Spring 5-2005

Measuring inter-rater agreement among pathologists in cancer screening studies

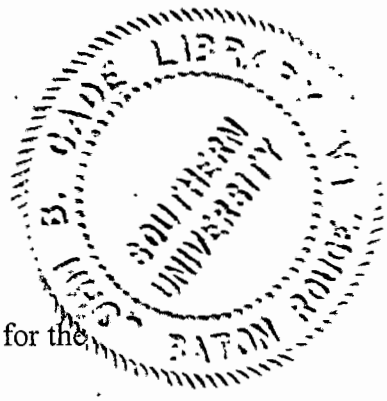
Krystle S. Oates

MEASURING INTER-RATER AGREEMENT AMONG PATHOLOGISTS IN CANCER SCREENING STUDIES

A THESIS

Presented to the

Honors College at Southern University
Baton Rouge, Louisiana



In Partial Fulfillment of the Requirements for the
Honors College Degree

By

Krystle S. Oates

May 2005

Honors College

Southern University and A and M College
Baton Rouge, Louisiana

CERTIFICATE OF APPROVAL

HONORS THESIS

This is to certify that the Honors Thesis of Krystle S. Oates
has been approved by the examination committee for the thesis requirement for the
Honors College degree in Mathematics

Lawrence A. Courellas
Advisor

Joyce W. Rourke
Chairman, Honors Advisory Committee

Dean, Honors College

ACKNOWLEDGMENT OF RESEARCH

This research was not supported by any local, state, regional, or national organization or agencies.

MEASURING INTER-RATER AGREEMENT AMONG PATHOLOGISTS IN CANCER SCREENING STUDIES

An abstract of a Thesis

Presented to the

Honors College at Southern University
Baton Rouge, Louisiana

In Partial Fulfillment of the Requirements for the
Honors College Degree

By

Krystle S. Oates

May 2005

ABSTRACT

Although cervical cancer is one of the most preventable forms of cancer in the United States, it still remains the leading cause of cancer mortality in El Salvador. This is attributed to poor cancer screening in El Salvador. The Pap smear test is the traditional screening method that pathologists use for the detection of cervical cancer in women, but the Thin Prep test is said to be an improvement to the Pap smear test. In this project, we will use the data from pathologists in the United States and El Salvador who analyzed the slides of 471 Salvadoran women who were given both the Pap smear and Thin Prep tests. With this data the Kappa statistic will be used to compare the level of agreement between the pathologists in the United States and pathologists in El Salvador. Their level of agreement will affect the outcome of which test, Thin Prep or Pap smear, is more sensitive in detecting cervical cancer. Finding results to this problem could help to terminate the growing number of women in El Salvador who are at risk or have cervical cancer.

AUTHOR'S ACKNOWLEDGMENTS

First and foremost I would like to thank God, for he is the source from whom all blessings flow. He is my strength, and I could not have made it through college without Him. Through Him I am able to end this chapter of my life successfully and begin a new one with expectations of even greater things to unfold. I thank my parents for being reflections of God's love and support. It is because of them that I have maintained the will to press on and strive towards the standards of excellence. I also thank my extended family members for their support throughout these important years.

I thank all the wonderful faculty members at Southern University who have been of great support and inspiration to me. I would especially like to thank Dr. Oswald D'Auvergne for taking time out to help me complete this project. I thank the staff of the LS-LAMP program who have helped me to flourish academically, and who have been another great support system at Southern University.

Last, but not least I thank the Southern University Honors College for providing me with this opportunity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I
BACKGROUND OF STUDY

Introduction

Invasive cervical cancer is abnormal cell growth on the surface of the cervix and beneath the lining of the cervix. It is one of the most preventable forms of cancer, and its mortality rate within the United States has decreased by 70% within the past 50 years. However, it still remains the leading cause of cancer mortality in Latin America. This factor is due to poor cancer screening in Latin America. Some aspects of poor cancer screening in Latin America include women at risk who are uninformed of the vitality of cancer screening, analysis of samples, sample collection processing and transportation, and informing patients of results and providing access to adequate treatment. One of the biggest atrocities in cancer screening is the insensitivity of the conventional Pap smear, which is commonly used. As a result, many women who have cancer or abnormal cell tissue are left undiagnosed. A way that this problem can be treated is to improve the quality of screening treatments that take place in Latin America. Not only could improved screening decrease the rate of mortality for cervical cancer in Latin America, but it could also be more cost effective. In developing countries such as El Salvador yearly screening is a rather costly procedure that requires many resources. If screening sensitivity is increased, and more women are diagnosed screening intervals could increase, which would minimize the cost.

To help terminate this serious problem, pathologists at the University of Wisconsin, Madison teamed up with pathologist in El Salvador. In 1996 health care promoters in Arcatao, El Salvador expressed concern to pathologists in its sister city in

Madison, Wisconsin about the growing number of women in El Salvador who were diagnosed with cervical cancer. As a pilot project over 150 women were screened through the Salvadoran system. The slides were then read by pathologists at the University of Wisconsin-Madison where it was discovered that the sample quality of the slides were very poor, and that the conventional Pap smear was not as sensitive as it was known to be. In order to improve sample quality cells were place in a liquid-based solution called Thin Prep, and were compared to cells obtained by Pap smears. In order to further the study, and prove the effectiveness of the Thin Prep solution, as opposed to the conventional Pap smear the institutional Review Board at the University of Wisconsin-Madison approved a screening study that took place in five rural communities in El Salvador between April of 1998 and May of 1999. Health care promoters from each of the five communities were trained in obtaining cervical cytology, and then returned to El Salvador where they conducted a health fair for cervical cancer screening. Conventional Pap smears were obtained from 471 women, and their residual cells were placed in the Thin Prep solution. Slides obtained from the Pap smears were analyzed by Salvadoran cytotechnologists and pathologists, while Thin Prep samples were being read at the Wisconsin State Lab of Hygiene. Then American pathologist and Salvadoran pathologist traded and Salvadoran pathologists read Thin Prep slides while Pap smear slides where being read at the Wisconsin State Lab of Hygiene. In order to eliminate any biases pathologists at the Wisconsin State Lab of Hygiene were blinded to the observations of Salvadoran pathologists, and Salvadoran pathologist blinded to US observations. During this process the pathologists in Wisconsin and El Salvador rated the cell tissue of the

women on a scale of one to five, one being no cancer and five being invasive cancer. Using the Kappa test for intraobserver variability the pathology readings from Wisconsin were compared to those of El Salvador in order to measure their level of agreement. A P-value of ≤ 0.5 defined the statistical significance. From the results that were found Wisconsin's Thin Prep analysis detected 5.8% LGSIL, 3.0% HGSIL, and .21%CIS. Their readings for Pap smear were 5.5% LGSIL, 0% HGSIL, and 0.4%CIS. In readings from El Salvador conventional Pap smears had readings of 7.7% LGSIL, 3.2% HGSIL, 0%CIS, and 0.6%CIS, and 0.6% invasive cancer were found using Thin Prep. The Kappa test showed the rate of cytological concordance was significantly higher when Thin Prep was used than when the Pap smear was used. The Kappa values between pathologists for Pap smear was 20.9% while the agreement level was 61.1% with Thin Prep [5].

Statement of the Problem

A way of comparing the accuracy of two tests is to measure the level of agreement between two separate groups in analyzing data using each test. If the results from both groups agree more with test 1 than test 2, then test 1 is generally deemed more accurate. In this project the slide results of 471 women in El Salvador who received both the Pap test and Thin Prep test were analyzed by pathologists in Madison, Wisconsin and El Salvador. The level of agreement between the two pathology groups in analyzing Pap smear slides and Thin Prep slides were tested using the Kappa statistic.

Our primary objective in this research project is to re-examine the inter-rater agreement between Thin Prep and Pap smear pathology ratings. This will allow us to investigate if Thin Prep is needed to improve the effectiveness of the conventional Pap

smear test for detecting cervical cancer. This study will in turn help to accommodate the need of sufficient treatment for cervical cancer to the women in El Salvador. We will also determine the statistical power of the two tests, the Kappa statistic and polychoric correlation, that are used to measure rater agreement.

Hypothesis

Through statistical analysis it is projected that the level of agreement between pathologists when analyzing slides is different when using Thin Prep than when using the conventional Pap smear.

CHAPTER II
REVIEW OF LITERATURE

Preventative measures for cervical cancer are being researched in order to reduce the risk of the disease among women living in underdeveloped countries such as El Salvador. One of the most current issues in cancer research is the reliability of the conventional Pap smear for the detection of cervical cancer. A liquid-based solution known as Thin Prep is said to enhance the quality of smears thereby improving the viability and detection of abnormal cells. This could help to identify the early stages of cervical cancer, and the stop the progression of the disease in women. Although Thin Prep seems to be a solution to the effectiveness of the Pap smear sufficient evidence has not yet proven this. Also, studies on the cost effectiveness of the Pap smear versus the Thin Prep need to be studied in order to come to the conclusion of which test is sufficient for terminating the spread of cervical cancer.

Cervical Cancer

The cervix is the lower part of the uterus that forms the canal that leads to the vagina. When abnormal cells and cell growth are present above and below the lining of the cervix, cancer is formed. Most cancers of the cervix undergo their first stage of cancer in squamos cells. Squamos cells are thin, flat cells that form on the surface of the cervix. When cervical cancer is in its first stage it is usually called Low Grade SIL. This stage refers to the early stages in the size, shape, and number of cells that form on the surface of the cervix [16]. Lesions that develop during this stage are called mild dysplasia, or cervical intraepithelial neoplasia 1 (CIN 1). When a number of precancerous cells are formed on the surface of the cervix it is called High-grade SIL. These cells have the

potential of invading deeper layers of the cervix, though it generally will not happen for many months. This stage of abnormal cell growth is also known as moderate or severe dysplasia, CIN 2 or CIN 3. However, when cells spread deeper into the cervix and to other organs of the body, invasive cancer is formed. Symptoms of cervical cancer are usually not evident until cervical cancer cells affect other tissue surrounding the cervix. When this occurs abnormal vaginal discharge or bleeding may occur, but these symptoms may be caused by other health problems as well. The diagnoses of these symptoms are generally received by getting a pelvic exam or a Pap test. [These tests] allow doctors to examine any changes or abnormalities in the cervix [16]. If a patient requires a more thorough examination for abnormalities in the cervix a coloscopy is used. With this treatment the doctor applies a vinegar-like solution to the cervix then uses an instrument called a coloscope (similar to a microscope) to magnify the view of the cervix. The cervix is then coated with an iodine solution to perform the Schiller test. This test contrasts healthy cells with cancerous cells. When this solution is placed on the cervix healthy cells turn brown, and infected cells turn yellow or white. For an even more thorough exam a process called conization is used. Conization allows pathologists to see if cancerous tissue has spread below the lining of the cervix. With this process a cone shaped sample of tissue is removed from the cervix for examination. Staging is a very important procedure that is used to detect whether or not the cancer has spread to other parts of the body. If the cancer has spread the doctor will usually x-ray the area, and examine other parts of the body as well. However, after the cancer is diagnosed it can usually be treated. The most common method of treatment for cervical cancer is surgery or radiation

therapy. When surgery is conducted on a cervical cancer patient abnormal tissue is removed from the cervix. For extensive surgery a women may need or make the decision to have a hysterectomy, depending on whether she wants to have children or not. With radiation therapy high-energy rays damage cancer cells, and stop them from growing. Treatments such as chemotherapy and biological therapy are also used to treat cervical cancer.

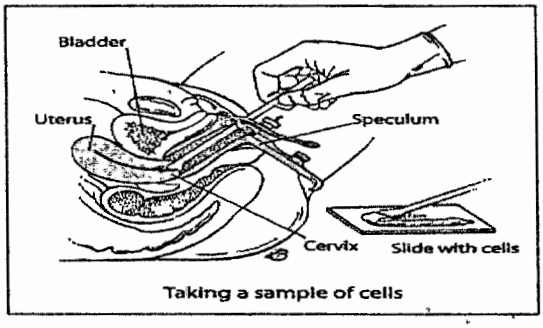
The cause and prevention for developing cervical cancer have been identified through studies conducted on women. From research studies it is shown that women who have sexual intercourse before the age of 18, and who engage in sex with a male who has had many sexual partners have an increased risk for developing cervical cancer. Smoking also increases the risks of developing cervical cancer. Studies have also suggested that women who use oral contraceptives such as birth control are also at risk for developing cervical cancer.

Papanicolaou Test

The Pap test is a widely known and commonly used test for detecting abnormalities in the reproductive systems of women. It is a test that every woman who is over the age of 18 or is sexually active is recommended to have once per year. The Pap test, also called a Pap smear, targets changes in the cells of the cervix. It can find cancer of the cervix before it moves to other parts of the body and becomes more invasive [15]. Pap tests can also pick up infections and inflammation that can eventually change into change into cancer cells [15]. The test is performed with an instrument called a speculum. Doctors use a speculum to open the vagina, and gently scrap cells off of the lining of the

cervix. These cells are then placed on a glass slide and sent to the lab for testing. The cells on the slide are checked for signs that they are changing from normal. Cells undergo a series of changes before they turn into cancer. If the cells on the slides are found to be normal then the Pap smear is said to be negative, but if the cells are abnormal the Pap smear is positive. However, the Pap smear tests are not always accurate. According to a recent meta-analysis the sensitivity of conventional PAP smears for detecting CIN ranges 30%-87% with a mean of 47% [10]. False positive and false negative results do occur, but the test is still reliable. Sometimes the tests may need to be redone because there were not enough cells on the slide [3].

Figure 1.Taking a Sample of Cells from the Cervix



Thin Prep Test

The Thin Prep test, also known as the Thin Prep Pap test, is performed similar to the conventional Pap test. Cells from the cervix are obtained in the same manner as the conventional Pap smear, but instead of being put directly onto a slide they are first placed in a liquid based solution that serves as a fluid preservative [4]. This solution is known to maintain the sample quality of the cells. As a result, the laboratory is able to produce a

higher quality slide for interpretation, thus providing a more accurate result. Although the Thin Prep Pap test is known to be the first improvement to the Pap smear in over 50 years, even under the best circumstances a very small number of abnormal Thin Prep Pap tests may be read as a false negative or a false positive[4][13].

Studies Comparing the Pap Test and Thin Prep Test

Lack of supporting evidence in favor of the Thin Prep test is so scarce that physicians are unable to make conclusions regarding the accuracy of the test. According to the Journal of Family Practice, "Evidence supports that Thin Prep is more sensitive than the conventional Pap smear at detecting cervical cancer, but there is insufficient evidence to recommend one preparation over the other [9]." Statements made about the wavering reputation of Thin Prep have stemmed from clinical trials involving women. A study involving 8636 women who had both tests performed on them reported that the Thin Prep test had higher sensitivity rates of 92.9% and 100% for detecting HGSIL and cancer, opposed to the 77.8% and 90.9% rates of the Pap test [9].

Another study comprised of 21, 752 patients compared the sensitivity and specificity rates of the Thin Prep test to the Pap smear test. The outcome of the study proved that although the sensitivity and specificity rates of the Thin Prep were higher than that of the Pap smear data was still insufficient to draw the conclusion that the Thin Prep was more efficient than the Pap smear. Sensitivity rates were 75% for the Thin Prep and 68% for the Pap smear [9]. However, the difference only met statistical significance in only two of the cases. The specificity for the Thin Prep was 86% while it was 79% for

the Pap smear, but the difference in percentage of the specificity for the two tests did not usually meet statistical significance [9].

Cervical Cancer Screening in Latin America

Although there is room for improvement for the conventional Pap smear in detecting cervical cancer, many other factors heavily impact the incidence and mortality rates for underdeveloped countries such as Latin America. A majority of this problem lies in inadequate health care coverage for poverty stricken countries. Evidence from inadequate screening procedures for cervical cancer in Latin America are shown in its mortality and incidence rates. Incidence rates of cervical cancer in Mexico are estimated to be the highest in the world, and the leading cause of death for Mexican women [8]. In order to decrease the burden of this problem many sources have suggested that more screening programs in these areas be implemented in order to target women who are at the highest risk for the development of cervical cancer [12]. Cervical cancer screening in Mexico does not facilitate even the majority of its women residents with Pap smears. Only 20% of the women who lived there are screened annually, and only 40% are screened over the three year period [8]. According to the Costa Rican National Health Authorities, "(Pap smears are recommended) every two years for women at low risk, and yearly for women at increased risk for cytological abnormalities [12]." Factors that deter many women in Latin America from getting Pap tests are the uneasiness they experience with the pelvic exam, the financial expenses, and the prolonged amount of time it takes to get tested and receive results. A study conducted by the Pan American Health Organization studied the benefits and setbacks of cancer screening in Latin America

through the feedback of patients. The main barriers that the participants identified were "availability and accessibility of quality services, facilities that lack comfort and privacy, costs, and courtesy of providers [2]." The most problematic barrier that prevent women in Latin America from seeking pap examinations are their lack of knowledge of sexual health. All but one of the participants in a sexual health study confessed that they did not have an adult in their lives to discuss sexual heath issues [12]. Many adult women also made claims that they were not aware of measures they could take to sustain their reproductive health [12]. However, effective measures for cervical cancer prevention in Latin America can be set into place if health service delivery was enhanced, and awareness was raised on sexual and reproductive health.

Cost Effectiveness of Cervical Cancer Screening

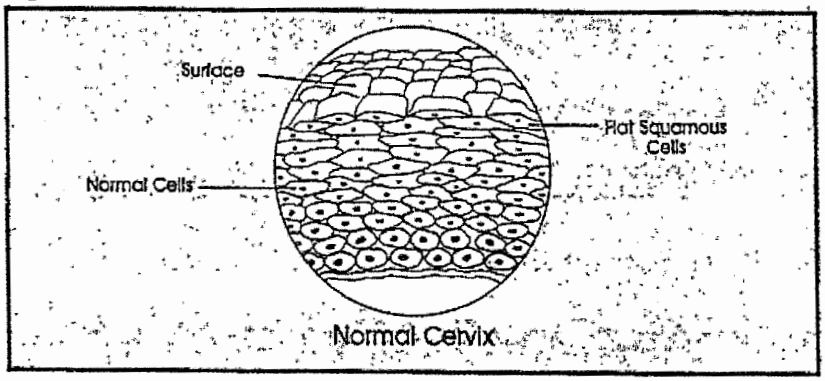
The cost effectiveness of a cancer screening test is very vital in its use in an underdeveloped country with limited financial resources. If a test has a high sensitivity rate, then intervals between screening examinations are able to increase because the chances of false-negative or false-positive results are decreased, thus enabling the patient to go longer without examinations [5]. This cost effective measure would allow more Latin American women the opportunity for screening than usual. According to the Journal of Family Practice, "The Thin Prep test is a cost effective screening tool if used at 3-year intervals [9]." However, a proper assessment has not yet been conducted to compare if the Thin Prep test is more economical than the Pap test.

CHAPTER III
METHODS SECTION

Pathologists' Procedures

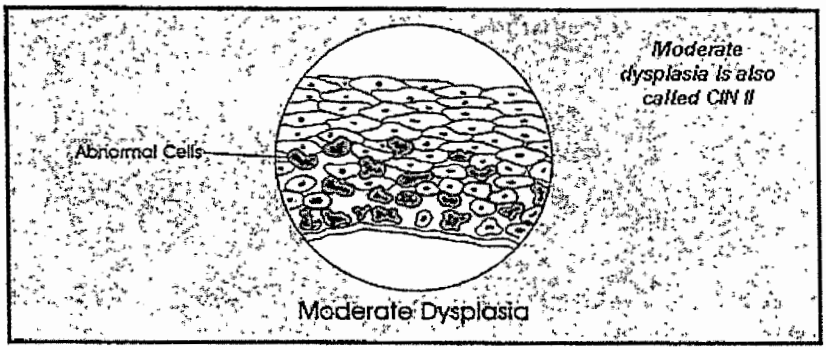
Both United States and Salvadoran pathologists analyzed each of the 471 Pap smear and Thin Prep slides. United States pathologists analyzed the 471 Pap smear slides while the Salvadoran pathologists analyzed 471 Thin Prep slides. The teams then switched slides and the United States pathologists analyzed the Thin Prep slides while the Salvadoran pathologists analyzed the Pap smear slides. Both teams of pathologists analyzed each Pap smear slide and Thin Prep slide separately and independently. The method they used in characterizing each slide was based on a rating scale of one to five, one being a normal cell tissue, and five being invasive cancer. An illustration of the slide ratings are shown below:

Figure 2.Normal Cervix



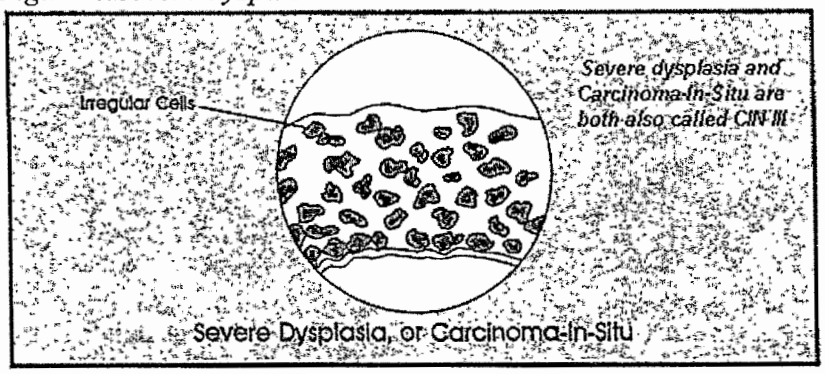
This is a slide of rating one where cells on the outer layer of the cervix are normal.

Figure 3.A cervix with moderate dysplasia



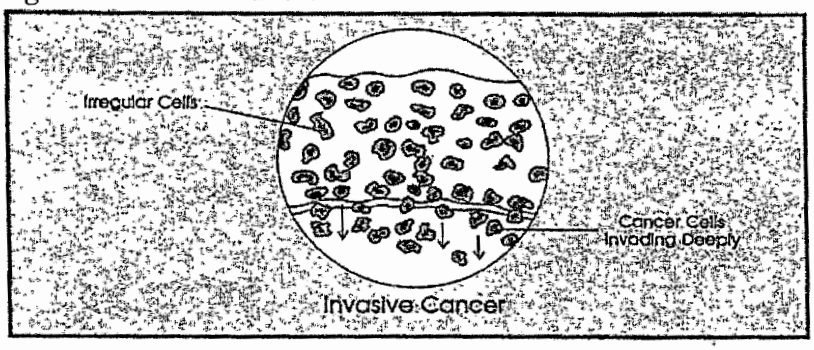
Moderate dysplasia is abnormal cells on the surface of the cervix. In this illustration, the brown cells represent abnormal cells. According to the pathologists' rating system, this slide would receive a rating of level three.

Figure 4.Severe dyspalsia



This slide illustrates severe dysplasia or carcinoma-in-situ where abnormal cells have spread throughout the surface of the cervix, but have not penetrated the lining of the cervix or spread to other parts of the body. According to the pathologists' rating system, this slide would receive a rating of level four.

Figure 5.Invasive Cancer



This slide illustrates invasive cancer where irregular cells have spread all throughout the lining of the cervix, and have deeply invaded below the surface of the cervix. According to the pathologists' rating systems, this slide would receive a rating of level five.

Tools

Calculations were computed utilizing a statistical software program called "R" version 1.9. With the use of computer programming, we were able to use "R" to calculate the kappa statistics, polychoric correlation, and the statistical power of both tests.

Experiment Procedures

In our analysis of agreement, the Pap smear and Thin Prep ratings were reduced from the five-category rating system into two categories: normal tissue vs. non-normal tissue.

Hypothesis Testing

One objective of this project is to compare the level of agreement between Thin Prep and Pap smear ratings. Hypothesis testing was conducted in order to prove whether the level of agreement between pathologists is different when using the Pap smear test than when using the Thin Prep test. In this case our null hypothesis (H_0) states that the level of agreement between pathologists in analyzing Pap smear and Thin Prep slides are the

Given a sample population that is rated on a fixed categorical scale, and a subject randomly selected denote the probability that two observer agree as

$$\prod_o = \sum \pi_{ii}.$$

Where the probability of classification in the ith category is by the first observer and the probability of classification in the jth category is by the second observer. Perfect agreement corresponds to $\prod_o = 1$.

When the observer's ratings are statistically independent

$$\prod_{ii} = \prod_i + \prod_{+i}.$$

And the probability of agreement equals

$$\prod_e = \sum \pi_i + \pi_{+i}.$$

Subtracting $\prod_o - \prod_e$ gives the access of observer agreement over that expressed by chance. Therefore the Kappa statistic equals

$$\frac{K = \sum \pi_{ii} - \sum \pi_i + \pi_{+i}}{1 - \sum \pi_i + \pi_{+i}} = \frac{\prod_o - \prod_e}{1 - \prod_e}.$$

Where the denominator replaces $\prod_o$ by its maximum value of one, which is perfect agreement. When the agreement is expected by chance K=0, but when there is perfect agreement K=1.)

The Kappa Statistic has been widely ostracized for its usefulness in testing rater agreement. It is known as a rather implicit test that makes theoretical assumptions that are arbitrary and untested. In the Kappa statistic different components of disagreement are not separated, and agreement is not expressed in terms that are especially useful.

Polychoric Correlation

Using "R' the polychoric correlation was utilized to find the level of agreement between United States and El Salvadoran pathologists in using both the Pap smear and Thin Prep tests. Polychoric correlations measure the dependence between two ordinal variables assuming an underlying, unobserved latent variable structure [6]. That is, polychoric correlations can be used to measure rater agreement for ordered category ratings. They estimate what the correlation between raters would be if ratings were made on a continuous scale. It is assumed that the underlying continuous variables have a standard normal distribution. Using this function both components, agreement on trait definition and agreement on definitions of specific categories, can be assessed. The polychoric correlation can be estimated using the maximum likelihood function. It estimates the correlation between the latent continuous indicators. The log likelihood is computed as follows:

For a c×d table of two ordinal variables x and y, the log likelihood is

$$\ln L = A + \sum_{i=1}^{c}\sum_{j=1}^{d} N_{ij} \ln(\Pi_{ij}).$$

Where c and d are the number of categories for the first and second ordinal indicators, A is an irrelevant constant, and N_{ij} is the frequency of observations in the ith and jth categories. The thresholds for y and b_j, j=0,1,….d, where $a_0 = b_0 = -\infty$ and $a_c = b_d = -\infty$, and $a_c = b_d = +\infty$. Also

$$\pi_{ij} = \Phi_2(a_i, b_j) - \Phi_2(a_{i-1} b_j) - \Phi_2(a_i, b_{j-1}) + \Phi_2(a_{i-1}, b_{j-1}).$$

Where $\Phi_2(.,.)$ is the bivariate normal distribution function with correlation ρ. Note that the computation of the log-likelihood function above requires evaluations of bivariate integrals. In order to maximize the log-likelihood function, numerical procedures such as quasi-Newton-Raphson algorithm can be utilized.

Simulation Study

A simulation study was performed to compare the polychoric correlation with the Kappa statistic. Two different experimental conditions were considered:

Condition 1: Independence assumption

Condition 2: Independence violation

The simulation repeated itself 100 times for each condition, and the power of each test was computed to see which test had a higher rate of correctly rejecting the null hypothesis.

Kappa Statistic

Table 4: Kappa values and 95% confidence intervals

	Pap smear (US-ES)	Thin Prep (US-ES)
Kappa Statistic	0.211	0.611
Upper Bound Confidence	0.275	0.691
Lower Bound Confidence	0.147	0.531

These are the agreement values between United States and El Salvador pathologists for each test using the Kappa statistic. This table shows that there is, in fact, some level of agreement between United States and Salvadoran pathologists with using the Pap smear and Thin Prep tests. This is due to the 95% confidence interval, which does not include zero. The Kappa values for the Thin Prep test are significantly higher which means that pathologists agreed more with using the Thin Prep test than when using the Pap smear test. The upper bound confidence interval shows the highest value for the Kappa statistic, while the lower bound confidence interval shows the lowest value for the Kappa statistic.

Polychoric Correlation

Table 5: Polychoric Correlations and 95% confidence intervals

	Pap smear (US-ES)	Thin Prep (US-ES)
Polychoric Correlation	0.457	0.850
Upper Bound Confidence	0.635	0.980
Lower Bound Confidence	0.270	0.780

These are the agreement values between United States and El Salvador Pathologists for each test using the polychoric correlation. These statistics show that there is some level of agreement between raters in the United States and El Salvador. This is due to the .95 confidence interval (upper bound and lower bound), which does not include zero. The level of agreement is shown to be significantly greater when using the polychoric correlation than the Kappa statistic.

Simulation Study

Table 6: Results simulation study 1 (Observations are independent)

Simulation 1		Polychoric		Kappa	
		Mean	Power	Mean	Power
ρ=0.1	N=50	0.088	0.070	0.550	0.050
	N=100	0.083	0.110	0.053	0.110
ρ=0.2	N=50	0.202	0.180	0.128	0.150
	N=100	0.185	0.240	0.118	0.220
ρ=0.4	N=50	0.429	0.560	0.284	0.510
	N=100	0.432	0.830	0.284	0.810

This table displays results from our simulation study where independence was assumed in the scenario where pathologists actually rated each slide separately and independently. The mean represents the value that we obtained using the polychoric correlation and the Kappa Statistic closest to the true value of rho (ρ), and mean is the number of times the null hypothesis was correctly rejected by each test. In most cases the values of the polychoric correlation are significantly higher than that of the Kappa statistic.

Table 7: Results simulation study 2 (Independence assumption violated)

Simulation 2		Polychoric		Kappa	
		Mean	Power 1	Mean	Power 2
ρ=0.1	N=50	0.094	0.090	0.060	0.070
	N=100	.0830	0.070	0.052	0.070
ρ=0.2	N=50	0.220	0.250	0.141	0.240
	N=100	0.192	0.260	0.122	0.250
ρ=0.4	N=50	0.363	0.510	0.235	0.450
	N=100	0.379	0.730	0.246	0.720

In both simulations, the polychoric correlation is compared to the Kappa statistic in terms of statistical power. In most cases the polychoric correlation has more statistical power than the Kappa statistic. In this simulation, the independence assumption between pathologists is violated. The data is simulated assuming an underlying latent normal random variable, which modeled as an autoregressive process. This is the case in many

applications where the same pathologist reviewed a series of slides. However, our results show that the dependent structure gives reasonable results that are close to the true value. As a result, dependence does not violate results of the structure.

CHAPTER V
DISCUSSION

In this project rater agreement for cervical cancer between pathology ratings in the United States and pathology ratings in El Salvador were compared in order to find a more efficient method of detecting cervical cancer. Using a test that is more dependable in reporting accurate results could reduce the number of women in El Salvador who are affected by the disease. A test that is more sensitive in the detection of abnormal cells growth on the cervix could allow women to be treated in earlier stages which would prevent the development or spread of cervical cancer. Pathologists found that the level of rater agreement between pathologist when using the Thin Prep test was much higher than when using the PAP smear test for examining cells for cervical cancer. The test that was originally used to measure the rater agreement between pathologists was the Kappa statistic. However, we incorporated polychoric correlations for measuring rater agreement between pathologist in El Salvador and the United States. In order to compare the polychoric correlations hypothesis testing was conducted. The result from the hypothesis testing using the likelihood ratio test to compute the p-value proved that the level agreement for Thin Prep was greater than the level of agreement for the Pap smear. This factor conveys that in this case Thin Prep is better in detecting cervical cancer than Pap smear.

In a simulation that compared the polychoric correlation to the Kappa statistic it was proven that the polychoric correlation had more statistical power than the Kappa statistic. With this factor in mind it is likely that the polychoric correlation measures higher rater agreement than the Kappa statistic. Using this simulation the dependency of

the data was manipulated in order to see if dependent data that was assumed to be independent data would still give results close to the true value. Dependency often happens when a few pathologist review massive amounts of data. However, the results showed that even if dependent data was assumed to be independent data it does not violate the results of the latent random variable which is an autoregressive process.

Bibliography

1.Agresti, A. *Categorical Data Analysis*. New York Wiley & Sons, Inc. 1990.

2. Agurto, I., Bishop, A., Sanchez, G., Betancourt, Z., Robles, S. "Perceived Barriers and Benefits to Cervical Cancer Screening in Latin America." *Pan American Health Organization.* 28:188-97 (2001)

3. American Academy of Family Physicians. Pap Smears: What They are and What the Results Mean. July 2003. 10 Nov. 2004 http://www.familydoctor.org/

4. AmeriPath. Disease Information. <http://www.ameripathsw.com/>

5. Ballegooijen, Marjolein van, Boer, Rob, Elske van den Akker-van Marle, M., Habbema, J. Dik F., Oortmarssen, Gerrit J. van. Cost-Effectiveness of Cervical Cancer Screening: Comparison of Screening Policies. *Journal of the National Cancer Institute. Vol. 94.3:194-204 (2002)*

6. Bollen, KA. *Structural Equations With Latent Variables*. New York: John Wiley& Sons, Inc, 1989.

7. Cremer, M, DeLasCasas L, Hoerl D, Kurtycz D, Kim K, Eickhoff JC, and Schink, LC. Evaluation of Liquid-Based Cytology to Screen for Cervical Cancer Screenings in Rural El Salvado. *Submitted 2002*

8. Flores, Yvonne, PhD, Bishai, David, PhD, MD, Hernandez, Maurcio, MD, DrSc, Salmeron, Jorge, MD, DrSc, The Morelos HPV Study Collaborators. "Improving Cervical Cancer Screening in Mexico: Results from the Moreles HPV Study." *Health Publishers of Mexico.* Vol. 45, supplement 3. (2003)

9. Heher, Jon O. "Is the Thin Prep Better Than the Conventional Pap Smear at Detecting Cervical Cancer?" *The Journal of Family Practice*. 53.4 (2004)

10. Nanda K, McCrory, Myers ER, Bastain LA, Hasslebald V, Hickey JD, Matchar DB. Accuracy of the Papanicolaou Test in Screening for and Follow-up of Cervical Cytological Abnormalities: A System Review, Ann Intern Med; 2000; 1132: 810-819

11. Palo Alto Medical Foundation. Health Maintenance Guidelines. 2 Oct. 2003. 8 Nov. 2004. http://www.pamf.org/

VITA

Krystle S. Oates is a 22-year old graduating senior matriculating in mathematics at Southern University and A&M College. She is a very vibrant young woman who loves life, and the lessons that each day bring. She is a member of the Southern University Honors College, a LS-LAMP Scholar, a member of Beta Kappa Chi, the National Institute of Science, and President of the Math Club. She has conducted research at Louisiana State University, and the University of Wisconsin-Madison. After graduation, she plans on working at the National Institute of Health where she will conduct biomedical research. The following year she plans on attending graduate school at the University of Michigan.

12. Polania, Laura. Cervical Cancer Prevention in Latin America: Are Pap Tests Enough? International Health Central American Institute Foundation. *Submitted 26 March 2002.*

13. Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). *Reliable estimation of generalized linear mixed models using adaptive quadrature*. The Stata Journal, 2, 1—21.

14. Renshaw, Andrew A., MD, Young, Nancy A., MD, Birdsong, George G., MD, Styer, Patricia E., PhD, Davey, Diane D., MD, Mody, Dina R., MD, Colgan, Terrence J., MD. "Comparison of Performance of Conventional and Thin Prep Gynecological Preparations in the College of American Pathologists Gynecologic Cytology Program." *Archives of Pathology and Laboratory Medicine*: 128.1 (2004): 17-22.

15. The National Women's Health Information Center. Homepage. July 2002. National Institutes of Health. 10 Nov. 2004. http://www.4women.gov/

16. Women's Health Care Associates. Cervical Cancer. Feb. 2003. National Institutes of Health. 10 Nov. 2004.

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants Southern University Honors College the right to reproduce, by appropriate methods, any or all portions of this thesis.

It is understood that "request" consists of the agreement, on part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this thesis.

The author of this thesis reserves the right to publish freely, in the literature, at any time, any or all portions of this thesis.

Kyle D. Gates
Author

May 4, 2005
Date