

Cervical Cancer Detection and Prevention in Haiti: A Comparison of Pap Smear and
Liquid-Based Cytology Detection Methods

by

Genevieve Wolpert

Duke Global Health Institute
Duke University

Date: _____

Approved:

Christopher Woods, Supervisor

David Walmer

Larry Park

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Duke Global Health Institute
in the Graduate School of Duke University

2015

ABSTRACT

Cervical Cancer Detection and Prevention in Haiti: A Comparison of Pap Smear and

Liquid-Based Cytology Detection Methods

by

Genevieve Wolpert

Duke Global Health Institute

Duke University

Date: _____

Approved:

Christopher Woods, Supervisor

David Walmer

Larry Park

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Duke Global Health Institute in the Graduate School of
Duke University

2015

Copyright by
Genevieve Wolpert
2015

Abstract

Human Papillomavirus-induced cervical cancer represents one of the most significant causes of female morbidity and mortality from cancer worldwide. Detecting HPV-related cervical disease during the premalignant treatable stage of development is critical to reduce the burden of this disease. Cervical cytology has been the primary screening tool for cervical dysplasia in the United States for decades. However in Haiti, early attempts to identify cervical dysplasia were thwarted by a high incidence of obscuring inflammation on conventional Pap smears. This study seeks to determine if liquid-based cytology screening can increase the detection of cervical dysplasia over conventional Pap smears when obscuring inflammation is present. The study population was recruited in Haiti and women underwent both types of cervical dysplasia testing; those for whom it was indicated underwent follow-up cervical biopsy. The cervical dysplasia tests were compared to each other using kappa agreement statistics with cervical biopsy as the gold standard for diagnosis. Both tests showed comparable sensitivity for dysplasia with and without inflammation-containing samples. The Pap test showed superior specificity by greater agreement with the gold standard biopsy, though the sample size was small. Interpretation of these results and application to a

low-resource setting for implementation of a standardized screening regimen would require a larger sample size and cost/benefit analysis.

Contents

Abstract	iv
List of Tables	viii
List of Figures	ix
1. Project Overview	1
1.1 Cervical Cancer	1
1.1.1 Epidemiology	1
1.1.2 Pathophysiology	2
1.1.3 Screening and Treatment	5
1.2 Fieldwork Site	7
1.3 Study Design	9
1.3.1 Hypothesis	9
1.3.2 Selection of Participants	9
1.3.3 Measures and Data Collection	10
1.3.3.1 Measures	10
1.3.3.2 Data Collection Methods	10
1.3.3.3 Bias Control in Design	12
1.3.4 Analysis	12
2. Results and Interpretation	14
2.1 Screening test comparisons (Pap vs LBC)	14
2.1.1 Dysplasia characteristics of entire study sample	14

2.1.2 Comparison of dysplasia characteristics among all dual-study samples.....	15
2.1.3 Comparison of dysplasia characteristics among dual-study samples without inflammation.....	17
2.1.4 Agreement statistics for screening test comparisons	18
2.2 Screening tests compared to biopsy results.....	20
2.2.1 Comparison of Pap results to biopsy results.....	20
2.2.2 Comparison of Liquid-Based Cytology results to biopsy results.....	21
2.2.3 Summary comparison of both screening tests to biopsy	22
3. Discussion	26
3.1 Discussion of results.....	26
3.2 Discussion of possible clinical implications.....	28
3.3 Limitations.....	30
3.4 Ethical Considerations	31
3.5 Lessons Learned	32
3.6 Implications for Future Research	32
3.7 Conclusion.....	33

List of Tables

Table 1: Kappa Agreement Strength Benchmarks (Landis & Koch, 1977).....	13
Table 2: Study sample dichotomized by presence or absence of dysplasia, reported with and without those samples with obscuring inflammation.	14
Table 3: Study samples with dual results	16
Table 4: Study samples with dual results dichotomized by dysplasia presence	17
Table 5: Samples with dual results without inflammation dichotomized by dysplasia presence	18
Table 6: Kappa agreement statistics for comparison of Pap and LBC tests with and without inflammation	19
Table 7: Pap vs biopsy	20
Table 8: Agreement statistics for Pap compared to biopsy results	20
Table 9: LBC vs biopsy	21
Table 10: Agreement statistics for LBC compared to biopsy results	21
Table 11: Agreement comparisons for all test types	23

List of Figures

Figure 1: Map of Haiti (Online, 2013).....	8
Figure 2: Leogane clinic wall three months after the earthquake (Photo by Genevieve Wolpert).....	30

1. Project Overview

1.1 Cervical Cancer

Cervical cancer is a leading cause of female death worldwide. It affects women all over the world, and the degree of morbidity and mortality caused by the Human Papillomavirus and subsequent cervical cancer has brought this disease to the forefront.

1.1.1 Epidemiology

Cervical cancer is an extremely prevalent disease worldwide, representing 13% of female cancers. The incidence is significant, with approximately 529,000 new cases diagnosed in 2008 (WHO/ICO, 2010c). The incidence of cervical cancer is second only to breast cancer among cancers affecting women (WHO/ICO, 2010c). The mortality rate is lower than the incidence, with approximately 275,000 deaths in 2008 worldwide, so the prevalence is consistently high, especially in areas without treatment resources. The incidence of cervical cancer has dropped by greater than 50% in areas where routine cytologic cervical screening has been implemented. The rates of decrease are disproportionately split along race and poverty lines, with higher rates correlating with underserved areas in the United States (Scarinci et al., 2010). This disparity indicates that social variables play a role in which patients receive screening and treatment.

This pattern of disparity is repeated on the global scale as well. There is a large discrepancy in disease burden between developed and developing countries, with 86%

of cervical cancer cases occurring in developing countries (WHO/ICO, 2010c). The United States is a leader in cancer research and screening, and has a comparatively low rate, with a cervical cancer incidence rate of 7.0 per 100,000 women as compared to 15.8 per 100,000 worldwide (WHO/ICO, 2010b). Haiti represents a severely underserved population from a screening perspective, with an incidence rate of 11.4 per 100,000, and an age-standardized incidence rate of 16.0 per 100,000 (WHO/ICO, 2010a). When compared to the United States, Haiti has a very high rate, though it is a current leader among Caribbean countries, where the incidence overall is 22.5 per 100,000 women. Haiti sits between Puerto Rico at an age-adjusted incidence rate of 7.5, and Jamaica at a rate of 45.7 per 100,000 (WHO/ICO, 2010a). Haiti represents the inequity of burden of disease in underserved areas, but it also represents an opportunity to move forward.

1.1.2 Pathophysiology

Cervical cancer is caused by the Human Papillomavirus (HPV). The discovery of this link resulted in a Nobel prize and allowed the development of vaccines against HPV that reduce cervical cancer rates (Crosbie, Einstein, Franceschi, & Kitchener). The natural history of this disease spans from cervical infection with HPV to development of squamous cell carcinoma.

The disease process starts with infection with one or more of several strains of HPV. Human Papillomavirus is a member of the papillomavirus family, with over 100

different strains. 13 high risk strains have been identified, which contribute to squamous cell carcinoma of the cervix, anus and oropharynx (Crosbie et al.). HPV is transmitted sexually, and it is very common. The prevalence in the United States is highest in women aged 20-24 years old at 44.8%, and it tapers to 19.6% in women aged 50-59 years old (Dunne Ef & et al., 2007). The strains considered high risk are those that have been associated with increased risk for development of squamous cell carcinoma. Strains 16 and 18 are considered the highest risk as they are associated with approximately 70% of cervical cancers (Dunne Ef & et al., 2007). Other strains are associated with anogenital warts, and many strains are asymptomatic.

Regardless of strain of HPV, the vast majority of infections are cleared by the immune system. Over 90% of HPV infections are cleared within 2 years of infection, mostly asymptotically. Persistent, non-clearing infection is associated with a higher risk of cervical dysplasia. Infection with a high risk strain of HPV (notably 16 or 18) and immunodeficiency are the two best-studied risk factors for failure to spontaneously clear HPV and progression from HPV infection to cervical dysplasia (Dunne Ef & et al., 2007).

HPV infection is most likely introduced into epithelial cells through passage into the basal cells that represent the base layer of the cervical tissue. The surface of the cervix is lined with squamous epithelium, and the endocervical canal is lined with columnar epithelium. The cervical surface meets the endocervical canal surface at the

transition zone or the squamo-columnar junction. At this junction the vulnerable basal cells are closer to the surface, where the HPV virus can enter and infect the cells.

One theory for why HPV may be more prevalent in Haiti relates to microabrasions, or very small scrapes or cuts. These are usually created by friction during sexual activity, which may be more common due to certain cultural practices in the Haitian population. Many Haitian women engage in a practice called “dry sex” where they place leaves in the vagina to induce dryness with the goal of increasing friction for the male partner’s benefit during sexual activity (Halperin, 1999). This practice may contribute to the incidence of microabrasions and therefore may play a role in HPV transmission. The body’s wound healing process at the site of a microabrasion involves lateral spreading of basal epithelial cells, and it is during this process of cellular generation and spread that the virus gains entry (Woodman, Collins, & Young, 2007).

Once it has entered the host, the HPV virus begins gene expression and viral activity in two phases. The virus initially expresses early genes E1, E2, E4, E5, E6 and E7 which enable replication using episomal DNA. Next the late genes L1, L2 and E4 are expressed, allowing the infected cells to proliferate (Woodman et al., 2007). The late genes enable viral shedding and further infection. The genes E6 and E7 encode special viral proteins that promote host cell viability and prolong the cell’s ability to allow viral replication and avoid apoptosis, or programmed cell death (Crosbie et al.). Both E6 and

E7 reduce the cell's ability to present cell surface markers of infection that would normally stimulate the immune system to mount an inflammatory response to the infected cells (Crosbie et al.). Without that response, the virus can propagate undetected for prolonged periods of time.

Once infection with a high risk strain of HPV has been established, the infected cells of the cervix begin to show detectable morphologic changes. These cells can be visualized using brush or spatula collection and microscopy, and the degree of dysplasia is classified as undetermined, low grade, or high grade according to the Bethesda system of classification (Solomon D & et al., 2002).

1.1.3 Screening and Treatment

Screening for the premalignant cervical changes described above has been the mainstay of cervical cancer prevention over the last several decades. The advent and implementation of vaccines against the highest risk strains of HPV will likely drastically change the landscape of cervical cancer in areas where the vaccines are widely used. At this time, however, secondary prevention of cervical cancer is still very important, especially in underserved areas where vaccine enactment is not yet feasible.

Cervical screening has traditionally been done via the Papanicolaou (Pap) smear, which involves cervical cells collected on a spatula or brush and directly transferred to a slide for microscopy. In developed countries, the alternative liquid-based cytology

(LBC) technique has largely taken over as the primary method of sampling. This technique involves sampling the cervix transformational zone with a brush, then suspending the collected cells in a liquid medium, which allows the removal of obscuring debris, including blood, discharge and inflammatory cells from the sample before viewing (Arbyn, 2008). The gold standard for grading dysplasia is visualization with colposcopy and biopsy. HPV testing has become an important part of this process as well. In developed countries, the common approach is initial screening with cervical sampling either by Pap or LBC techniques; both are considered acceptable by the American College of Obstetrics and Gynecology (ACOG, 2012). Any sample that shows atypical squamous cells of undetermined significance (ASCUS) should be tested for HPV to stratify risk based on HPV strain, and HPV testing may also be used as a co-test method with cervical sampling in women over 30 years old (ACOG, 2012). Samples that show an oncogenic HPV strain are then followed with colposcopy and biopsy as needed.

Over the last several years the presumed superior sensitivity and specificity of the liquid-based technique has come into question. The benefits of liquid suspension prior to visualization include removal of obscuring debris such as blood, discharge, inflammatory cells or other foreign matter including lubricant or douche material (ACOG, 2012). The other major benefit of liquid-based cytology is the residual cellular material not used for cytology, which can be used for reflex HPV testing or other testing

as needed (Siebers Ag & et al., 2009). The first issue, that of whether LBC technology truly improves sensitivity and specificity of testing over the Pap smear, has been questioned. A meta-analysis of papers published between 1991 and 2007 comparing Pap and LBC test characteristics against gold standard colposcopy and biopsy results showed no statistically significant difference in sensitivity between the methods. The only difference found was that the LBC technique had a slightly lower specificity with the outcome measure of ASCUS, and this difference was not considered to be clinically significant (Arbyn, 2008). A 2009 Dutch study reporting a cluster randomized controlled trial involving approximately 90,000 women compared Pap smear and LBC techniques with colposcopy/biopsy corroboration to assess positive predictive value and diagnostic agreement with gold standard. The results showed no statistically significant difference between Pap and LBC techniques (Siebers Ag & et al., 2009).

This controversy opens the door for continued investigation of these two methods of cervical sampling and it is particularly relevant in areas like Haiti where high degrees of obscuring inflammation have been observed. This paper aims to continue this debate in the setting of a developing country.

1.2 Fieldwork Site

Data were collected at two clinic sites in Haiti. The sites are Blanchard, a residential community near Port-au-Prince, and Leogane. The clinics are operated by

Family Health Ministries, a nonprofit organization dedicated to cooperative, community-centered efforts to improve the health of women and families in Haiti. The organization has been active in Haiti for over 20 years, supporting cervical cancer prevention and education as well as expanding to other forms of health support (FHM, 2013).



Figure 1: Map of Haiti (Online, 2013)

Cervical cancer represents a large portion of the cancer burden of women in Haiti, making this site appropriate for this study. Family Health Ministries has a legacy of community involvement and a reputation for providing excellent care, enabling convenience sampling and high quality sample collection.

1.3 Study Design

1.3.1 Hypothesis

This study hypothesizes that there is greater agreement between liquid-based cytology and biopsy in detection of cervical dysplasia, especially in the presence of inflammation, than the level of agreement between pap smears and biopsy. This leads to a positive relationship between increased use of LBC testing over Pap smear testing and increased detection of cervical dysplasia among Haitian women.

1.3.2 Selection of Participants

The study population was selected from female patients of the Family Health Ministries clinic in Haiti. Patients were recruited through local radio announcements and word of mouth. Family Health Ministries enrolled patients at clinics located in the towns of Leogane and Blanchard. Women were deemed eligible for the study if they were willing to receive continuing gynecological care and undergo cervical screening by both study methods at the clinic. All participants were between 25 and 60 years old, had at least one lifetime sexual partner, were not pregnant, had not had a hysterectomy, and were not menstruating at the time of sample collection. Participants were recruited via non-probability convenience sampling, and the study population was predominantly made up of ethnically Haitian women, and all participants were Haitian residents.

1.3.3 Measures and Data Collection

1.3.3.1 Measures

The primary measure was degree of cervical dysplasia on the two different screening tests, the traditional pap and liquid-based cytology. The measures of cervical dysplasia were based on the Bethesda system of pathological grading: normal, ASC-US (atypical squamous cells of undetermined significance), LSIL (low grade squamous intraepithelial lesion), HSIL (high grade squamous intraepithelial lesion), and squamous cell carcinoma (Solomon D & et al., 2002). The dysplasia variable was dichotomized during analysis to “yes” including LSIL, HSIL, and cancer, and “no” including none and ASCUS.

Biopsy results include measures on the same scale for dysplasia. Biopsy results are included only for those women who needed biopsy based on their screening test results.

Other measures include presence of inflammation on pap, LBC and biopsy as well as demographics including patient age.

1.3.3.2 Data Collection Methods

Cervical swabs were collected from female patients receiving gynecological care at the Family Health Ministries clinics. Study participants had cervical cells collected during routine gynecological screening, and swabs were done for both Pap smear

testing and liquid-based cytology (LBC) testing for each patient at the same visit. The samples were analyzed at a certified US pathology lab. Results testing positive for dysplasia on either screening test were followed by biopsy to evaluate the degree of advancement of disease, and those patients were treated accordingly.

Samples testing positive for dysplasia were also reflexively tested for HPV DNA to distinguish precancerous changes from other causes of atypical cytology that can be read as dysplasia (usually cellular abnormalities caused by vaginitis/vaginosis). Pathology results from both sample methods were communicated to the patient in follow-up.

Patients who were symptomatic of non-HPV vaginal infection either by patient complaint or physician impression were treated appropriately and had cervical samples collected approximately one week later. Therefore samples with inflammation noted had either clinically unrecognized inflammation or residual inflammation following treatment.

HPV DNA and biopsy analyses were completed along with comparison of Pap and LBC methods of detecting precancerous dysplasia for samples that show dysplasia on either testing method. This corroboration was used to determine whether either or both tests were consistently returning false positives for dysplasia due to other

causes of cell abnormalities, primarily vaginitis. Demographic and clinical data were collected for each patient.

1.3.3.3 Bias Control in Design

This dual-sample method of collecting two concurrent cervical swabs allows direct comparison of samples for each patient. This method also helps to avoid measurement error, as both samples are collected at the same time, by the same physician. Sampling in the medical clinic and conducting pathology analysis in a certified lab ensures a high level of quality of collection and analysis. Screening as part of regular treatment adds no additional time or financial burden for patients participating in the study.

The convenience method of sampling does introduce a level of bias, though this was felt to be the most feasible method of recruitment.

1.3.4 Analysis

Data analysis was conducted using Stata version 11.0. The agreement statistic used is kappa, which allows comparison of results from two different observers, or test methods in this case (Anthony J. Viera, 2005). The results below show the percent agreement between two test methods as well as the “expected agreement” or level of agreement that would be expected by chance alone. The kappa statistic reports how meaningful that percent agreement is given the expected level of agreement. A kappa

value of zero indicates agreement that is no better than chance, while a kappa value approaching 1.0 indicates almost perfect agreement. The table below shows commonly used benchmarks for interpretation of kappa agreement levels as described in a 1977 paper by Richard Landis and Gary Koch (Landis & Koch, 1977).

Table 1: Kappa Agreement Strength Benchmarks (Landis & Koch, 1977)

Kappa statistic value range	Agreement strength
<0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

95% confidence intervals are calculated for the kappa statistic using a Stata add-on program (Reichenheim, 2004).

2. Results and Interpretation

2.1 Screening test comparisons (Pap vs LBC)

2.1.1 Dysplasia characteristics of entire study sample

Table 2 represents the study samples broken down by study type (LBC or Pap) and by the presence of dysplasia on that study. This table includes all samples that were collected, which includes several where results were only available for one study type; therefore the total numbers of Pap and LBC samples do not match. This table investigates how each test individually detects dysplasia, both with and without the presence of inflammation. Dysplasia on a sample was dichotomized to “no” if the result was zero, ASCUS, or LSIL, and to “yes” if the result was HSIL or carcinoma. A second dichotomization was performed for each study type after those samples showing obscuring inflammation were removed. The percentage of samples showing dysplasia was calculated for each study type, both with and without the inflammation samples.

Table 2: Study sample dichotomized by presence or absence of dysplasia, reported with and without those samples with inflammation.

	Dysplasia present?		
	No	Yes	Percentage with dysplasia
All Pap samples	298	70	19.0%
Pap samples without inflammation	32	6	15.7%
All LBC samples	365	78	17.6%
LBC samples without inflammation	307	73	19.2%

The results from this table show that both Pap and LBC testing methods appear to detect dysplasia in a similar percentage of studies. Both tests detected dysplasia in 17-19% of samples when inflammation was not excluded. When those samples with inflammation were excluded, the two tests did not appreciably change their detection rate of dysplasia; Pap samples appeared to detect slightly less dysplasia when inflammation was removed, though by a very small amount. LBC samples appeared to detect slightly more dysplasia when inflammation was removed, though this is also a very small change.

The striking difference in this comparison is that obscuring inflammation was found in 90% of Pap samples, and in only 16% of LBC samples. This observation highlights the ability of the LBC technique to removing obscuring cells, though this comparison does not prove a benefit to that ability as both tests appear to detect dysplasia at similar rates. However, the very small size of samples remaining when inflammation is removed from Pap samples make these results somewhat difficult to interpret. This is explored further in the discussion section below.

2.1.2 Comparison of dysplasia characteristics among all dual-study samples

Tables 3 and 4 include all samples where results for both Pap and LBC tests were available. This comprises 303 total samples. Table 3 shows all samples stratified by dysplasia level as read by the pathologist, and Table 4 shows the same information after dichotomization into presence or absence of dysplasia. Dysplasia was categorized as

“no” for NILM, ASCUS, and LSIL, while “yes” was HSIL or cancer. The reading used was the highest grade given for a sample in the case of multiple interpretations. For example, a sample that could be read as LSIL or HSIL based on pathological characteristics would be reported as HSIL. A consistent standard was used in order to standardize the readings, and the highest reading was chosen to increase the specificity of the detection method (see discussion section). These tables allow comparison of how frequently dysplasia is detected by each test, and Table 4 combines those data into a 2x2 table for comparison.

Table 3: Study samples with dual results

All samples with full results		LBC cytology read				
		Negative	ASCUS	LSIL	HSIL	Total
Pap cytology read	Negative	132	27	17	6	182
	ASCUS	19	6	2	1	28
	LSIL	15	8	3	11	37
	HSIL	7	3	5	24	39
	Cancer	0	0	0	17	17
	Total	173	44	27	59	303

Table 4: Study samples with dual results dichotomized by dysplasia presence

All samples with full results		Dysplasia N/Y on LBC		Total
		No	Yes	
Dysplasia N/Y on Pap	No	229	18	247
	Yes	15	41	56
	Total	244	59	303

This table shows that among all samples with results for both Pap and LBC tests (303 total) there was dysplasia detected by both tests on 41 of them, and 229 were read as negative by both tests. 33 tests were interpreted differently by the two test methods.

Calculated agreement statistics are reported in Table 6 below.

2.1.3 Comparison of dysplasia characteristics among dual-study samples without inflammation

Table 5 shows the same setup as table 4, comparing dysplasia characteristics of samples that have results for both test types. The difference is that table 5 has only samples with no inflammation. This leaves just 33 samples that were clear of inflammation on both Pap and LBC testing.

Table 5: Samples with dual results without inflammation dichotomized by dysplasia presence

All inflammation removed		Dysplasia N/Y on LBC		Total
		No	Yes	
Dysplasia N/Y on pap	No	26	1	27
	Yes	2	6	6
	Total	26	7	33

The results in this table indicate that when the samples with inflammation are removed, the LBC and the Pap samples agree much more closely; only 3 samples of the 33 included in this table were read differently between the two tests. This level of agreement is higher than when all samples (including those with inflammation) are included. This is explored further in Table 6.

2.1.4 Agreement statistics for screening test comparisons

Table 6 shows the calculated agreement of test results for the Pap and LBC test methods, both with and without the inflammation samples included.

Table 6: Kappa agreement statistics for comparison of Pap and LBC tests with and without inflammation

	Agreement	Expected agreement	Kappa	Standard error	Z	95% CI
Pap vs LBC (all samples)	89.11%	69.24%	0.6459	0.0574	11.25	(0.535-0.757)
Pap vs LBC (without inflammation)	96.97%	68.32%	0.9043	0.1733	5.22	(0.721-1.0)

The first row shows that among all samples, the Pap and LBC tests shows 89.11% agreement with a kappa of 0.6459. The level of agreement increased to 96.97% with a kappa of 0.9043 when samples with inflammation were removed. This level of agreement is validated by a kappa approaching 1 and a small 95% confidence interval. The agreement appears more significant with a kappa closer to 1 when inflammation is removed, indicating that in the absence of inflammation, the two test methods agree very closely. This aligns with expectations and indicates that inflammation may in fact play a statistically significant role in the screening tests., despite the inconclusive results in table 2.

This table shows the level of agreement between the two screening tests, and the next section explores how each screening test compares to the gold standard of cervical biopsy.

2.2 Screening tests compared to biopsy results

2.2.1 Comparison of Pap results to biopsy results

Table 7: Pap vs biopsy

All samples		Dysplasia N/Y on biopsy		
		No	Yes	Total
Dysplasia N/Y on Pap	No	35	5	40
	Yes	4	15	19
	Total	39	20	59

Table 8: Agreement statistics for Pap compared to biopsy results

	Agreement	Expected agreement	Kappa	Standard error	Z	95% CI
Pap vs biopsy (all samples)	84.75%	55.73%	0.6554	0.1301	5.04	(0.450- 0.861)
Pap vs biopsy (without inflammation)	<i>Unable to calculate (4 total samples)</i>					

Tables 7 and 8 show the samples that have results for both Pap testing and biopsy pathology reading. Table 7 details whether dysplasia was detected by either method for each sample, and Table 8 shows the agreement between the Pap screening test for dysplasia and the biopsy gold standard measure. Table 8 shows a 84.75% agreement between the two measures, with a kappa of 0.6554. This kappa is corroborated by a 95% CI which is wider than that comparing Pap and LBC, indicating

that the screening tests agree more closely with each other than the Pap test does with the diagnostic test (biopsy).

2.2.2 Comparison of Liquid-Based Cytology results to biopsy results

Table 9: LBC vs biopsy

All samples		Dysplasia N/Y on biopsy		
		No	Yes	Total
Dysplasia N/Y on LBC	No	25	3	28
	Yes	6	13	19
	Total	31	16	47

Table 10: Agreement statistics for LBC compared to biopsy results

	Agreement	Expected agreement	Kappa	Standard error	Z	95% CI
LBC vs biopsy (all samples)	80.85%	53.06%	0.5921	0.1445	4.10	(0.356-0.828)
LBC vs biopsy (without inflammation, n=43)	81.40%	53.16%	0.6028	0.1517	3.97	(0.357-0.848)

Tables 9 and 10 show the degree to which LBC testing agrees with the gold standard of biopsy. Table 9 shows that among the 47 samples where there were results for both LBC screening and gold standard biopsy, 38 out of the 47 samples were read the

same way, while 9 of the samples were interpreted differently for presence or absence of dysplasia.

Table 10 shows the calculated agreement between the LBC samples and the biopsy results. The LBC screening test had 80.85% agreement with biopsy results among all samples and a kappa value of 0.5921 indicating weaker agreement than the Pap vs biopsy comparison (84.75% agreement with kappa 0.6554).

2.2.3 Summary comparison of both screening tests to biopsy

Table 11 shows the agreement statistics for each type of test comparison, summarizing the previous tables. The results include comparisons of the two screening tests to each other (rows 1 and 2), of the Pap test to biopsy (rows 3 and 4) and of the LBC test to biopsy (rows 5 and 6). This table includes subsets of samples with inflammation removed.

Table 11: Agreement comparisons for all test types

	Agreement	Expected agreement	Difference between actual and expected agreement	Kappa	Standard error	Z	95% CI
Pap vs LBC (all samples)	89.11%	69.24%	19.87	0.6459	0.0574	11.25	(0.535-0.757)
Pap vs LBC (without inflammation)	96.97%	68.32%	28.65	0.9043	0.1733	5.22	(0.721-1.0)
Pap vs biopsy (all samples)	84.75%	55.73%	29.02	0.6554	0.1301	5.04	(0.450-0.861)
Pap vs biopsy (without inflammation)	<i>Unable to calculate</i>						
LBC vs biopsy (all samples)	80.85%	53.06%	27.79	0.5921	0.1445	4.10	(0.356-0.828)
LBC vs biopsy (without inflammation)	81.40%	53.16%	28.24	0.6028	0.1517	3.97	(0.357-0.848)

The results of comparisons that include all samples are more useful in this comparison given the small starting sample size. The screening tests agree with each other 89.11% of the time, which increased to 96.97% of the time when inflammation is removed. The Pap test agrees with subsequent biopsy results 84.75% of the time, while the LBC test agrees with the biopsy results 80.85% of the time. Table 11 also allows for

comparison of the strength of the kappa values for each comparison using the 95% confidence intervals.

One key finding discussed above is that the two screening tests achieve almost perfect agreement based on Table 1 cutoffs when the samples with inflammation are removed, with a 95% confidence interval approaching 1.0. Table 2 appears to indicate that neither test's ability to detect dysplasia is particularly affected by the presence of inflammation, though this table does not compare the test agreement directly, just the percentage that each test called dysplasia. When the tests are compared on sample-by-sample agreement, the presence of inflammation does in fact appear to alter the ability of the tests to detect dysplasia in *the same* samples. When inflammation is removed, the tests agree 96% of the time. One weakness of this study is that there are not enough study samples to perform a meaningful comparison of each screening test to biopsy results on samples with no inflammation.

When comparing each screening test to biopsy results, the key comparison in this table is row 3, Pap vs biopsy results among all samples. This result shows 84.75% agreement with a kappa of 0.6554, achieving *substantial* agreement (Table 1), with a 95% CI of (0.450-0.861). The second comparison is row 5, where the LBC test is compared to biopsy results, achieving 80.85% agreement with a kappa of 0.5921, or just *moderate* agreement. That kappa has a 95% CI of (0.356-0.828) which indicates that the kappa values are similar in strength. This comparison indicates that for this sample set

the Pap test agrees more strongly with the gold standard biopsy results among all samples, and the strength of that agreement is stronger based on the kappa in the setting of similar 95% confidence intervals.

3. Discussion

3.1 Discussion of results

The goal of this project was to investigate the hypothesis that there is greater agreement between the LBC screening test and biopsy results than between the Pap screening test and biopsy results. The study included analysis of the presence of inflammation in cervical samples and what role that may play.

The screening tests were compared to each other (Table 11) and, among all samples, they agree 89.11% of the time. Each test was compared to the gold standard of biopsy, and the LBC test agreed 80.85% of the time while the Pap agreed 84.75% of the time. These results indicate that among all samples, including those with inflammation, the Pap test agrees more often with the biopsy results. These results show that the Pap test is the more specific test for a cervical dysplasia screening tool in this population.

The screening tests were also evaluated for the influence of inflammation on the results (Table 2). Neither test's ability to detect dysplasia (without a comparison method) changed significantly in the presence or absence of inflammation. However, the presence of inflammation did affect the *agreement* of the tests when detecting dysplasia in matched samples (Table 6). These results indicate that the presence of inflammation affects the tests' ability to match results.

The conclusion that inflammation affects the screening tests' agreement with each other begs the question of how inflammation could affect those results. The most straightforward hypothesis as disused in the introduction is that perhaps inflammatory cells obscure the cervical cells on a Pap smear in a way that can be avoided on LBC. However, these data show that the Pap test agrees more often with the biopsy result than does the LBC test. These data do not offer a clear explanation for the difference in agreement.

Table 11 also includes results of agreement statistics for comparisons when inflammation was removed, though this is difficult to interpret in a meaningful way when comparing to biopsy results because there were just 4 samples that had results for Pap testing and biopsy that also had zero inflammation. The comparison is meaningful for rows 1 and 2, which demonstrate that the two screening tests agree with each other 96.97% of the time when inflammation is not present, which is significantly higher than the 89.11% agreement among all samples. This result speaks to the effect of inflammation on the sensitivity of the Pap test.

3.2 Discussion of possible clinical implications

The clinical importance of an analysis like this stems from its possible contribution to confirming or amending cervical cancer screening protocols. The larger goal of this study is to help determine the best practices for cervical cancer screening in this population. The results of this analysis show that the Pap screening test appears to be more specific for cervical dysplasia among all samples, and the sensitivity of both screening tests are similar among samples without cervical inflammation. The Pap test showed greater specificity for dysplasia when compared to cervical biopsy as a gold standard measure.

Choosing an appropriate screening test means balancing sensitivity and specificity, and the results of this study do not make that choice clear. In a population with very low levels of inflammation, the Pap test would be superior given its higher specificity in all samples and its almost equivalent sensitivity (when compared to LBC) in the setting of zero inflammation. These findings, however, are very limited by the small number of samples collected by Pap that did not have inflammation present. This study population has a very high level of cervical inflammation (one reason that sample is so small), and a larger sample size could shed light on this issue. This study does not clearly elucidate the effect of inflammation on either test, though it does suggest that inflammation plays a role in test interpretation.

Choosing the right screening test is important as the potential for false positive results may lead to unnecessary intervention that can be costly financially, physically and emotionally, and false negative results can be devastating when the window of opportunity for treatment is missed. A key in this population is detecting as many patients as possible who might benefit from early intervention, especially among women with little access to resources for advanced-stage cancer care, while avoiding costly and unnecessary treatment in women with low-risk lesions who are likely to clear the HPV infection without treatment. The sample sizes in this study are small enough to make that decision difficult.

Another aspect of real-world implementation of screening protocols is cost. The Liquid-Based Cytology test requires test-specific sample collection equipment and pathology interpretation equipment, whereas the Pap test requires less expensive collection materials and a light microscope. These costs may make a significant difference in some settings, and may not make a difference in others.

A third important aspect of choosing a testing method is feasibility. In resource-poor settings, finding a clinician with the necessary time and expertise to interpret the results is important. LBC sample collected, storage, transport and reading is costly and requires a supply chain that is not present in many rural areas. This aspect of testing may also make a difference in some settings and may not in others.

3.3 Limitations

There were several limitations encountered during this study. The most significant was the devastating earthquake that hit Haiti in January of 2010. This study was ongoing at the time, and data collection was suspended during some rebuilding. The Leogane clinic was destroyed, and the clinicians and patients involved in the study had extremely serious personal concerns to attend to.



Figure 2: Leogane clinic wall three months after the earthquake (Photo by Genevieve Wolpert)

A consequence of the earthquake was a small sample size for this analysis. A large proportion of samples had inflammation, which was expected, but the size of the cohort was not large enough to generate significant numbers for comparison when

inflammation samples were removed, especially among the samples with biopsy results. Future analyses would be more meaningful with a larger sample size.

3.4 Ethical Considerations

The primary ethical consideration during conduction of this study was maintaining cultural sensitivity and awareness of our study population. Every effort was made to ensure that participants were respected in terms of cultural sensitivity, privacy, and adherence to medical standards. Participation in the study was completely voluntary, and clinic patients who declined to participate received appropriate gynecologic care without participation.

Gynecological examination and treatment can be a socially taboo subject for open discussion, and sexually transmitted diseases such as HPV carry a stigma in society. It is therefore necessary to recruit patients in a culturally sensitive way. Education campaigns in the region have spread awareness and understanding of HPV and cervical cancer, and past experience of the researchers has shown that Haitian women who understand the link between HPV and cervical cancer are willing to undergo testing. The testing is always done with a female Family Health Ministries staff member present. Family Health Ministries provided treatment identified as necessary during the study.

In Haitian society women are the primary supporters of the family, and study participants must spend time and money to travel to and from the clinic. This

burden may have proven to be an insurmountable barrier for some eligible women, and while that problem was recognized, the investigators were unable to avoid this element of convenience sampling. There is an element of selection bias due to this financial limitation.

3.5 Lessons Learned

Working with Haitian women in a post-disaster environment was extremely educational. This project expanded my views on public health and the importance of working with underserved populations. The women with whom I worked made their own health a priority in situations where that is extremely difficult, and it was a true privilege to work alongside them. The experience inspired me to pursue a career in the field of women's health, and beyond that I feel a deep sense of commitment to working to decrease barriers to care. This project brought the issues of access to basic care into focus for me.

3.6 Implications for Future Research

As discussed above, directions for future research could include larger sample sizes for more significant analysis, as well as a location-specific cost/benefit analysis for the screening test options.

Given the controversy over cervical sampling methods and the recent increase in availability as well as sensitivity and specificity of HPV testing, screening in the developing world has trended toward HPV testing as the first line screen for cervical

cancer. In regions where this is feasible, HPV testing is used as a primary screening tool, followed by colposcopy and biopsy as necessary. An advantage to this approach is the necessary link between oncologic HPV strains and cervical cancer, so testing for HPV may allow clinicians to bypass the pathology read involved in Pap or LBC testing. The future of cervical cancer screening may move away from the screening tests discussed in this analysis. Future research will involve discussion of how these screening tests compare to the sensitivity and specificity of HPV testing.

3.7 Conclusion

The results of this study demonstrate that the presence of inflammation affects the two screening tests' ability to agree on sample results, though the mechanism remains unclear. In this analysis the Pap screening test is a more specific test for cervical dysplasia in all samples including those with inflammation. The sensitivity of both screening tests appear similar among samples without cervical inflammation. The Pap test showed greater agreement with cervical biopsy as a gold standard measure for detection of cervical dysplasia than did the LBC test, though this conclusion is limited by small sample size. In this analysis both the Pap and the LBC tests performed consistently with or without the presence of cervical inflammation. Further research with larger sample sizes would lend more statistical significance to these results, and more investigation into costs and feasibility of each method would be instructive. Addition of a standardized screening method to HPV screening as a first line screening

tool in this population will advance knowledge on the subject and lead to an advancement in the standard of care for women in this study population.

References

- ACOG, A. C. o. O. a. G. (2012). Practice Bulletin No.131: Screening for cervical cancer. *Obstetrics & Gynecology*, 120, 1222-1238.
- Anthony J. Viera, J. M. G. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, 37(5), 360-363.
- Arbyn, M. B., Christine; Klinkhamer, Paul; Martin-Hirsch, Pierre; Siebers, Albertus; Bulten, Johan. (2008). Liquid Compared With Conventional Cervical Cytology: A Systematic Review and Meta-analysis. [Review]. *Obstetrics & Gynecology*, 11(1), 167-177.
- Crosbie, E. J., Einstein, M. H., Franceschi, S., & Kitchener, H. C. Human papillomavirus and cervical cancer. *The Lancet*(0). doi: [http://dx.doi.org/10.1016/S0140-6736\(13\)60022-7](http://dx.doi.org/10.1016/S0140-6736(13)60022-7)
- Dunne Ef, U. E. R. S. M., & et al. (2007). PRevalence of hpv infection among females in the united states. *JAMA*, 297(8), 813-819. doi: 10.1001/jama.297.8.813
- FHM. (2013, 5/23/13). Family Health Ministries Retrieved 7/17/13, 2013, from <http://familyhm.org/>
- Halperin, D. T. (1999). Dry sex practices and HIV infection in the Dominican Republic and Haiti. [Journal]. *Sexually Transmitted Diseases*, 75(6), 445-446.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. doi: 10.2307/2529310
- Online, O. W. N. (2013). Haiti Map.
- Reichenheim, M. E. (2004). Confidence intervals for the kappa statistic. *The Stata Journal*, 4(4), 421-428.
- Scarinci, I. C., Garcia, F. A. R., Kobetz, E., Partridge, E. E., Brandt, H. M., Bell, M. C., . . . Castle, P. E. (2010). Cervical cancer prevention. *Cancer*, 116(11), 2531-2542. doi: 10.1002/cncr.25065
- Siebers Ag, K. P. M. G. J. M., & et al. (2009). Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: A randomized controlled trial. *JAMA*, 302(16), 1757-1764. doi: 10.1001/jama.2009.1569

- Solomon D, D. D. K. R., & et al. (2002). The 2001 Bethesda system: Terminology for reporting results of cervical cytology. *JAMA*, 287(16), 2114-2119. doi: 10.1001/jama.287.16.2114
- WHO/ICO. (2010a). Human Papillomavirus and Related Cancers in Haiti: Summary Report 2010: WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre).
- WHO/ICO. (2010b). Human Papillomavirus and Related Cancers in United States of America: Summary Report 2010: WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre).
- WHO/ICO. (2010c). Human Papillomavirus and Related Cancers in World: Summary Report 2010: WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre).
- Woodman, C. B. J., Collins, S. I., & Young, L. S. (2007). The natural history of cervical HPV infection: unresolved issues. [10.1038/nrc2050]. *Nat Rev Cancer*, 7(1), 11-22. doi: http://www.nature.com/nrc/journal/v7/n1/supinfo/nrc2050_S1.html