
Total	228	124	352
Risk	.4561404	.5887097	.5028409

	Point estimate	[95% Conf.	Interval]	
Risk difference	-.1325693	-.2406471	-.0244916	
Risk ratio	.7748137	.6316522	.9504222	
Prev. frac. ex.	.2251863	.0495778	.3683478	
Prev. frac. pop	.1458593			
Odds ratio	.5859479	.3768586	.9110957	(Cornfield)

chi2(1) = 5.65 Pr>chi2 = 0.0175

```
. tab faminc_bin highdx_tmbin
```

faminc_bin	highdx_tmbin 0	1	Total
0	44	80	124
1	112	116	228
Total	156	196	352

```
. csi 116 80 112 44, or
```

	Exposed	Unexposed	Total	
Cases	116	80	196	
Noncases	112	44	156	
Total	228	124	352	
Risk	.5087719	.6451613	.5568182	
	Point estimate		[95% Conf.	Interval]
Risk difference	-.1363894		-.2427045	-.0300742
Risk ratio	.7885965		.6570465	.9464847
Prev. frac. ex.	.2114035		.0535153	.3429535
Prev. frac. pop	.1369318			
Odds ratio	.5696429		.3636487	.8924561 (Cornfield)

chi2(1) = 6.05 Pr>chi2 = 0.0139

Final models – varying by diagnostic method

```
. xi: logit colpo_diagbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -243.62405
Iteration 1:   log likelihood = -235.47913
Iteration 2:   log likelihood = -235.45192
Iteration 3:   log likelihood = -235.45191

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      16.34
                                                  Prob > chi2     =     0.0026
Log likelihood = -235.45191                       Pseudo R2       =     0.0335
```

colpo_diag~n	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age_bin_30	1.076532	.2405522	0.33	0.741	.6947441	1.668126
edlev_bin	.5764206	.1832408	-1.73	0.083	.3091344	1.07481
faminc_bin	.6094735	.1441041	-2.09	0.036	.3834391	.9687534
vag_horm	.3470639	.1639614	-2.24	0.025	.1374932	.8760676

```
. xi: logit conv_diag_tmbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -242.70785
Iteration 1:   log likelihood = -235.21107
Iteration 2:   log likelihood = -234.88588
Iteration 3:   log likelihood = -234.87968
Iteration 4:   log likelihood = -234.87967

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      15.66
                                                  Prob > chi2     =     0.0035
Log likelihood = -234.87967                       Pseudo R2       =     0.0323
```

conv_di~mbin	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age_bin_30	1.070207	.2377339	0.31	0.760	.6924421	1.654064
edlev_bin	.6858408	.2202002	-1.17	0.240	.365537	1.286812
faminc_bin	.998701	.2342002	-0.01	0.996	.6307025	1.581417
vag_horm	.1460751	.0922966	-3.04	0.002	.0423399	.5039676

```
. xi: logit molec_c16_tmbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -243.98213
Iteration 1:   log likelihood = -232.84189
Iteration 2:   log likelihood = -232.47937
Iteration 3:   log likelihood = -232.47208
Iteration 4:   log likelihood = -232.47207

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      23.02
                                                  Prob > chi2     =     0.0001
Log likelihood = -232.47207                       Pseudo R2       =     0.0472

------------------------------------------------------------------------------
mole~6_tmbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |    1.03346   .2312187     0.15   0.883     .6665799    1.602268
   edlev_bin |     .771903    .245934    -0.81   0.416     .4133934    1.441325
  faminc_bin |    .6362983   .1510421    -1.90   0.057     .3995819    1.013248
    vag_horm |    .1247838   .0790165    -3.29   0.001     .0360706    .4316802
------------------------------------------------------------------------------


. xi: logit molec_c67_tmbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -243.02671
Iteration 1:   log likelihood = -230.68451
Iteration 2:   log likelihood = -230.33998
Iteration 3:   log likelihood = -230.33334
Iteration 4:   log likelihood = -230.33334

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      25.39
                                                  Prob > chi2     =     0.0000
Log likelihood = -230.33334                       Pseudo R2       =     0.0522

------------------------------------------------------------------------------
mole~7_tmbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |   .9187903   .2065286    -0.38   0.706     .5913991    1.427421
   edlev_bin |   .8783234   .2787575    -0.41   0.683      .471525    1.636079
  faminc_bin |   .6297636   .1513318    -1.92   0.054     .3932189    1.008604
    vag_horm |   .1114173   .0704795    -3.47   0.001     .0322482    .3849455
------------------------------------------------------------------------------


. xi: logit conv_diag_rkbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -242.70785
Iteration 1:   log likelihood = -231.75563
Iteration 2:   log likelihood = -231.57835
Iteration 3:   log likelihood = -231.57706
Iteration 4:   log likelihood = -231.57706

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      22.26
                                                  Prob > chi2     =     0.0002
Log likelihood = -231.57706                       Pseudo R2       =     0.0459

------------------------------------------------------------------------------
conv_di~kbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |   .9646336    .216699    -0.16   0.873     .6210766    1.498234
   edlev_bin |   .6840032    .216577    -1.20   0.230       .36774    1.272259
  faminc_bin |   .6810441   .1630016    -1.60   0.109     .4260377    1.088685
    vag_horm |   .1524771    .086208    -3.33   0.001     .0503442    .4618062
------------------------------------------------------------------------------
```

```
. xi: logit molec_c16_rkbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -233.86772
Iteration 1:   log likelihood = -221.35475
Iteration 2:   log likelihood = -221.31405
Iteration 3:   log likelihood = -221.31402

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      25.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -221.31402                       Pseudo R2       =     0.0537

------------------------------------------------------------------------------
mole~6_rkbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |   .9565997    .222068    -0.19   0.848     .6069176    1.507755
   edlev_bin |   .6670856     .212017   -1.27   0.203     .3578061    1.243699
  faminc_bin |   .6856657   .1715279    -1.51   0.131     .4199268     1.11957
    vag_horm |   .1432008   .0748746    -3.72   0.000     .0513905    .3990325
------------------------------------------------------------------------------

. xi: logit molec_c67_rkbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -233.37503
Iteration 1:   log likelihood =  -221.1291
Iteration 2:   log likelihood = -221.09385
Iteration 3:   log likelihood = -221.09383

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      24.56
                                                  Prob > chi2     =     0.0001
Log likelihood = -221.09383                       Pseudo R2       =     0.0526

------------------------------------------------------------------------------
mole~7_rkbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |   .9300982   .2159323    -0.31   0.755     .5900831    1.466035
   edlev_bin |   .7232173   .2304538    -1.02   0.309     .3872868    1.350532
  faminc_bin |   .6892851   .1725489    -1.49   0.137     .4220049     1.12585
    vag_horm |   .1423388   .0743398    -3.73   0.000     .0511403    .3961718
------------------------------------------------------------------------------

. xi: logit highdx_tmbin age_bin_30 edlev_bin faminc_bin vag_horm, or

Iteration 0:   log likelihood = -241.71016
Iteration 1:   log likelihood = -229.63172
Iteration 2:   log likelihood =  -229.4578
Iteration 3:   log likelihood = -229.45661
Iteration 4:   log likelihood = -229.45661

Logistic regression                               Number of obs   =        352
                                                  LR chi2(4)      =      24.51
                                                  Prob > chi2     =     0.0001
Log likelihood = -229.45661                       Pseudo R2       =     0.0507

------------------------------------------------------------------------------
highdx_tmbin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_bin_30 |   .9084133   .2051153    -0.43   0.671     .5835611    1.414102
   edlev_bin |   .7206968   .2281246    -1.03   0.301     .3875421    1.340252
  faminc_bin |   .6298277   .1523644    -1.91   0.056      .392016    1.011905
    vag_horm |   .1462897   .0827572    -3.40   0.001     .0482706    .4433479
------------------------------------------------------------------------------
```

References

1. Gustafsson L, Ponten J, Zack M, Adami HO. International incidence rates of invasive cervical cancer after introduction of cytological screening. Cancer Causes & Control. 1997 Sep;8(5):755-63.

2. Wright TC,Jr, Massad LS, Dunton CJ, Spitzer M, Wilkinson EJ, Solomon D, et al. 2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ.[see comment]. American Journal of Obstetrics & Gynecology. 2007 Oct;197(4):340-5.

3. Szklo M, Nieto FJ. Epidemiology: Beyond the basics. 2nd ed. Sadbury, Massachusetts: Jones & Bartlett Publishers; 2006.

4. Tavassoli FA, Devilee P, editors. WHO classification of tumors of the breast & female genital tract. Lyon: IARC Press; 2003.

5. Ceballos KM, Chapman W, Daya D, Julian JA, Lytwyn A, McLachlin CM, et al. Reproducibility of the histological diagnosis of cervical dysplasia among pathologists from 4 continents. International Journal of Gynecological Pathology. 2008 Jan;27(1):101-7.

6. Cai B, Ronnett BM, Stoler M, Ferenczy A, Kurman RJ, Sadow D, 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 Dec;31(12):1854-60.

7. Carreon JD, Sherman ME, Guillen D, Solomon D, Herrero R, Jeronimo J, et al. CIN2 is a much less reproducible and less valid diagnosis than CIN3: Results from a histological review of population-based cervical samples. International Journal of Gynecological Pathology. 2007 Oct;26(4):441-6.

8. Castle PE, Stoler MH, Solomon D, Schiffman M. The relationship of community biopsy-diagnosed cervical intraepithelial neoplasia grade 2 to the quality control pathology-reviewed diagnoses: An ALTS report. Am J Clin Pathol. 2007 May;127(5):805-15.

9. Malpica A, Matisic JP, Niekirk DV, Crum CP, Staerkel GA, Yamal JM, et al. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: Qualitative histopathologic analysis and methodologic issues. Gynecol Oncol. 2005 Dec;99(3 Suppl 1):S38-52.

10. Parker MF, Zahn CM, Vogel KM, Olsen CH, Miyazawa K, O'Connor DM. Discrepancy in the interpretation of cervical histology by gynecologic pathologists. Obstetrics & Gynecology. 2002 Aug;100(2):277-80.

11. 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.

12. Horn LC, Reichert A, Oster A, Arndal SF, Trunk MJ, Ridder R, et al. Immunostaining for p16INK4a used as a conjunctive tool improves interobserver agreement of the histologic diagnosis of cervical intraepithelial neoplasia. Am J Surg Pathol. 2008 Apr;32(4):502-12.

13. Sayed K, Korourian S, Ellison DA, Kozlowski K, Talley L, Horn HV, et al. Diagnosing cervical biopsies in adolescents: The use of p16 immunohistochemistry to improve reliability and reproducibility. Journal of Lower Genital Tract Disease. 2007 Jul;11(3):141-6.

14. Castle PE, Schiffman M, Wheeler CM, Wentzensen N, Gravitt PE. Impact of improved classification on the association of human papillomavirus with cervical precancer. Am J Epidemiol. 2010 Jan 15;171(2):155-63.

15. McInturff P, Johnson WO, Cowling D, Gardner IA. Modelling risk when binary outcomes are subject to error. Stat Med. 2004 Apr 15;23(7):1095-109.

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.

17. Klaes R, Friedrich T, Spitkovsky D, Ridder R, Rudy W, Petry U, et al. Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. International Journal of Cancer. 2001 Apr 15;92(2):276-84.

18. Gurrola-Diaz CM, Suarez-Rincon AE, Vazquez-Camacho G, Buonocunto-Vazquez G, Rosales-Quintana S, Wentzensen N, et al. P16INK4a immunohistochemistry improves the reproducibility of the histological diagnosis of cervical intraepithelial neoplasia in cone biopsies. Gynecol Oncol. 2008 Oct;111(1):120-4.

19. Agoff SN, Lin P, Morihara J, Mao C, Kiviat NB, Koutsky LA. p16(INK4a) expression correlates with degree of cervical neoplasia: A comparison with ki-67 expression and detection of high-risk HPV types. Modern Pathology. 2003 Jul;16(7):665-73.

20. Baak JP, Kruse AJ, Robboy SJ, Janssen EA, van Diermen B, Skaland I. Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. J Clin Pathol. 2006 Oct;59(10):1017-28.

21. Negri G, Vittadello F, Romano F, Kasal A, Rivasi F, Girlando S, et al. p16INK4a expression and progression risk of low-grade intraepithelial neoplasia of the cervix uteri. Virchows Archiv. 2004 Dec;445(6):616-20.

22. Pinto AP, Schlecht NF, Woo TY, Crum CP, Cibas ES. Biomarker (ProEx C, p16(INK4A), and MiB-1) distinction of high-grade squamous intraepithelial lesion from its mimics. Modern Pathology. 2008 Sep;21(9):1067-74.

23. Shi J, Liu H, Wilkerson M, Huang Y, Meschter S, Dupree W, et al. Evaluation of p16INK4a, minichromosome maintenance protein 2, DNA topoisomerase IIalpha, ProEX C, and p16INK4a/ProEX C in cervical squamous intraepithelial lesions. Hum Pathol. 2007 Sep;38(9):1335-44.

24. Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, Patel SS, et al. Prevalence of HPV infection among females in the united states. JAMA. 2007 Feb 28;297(8):813-9.

25. de Sanjose S, Diaz M, Castellsague X, Clifford G, Bruni L, Munoz N, et al. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: A meta-analysis. The Lancet Infectious Diseases. 2007 Jul;7(7):453-9.

26. Baseman JG, Koutsky LA. The epidemiology of human papillomavirus infections. Journal of Clinical Virology. 2005 Mar;32(Suppl 1):S16-24.

27. Revzina NV, Diclemente RJ. Prevalence and incidence of human papillomavirus infection in women in the USA: A systematic review. Int J STD AIDS. 2005 Aug;16(8):528-37.

28. Cates W,Jr. Estimates of the incidence and prevalence of sexually transmitted diseases in the united states. american social health association panel. Sex Transm Dis. 1999 Apr;26(4 Suppl):S2-7.

29. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans. *IARC working group on the evaluation of carciongenic risk to humans* Human papillomaviruses. Lyon, France: International Agency for Research on Cancer; 2007.

30. Bosch FX, Lorincz A, Munoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer.[see comment]. J Clin Pathol. 2002 Apr;55(4):244-65.

31. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003 Feb 6;348(6):518-27.

32. Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. Vaccine. 2006 Mar 30;24(Suppl 1):S1-15.

33. Kalantari M, Karlsen F, Johansson B, Sigurjonsson T, Warleby B, Hagmar B. Human papillomavirus findings in relation to cervical intraepithelial neoplasia grade: A study on 476 stockholm women, using PCR for detection and typing of HPV. Hum Pathol. 1997 Aug;28(8):899-904.

34. Khan MJ, Castle PE, Lorincz AT, Wacholder S, Sherman M, Scott DR, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst. 2005 Jul 20;97(14):1072-9.

35. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999 Sep;189(1):12-9.

36. zur Hausen H. Roots and perspectives of contemporary papillomavirus research. Journal of Cancer Research & Clinical Oncology. 1996;122(1):3-13.

37. zur Hausen H. Papillomaviruses and cancer: From basic studies to clinical application. Nature Reviews.Cancer. 2002;2(5):342-50.

38. Jordan J, Singer A, Jones H, Shafi M, editors. The cervix. Wiley-Blackwell; 2006.

39. Shai A, Brake T, Somoza C, Lambert PF. The human papillomavirus E6 oncogene dysregulates the cell cycle and contributes to cervical carcinogenesis through two independent activities. Cancer Res. 2007 Feb 15;67(4):1626-35.

40. Sano T, Oyama T, Kashiwabara K, Fukuda T, Nakajima T. Expression status of p16 protein is associated with human papillomavirus oncogenic potential in cervical and genital lesions. Am J Pathol. 1998 Dec;153(6):1741-8.

41. Nakao Y, Yang X, Yokoyama M, Ferenczy A, Tang SC, Pater MM, et al. Induction of p16 during immortalization by HPV 16 and 18 and not during malignant transformation. Br J Cancer. 1997;75(10):1410-6.

42. Maufort JP, Shai A, Pitot HC, Lambert PF. A role for HPV16 E5 in cervical carcinogenesis. Cancer Res. 2010;70(7):2924,2924-2931.

43. Jabbar SF, Abrams L, Glick A, Lambert PF. Persistence of high-grade cervical dysplasia and cervical cancer requires the continuous expression of the human papillomavirus type 16 E7 oncogene. Cancer Res. 2009 May 15;69(10):4407-14.

44. Balsitis S, Dick F, Dyson N, Lambert PF. Critical roles for non-pRb targets of human papillomavirus type 16 E7 in cervical carcinogenesis. Cancer Res. 2006 Oct 1;66(19):9393-400.

45. Shai A, Pitot HC, Lambert PF. p53 loss synergizes with estrogen and papillomaviral oncogenes to induce cervical and breast cancers. Cancer Res. 2008 Apr 15;68(8):2622-31.

46. Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human papillomavirus DNA versus papanicolaou screening tests for cervical cancer. N Engl J Med. 2007 Oct 18;357(16):1579-88.

47. Moscicki AB. Impact of HPV infection in adolescent populations. Journal of Adolescent Health. 2005 Dec;37(6 Suppl):S3-9.

48. Crum CP, Nucci MR, Lee KR. Sternberg's diagnostic surgical pathology. In: Mills SE, Carter D, Greenson JK, Oberman HA, Reuter V, Stoler MH, editors. 4th ed. Lippincott Williams & Wilkins; 2004. p. 2377-91.

49. Remoue F, Jacobs N, Miot V, Boniver J, Delvenne P. High intraepithelial expression of estrogen and progesterone receptors in the transformation zone of the uterine cervix. American Journal of Obstetrics & Gynecology. 2003 Dec;189(6):1660-5.

50. Plummer M, Schiffman M, Castle PE, Maucort-Boulch D, Wheeler CM, ALTS G. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. J Infect Dis. 2007 Jun 1;195(11):1582-9.

51. Moscicki AB, Hills N, Shiboski S, Powell K, Jay N, Hanson E, et al. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. JAMA. 2001 Jun 20;285(23):2995-3002.

52. Franco EL, Villa LL, Sobrinho JP, Prado JM, Rousseau MC, Desy M, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. J Infect Dis. 1999 Nov;180(5):1415-23.

53. Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med. 1998 Feb 12;338(7):423-8.

54. Kanodia S, Fahey LM, Kast WM. Mechanisms used by human papillomaviruses to escape the host immune response. Current Cancer Drug Targets. 2007 Feb;7(1):79-89.

55. Schiffman M, Kjaer SK. Chapter 2: Natural history of anogenital human papillomavirus infection and neoplasia. J Natl Cancer Inst. 2003;Monographs(31):14-9.

56. Mitchell MF, Tortolero-Luna G, Wright T, Sarkar A, Richards-Kortum R, Hong WK, et al. Cervical human papillomavirus infection and intraepithelial neoplasia: A review. J Natl Cancer Inst. 1996;Monographs(21):17-25.

57. Ostor AG. Natural history of cervical intraepithelial neoplasia: A critical review. International Journal of Gynecological Pathology. 1993 Apr;12(2):186-92.

58. Chan JK, Monk BJ, Brewer C, Keefe KA, Osann K, McMeekin S, et al. HPV infection and number of lifetime sexual partners are strong predictors for 'natural' regression of CIN 2 and 3. Br J Cancer. 2003 Sep 15;89(6):1062-6.

59. McCredie MR, Sharples KJ, Paul C, Baranyai J, Medley G, Jones RW, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: A retrospective cohort study. Lancet Oncology. 2008 May;9(5):425-34.

60. Woodman CB, Collins S, Winter H, Bailey A, Ellis J, Prior P, et al. Natural history of cervical human papillomavirus infection in young women: A longitudinal cohort study. Lancet. 2001 Jun 9;357(9271):1831-6.

61. Insinga RP, Dasbach EJ, Elbasha EH, Liaw KL, Barr E. Incidence and duration of cervical human papillomavirus 6, 11, 16, and 18 infections in young women: An evaluation from multiple analytic perspectives. Cancer Epidemiology, Biomarkers & Prevention. 2007 Apr;16(4):709-15.

62. Winer RL, Kiviat NB, Hughes JP, Adam DE, Lee SK, Kuypers JM, et al. Development and duration of human papillomavirus lesions, after initial infection. J Infect Dis. 2005 Mar 1;191(5):731-8.

63. Fuchs K, Weitzen S, Wu L, Phipps MG, Boardman LA. Management of cervical intraepithelial neoplasia 2 in adolescent and young women. Journal of Pediatric & Adolescent Gynecology. 2007 Oct;20(5):269-74.

64. Wright TC,Jr. Chapter 3. pathology of HPV infection at the cytologic and histologic levels: Basis for a 2-tiered morphologic classification system. International Journal of Gynecology and Obstetrics. 2006;94(Supplement 1):S22-S31.

65. Fletcher CD, editor. Diagnostic histopathology of tumors. Churchill Livingston Elsevier; 2007.

66. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. Fam Med. 2005 May;37(5):360-3.

67. de Vet HC, Koudstaal J, Kwee WS, Willebrand D, Arends JW. Efforts to improve interobserver agreement in histopathological grading. J Clin Epidemiol. 1995 Jul;48(7):869-73.

68. Dalla Palma P, Giorgi Rossi P, Collina G, Buccoliero AM, Ghiringhello B, Lestani M, et al. The risk of false-positive histology according to the reason for colposcopy referral in cervical cancer screening: A blind revision of all histologic lesions found in the NTCC trial. Am J Clin Pathol. 2008 Jan;129(1):75-80.

69. Samson SL, Bentley JR, Fahey TJ, McKay DJ, Gill GH. The effect of loop electrosurgical excision procedure on future pregnancy outcome. Obstetrics & Gynecology. 2005 Feb;105(2):325-32.

70. Suh-Burgmann EJ, Whall-Strojwas D, Chang Y, Hundley D, Goodman A. Risk factors for cervical stenosis after loop electrocautery excision procedure. Obstetrics & Gynecology. 2000 Nov;96(5 Pt 1):657-60.

71. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: Systematic review and meta-analysis. Lancet. 2006 Feb 11;367(9509):489-98.

72. Tyler LN, Andrews N, Parrish RS, Hazlett LJ, Korourian S. Significance of margin and extent of dysplasia in loop electrosurgery excision procedure biopsies performed for high-grade squamous intraepithelial lesion in predicting persistent disease. Arch Pathol Lab Med. 2007 Apr;131(4):622-4.

73. Kalliala I, Dyba T, Nieminen P, Hakulinen T, Anttila A. Mortality in a long-term follow-up after treatment of CIN. International Journal of Cancer. 2010 Jan 1;126(1):224-31.

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

75. Kruse AJ, Skaland I, Janssen EA, Buhr-Wildhagen S, Klos J, Arends MJ, et al. Quantitative molecular parameters to identify low-risk and high-risk early CIN lesions: Role of markers of proliferative activity and differentiation and rb availability. International Journal of Gynecological Pathology. 2004 Apr;23(2):100-9.

76. Saha B, Chaiwun B, Tsao-Wei DD, Groshen SL, Naritoku WY, Atkinson RD, et al. Telomerase and markers of cellular proliferation are associated with the progression of cervical intraepithelial neoplasia lesions. International Journal of Gynecological Pathology. 2007 Jul;26(3):214-22.

77. Badr RE, Walts AE, Chung F, Bose S. BD ProEx C: A sensitive and specific marker of HPV-associated squamous lesions of the cervix. Am J Surg Pathol. 2008 Jun;32(6):899-906.

78. Dabbs D, editor. Diagnostic immunohistochemistry. Churchill Livingston Elsevier; 2006.

79. Kim JW, Namkoong SE, Ryu SW, Kim HS, Shin JW, Lee JM, et al. Absence of p15INK4B and p16INK4A gene alterations in primary cervical carcinoma tissues and cell lines with human papillomavirus infection. Gynecol Oncol. 1998 Jul;70(1):75-9.

80. Fujii T, Tsukazaki K, Kiguchi K, Kubushiro K, Yajima M, Nozawa S. The major E6/E7 transcript of HPV-16 in exfoliated cells from cervical neoplasia patients. Gynecol Oncol. 1995 Aug;58(2):210-5.

81. Ishikawa M, Fujii T, Saito M, Nindl I, Ono A, Kubushiro K, et al. Overexpression of p16 INK4a as an indicator for human papillomavirus oncogenic activity in cervical squamous neoplasia. International Journal of Gynecological Cancer. 2006 Jan-Feb;16(1):347-53.

82. Yildiz IZ, Usubutun A, Firat P, Ayhan A, Kucukali T. Efficiency of immunohistochemical p16 expression and HPV typing in cervical squamous intraepithelial lesion grading and review of the p16 literature. Pathology, Research & Practice. 2007;203(6):445-9.

83. Keating JT, Cviko A, Riethdorf S, Riethdorf L, Quade BJ, Sun D, et al. Ki-67, cyclin E, and p16INK4 are complimentary surrogate biomarkers for human papilloma virus-related cervical neoplasia. Am J Surg Pathol. 2001 Jul;25(7):884-91.

84. Guedes AC, Brenna SM, Coelho SA, Martinez EZ, Syrjanen KJ, Zeferino LC. p16(INK4a) expression does not predict the outcome of cervical intraepithelial neoplasia grade 2. International Journal of Gynecological Cancer. 2007 Sep-Oct;17(5):1099-103.

85. Hariri J, Oster A. The negative predictive value of p16INK4a to assess the outcome of cervical intraepithelial neoplasia 1 in the uterine cervix. International Journal of Gynecological Pathology. 2007 Jul;26(3):223-8.

86. Redman R, Rufforny I, Liu C, Wilkinson EJ, Massoll NA. The utility of p16(Ink4a) in discriminating between cervical intraepithelial neoplasia 1 and nonneoplastic equivocal lesions of the cervix. Arch Pathol Lab Med. 2008 May;132(5):795-9.

87. Qiao X, Bhuiya TA, Spitzer M. Differentiating high-grade cervical intraepithelial lesion from atrophy in postmenopausal women using ki-67, cyclin E, and p16 immunohistochemical analysis. Journal of Lower Genital Tract Disease. 2005 Apr;9(2):100-7.

88. Regauer S, Reich O. CK17 and p16 expression patterns distinguish (atypical) immature squamous metaplasia from high-grade cervical intraepithelial neoplasia (CIN III). Histopathology. 2007 Apr;50(5):629-35.

89. Aoyama C, Liu P, Ostrzega N, Holschneider CH. Histologic and immunohistochemical characteristics of neoplastic and nonneoplastic subgroups of atypical squamous lesions of the uterine cervix. Am J Clin Pathol. 2005 May;123(5):699-706.

90. Brown DC, Gatter KC. Ki67 protein: The immaculate deception? Histopathology. 2002 Jan;40(1):2-11.

91. Kruse AJ, Baak JP, de Bruin PC, Jiwa M, Snijders WP, Boodt PJ, et al. Ki-67 immunoquantitation in cervical intraepithelial neoplasia (CIN): A sensitive marker for grading. J Pathol. 2001 Jan;193(1):48-54.

92. Kruse AJ, Baak JP, Janssen EA, Bol MG, Kjellevold KH, Fianne B, et al. Low- and high-risk CIN 1 and 2 lesions: Prospective predictive value of grade, HPV, and ki-67 immuno-quantitative variables. J Pathol. 2003 Apr;199(4):462-70.

93. Panjkovic M, Ivkovic-Kapicl T. Ki-67 expression in squamous intraepithelial lesions of the uterine cervix. Arch Oncol. 2006;14(1):23-5.

94. Harris TG, Kulasingam SL, Kiviat NB, Mao C, Agoff SN, Feng Q, et al. Cigarette smoking, oncogenic human papillomavirus, ki-67 antigen, and cervical intraepithelial neoplasia. Am J Epidemiol. 2004 May 1;159(9):834-42.

95. Castellsague X, Munoz N. Chapter 3: Cofactors in human papillomavirus carcinogenesis--role of parity, oral contraceptives, and tobacco smoking. J Natl Cancer Inst. 2003;Monographs(31):20-8.

96. Castle PE, Giuliano AR. Chapter 4: Genital tract infections, cervical inflammation, and antioxidant nutrients--assessing their roles as human papillomavirus cofactors. J Natl Cancer Inst. 2003;Monographs(31):29-34.

97. Wang SS, Hildesheim A. Chapter 5: Viral and host factors in human papillomavirus persistence and progression. J Natl Cancer Inst. 2003;Monographs(31):35-40.

98. Kahn JA, Lan D, Kahn RS. Sociodemographic factors associated with high-risk human papillomavirus infection. Obstetrics & Gynecology. 2007 Jul;110(1):87-95.

99. Castle PE, Walker JL, Schiffman M, Wheeler CM. Hormonal contraceptive use, pregnancy and parity, and the risk of cervical intraepithelial neoplasia 3 among oncogenic HPV DNA-positive women with equivocal or mildly abnormal cytology. International Journal of Cancer. 2005 Dec 20;117(6):1007-12.

100. Bauer HM, Hildesheim A, Schiffman MH, Glass AG, Rush BB, Scott DR, et al. Determinants of genital human papillomavirus infection in low-risk women in portland, oregon. Sex Transm Dis. 1993 Sep-Oct;20(5):274-8.

101. Deacon JM, Evans CD, Yule R, Desai M, Binns W, Taylor C, et al. Sexual behaviour and smoking as determinants of cervical HPV infection and of CIN3 among those infected: A case-control study nested within the manchester cohort. Br J Cancer. 2000 Dec;83(11):1565-72.

102. Khan MJ, Partridge EE, Wang SS, Schiffman M. Socioeconomic status and the risk of cervical intraepithelial neoplasia grade 3 among oncogenic human papillomavirus DNA-positive women with equivocal or mildly abnormal cytology. Cancer. 2005 Jul 1;104(1):61-70.

103. Castle PE, Wacholder S, Lorincz AT, Scott DR, Sherman ME, Glass AG, et al. A prospective study of high-grade cervical neoplasia risk among human papillomavirus-infected women. J Natl Cancer Inst. 2002 Sep 18;94(18):1406-14.

104. Moreno V, Bosch FX, Munoz N, Meijer CJ, Shah KV, Walboomers JM, et al. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: The IARC multicentric case-control study. Lancet. 2002 Mar 30;359(9312):1085-92.

105. Maucort-Boulch D, Plummer M, Castle PE, Demuth F, Safaeian M, Wheeler CM, et al. Predictors of human papillomavirus persistence among women with equivocal or mildly abnormal cytology. International Journal of Cancer. 2010 Feb 1;126(3):684-91.

106. Parikh S, Brennan P, Boffetta P. Meta-analysis of social inequality and the risk of cervical cancer. International Journal of Cancer. 2003 Jul 10;105(5):687-91.

107. Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst. 1993 Jun 16;85(12):958-64.

108. Kjellberg L, Hallmans G, Ahren AM, Johansson R, Bergman F, Wadell G, et al. Smoking, diet, pregnancy and oral contraceptive use as risk factors for cervical intra-epithelial neoplasia in relation to human papillomavirus infection. Br J Cancer. 2000 Apr;82(7):1332-8.

109. Munoz N, Franceschi S, Bosetti C, Moreno V, Herrero R, Smith JS, et al. Role of parity and human papillomavirus in cervical cancer: The IARC multicentric case-control study. Lancet. 2002 Mar 30;359(9312):1093-101.

110. Belinson S, Smith JS, Myers E, Olshan A, Belinson J, Pretorius R, et al. Descriptive evidence that risk profiles for cervical intraepithelial neoplasia 1, 2, and 3 are unique. Cancer Epidemiology, Biomarkers & Prevention. 2008 Sep;17(9):2350-5.

111. Becker TM, Wheeler CM, McGough NS, Parmenter CA, Stidley CA, Jamison SF, et al. Cigarette smoking and other risk factors for cervical dysplasia in southwestern hispanic and non-hispanic white women. Cancer Epidemiology, Biomarkers & Prevention. 1994 Mar;3(2):113-9.

112. Plummer M, Herrero R, Franceschi S, Meijer CJ, Snijders P, Bosch FX, et al. Smoking and cervical cancer: Pooled analysis of the IARC multi-centric case--control study. Cancer Causes & Control. 2003 Nov;14(9):805-14.

113. McIntyre-Seltman K, Castle PE, Guido R, Schiffman M, Wheeler CM, ALTS G. Smoking is a risk factor for cervical intraepithelial neoplasia grade 3 among oncogenic human papillomavirus DNA-positive women with equivocal or mildly abnormal cytology. Cancer Epidemiology, Biomarkers & Prevention. 2005 May;14(5):1165-70.

114. Delgado-Rodriguez M, Llorca J. Bias. Journal of Epidemiology & Community Health. 2004 Aug;58(8):635-41.

115. Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. Journal of the American Statistical Association. 1975;70(351):561-7.

116. Birkett N. Effect of nondifferential misclassification on estimates of odds ratios with mulitiple levels of exposure. American Journal of Epidemiology. 1992;136(3):356-62.

117. Bross I. Misclassification in 2x2 tables. Biometrics. 1954;10:478-86.

118. Xue X, Gange SJ, Zhong Y, Burk RD, Minkoff H, Massad LS, et al. Marginal and mixed-effects models in the analysis of human papillomavirus natural history data. Cancer Epidemiology, Biomarkers & Prevention. 2010 Jan;19(1):159-69.

119. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol. 1992 Dec 1;136(11):1400-13.

120. David O, Cabay RJ, Pasha S, Dietrich R, Leach L, Guo M, et al. The role of deeper levels and ancillary studies (p16(Ink4a) and ProExC) in reducing the discordance rate of papanicolaou findings of high-grade squamous intraepithelial lesion and follow-up cervical biopsies. Cancer Cytopathology. 2009 Jun 25;117(3):157-66.