McMaster University
DigitalCommons@McMaster

Open Access Dissertations and Theses | Open Dissertations and Theses

Fall 2011

EFFECTS OF COVARIATES ON THE PERFORMANCE OF CERVICAL CANCER SCREENING TESTS: LOGISTIC REGRESSION AND LATENT CLASS MODELS

Amidu O. Raifu
McMaster University, amidu.raifu@gmail.com

Follow this and additional works at: http://digitalcommons.mcmaster.ca/opendissertations

Part of the Biostatistics Commons

Recommended Citation
Raifu, Amidu O., "EFFECTS OF COVARIATES ON THE PERFORMANCE OF CERVICAL CANCER SCREENING TESTS: LOGISTIC REGRESSION AND LATENT CLASS MODELS" (2011). *Open Access Dissertations and Theses.* Paper 6113.

This Thesis is brought to you for free and open access by the Open Dissertations and Theses at DigitalCommons@McMaster. It has been accepted for inclusion in Open Access Dissertations and Theses by an authorized administrator of DigitalCommons@McMaster. For more information, please contact scom@mcmaster.ca.

PERFORMANCE OF CERVICAL CANCER SCREENING TESTS

EFFECTS OF COVARIATES ON THE PERFORMANCE OF CERVICAL CANCER SCREENING TESTS: LOGISTIC REGRESSION AND LATENT CLASS MODELS

By AMIDU OLALEKAN RAIFU, MSc, CStat

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Masters of Science

McMaster University © Copyright by Amidu Olalekan Raifu, August 2011

McMaster University MASTER OF SCIENCE (2011) Hamilton, Ontario (Statistics)

TITLE: Effects of Covariates on the Performance of Cervical Cancer Screening Tests: Logistic Regression and Latent Class Models

AUTHOR: Amidu Olalekan Raifu, MSc (Belgium)

SUPERVISOR: Prof. Stephen Walter

NUMBER OF PAGES: 1 to 53

To my mother, wife, and daughters

Acknowledgement

My sincere appreciation goes to my supervisor, Prof. Stephen Walter for his support, guidance, and suggestions that lead to successful completion of this thesis. My thanks also go to Prof. Eduardo Franco and Dr. Vekanta Ramana-Kumar Agnihotran of Department Clinical Epidemiology and Biostatistics, McGill University, Montreal for their suggestions and providing the cervical cancer data set.

I would also like to thank my defense committee members, Prof. Peter Macdonald, and Prof. Lehana Thabane for their suggestions and constructive criticism on the thesis. I would not forget to thank my fellow colleagues and most especially Dr. Conrad Kabali and Olga Melnichouk at Population Health Research Institute (PHRI) in Hamilton for proofing reading and correcting the typo-errors in the draft copy of this thesis. I would like to express my gratitude to my friend, Mr. Lameed Asafe Mubashir for the hints on how to format the thesis.

Finally, my special thanks go to my dear wife, Mrs. Iyabode Bolade Raifu for her caring, love, and sacrifice since we both left our country of origin up till present time here in Canada. I would also thank my daughters, Aishat, Mariam, and my little baby Khadijah for their love, caring and understanding. My sincere gratitude goes to my mother, Mrs. Wosilat Abike Raifu for her moral support and parental advice every time.

Abstract

In diagnostic accuracy studies, sensitivity and specificity are the most common measures to assess the performance of diagnostic or screening tests. The estimation of these measures can be done using empirical or model-based methods. The primary objective of this thesis is to use both the empirical and the model-based (logistic regression) approach to assess the effects of covariates on the performance of the visual inspection with acetic acid (VIA) and lugol iodine (VILI) tests using the data from women screened for cervical cancer in Kinshasa, the Democratic Republic of Congo. The secondary objectives are: first, to adjust for the false negative and false positive error rates by the two tests through latent class models (LCM), and second, to evaluate the effects of covariates on the agreement between the measurements of the two tests taken by nurse and physician through Kappa statistic.

No particular pattern could be observed in the trend of empirically estimated sensitivity and specificity of the VIA and VILI tests measured by the nurse and by the physician across age and parity categories. From the logistic regression models, age, parity, and the quadratic term of age have shown significant effects on the probability of VIA and VILI tests to detect cervical cancer. For other covariates such as marital status, smoking, and hybrid capture2 (HPV DNA), there is no significant effect on the probability of VIA and VILI tests measured by nurse to detect cervical cancer. However, only HPV DNA has shown significant effects on the probability of VIA and VILI tests measured by physician to detect cervical cancer. The trend of the estimated sensitivity of VIA and VILI tests measured by the nurse is not different across age groups but the specificity does vary. The trend of both the sensitivity and specificity of VIA and VILI tests are significantly different across parity groups. The reverse is the case for the sensitivity and specificity of VIA and VILI tests measured by physician across age and parity groups. The false negative and false positive error rates in the sensitivity and specificity of VIA and VILI tests measured by nurse are higher compared to that of physician. With Kappa statistic results, there is almost perfect agreement between the ratings by the nurse and physician for the dichotomized VIA and VILI test outcomes.

In conclusion, there is a significant effects of age, parity and the quadratic term of age on the performance of VIA and VILI tests outcomes measured by nurse. On the VIA and VILI test outcomes measured by physician, age, parity, HPV DNA and quadratic term of age have shown significant effects on the performance of VIA and VILI tests outcomes measured by physician alone.

Table of Contents

ACKNOWLEDGEMENT ..... **V**

ABSTRACT ..... **VI**

LIST OF TABLES ..... **VIII**

LIST OF TABLES ..... **VIII**

LIST OF FIGURES ..... **IX**

LIST OF ABBREVIATIONS ..... **X**

INTRODUCTION TO CERVICAL CANCER DISEASE AND THE SCREENING STUDY IN KINSHASA **1**

1.1 BRIEF INTRODUCTION TO CERVICAL DISEASE 1
1.2 CERVICAL CANCER SCREENING STUDY IN KINSHASA 2

THE DATA SET **5**

STATISTICAL METHODOLOGY **7**

3.1. ESTIMATION OF SENSITIVITY AND SPECIFICITY: EMPIRICAL METHOD 7
3.2. ESTIMATION OF SENSITIVITY AND SPECIFICITY: LOGISTIC REGRESSION METHOD 8
3.2.1. Logistic regression model for estimating the sensitivity and specificity *8*
3.2.2. Model building *10*
3.2.3. Goodness-of-fit of the model *12*
3.3 LATENT CLASS MODELING 13
3.3.1 Latent Class Model Diagnosis *14*
3.4 MEASURE OF AGREEMENT BETWEEN THE SCREENING TESTS RATERS 16

RESULTS **18**

4.1 DATA EXPLORATION 18
4.2 ESTIMATION OF SENSITIVITY AND SPECIFICITY: EMPIRICAL METHOD 20
4.3 ESTIMATION OF SENSITIVITY AND SPECIFICITY: LOGISTIC REGRESSION METHOD 24
4.3.1 Model building *24*
4.3.2 Logistic regression model for estimating the sensitivity and specificity *28*
4.4 LATENT CLASS MODELING 33
4.5 MEASURE OF AGREEMENT BETWEEN THE SCREENING TESTS RATERS: KAPPA STATISTIC 37

DISCUSSION **39**

CONCLUSION **42**

REFERENCES **43**

APPENDIX **46**

List of Tables

Table 1: Description of the variables in the data set 5
Table 2: Binary outcomes of a diagnostic test 7
Table 3: Interrater variation 16
Table 4: Distribution of the screened women across categories of variables 18
Table 5: Univariate analyses results using Pearson Chi-square test statistic 25
Table 6: Parameter estimates of the final logistic regression models for VIA and VILI tests measured by nurse 27
Table 7: Parameter estimates of the final logistic regression models for VIA and VILI tests measured by physician 27
Table 8: Results of classical LCM for VIA and VILI test outcomes measured by nurse and physician 34
Table 9: Estimated sensitivity and specificity of VIA and VILI test outcomes measured by nurse and physician with false positive and false negative error rates using classical LCM 34
Table 10: Results of the diagnostic test for conditional dependence using classical LCM 35
Table 11: Results of loglinear LCM for VIA and VILI test outcomes measured by nurse and physician 35
Table 12: Estimated sensitivity and specificity of VIA and VILI test outcomes measured by nurse and physician with false positive and false negative error rates using log-linear LCM 36
Table 13: Results of the diagnostic test for conditional dependence using log-linear LCM 36
Table 14: Estimated Kappa statistic for the agreement between nurse and physician on VIA test outcomes by covariates 37
Table 15: Estimated Kappa statistic for the agreement between nurse and physician on VILI test outcomes by covariates 38
Table 16: Results of the univariate logistic regression for model selection 46
Table 17: Results of the univariate logistic regression for model selection 47
Table 18: Estimated sensitivity and specificity of VIA test measured by the Nurse with 95% confidence interval using empirical method 48
Table 19: Estimated sensitivity and specificity of VILI test measured by the Nurse with 95% confidence interval using empirical method 49
Table 20: Estimated sensitivity and specificity of VIA test measured by the Physician with 95% confidence interval using empirical method 50
Table 21: Estimated sensitivity and specificity of VILI test measured by the Physician with 95% confidence interval using empirical method 51
Table 22: Estimated sensitivity and specificity from logistic regression model using VIA and VILI test outcomes measured by the Nurse with their 95% confidence interval52
Table 23: Estimated sensitivity and specificity from logistic regression model using VIA and VILI test outcomes measured by the Physician with their 95% confidence interval 53

List of Figures

Figure 1: Flow chart describing the design of the cervical cancer screening in Kinshasa..4
Figure 2: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across age group.....19
Figure 3: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across age groups with 95% CI 20
Figure 4: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across parity groups with 95% CI 21
Figure 5: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across parity groups with 95% CI.....22
Figure 6: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across age groups with 95% CI through logistic regression.....29
Figure 7: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across parity groups with 95% CI through logistic regression 30
Figure 8: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across age groups with 95% CI through logistic regression.....31
Figure 9: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across parity groups with 95% CI through logistic regression 32

List of Abbreviations

VIA	Visual Inspection with Acetic acid
VILI	Visual Inspection with Lugol Iodine
LCM	Latent Class Models
CI	Confidence Interval
HL	Hosmer and Lemeshow
CIN	Cervical Intraepithelial Neoplasia
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
LORC	Log-Odds Ratio Check
EM	Expectation-Maximization
OR	Odds Ratio

CHAPTER 1

Introduction to Cervical Cancer Disease and the Screening Study in Kinshasa

1.1 Brief Introduction to Cervical Disease

Cervical cancer is the second most common cancer among women worldwide with annual incidence and death of 471,000 and 233,000 respectively [1]. In less developed countries, almost 80% of the cases occur, with highest incidence rate in Latin America and the Caribbean, sub-Saharan Africa, and south and south-east Asia [1]. Despite this burden, most developing countries lack well organized cervical cancer screening programmes. In few of the countries that organized the screening, there has been no significant reduction in the cervical cancer burden. Two studies [2, 3], have attributed low level of success in cervical cancer screening in developing countries to poverty, disenfranchisement of women, lack of resources and infrastructures for cytology-based screening, well-trained technical personnel, and a good organization for population-based screening and follow-up. However, in some of the developed countries such as Finland, Netherlands and Belgium, there has been significant reduction in the incidence and mortality of cervical cancer over more than 4 decades due to well organized population-based cervical cancer screening programmes through pap smears and proper follow-up [1, 4].

According to Denny et al, [2], for cervical cancer screening to be successful in low-resource setting, the following should be done: 1) the screening, diagnosis and treatment must be provided on-site, or in clinics accessible to the majority of women at risk, 2) low-cost, low-technology screening test that can lead to immediate treatment of abnormalities, 3) wide coverage of at-risk women, 4) appropriate educational programmes for both the health providers and the women for correct implementation and high participation, 5) built-in mechanism for evaluation of the screening programme.

In the search for alternative screening methods for cervical cancer prevention that are cost-effective with low-technology, visual inspection with acetic acid (VIA) or lugol iodine (VILI) have been tested and considered as alternative screening tests to conventional Pap smear in a low-resource setting. Different studies [2, 3, 5, 8] have investigated the performance of VIA and VILI in detecting high-grade precursor lesions and invasive cervical cancer in some low-resource settings. The results from these studies showed that the sensitivity of VIA varied from 55% to 96% and the specificity varied from 49% to 98% and the sensitivity of VILI varied from 44% to 98% and the specificity varied from 75% to 91%.

1.2 Cervical Cancer Screening Study in Kinshasa

In diagnostic accuracy studies, sensitivity and specificity are the most common measures used to assess the performance of diagnostic or screening tests. The sensitivity is defined as the probability of a screening test being positive given that the disease is present [6]. The specificity is defined as the probability of a screening test being negative given that the disease is not present [6]. That is, the former is the probability of correctly classifying the diseased patients while the latter is the probability of correctly classifying non-diseased patients. The reference or gold standard test determines true status of the disease.

An example of diagnostic accuracy studies is a community-based cervical cancer screening study conducted in a primary health care setting in the suburbs of Kinshasa, Democratic Republic of Congo between 2003 and 2004. In this study, a total of 1,699 women aged 30 years and above were invited to participate in the study. Out of 1,699 women invited, 1,571 participated in the education session at the beginning of the study. Among the 1,571 women, 43 were ineligible for the study because they were pregnant or had hysterectomy (n= 10) or vaginal atresia or malformation (n= 9) and the 24 of them failed to return after their menstruations. The applicable screening tests for the low-resource setting like Kinshasa were visual inspection with acetic acid (VIA) and lugol iodine (VILI) and colposcopy tests which were completed for the remaining 1,528 women. The gold standard test for this study was the histology supplemented with colposcopy if histology is not available. The sample processing and screening tests (VIA and VILI) were performed independently by trained nurse and gynecologist who were blinded to each other's VIA/VILI scores. The findings from VIA and VILI tests were classified as 1 = Normal, 2 = Suspicious, 3 = Abnormal, and 4 = Cancer. In order to correct for verification bias which is a problem if the gold standard tests are performed only among the patients with positive results from the index tests, cervical biopsies were performed randomly in 20% sample of women who had normal colposcopy findings. The verification bias occurs when a study selects only the patients with positive results or excludes the patients with negative results from index test(s) for disease verification using gold standard test. However, when application of gold standard to individuals who have negative test results is infeasible or costly or risky, the gold standard test is limited to individuals with positive results due to ethical reasons [7]. In other words, it is ethically unacceptable to carry further confirmatory tests on individuals with initial negative results. The detailed description of the design of the study is presented in a flowchart shown in Figure 1.

An epidemiological analysis on the data set from Kinshasa cervical cancer screening study was conducted by Sangwa-Lugoma et al, 2006 [8]. The results of the analysis showed that the sensitivity and specificity of VIA for detecting high-grade cervical intraepithelial neoplasia (CIN2+) performed by the nurse were 55.5% (95% CI: 34.7-76.2) and 64.7% (95% CI: 62.0-67.1) respectively and that of VILI were 44.0% (95% CI: 24.2-63.8) and 74.6% (95% CI: 72.3-76.9) respectively. For the physician, the sensitivity and specificity of VIA for detecting CIN2+ were 71.1% (95% CI: 46.7-95.5), 71.3% (95% CI: 68.9-73.6) respectively and that of VILI were 68.3% (95% CI: 42.5-94.0),

76.2% (95% CI: 74.0-78.4) respectively. However, none of the studies mentioned above and in Chapter 1.1 ever investigated how the covariates measured on the patients affected the performance of the two tests.

In this thesis, the primary objective of this thesis is to use both the empirical and the model-based (logistic regression) approach to assess the effects of covariates on the performance of VIA and VILI tests using the data from women screened for cervical cancer in Kinshasa, the Democratic Republic of Congo. The secondary objectives are: first, to adjust for the false negative and false positive error rates by the two tests through latent class models (LCM), and second, to evaluate the effects of covariates on the agreement between the measurements of the two tests taken by nurse and physician through Kappa statistic.

The statistical techniques applied to achieve the primary objective include the empirical method and extension of logistic regression used in Coughlin et al, 1992 [9] for estimating the sensitivity and specificity of the screening tests. For the secondary objectives, latent class models (LCM) and Kappa statistics test were used. These statistical techniques were carried out in SAS 9.1 version [10], STATA 11.0 [11] version, LEM [12], and CONDEP [13].

The description of the data set from the study is presented in Chapter 2 of this thesis. In Chapter 3, statistical methods applied to analyze the data are described. Results are presented in Chapter 4. In Chapter 5 and Chapter 6, the discussion and conclusions are presented respectively.

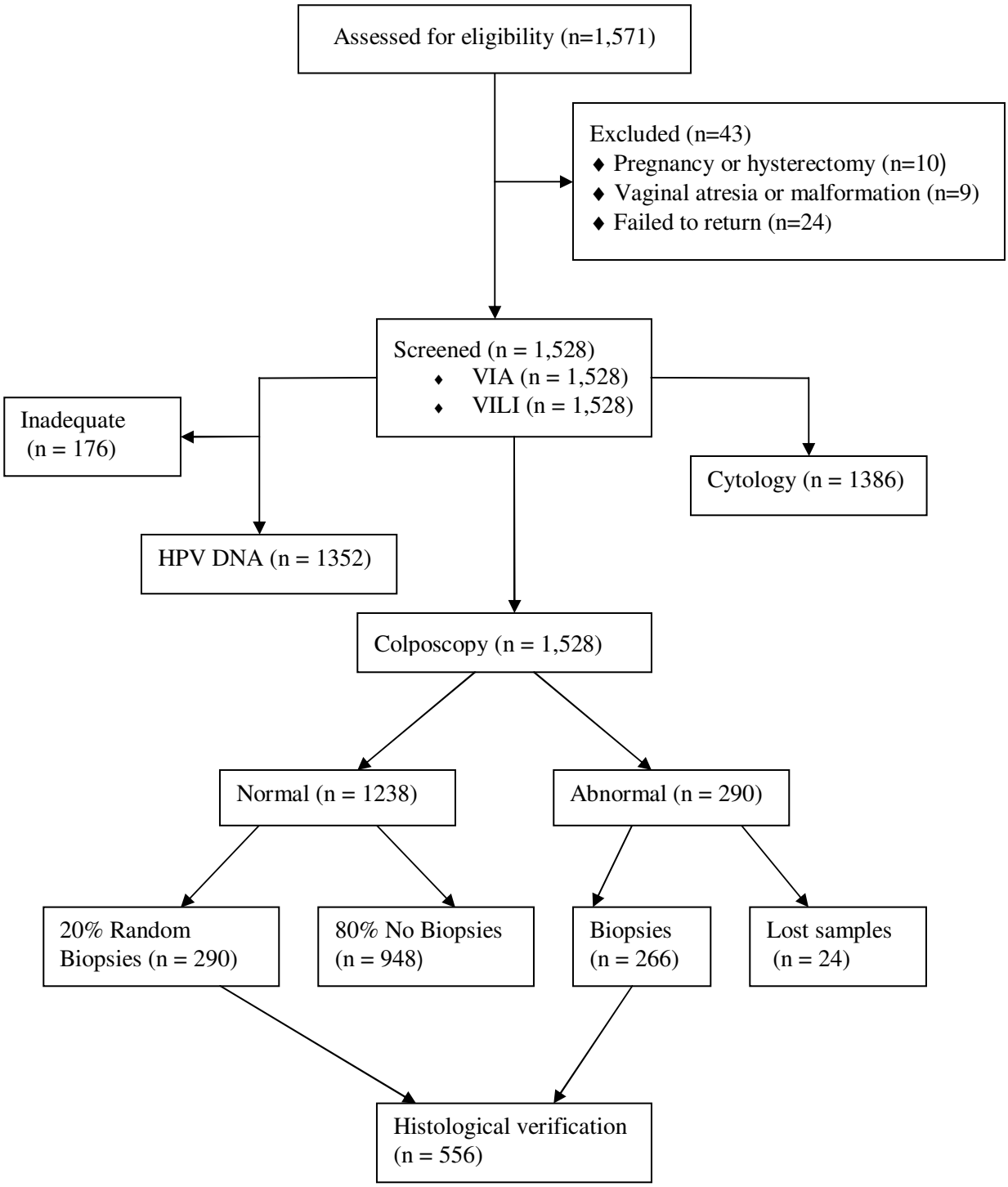

Figure 1: Flow chart describing the design of the cervical cancer screening in Kinshasa

CHAPTER 2

The Data Set

The data set for this research thesis was obtained from the study conducted by Sangwa-Lugoma, et al 2006 [8]. This study was based on cervical cancer screening at a primary health care setting in the suburbs of Kinshasa, Democratic Republic of Congo between 2003 and 2004 as earlier described in Chapter 1 and in the flowchart presented in Figure 1. The data set consists of 1,528 observations and 20 variables. The description of the variables in the data set is presented in Table 1.

Table 1: Description of the variables in the data set

Type of variable	Variable name	Categories	Description
Gold standard	*histcolpo*	1 = WNL 2 = CIN1 3 = CIN2+	Gold standard test outcome
Outcome variables	*via_nurse*	1 = Normal 2 = Suspicious	VIA screening test outcome measured by the Nurse
	vili_nurse	1 = Normal 2 = Suspicious 3 = Abnormal	VILI screening test outcome measured by the Nurse
	via_md	1 = Normal 2 = Suspicious 3 = Abnormal 4 = Cancer	VIA screening test outcome measured by the Physician
	vili_md	1 = Normal 2 = Suspicious 3 = Abnormal 4 = Cancer	VILI screening test outcome measured by the Physician
Other variables	*agecat*	1 = 30-34 2 = 35-39 3 = 40-44 4 = 45-49 5 = 50-54 6 = 55-59 7 = 60-64 8 = 65-69 9 = 70+	Age group
	province	1 = BAND 2 = Others	Provinces group
	religion	1 = Catholic 2 = Protestants 3 = Others	Religion group
	kinshasa	Continuous	Number of years living in Kinshasa

	education	1 = None 2 = Elementary 3 = Secondary 4 = High school 5 = University	Education group
	schoolyrs	Continuous	Number of school years
	profession	0 = None 1 = Manual 2 = Skilled workers 3 = Professional	Profession group
	SLI	1 = 0-18 (Low) 2 = 19-24 (Medium) 3 = 25-45 (High)	Standard of living index group
	smoking	1 = No 2 = Yes	Smoking group
	marital	1 = Married, cohabitating 2 = Otherwise	Marital status group
	co-wives	1 = None 2 = 1+	Number of wives group
	parity	1 = 0-1 2 = 2-5 3 = 6-9 4 = 10+	Parity group
	contragrp	0 = None 1 = Natural 2 = Medical	Contraceptive use group
	HC2	0 = Negative 1 = Positive	Hybrid Capture 2 (RLU> 1)
	HC2a	0 = ≤ 0.75 1 = 0.76-1.25 2 = 1.26-79.99 3 = 80+ 9 = Not done	Semi-quantitative group
	sumhist	0 = WNL 1 = CIN1 2 = CIN2+ 3 = CIN3+ 9 = Not done	Summary histology group
	colpo	1 = Normal 2 = Low-grade 3 = High-grade 4 = Cancer	Colposcopy test outcome

WNL = Within normal limit; CIN1 = Low-grade cervical intraepithelial neoplasia; CIN2+ = High-grade cervical intraepithelial neoplasia; RLU = Relative light unit; SLI = Standard living index; HC2 = Hybrid capture 2

CHAPTER 3

Statistical Methodology

This chapter describes the statistical methods used to achieve the primary and secondary objectives of this research thesis. The methods include empirical and logistic regression methods for estimating the sensitivity and specificity of the tests, latent class models for adjusting for the misclassification error by the index test, and the test for measuring the agreement between the results of the tests as affected by the covariates using Kappa statistic. Each of these methods is presented as follows.

3.1. Estimation of sensitivity and specificity: empirical method

Suppose there are two groups of patients (diseased and non-diseased) undergoing screening test A and gold standard (test B) with sample size n_1 and n_2. The outcome of the test A and gold standard (test B) which is either positive (+ve) or negative (-ve) is described in Table 2.

Table 2: Binary outcomes of a diagnostic test

Screening test A	Gold standard (test B)	
	+ve	-ve
+ve	a	b
-ve	c	d
Total	n_1	n_2

Where a is the number of true-positive (TP), b is the number of false-positive (FP), c is the number of false-negative (FN), and d is the number of true-negative (TN).

The empirical way of estimating the sensitivity and specificity of a screening test, for instance, test A as shown in Table 1 can be expressed as (1) and (2).

$$Sensitivity = P(test\ A = +ve \mid test\ B = +ve)$$
$$= \frac{a}{n_1} = \frac{TP}{TP + FN} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The sensitivity can be referred to as the true positive rate.

$$Specificity = P(test\ A = -ve \mid test\ B = -ve)$$
$$= \frac{d}{n_2} = \frac{TN}{TN + FP} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The specificity can also be referred to as the true negative rate.

The confidence intervals for the two measures (sensitivity and specificity) can be obtained using Clopper-Pearson interval (exact binomial confidence intervals) as shown in (3) and (4).

$$\left\{\theta_1 \mid P\left[Bin(n_1;\theta_1) \le a\right] \ge \alpha/2\right\} \cap \left\{\theta_1 \mid P\left[Bin(n_1;\theta_1) \ge a\right] \ge \alpha/2\right\} \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where a is the number of true positives (TP), and $Bin(n_1;\theta_1)$ is a binomial random variable with n_1 number of success (number of diseased patients) and θ_1 is the conditional probability of Test A being positive given that Test B is positive (sensitivity).

$$\left\{\theta_2 \mid P\left[Bin(n_2;\theta_2) \le d\right] \ge \alpha/2\right\} \cap \left\{\theta_2 \mid P\left[Bin(n_2;\theta_2) \ge d\right] \ge \alpha/2\right\} \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where d is the number of true negatives (TP), and $Bin(n_2;\theta_2)$ is a binomial random variable with n_2 number of success (number of non-diseased patients) and θ_2 is the conditional probability of Test A being negative given that Test B is negative (specificity).

3.2. Estimation of sensitivity and specificity: logistic regression method

The outcome of a diagnostic or screening test is usually dichotomous and the empirical way of estimating the sensitivity and specificity of a test has been described in Chapter 3.1. The model-based approach for estimating the sensitivity and specificity of the screening tests was based on the model derived by Cornfield (1962) [14] to estimate a subgroup risks of coronary heart disease and was applied in Coughlin et al (1992) [9] to estimate the sensitivity and specificity of the screening tests. In the following sub-chapters, the general brief description of the models, the model building and goodness-of-fit of the models are presented.

3.2.1. Logistic regression model for estimating the sensitivity and specificity

Different models have been proposed for analyzing binary data. A common model for analyzing binary data is logistic regression model. Due to flexibility in use and clinically meaningful interpretation, logistic regression model has been widely accepted for analyzing dichotomized data [15, 16]. According to Coughlin et al [9], the use of logistic regression model for estimating the sensitivity and specificity of the screening tests gives the opportunity to evaluate and adjust for the explanatory variables. In the paper by Cornfield (1962) [14], the gold standard variable was included in the model and the whole data set was used to incorporate the information about the covariates from both the diseased and non-disease group of patients. Thus, the model only incorporates the main factor of gold standard variable and other covariates in the model without any interaction terms. However, Coughlin et al [9] suggested inclusion of disease-by-covariate interaction term to the model. This is to have different slopes for the disease and non-disease groups and to know if there is difference between the disease and no-diseased group effects.

In this thesis, I further extend the Coughlin et al [9] logistic regression model by including the disease-by-covariates interaction terms and the square of age and parity with disease-by-square of age and parity interaction terms.

The general form of the extended logistic regression model is given in equation (5) where **Y** is the column vector of binary (0, 1) test outcomes. The outcome of a test being 1 represents positive result and 0 represents negative result. That is,

$$\pi(\mathbf{x}) = P(\mathbf{Y} = 1 \mid \mathbf{X} = \mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T \boldsymbol{\beta})} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

$$Logit[\pi(\mathbf{x})] = \log\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \mathbf{x}^T \boldsymbol{\beta} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

where $\mathbf{x}^T = (1, x_1, x_2, ..., x_p)$ is a vector of the constant and a set of values for p-covariates (including the second degree order and the interaction terms), $\pi(x)$ is the conditional probability of an index test outcome given the gold standard outcome, and $\boldsymbol{\beta}^T = (\beta_o, \beta_1, \beta_2, ..., \beta_p)$ represents the vector of the $p + 1$ coefficients.

The log odds or logit is linear in its parameters as link function to the linear predictor shown in (5) which may be continuous and range from $-\infty$ to $+\infty$ depending on the range of **x**. Since the outcome variable **Y** is dichotomous, the estimated probability $\pi(\mathbf{x})$ is bounded between 0 and 1 using the maximum likelihood estimation method.

For the variance of the sensitivity and specificity, in general, let $\hat{g}(\mathbf{x}) = \mathbf{x}^T \hat{\boldsymbol{\beta}}$ be the vector of the estimator of the logit of the sensitivity and specificity. The estimated variance of the logit of the logistic regression could be obtained as shown as follows;

$$V\hat{a}r\left[\ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right)\right] = V\hat{a}r[\hat{g}(\mathbf{x})] = \sum_{i=0}^{p} x_i^2 V\hat{a}r(\hat{\beta}_i) + \sum_{i=0}^{p}\sum_{j=i+1}^{p} 2x_i x_j C\hat{o}v(\beta_i, \beta_j) \ \ldots\ldots\ldots\ldots (7)$$

where the estimated variance-covariance matrix of the coefficients could also be obtained as;

$$C\hat{o}v(\hat{\boldsymbol{\beta}}) = (\mathbf{x}^T \mathbf{V} \mathbf{x})^{-1} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where $\mathbf{V} = \mathbf{diag}[n_i \hat{\pi}_i (1 - \hat{\pi}_i)]$ which denotes the $N \times N$ diagonal matrix having $[n_i \hat{\pi}_i (1 - \hat{\pi}_i)]$ on the main diagonal and $\hat{\pi}_i$ is the maximum likelihood estimate of $\pi(\mathbf{x}_i)$.

Therefore, the expression in (7) could be equivalent to

$$\begin{aligned} V\hat{a}r[\hat{g}(\mathbf{x})] &= \mathbf{x} C\hat{o}v(\hat{\boldsymbol{\beta}}) \mathbf{x}^T \\ &= \mathbf{x}(\mathbf{x}^T \mathbf{diag}[n_i \hat{\pi}_i (1 - \hat{\pi}_i)] \mathbf{x})^{-1} \mathbf{x}^T \end{aligned} \ \ldots\ldots\ldots\ldots (9)$$

With the estimated variance-covariance matrix $C\hat{o}v(\hat{\boldsymbol{\beta}})$ expressed above, one could obtain inference about $\boldsymbol{\beta}$ and the confidence intervals for the response probabilities $\pi(\mathbf{x})$ at

particular settings **x.** This variance-covariance matrix could also be obtained from the outputs of most logistic regression statistical software programs.

For large samples, a confidence interval (CI) for the true logit can be obtained as in (10). This can then be transformed through exponential as shown in equation (5) to obtain the corresponding confidence interval for $\pi(\mathbf{x})$.

$$Logit[\hat{\pi}(\mathbf{x})] \pm z_{\alpha/2}\sqrt{\mathbf{x}\widehat{Cov}(\hat{\boldsymbol{\beta}})\mathbf{x}^T} \quad \ldots\ldots\ldots\ldots (10)$$

3.2.2. Model building

Model building is usually motivated by different purposes. The purpose of building a model could be for description, control or prediction. Thus, any selected model depends on the purpose. For instance, model selected for descriptive purpose may not be suitable for control or predictive purposes. Therefore, the model building process should not be wholly mechanical. Instead, it should involve some subjective judgment [17]. In model building, the main goal is to achieve the most parsimonious model that is easy to interpret. The higher the number of explanatory variables included in the model, the more difficult the model building process could be. Thus, the higher the number of possible effects and interaction terms, the higher the number of estimated standard errors and dependence of the model on the observed data [15, 16].

To avoid this difficulty, different methods have been proposed in the literature for variable selection in model building. These methods include univariate analyses, forward stepwise regression, backward stepwise regression and the combination of both the forward and backward selection procedures usually referred to as stepwise regression method. Another variable selection method is through likelihood ratio test for nested models. Each of these selection methods is computer-based iterative methods. Most of them are available in most statistical software programs for selecting variables to be included in the model.

The univariate analyses are like descriptive analysis that gives an idea of possible effect of each covariate or the relationship between a response variable and each covariate. This method involves conducting separate analyses between dependent variable and each predictor variable. Any significant predictor variable in each of the analyses based on pre-specified p-value would be selected.

The forward stepwise regression involves fitting separate univariate analyses between a depended variable and each predictor variables. The most significant predictor variable(s) is first entering into the model followed by the next significant variable(s). This selection is usually based on pre-specified "p-value of entry" into the model. In this selection procedure, any predictor variable that enters into the model would never be removed in the next stage.

The backward stepwise regression selection procedure involves fitting a model that contain all the candidate covariates in the model and testing them individually to drop the

non-significant variable(s) from the model based on pre-specified "p-value of exit". With this procedure only the significant covariates based on the "p-value of exit" are kept in the final model.

The selection procedure that combines both the forward and backward stepwise selection procedures is referred to as stepwise selection procedure. In this procedure, the most significant covariate is first entering into the model. In the second stage, the second significant covariate is also entering into the model making two covariates in the model and a multivariable model is fitted. The non-significant covariate between the two covariates is dropped from the model and so on. Thus, a variable that first enters the model in the first stage could be removed in the second stage based on the pre-specified "p-value of exit".

The likelihood ratio test method for variable selection involves dropping a non-significant variable from the full model to test its significance using Chi-square test with a number of degree of freedom. The Chi-square value of the test is obtained by multiplying the difference between the log-likelihood of the reduced and full model by 2. The number of degree of freedom is also obtained by finding the difference between the degree of freedom of the reduced and full model.

In this thesis, the univariate analyses were first conducted to select the possible covariates that have effects or association with each of the 4 binary outcome variables. The univariate analyses involve contingency tables with Pearson Chi-square test statistic. Through the univariate analyses, the possible 6 covariates (main factors) were then selected out of the 16 covariates. The 6 possible covariates were carefully selected based on the suggestions from the Epidemiologist, Dr. Eduardo Franco who is one of the principal investigators of the cervical cancer screen study in Kinshasa, and their respective significance values. The results of the univariate analyses are presented in Table 4 in Chapter 4.1. The categorical variables age and parity were converted to continuous variables by using the mid-point value of their respective category. This is to avoid the computational problem due to small number of women in some categories of both age and parity (see Table 16 in the Appendix) during multivariable logistic regression model.

Since some of the variables suggested by the epidemiologist were not significant during univariate analyses, and in order to obtain more parsimonious models, likelihood ratio test and backward stepwise selection procedure were applied to reduce the full model. The final models obtained from both likelihood ratio test and backward stepwise selection procedure were the same for the VIA and VILI tests outcomes measured by physician. For the final models for VIA and VILI tests outcomes measured by nurse, different models were obtained from likelihood ratio test and backward stepwise selection procedure. Therefore, the final models from likelihood ratio test were chosen for VIA and VILI tests outcomes measured by nurse based on the fit of the two models.

To avoid multicollinearity between two or more covariates during model building, continuous variables age and parity were centered. By centering the age and parity, their

respective observations were replaced by the value of their deviation from their respective mean. The categories of the gold standard and VIA and VILI tests outcome variables were also re-categorized as binary due to different categories of VIA and VILI tests outcomes measured by nurse and the sparse data in the 4th categories of VIA and VILI tests outcomes measured by physician. That is, the second and third categories of gold standard were combined together as 1 = "Abnormal" and the first category as 0 = "Normal". Since the VIA test outcome measured by nurse has two observed categories, the first category was considered as 0 = "Normal" and the second category as 1 = "Abnormal". The first category of VILI test measured by nurse was regarded as 0 = "Normal" while the second and third categories were combined together as 1 = "Abnormal". The second, third, and fourth categories of VIA and VILI tests measured by physician were respectively combined together as 1 = "Abnormal" and the first category was regarded as 0 = "Normal".

3.2.3. Goodness-of-fit of the model

Goodness-of-fit test is a summary measure for assessing the fit or adequacy of the model. In statistical literatures, different summary measures had been proposed and widely accepted for assessing the fit of regression models. These measures are Pearson Chi-square, Deviance and Hosmer and Lemeshow (*HL*) test statistic which are available in most statistical software for assessing the fit of logistic regression model. Out of these goodness-of-fit tests statistic, *HL* test statistic was applied to assess the fit of the final four models selected in Chapter 3.2.2 and described briefly as follow.

The *HL* test statistic was proposed by Hosmer and Lemeshow, 1980 [18]. The test was based on grouping the values of the estimated probabilities into subgroups [15]. Assuming there were n columns of the estimated probabilities, the first column containing the smallest values and the nth column containing the largest values. There are two ways in which the groupings can be done. First, the table can be collapsed based on percentiles of the estimated probabilities. The second way was to collapse the table based on the fixed values of the estimated probabilities. The *HL* goodness-of-fit statistic can then be obtained as shown in (11) by calculating the Pearson chi-square statistic based on $g \times 2$ table where g is assumed to be 10 which correspond to the number of groups. According to Hosmer and Lemeshow, 1980 [18], when the number of parameters is the same as sample size, the null distribution approximate Chi-square distribution by g – 2 degree of freedom.

$$HL = \sum_{k=1}^{g} \frac{\left(O_k - n_k \bar{\pi}_k\right)^2}{n_k \bar{\pi}_k \left(1 - \bar{\pi}_k\right)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

where O_k is the observed number of events, n_k is the total number of subjects in the kth group and $\bar{\pi}_k$ is the average estimated probability.

3.3 Latent Class Modeling

Latent class model (LCM) is known for analyzing data containing one or more unobserved variables. Initially, it includes only the categorical variables but in the recent times, its model has been extended to include other variables such as nominal, continuous, and count variables [19, 20].

According to Lazarsfeld and Henry, 1968 [21] and Goodman, 1974 [22], the LCMs can be modeled as a probability. For instance, there are four diagnostic tests A, B, C, and D where test A represents the VIA test measured by nurse (*via_nurse*), test B represents VIA test measured by physician (*via_md*), test C represents VILI test measured by nurse (*vili_nurse*) and test D represents VILI test measured by physician (*vili_md*).These four diagnostic tests are indicators for latent variable X each with two categories, positive = 1 or negative = 0. Each of the tests are indexed by *a*, *b*, *c*, and *d*. The cell probabilities of a 2^4 multinomial distribution is π_{abcd} with *a*, *b*, *c*, *d* = 1, 0.

Let y_{abcd} and p_{abcd} be the observed cell frequency and probability respectively, e_{abcd} and π_{abcd} be expected cell frequency and probability respectively. This implies that $y_{abcd} = Np_{abcd}$ and $e_{abcd} = N\pi_{abcd}$ such that the total cell frequency $N = \sum_{abcd} y_{abcd} = \sum_{abcd} e_{abcd}$. The latent variable is X with subscript *x* and it is assumed to be the true disease status such that $X = 1$ is the latent class containing the diseased group and $X = 0$ is the latent class containing the non-diseased group. The cell probability π_{abcd} can be expressed as

$$\begin{aligned} \pi_{abcd} &= \sum_{x=0}^{1} \pi_{abcdx}^{ABCDX} = \sum_{x=0}^{1} \pi_{x}^{X} \pi_{abcdx}^{ABCD|X} \\ \sum_{x=0}^{1} \pi_{x}^{X} &= 1, \qquad \sum_{x=0}^{1} \pi_{abcdx}^{ABCD|X} = 1, \ a,b,c,d,x = 0,1 \end{aligned} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

where $\pi_{abcdx}^{ABCD|X}$ is the conditional response probability that a pattern of test results with respect to the four diagnostic tests will be (*a*, *b*, *c*, *d*) given the disease status *x*. π_{x}^{X} is the latent probability with true disease status *x*. The sensitivity and specificity of each of the four diagnostic tests (A, B, C, D) could be expressed as follow in (13) and (14);

$$Sensitivity = \begin{cases} P(a=1|x=1) = \pi_{1+++1}^{ABCD|X} \equiv \pi_{11}^{A|X} \\ P(b=1|x=1) = \pi_{+1++1}^{ABCD|X} \equiv \pi_{11}^{B|X} \\ P(c=1|x=1) = \pi_{++1+1}^{ABCD|X} \equiv \pi_{11}^{C|X} \\ P(d=1|x=1) = \pi_{+++11}^{ABCD|X} \equiv \pi_{11}^{D|X} \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$Specificity = \begin{cases} P(a=0|x=0) = \pi_{0+++0}^{ABCD|X} \equiv \pi_{00}^{A|X} \\ P(b=0|x=0) = \pi_{+0++0}^{ABCD|X} \equiv \pi_{00}^{B|X} \\ P(c=0|x=0) = \pi_{++0+0}^{ABCD|X} \equiv \pi_{00}^{C|X} \\ P(d=0|x=0) = \pi_{+++00}^{ABCD|X} \equiv \pi_{00}^{D|X} \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

The + subscript denotes the summation over the replaced subscript. The probability on the right hand side of the expressions are equivalents of the term in the middle after suppressing the + subscripts for notational convenience.

Due to basic assumption in a classical LCM that given a true disease status, the manifest variables A, B, C, D are mutually independent, the $\pi_{abcdx}^{ABCD|X}$ can be written as (15).

$$\pi_{abcdx}^{ABCD|X} = \pi_{ax}^{A|X} \pi_{bx}^{B|X} \pi_{cx}^{C|X} \pi_{dx}^{D|X}; a,b,c,dx = 0,1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

With the basic assumption of LCM, it is believed that diagnostic tests outcomes on unobserved disease status are imperfect and the association between diagnostic tests is only explained by the latent variable [23]. In another way, the responses within each latent class are independent and the level of the manifest variables is independent of the levels of all other observed variables [13]. This assumption is known as local independence. However, when the manifest variables have multiple indicators or similar content or the same item that are repeated different times, the assumption may fail to hold [13]. In some studies, the manifest variables may be associated within latent classes and it is otherwise called conditional dependence. This is a serious problem with studies that apply LCM without taking the dependence into account in the model. The problem could cause the model fit statistic to be very high [13]. This may leads to adding another latent class in order to improve the fit the model.

The maximum likelihood estimation of the sensitivity and specificity of the manifest variables in LCM can be obtained through EM algorithm. Detailed information about the implementation of the LCM can be found elsewhere [7, 19, 20, 22, 24-29].

3.3.1 Latent Class Model Diagnosis

Different models have been fitted by different authors [24-28] to increase the number of latent classes to provide a good fit for the model in order to account for the dependence between the manifest variables.

In the recent time, new statistical software packages [13, 20] have been developed to take dependence between the manifest variables into account. The new statistical software packages were developed since previous literatures have underemphasized this local dependence and these theories were incorporated into some known statistical software. Detailed technical discussion on local dependence could be found elsewhere [13, 20, and 29].

For this thesis, the classical LCM model was initially fitted with local independent assumption. After fitting the model, a modified version of Garrett and Zeger's (2000) [30] Log-Odds Ratio Check (LORC) implemented in CONDEP [13] statistical program for diagnosing conditional dependence in latent class models was applied. The modified version involves constructing the observed and model-predicted two-way cross-classification frequency tables for the depended variables. Then calculate the log-odds ratio in both observed and expected two-way tables with standard error for the expected data. The observed data is then expressed as a z-score relative to the expected data. If the calculated z-value exceeds a z-critical value (+/- 1.645 or +/- 1.96) then there is evidence that the items are conditionally dependent. More details about the method is available on the manual for CONDEP.

The CONDEP [13] program is an easy to use software. The program was designed to be used only for dichotomous data. It utilizes the output data for observed and expected frequencies for each response pattern obtained from either of other LCM programs such as PAMARK, LEM [12] or MLLSA. In this analysis, the output from LEM [12] software was used. For the detailed information about the CONDEP program, see the user manual available for download with the program.

Having diagnosed the local dependence through CONDEP [13] software, a latent log-linear model was fitted based on the models fitted by Walter and Franco, 2008 [22] and Espeland and Handelman, 1989 [31] to account for the conditional dependence. In these log-linear models, joint cell probabilities were formulated by introducing the interaction terms through model parameterization. In Walter and Franco, 2008 [22], and Espeland and Handelman, 1989 [31], a four variable latent log-linear model fitted is shown in (16) and (17) respectively.

$$\ln\left(\pi_{abcdx}^{ABCDX}\right) = \lambda + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_d^D + \lambda_x^X + \lambda_{ax}^{AX} + \lambda_{bx}^{BX} + \lambda_{cx}^{CX} + \lambda_{dx}^{DX} \quad \ldots\ldots\ldots\ldots\ldots\ldots(16)$$

where λ is the overall mean frequency across all cells, $\lambda_a^A, \lambda_b^B, \lambda_c^C$, and λ_d^D are marginal constraints on the frequencies at each level of A, B, C, and D respectively. λ_{ax}^{AX}, λ_{bx}^{BX}, λ_{cx}^{CX}, and λ_{dx}^{DX} are the interaction terms which account for the associations between each A, B, C, D and X respectively.

$$\ln\left(\pi_{abcdx}^{ABCDX}\right) = \lambda_x^X + \lambda_{ax}^{AX} + \lambda_{bx}^{BX} + \lambda_{cx}^{CX} + \lambda_{dx}^{DX} + \lambda_{cdx}^{CDX} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

where X is the latent variable, A, B, C, and D are the four raters that their agreement/disagreement in diagnostic study were analyzed and λ_{cdx}^{CDX} is the interaction term representing the parameter for the dependence between C and D. The maximum likelihood estimation of the sensitivity and specificity of the manifest variables in log-linear model can also be obtained through EM algorithm. Detailed information about the implementation of the LCM can be found elsewhere [7, 19, 20, 22, 24-29].

3.4 Measure of agreement between the screening tests raters

In the previous chapters, different methods for estimating the sensitivity and specificity of the screening tests have been discussed. However, none of these methods takes into account the agreement or disagreement by chance between the observers or raters of the screening tests. In order to measure the degree of agreement or disagreement of the two raters of the cervical cancer screening study in Kinshasa, Kappa statistic was applied.

Kappa statistic is one of the statistical methods for measuring the agreement between two or more independent observers or raters of screening tests. It was first proposed by Cohen (1960) [32]. The generalization for weights of each disagreement was also proposed by Cohen (1968) [33]. The Kappa statistic compares the probability of observed agreement to the probability of agreement to be expected by chance. The expression for Kappa statistic (κ) is presented in (18) based on the information presented in Table 3.

Table 3: Interrater variation

	1st Rater		
2nd Rater	Yes (a)	No (b)	Total
Yes (a)	π_{aa}	π_{ab}	π_{a+}
No (b)	π_{ba}	π_{bb}	π_{b+}
Total	π_{+a}	π_{+b}	1

$$\kappa = \frac{(\pi_{aa} + \pi_{bb}) - (\pi_{a+}\pi_{+a} + \pi_{b+}\pi_{+b})}{1 - (\pi_{a+}\pi_{+a} + \pi_{b+}\pi_{+b})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

where π_{aa} and π_{bb} are the probabilities that the two raters agree, π_{ab} and π_{ba} are the probabilities that the two raters disagree. Thus, when the two raters always agree (perfect agreement), π_{ab} and π_{ba} equal to zero, then the observed probability of agreement $(\pi_{aa} + \pi_{bb})$ equals 1 but when they always disagree, π_{aa} and π_{bb} equal to zero. Therefore, the stronger the agreement, the higher is the value of κ. There could also be negative value for κ which indicates the agreement is weaker than expected by chance.

The confidence interval for the Kappa statistic can be obtained through analytical method for dichotomous variables or two raters and through bootstrap method for more than two raters or more complex situation [34]. The estimation of the variance for the Kappa statistic was based on asymptotic variance developed by Fleiss, Cohen, and Everitt (1969) [35].

$$Q = (1-\pi_e)^{-1}\left[\sum_a \pi_o\{(1-\pi_e) - (\pi_{+a} + \pi_{a+})(1-\pi_o)\}^2 + (1-\pi_o)^2 \sum\sum_{a \neq b} \pi_{ab}(\pi_{+b} + \pi_{b+})^2 - (\pi_o\pi_e - 2\pi_e + \pi_o)^2\right]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

where $\pi_e = \pi_{a+}\pi_{+a} + \pi_{b+}\pi_{+b}$ is the expected probability under null hypothesis of chance agreement, $\pi_o = \pi_{aa} + \pi_{bb}$ is the observed probability and Q is the variance of Kappa times the sample size. The expression for the standard error of Kappa statistic for the

confidence interval for the 2×2 table as in Table 3 follows procedure presented in Cantor (1996) [36] as shown in (20). An approximate 100 (1-α)% confidence interval for κ is shown in (21).

$$s.e.(\hat{\kappa}) = \sqrt{Q/N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (20)$$

$$\hat{\kappa} - c_{\alpha/2} s.e.(\hat{\kappa}) \leq \kappa \leq \hat{\kappa} + c_{\alpha/2} s.e.(\hat{\kappa}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

Some available statistical software provides the p-value and confidence intervals for the Kappa statistic. Both the p-value and confidence are informative but they are sensitive to the size of the data. For any large sample size, κ value greater than zero is tend to be significant [37]. The p-value reported in some of the statistical software does not test for the strength of the agreement. Rather, it tests if the estimated Kappa statistic is due to chance or not.

For the cervical cancer screening data in this thesis, the frequency of ratings by the two raters, nurse and physician in the categories of VIA and VILI were not equal as can be seen in Table 4. To account for this, I dichotomize the categories of the ratings. The results of the Kappa statistic for the agreement between nurse and physician measurement are presented in the next chapter.

CHAPTER 4

Results

4.1 Data Exploration

The distribution of the screened women across the categories of each of the 10 variables considered in this thesis is presented in Table 4.

Table 4: Distribution of the screened women across categories of variables

Variable	Categories	Frequency (Percentage)	Missing (Percentage)
Gold standard	1 = WNL	1450 (94.90)	-
	2 = CIN1	47 (3.08)	
	3 = CIN2+	31 (2.03)	
VIA test outcome measured by nurse	1 = Normal	973 (63.68)	-
	2 = Suspicious	555 (36.32)	
VILI test outcome measured by nurse	1 = Normal	1127 (73.76)	-
	2 = Suspicious	396 (25.92)	
	3 = Abnormal	5 (0.33)	
VIA test outcome measured by physician	1 = Normal	1066 (69.76)	-
	2 = Suspicious	427 (27.95)	
	3 = Abnormal	33 (2.16)	
	4 = Cancer	2 (0.13)	
VILI test outcome measured by physician	1 = Normal	1143 (74.80)	-
	2 = Suspicious	320 (20.94)	
	3 = Abnormal	64 (4.19)	
	4 = Cancer	1 (0.07)	
Age group	1 = 30-34	283 (18.52)	3 (0.20)
	2 = 35-39	246 (16.10)	
	3 = 40-44	263 (17.21)	
	4 = 45-49	290 (18.98)	
	5 = 50-54	185 (12.11)	
	6 = 55-59	109 (7.13)	
	7 = 60-64	69 (4.52)	
	8 = 65-69	33 (2.16)	
	9 = 70+	47 (3.08)	
Smoking	1 = No	1092 (71.47)	2 (0.13)
	2 = Yes	434 (28.40)	
Marital status	1 = Married, cohabitating	1146 (75.00)	-
	2 = Otherwise	382 (25.00)	
Parity	1 = 0-1	140 (9.16)	8 (0.52)
	2 = 2-5	489 (32.00)	
	3 = 6-9	702 (45.94)	
	4 = 10+	189 (12.37)	
Hybrid Capture 2	0 = Negative	1183 (77.42)	176 (11.52)
	1 = Positive	169 (11.06)	

WNL: Within normal limit; CIN: Cervical intraepithelial neoplasia

In Table 4, it could be observed that less than 40% of women screened with VIA and VILI tests were classified by nurse and by physician as being suspicious or abnormal or having cervical cancer while more than 60% of the women were classified as being normal. These classifications of VIA and VILI tests outcomes by nurse and physician may not be the true disease status among the screened women. However, the results from the gold standard test show that only a small proportion (about 5%) of the screened women has low squamous intraepithelial lesions (CIN1) or high squamous intraepithelial lesions (CIN2) while about 95% of the women were normal. These results show that less than 35% of the screened women were wrongly classified with false positive and false negatives by nurse and by physician based on gold standard test.

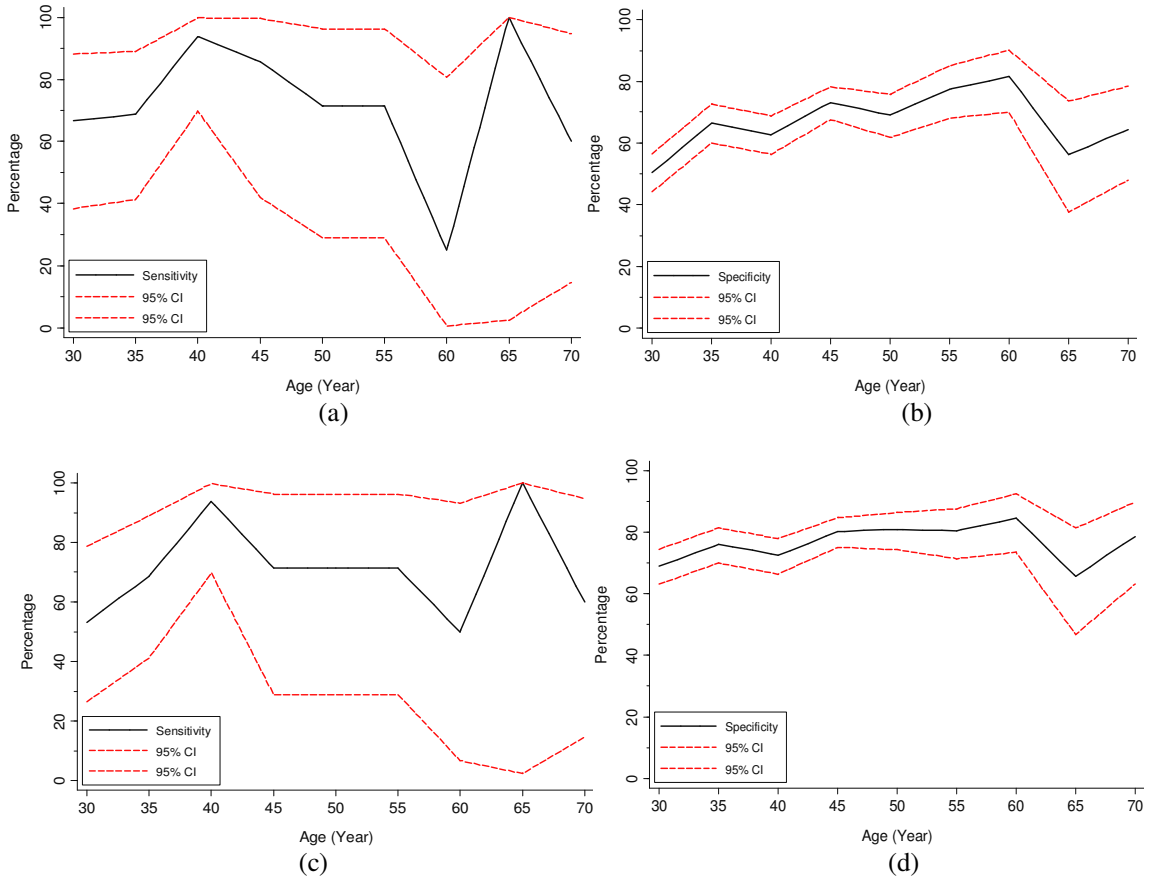

Figure 2: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across age group

The distribution of the screened women across different age groups as shown in Table 4 also varies from one age group to the other. The table shows that majority of the screened women were within the age range of 30 to 59 years old. For the distribution of the screened women across parity group, smoking, marital status and hybrid capture 2, the screened women belonging to parity group "6-9" has the highest percentage (about 46%) followed by those in parity group 2-5 (about 32%) while those in parity group 0-1 has the smallest percentage (about 9%). It can also be seen from Table 4 that the percentage of non-smokers is higher compared to the smokers.

The percentage of the "married/cohabiting" group of the screened women is higher compared to those in "otherwise" group. Also, the percentage of screened women with positive results of hybrid capture 2 is smaller than those with negative results.

4.2 Estimation of sensitivity and specificity: empirical method

The results of the estimated sensitivity and specificity of VIA and VILI tests measured by both nurse and physician across the categories of the 6 covariates are presented in Table 18 to Table 21 respectively in the Appendix. The plots of the results obtained using the empirical method for the estimated sensitivity and specificity of VIA and VILI tests measured by nurse across the age and parity groups are respectively shown in Figure 2 and Figure 3 and that of physician across the age and parity groups are also shown in Figure 4 and Figure 5 respectively.

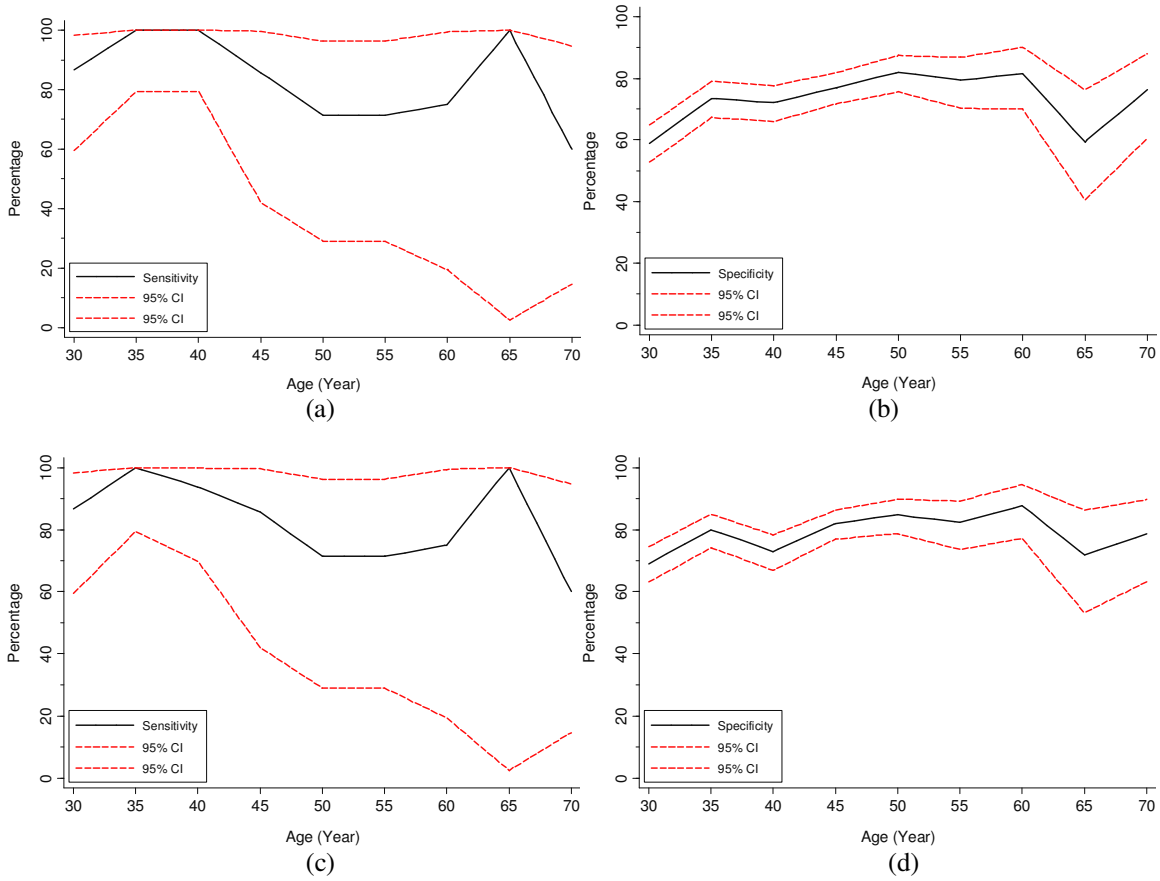

Figure 3: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across age groups with 95% CI

From Figure 2 and Figure 3 respectively, no particular trend pattern of sensitivity and specificity could be observed across different categories of age. The increase and decrease in the trend of the sensitivity and specificity across the age groups are non-monotonic. For instance in Figure 2b, the sensitivity of VIA measured by nurse increases from age group 30-34 until age group 40-44 then decreases from age group 40-44 to age

group 45-49. It then remains constant from age group 45-49 until age group 55-59. From age group 55-59, the sensitivity then decreases till age group 60-64 and then increases till age group 65-69 before decreasing till age group 70+.

The plots of the 95% confidence intervals shown in Figure 2a and Figure 2c for the sensitivity of VIA and VILI tests measured by nurse respectively are wider. Similarly, wider 95% confidence intervals are also shown in Figure 3a and Figure 3c for the sensitivity of VIA and VILI tests measured by physician respectively. However, the 95% confidence intervals shown in Figure 2b and Figure 3b for the specificity of VIA and VILI tests measured by nurse and by physician respectively are much narrower.

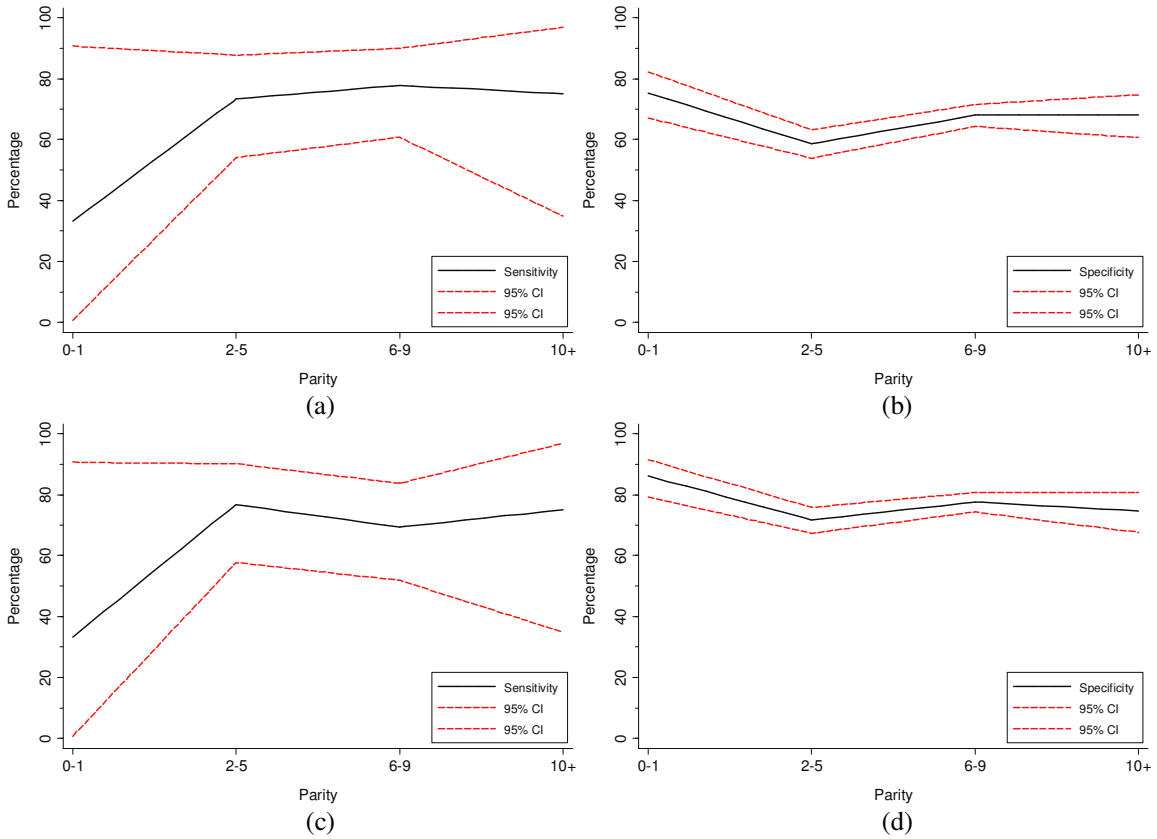

Figure 4: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across parity groups with 95% CI

The wider confidence intervals shown in Figure 2a, Figure 2c, Figure 3a and Figure 3c for the sensitivity of VIA and VILI tests respectively are not surprising. These could be attributed, first, to the variability in the number of screened women between the different age groups earlier seen during data exploration (Table 4). Secondly, it could also due to the small number of screened women with positive test results in some of the age groups as shown in Table 18 to Table 21 in the Appendix.

In the same vein, the narrower confidence intervals shown in Figure 2b, Figure 2d, Figure 3b and Figure 3d for the specificity of VIA and VILI tests respectively could be due to

the variability between different age groups and much number of screened women in some of the age groups

The trends of the sensitivity and specificity of VIA and VILI tests measured by nurse and physician are shown in Figure 4 and Figure 5 respectively across parity groups are non-monotonic. The upward and downward trends of both sensitivity and specificity could be observed. In addition, there are wider 95% confidence intervals for the sensitivity of VIA and VILI tests (Figure 4a, Figure 4c and Figure 5a, Figure 5c) and narrower 95% confidence intervals for their respective specificity (Figure 4b, Figure 4d and Figure 5b, Figure 5d).

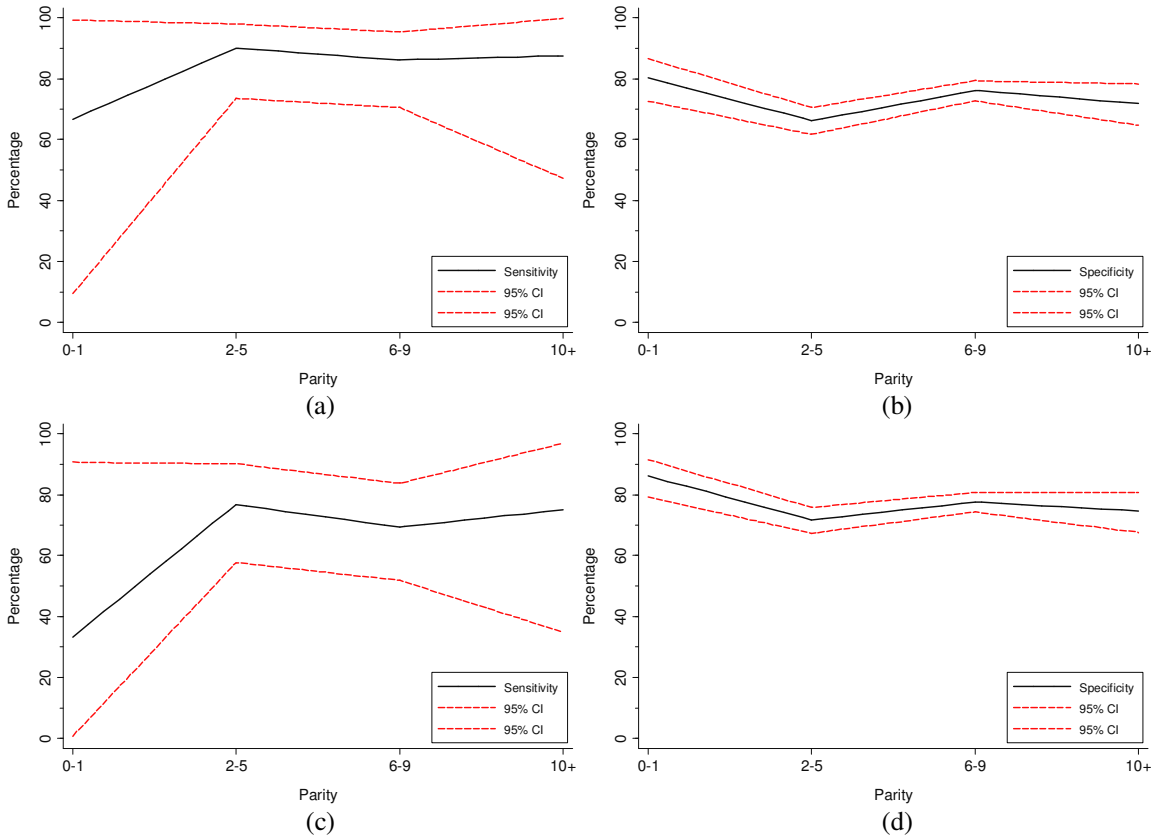

Figure 5: Plots of the empirically estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across parity groups with 95% CI

These wider confidence intervals for the sensitivity of both VIA and VILI tests could also be attributed to the small number of screened women with positive test results in the first and the last categories of parity groups. For the narrower confidence intervals for the specificity of VIA and VILI tests, this could also be due to more data that is available at the two categories.

Since what we have observed from the plots shown above is all informal statistical techniques in data analysis, a nonparametric trend test available in STATA [11] software was applied to confirm if the trends of the sensitivity and specificity across the different

age and parity groups are significant or not. This test was developed by Cuzick (1985) [38]. It is an extension of Wilcoxon rank-sum test and is a useful adjunct to the Kruskal-Wallis test of trend of ranks across ordered groups. The formula for the test statistic was given by Cuzick (1985) [38] and Altman (1991) [39]. Details on the application of the trend test can be found in Stepniewska and Altman (1992) [40].

The results of the nonparametric test showed no significant trend of sensitivity (p-value = 0.831) and specificity (p-value = 0.451) of VIA and sensitivity (p-value = 0.811) and specificity (p-value = 0.451) of VILI tests measured by nurse across the nine age groups. For the trend of the sensitivity and specificity of VIA and VILI tests measured by physician, the test also showed that no significant trends of the empirically estimated sensitivity and specificity is present across the age groups with. P-values of 0.142 and 0.322 for the sensitivity and specificity of VIA respectively and p-values of 0.175 and 0.451 for the sensitivity and specificity of VILI test respectively.

The non-parametric test for the trends also showed that there is no significant trend across parity ordered groups for the sensitivity (p-value = 0.115) and specificity (p-value = 0.480) of VIA test measured by nurse. No significant trends across parity ordered groups also confirmed from the nonparametric test for the sensitivity and specificity of VILI test measured by nurse with p-value of 0.480. The trend test results for the sensitivity (p-value = 0.480) and specificity (p-value = 0.480) of VIA test and the sensitivity (p-value = 0.211) and specificity (p-value = 0.550) of VILI test measured by physician showed no significant pattern of trend could be seen across the 4 groups of parity.

The empirical results of the VIA test measured by nurse in Table 18 and Table 19 (see Appendix) show that the detection rate of VIA test was higher with lower specificity among the “married/cohabiting” group of than in “other” group of screened women. The Z-test for independent proportions showed there is a significant difference in the sensitivity of VIA test measured by nurse between “married/cohabiting” group and “other” group of screened women with p-value < 0.001 but no difference in the specificity based on p-value of 0.318. For the smoking, the sensitivity (p-value = 0.005) and specificity (p-value = 0.001) of VIA test measured by nurse were significantly different between the smokers and non-smokers. In hybrid capture2 categories, there is significant difference (p-value < 0.001) in the sensitivity but no significant difference (p-value = 0.798) in the specificity between the women with positive results and those with negative results. The results of Z-test also show no significant difference in the sensitivity (p-value = 0.169) and the specificity (p-value = 0.619) of VILI tests measured by nurse between “married/cohabiting” and “other” groups of screened women. There is significant difference in the sensitivity (p-value < 0.001) and specificity (p-value = 0.043) of VILI test measured by nurse between the two categories of smoking. In the hybrid capture2 variable, there is significant difference in sensitivity (p-value < 0.001) and but not in specificity (p-value = 0.441) between the screened women with positive results and those with negative results.

The sensitivity and specificity of VIA and VILI tests measured by physician are presented in Table 20 and Table 21 (see Appendix). The result of the Z-test shows

significant difference in the sensitivity (p-value = 0.016) but not in the specificity (p-value = 0.879) of VIA test measured by physician between "married/cohabiting" group and "other" group of screened women. For smoking, there is significant difference in the sensitivity (p-value = 0.003) but not in specificity (p-value = 0.074) between smokers and non-smokers. In hybrid capture2, there is significant difference in the sensitivity (p-value < 0.001) and specificity (p-value = 0.027) of VIA test measured by physician between the women with positive and those with negative results. There is no significant difference in the sensitivity (p-value = 0.171) and specificity (p-value = 0.935) of VILI test measured by physician between the two categories of marital status of the screened women. There is also no significant difference in the sensitivity (p-value = 0.091) and specificity (p-value = 0.249) of VILI test between smokers and non-smokers. Between the positive and negative levels of hybrid capture2, the sensitivity (p-value < 0.001) is significantly different but specificity (p-value = 0.052) is not significantly different.

4.3 Estimation of sensitivity and specificity: logistic regression method

4.3.1 Model building

Table 5 shows the results of the Pearson Chi-square test used for selecting the possible covariates from the categorical covariates measured during the cervical cancer screening study in Kinshasa. The results of the univariate logistic regression for both the categorical and continuous covariates are presented in Table 16 and Table 17 in the Appendix. As could be seen from Table 5, the covariates with bold P-values are the variables that have significant association with each of the screening tests outcome. The covariates with asterisks are those suggested to be considered in the full model (17) by the Epidemiologist, Dr Eduardo Franco.

The initial full model for the covariates selected through univariate analyses for the multivariable logistic regression is shown in equation (17). The model contains the gold standard variable (*histcolpo*), smoking variable (*smoking*), marital status variable (*marital*), continuous centered age variable (*agec*), continuous centered parity variable (*parityc*), hybrid capture2 variable (*hc2*), interaction term between centered age and gold standard variable (*agec*histcolpo*), interaction term between centered parity and gold standard variable (*parityc*histcolpo*), and quadratic term of centered age ($agec^2$).

Others are interaction term between quadratic term of centered age and gold standard variable ($agec^2$**histcolpo*), quadratic term of parity (*parityc*), and interaction term between the quadratic term centered parity and gold standard variable ($parityc^2$**histcolpo*).

Table 5: Univariate analyses results using Pearson Chi-square test statistic

Outcome variable	Individual covariate	P-value	Outcome variable	Individual covariate	P-value
VIA test outcomes measured by nurse (*via_nurse*)	Agecat	**< 0.001***	VIA test outcomes measured by physician (*via_md*)	Agecat	**< 0.001***
	Province	0.416		Province	0.876
	Religion	0.533		Religion	0.551
	Education	0.195		Education	0.018
	Profession	0.151		Profession	0.952
	SLI	0.348		SLI	0.466
	Smoking	**0.036***		Smoking	0.055*
	Marital	0.282*		Marital	0.847*
	Cowives	0.567		Cowives	0.719
	Parity	**0.001***		Parity	**< 0.001***
	Contragrp	0.203		Contragrp	0.157
	HC2	**0.009***		HC2	**< 0.001***
	HC2a	0.107		HC2a	**< 0.001**
VILI test outcomes measured by nurse (*vili_nurse*)	Agecat	**0.009***	VILI test outcomes measured by physician (*vili_md*)	Agecat	**< 0.001***
	Province	0.904		Province	0.410
	Religion	0.714		Religion	0.842
	Education	0.693		Education	0.133
	Profession	0.699		Profession	0.515
	SLI	0.205		SLI	0.478
	Smoking	0.511*		Smoking	**0.048***
	Marital	0.524*		Marital	0.610*
	Cowives	0.893		Cowives	0.978
	Parity	**0.001***		Parity	0.085*
	Contragrp	0.624		Contragrp	0.237
	HC2	**0.007***		HC2	**< 0.001***
	HC2a	0.081		HC2a	**< 0.001**

* Suggested covariates

$$\begin{aligned} Logit[\pi(\mathbf{x})] = {} & \beta_0 + \beta_1 * histcolpo + \beta_2 * smoking + \beta_3 * marital + \beta_4 * age + \beta_5 * parity + \\ & \beta_6 * hc2 + \beta_7 * age * histcolpo + \beta_8 * parity * histcolpo + \beta_9 * age^2 + \\ & \beta_{10} * age^2 * histcolpo + \beta_{11} * parity^2 + \beta_{12} * parity^2 * histcolpo \qquad .....(22) \end{aligned}$$

The expressions for estimating the sensitivity and specificity of VIA and VILI tests using the full model (22) as an example are shown in equation (23) and (24) respectively with their corresponding confidence intervals in equation (25) and (26).

$$Sensitivity = \frac{1}{1 + \exp\left[-\left(\begin{array}{l} \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 * smoking + \hat{\beta}_3 * marital + \left(\hat{\beta}_4 + \hat{\beta}_7\right) * age + \\ \left(\hat{\beta}_5 + \hat{\beta}_8\right) * parity + \hat{\beta}_6 * hc2 + \left(\hat{\beta}_9 + \hat{\beta}_{10}\right) * age^2 + \\ \left(\hat{\beta}_{11} + \hat{\beta}_{12}\right) * parity^2 \end{array} \right) \right]} \qquad ..... (23)$$

where the gold standard variable $histcolpo = 1$ and

$$Specificity = 1 - \left\{ \frac{1}{1 + \exp\left[-\begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_2 * smoking + \hat{\beta}_3 * marital + \hat{\beta}_4 * age + \hat{\beta}_5 * parity + \\ \hat{\beta}_6 * hc2 + \hat{\beta}_9 * age^2 + \ \hat{\beta}_{11} * parity^2 \end{pmatrix} \right]} \right\}$$

…… (24)

where the gold standard variable $histcolpo = 0$.

The confidence intervals for the estimated sensitivity and specificity based on the estimated variance of the true logit in (9) is shown in (25) and (26) respectively.

$$\frac{1}{1 + \exp\left[-\begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 * smoking + \hat{\beta}_3 * marital + \left(\hat{\beta}_4 + \hat{\beta}_7\right) * age + \hat{\beta}_6 * hc2 \\ \left(\hat{\beta}_5 + \hat{\beta}_8\right) * parity + \left(\hat{\beta}_9 + \hat{\beta}_{10}\right) * age^2 + \ \left(\hat{\beta}_{11} + \hat{\beta}_{12}\right) * parity^2 \end{pmatrix} \pm 1.96\sqrt{\mathbf{x}\widehat{Cov}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}^T} \right]}$$

……. (25)

$$1 - \left\{ \frac{1}{1 + \exp\left[-\begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_2 * smoking + \hat{\beta}_3 * marital + \hat{\beta}_4 * age + \hat{\beta}_5 * parity + \\ \hat{\beta}_6 * hc2 + \hat{\beta}_9 * age^2 + \ \hat{\beta}_{11} * parity^2 \end{pmatrix} \pm 1.96\sqrt{\mathbf{x}\widehat{Cov}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}^T} \right]} \right\}$$

…….…….. (26)

Where $\mathbf{x}$ is the vector of the covariates for the expression for the sensitivity (23) and the specificity (24) respectively and $\hat{\boldsymbol{\beta}}$ is the vector of the estimated coefficients for the sensitivity (23) and the specificity (24) respectively.

Meanwhile, the parameter estimates of the final logistic regression models obtained through likelihood ratio test and backward selection procedure for VIA and VILI test outcomes measured by nurse are presented in Table 6 and for VIA and VILI test outcomes measured by physician are presented in Table 7.

The results in Table 6 show that there is significant effect of age and its quadratic term on the probability of VIA and VILI tests measured by nurse to detect cervical cancer. The significant effect of age as could be seen is negative. This shows that as age increases with one year, the chance of VIA and VILI tests measured by nurse to detect cervical cancer decreases with 4.1% and 3.0% respectively. The parameter estimate of the interaction term between the gold standard and the quadratic term of age is also negative but there is no significant difference between the squared age of the diseased and non-diseased screened women. For the effect of parity on the probability of VIA and VILI tests to detect cervical cancer, the results in Table 6 show that there is significant positive effect. This means for both VIA and VILI, the parity increases the chance of VIA and VILI tests measured by nurse to detect cervical cancer by 3.4% and 4.7% respectively.

Table 6: Parameter estimates of the final logistic regression models for VIA and VILI tests measured by nurse

Outcome variable	Covariate	Estimate	Standard Error	Z	P-value	95% Confidence Interval
VIA test outcomes measured by nurse (*via_nurse*)	Intercept	-0.772	0.087	-8.84	< 0.001	(-0.944, -0.601)
	Age	-0.041	0.007	-6.01	< 0.001	(-0.054, -0.027)
	Parity	0.034	0.019	1.75	0.080	(-0.004, 0.072)
	Goldstandard	2.009	0.344	5.84	< 0.001	(1.335, 2.683)
	Age^2	0.002	0.000	4.29	< 0.001	(0.001, 0.003)
	Goldstandard*Age^2	-0.002	0.001	-1.71	0.087	(-0.005, 0.000)
	$Parity^2$	-0.009	0.004	-2.11	0.035	(-0.018, -0.001)
VILI test outcomes measured by nurse (*vili_nurse*)	Intercept	-1.307	0.085	-15.43	< 0.001	(-1.474, -1.141)
	Age	-0.030	0.008	-3.99	< 0.001	(-0.045, -0.015)
	Parity	0.047	0.021	2.25	0.025	(0.006, 0.087)
	Goldstandard	2.646	0.389	6.8	< 0.001	(1.883, 3.409)
	Age^2	0.001	0.000	2.53	0.012	(0.000, 0.002)
	Goldstandard*Age	0.057	0.029	1.97	0.048	(0.000, 0.114)
	Goldstandard*Age^2	-0.004	0.002	-2.23	0.025	(-0.008, -0.001)

The *HL* test statistic for assessing the goodness-of-fit of VIA test outcomes measured by nurse shows that the fit of the model is good based on Chi-square value of 14.26 with 8 degrees of freedom and p-value of 0.075. For the final model for the VILI test measured by nurse, the *HL* shows the model has a good fit based on Chi-square value of 11.49 based on 8 degrees of freedom and p-value of 0.176.

In Table 7, it could be observed that similar models were obtained as final models for the VIA and VILI tests outcomes measured by physician. The parameter estimates in the two models are very close likewise the p-values except for the p-value for the parity. As could be seen, similar interpretation could be derived for the effects of age, parity and hybrid capture2 on the probability of VIA and VILI test to detect cervical cancer.

Table 7: Parameter estimates of the final logistic regression models for VIA and VILI tests measured by physician

Outcome variable	Covariate	Estimate	Standard Error	Z	P-value	95% Confidence Interval
VIA test outcomes measured by physician (*via_md*)	Intercept	-1.281	0.091	-14.03	< 0.001	(-1.460, -1.102)
	Age	-0.044	0.008	-5.66	< 0.001	(-0.059, -0.028)
	Parity	0.049	0.022	2.24	0.025	(0.006, 0.092)
	Goldstandard	2.861	0.391	7.31	< 0.001	(2.094, 3.628)
	Age^2	0.002	0.001	4.07	< 0.001	(0.001, 0.003)
	Hybrid capture 2	0.426	0.192	2.22	0.026	(0.050, 0.801)
VIA test outcomes measured by physician (*vili_md*)	Intercept	-1.508	0.098	-15.42	< 0.001	(-1.699, -1.316)
	Age	-0.037	0.008	-4.52	< 0.001	(-0.053, -0.021)
	Parity	0.044	0.023	1.9	0.058	(-0.001, 0.090)
	Goldstandard	3.023	0.374	8.07	< 0.001	(2.289, 3.757)
	Age^2	0.001	0.001	2.57	0.010	(0.000, 0.003)
	Hybrid capture 2	0.430	0.199	2.15	0.031	(0.039, 0.821)

From Table 7, there is significant effect of age and its quadratic term on the probability of VIA and VILI tests measured by physician to detect cervical cancer. The parameter estimate of age as could be seen is negative. This shows that as age increases with one year, the chance of VIA and VILI tests measured by nurse to detect cervical cancer decreases with 4.4% and 3.7% respectively. For the effect of parity on the probability of VIA and VILI tests to detect cervical cancer, there is significant positive effect. That is for both VIA and VILI, the parity increases the chance of VIA and VILI tests measured by nurse to detect cervical cancer by 4.9% and 4.4% respectively. In Table 6 and Table 7, the parameter estimate of the quadratic term of age is positive and significant. The significant effect of the quadratic term of age will influence the shape of the sensitivity and specificity of VIA and VILI tests across age groups.

The *HL* test statistic for assessing the goodness-of-fit of VIA test outcomes measured by physician also shows that the fit of the model is good based on Chi-square value of 13.84 with 8 degrees of freedom and p-value of 0.086. For the final model for the VILI test measured by physician, the *HL* also shows the model has a good fit based on Chi-square value of 4.30 based on 8 degrees of freedom and p-value of 0.829.

4.3.2 Logistic regression model for estimating the sensitivity and specificity

The results of the estimated sensitivity and specificity of VIA and VILI tests measured by nurse through logistic regression model with their respective 95% confidence interval (95% CI) are shown in Table 22 and Table 23 in the Appendix. The plots of the sensitivity and specificity of VIA and VILI tests measured by nurse are presented in Figure 6.

As can be seen in Figure 6, the plots of the sensitivity and specificity of VIA and VILI tests measured by nurse are smoother than the corresponding plots in Figure 2. The smoothness of the plots may be due to the shape of logistic regression curves and inclusion of the quadratic terms of age and parity in the model. The plot of the change in the trend of the sensitivity and specificity across the age groups in Figure 6 shows how the increase in age affects the detection rate and specificity of cervical cancer. For instance in Figure 6a, there was an increase in the sensitivity from age group 30-34 years old up to age group 45-49 years old before starting decreasing until age group 70 years and above. Similar trend could also be observed in Figure 6c for the sensitivity of VILI test. For Figure 6b and Figure 6d, the trend in specificity is monotonically increases from age group 30-34 years old until age group 55-59 years old before descending from age group 60-64 years old to age group 70 years old and above.

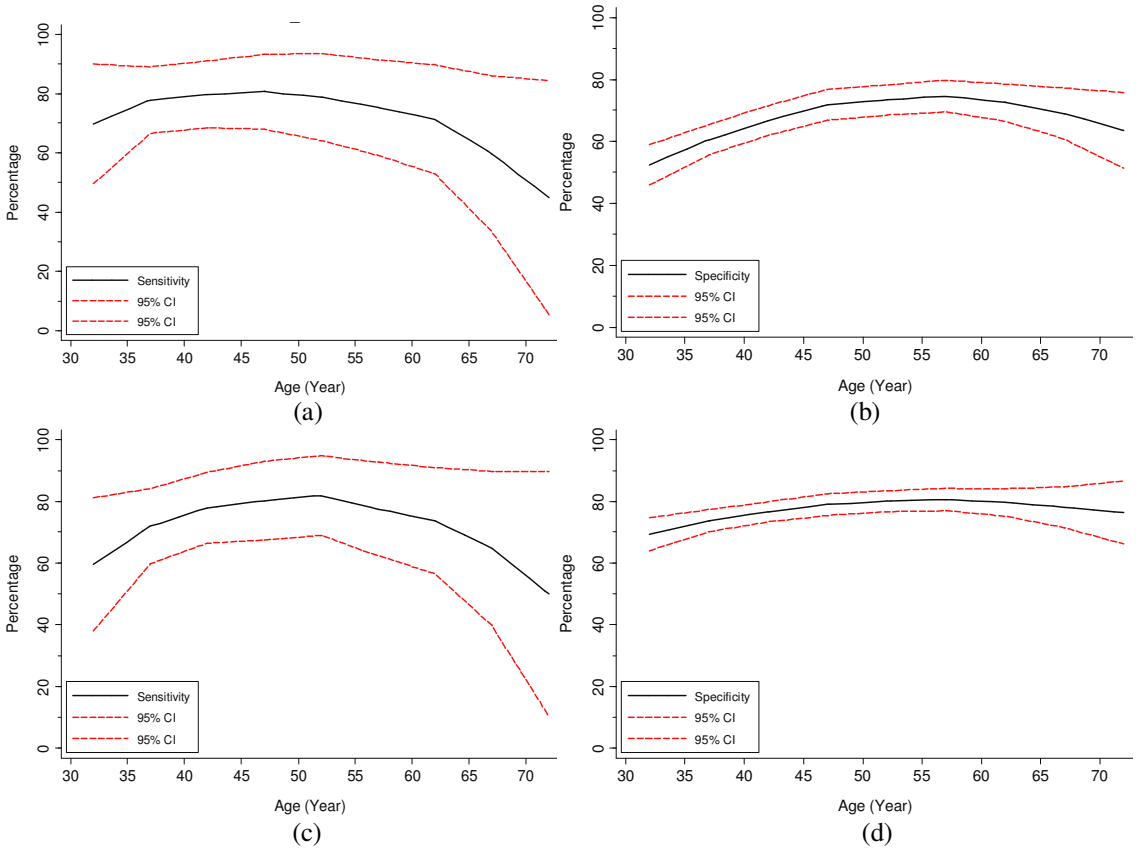

Figure 6: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across age groups with 95% CI through logistic regression

The nonparametric test conducted shows no significant trend in the sensitivity (p-value = 0.072) of VIA tests as measured by nurse (Figure 6a) but trend in the specificity was significant (p-value = 0.008) across the different age groups (Figure 6b). For the VILI tests measured by nurse, no significant trend is present in the sensitivity (p-value = 0.117) of VILI tests measured by nurse but significant trend in the specificity (p-value = < 0.001) across the different age groups. As earlier seen in the plots of the empirical sensitivity and specificity, there are wider confidence intervals in Figure 6a and narrower confidence intervals in Figure 6b respectively. The reason for the wider confidence intervals for the sensitivity is due to few data while the narrower confidence intervals for the specificity are due to more data.

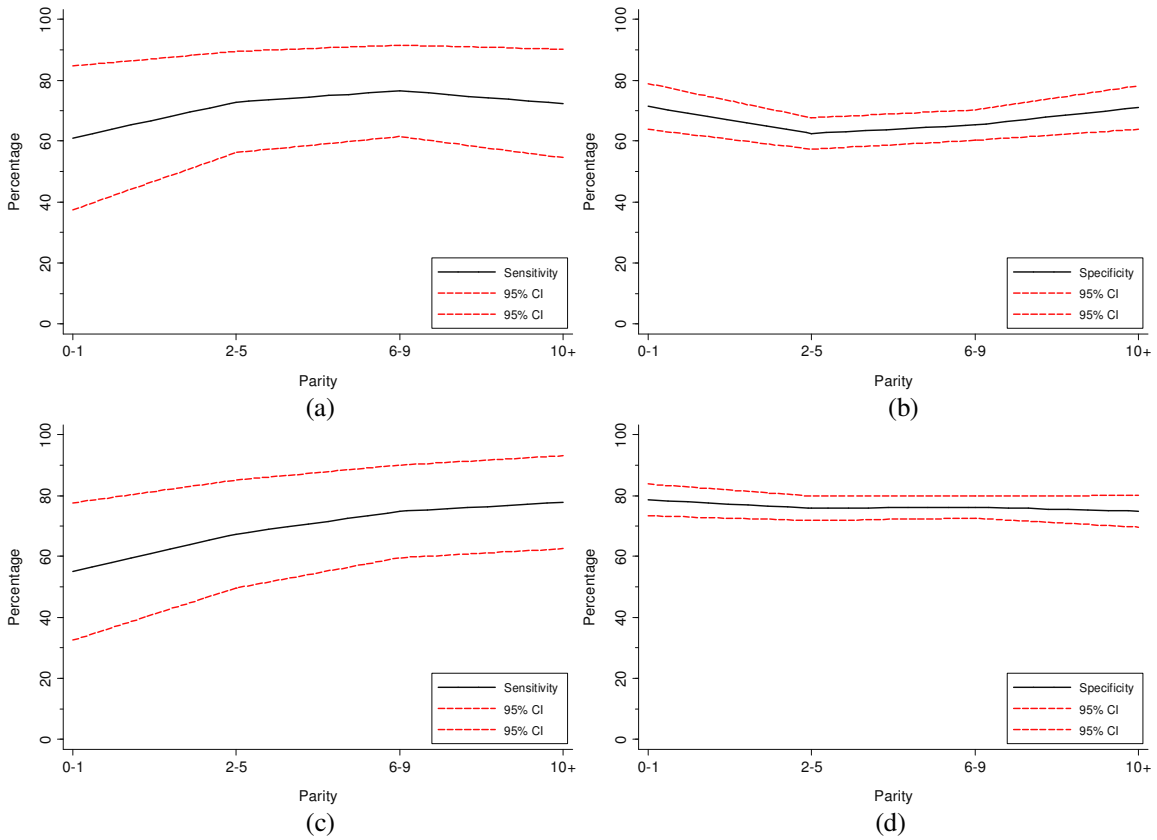

Figure 7: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by nurse across parity groups with 95% CI through logistic regression

The trends of the sensitivity and specificity of VIA and VILI tests across parity groups for the measurement taken by nurse are shown in Figure 7 respectively. From Figure 7a, the sensitivity of VIA test measured by nurse increases from 0-1 parity group up till 6-9 parity groups then decreases till 10+ parity group. In Figure 7b, the specificity of VIA test measured by nurse is in opposite direction of the sensitivity. For the VILI test, the trends of the sensitivity and specificity are similar to that of VIA test measured by nurse as could be seen in Figure 7c and Figure 7d.

The nonparametric trend tests conducted to know if there is change in the trend of the sensitivity (p-value = 0.008) and specificity (p-value < 0.001) of VIA and the sensitivity (p-value < 0.001) and specificity (p-value < 0.001) of VILI tests measured by nurse shows that there is significant difference in the trends across the parity groups.

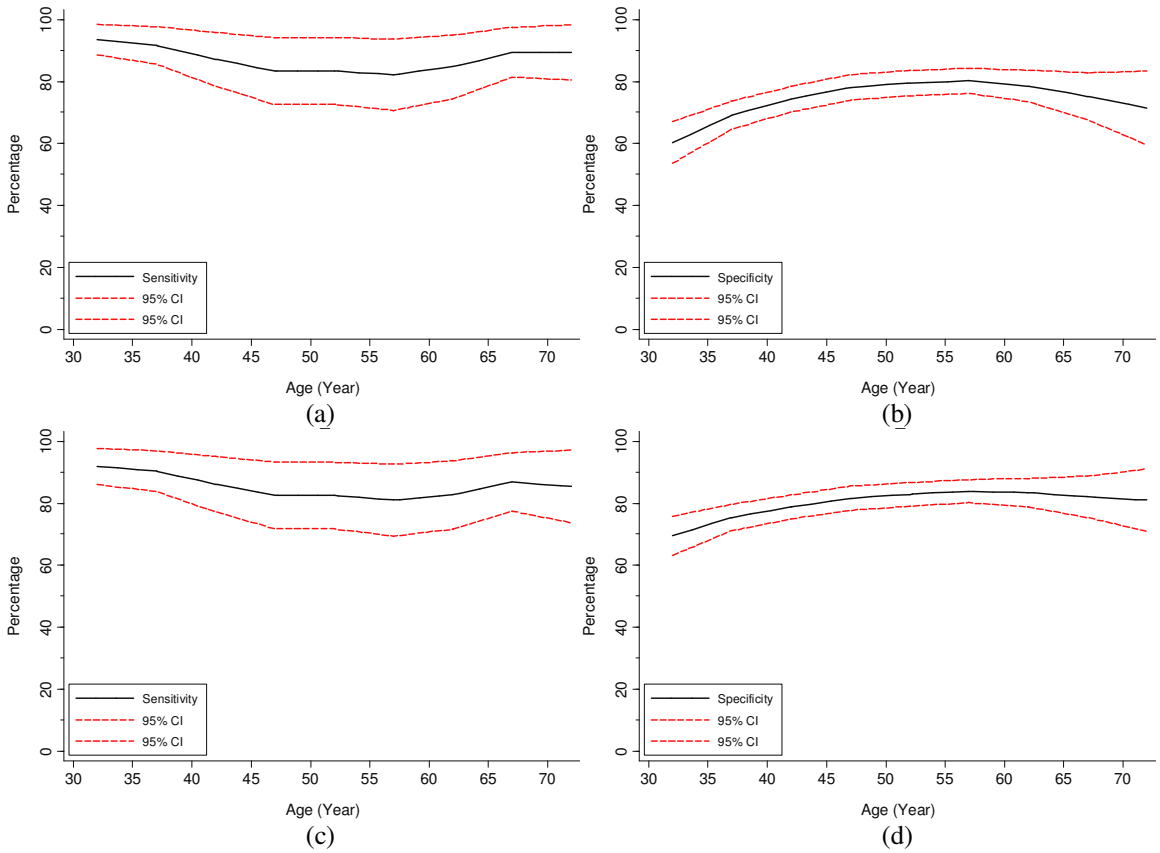

Figure 8: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across age groups with 95% CI through logistic regression

The results of other covariates presented in Table 22 and Table 23 in the Appendix show the sensitivity of VIA test measured by nurse is higher in the smokers than the sensitivity in the non-smokers categories of smoking. However, the student t-test showed that there is no significant difference between the sensitivity in the two categories based on p-value of 0.096. For marital status, the sensitivity of VIA test was also higher in the "married/cohabiting" than "other" categories with non-significant p-value of 0.051. The sensitivity of VIA test is higher in the positive category of hybrid capture2 than negative category with no significant difference (p-value = 0.259). Thus, smoking, marital status, and hybrid capture 2 measured in the cervical cancer study have no significant effects in the detection rate of VIA test measured by nurse.

The specificity of VIA test measured by nurse was significantly different between two categories of smoking and marital based on the student t-test p-value < 0.001 but not significantly different between the two levels of hybrid capture2 (p-value = 0.257). There was significant difference between sensitivity of VILI test measured by nurse in the two categories of smoking based on the student t-test p-value of 0.048. For marital and hybrid capture2, there was no significant difference between the sensitivity of VILI test with p-value of 0.261 and 0.101 respectively. The specificity of VILI test measured by nurse was also significantly different between two categories of smoking and marital based on the student t-test p-value < 0.001 but not significantly different between the two levels of

hybrid capture2 (p-value = 0.084). These t-test results confirm that there is effect of smoking on the detection rate of VILI test measured by nurse. In addition, there are also effects of smoking and marital on the specificity of VILI test measured by nurse.

The sensitivity and specificity of VIA and VILI tests measured by physician shown in Figure 8 is also smoother than the corresponding plots shown in Figure 3. As shown in Figure 8, the trend of both the sensitivity and specificity of VIA and VILI tests are in opposite directions. The trend in the sensitivity in Figure 8a and Figure 8c started high at age group 30-34 years old then gradually deceasing up till age group 55-59 years old before increasing again from age group 60-64 years old until age group 70 years old and above. The trend in specificity of VIA and VILI tests in Figure 8b and Figure 8d are similar to the trend seen in Figure 6b and Figure 6d above.

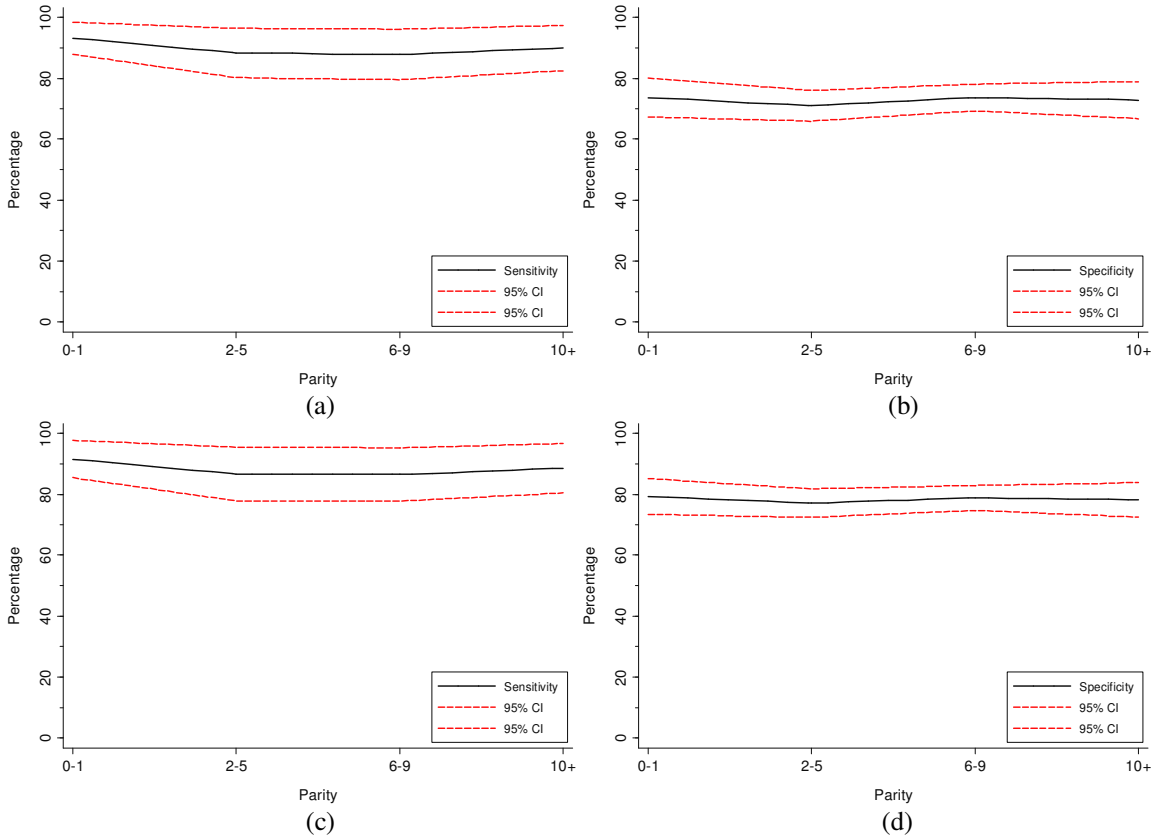

Figure 9: Plots of the estimated (a) sensitivity of VIA test, (b) specificity of VIA test, (c) sensitivity of VILI test, and (d) specificity of VILI test measured by physician across parity groups with 95% CI through logistic regression

The confidence intervals for the sensitivity are wider due to few numbers of screened women with positive test results while that of the specificity are narrower due much number of screened women with negative results in Table 20 and Table 21 (see Appendix). The nonparametric test trend test shows there is significant difference in the trends of both the sensitivity and specificity measured by physician across the age groups with p-value < 0.001.

Figure 9 shows how the trends of the estimated sensitivity and specificity of VIA and VILI tests measured by physician change across the parity groups. For the sensitivity of VIA and VILI tests, it started high from 0-1 parity group and then declined gradually until 6-9 parity group before risen up till 10+ parity group. The reverse is the case for the specificity of VIA and VILI tests as shown in Figure 9.

However, the nonparametric trend tests showed that no significant difference in the sensitivity (p-value = 0.351) and specificity (p-value = 0.535) of VIA test measured by physician. Also for the VILI test, there was no significant difference in the sensitivity (p-value = 0.739) and specificity (p-value = 0.500) across the parity groups.

The results of other covariates show that the sensitivity of VIA test measured by physician is significantly higher in the smokers than the sensitivity in the non-smokers categories of smoking with p-value of 0.024. For marital status, the sensitivity of VIA test was also higher in the "married/cohabiting" than "other" categories with non-significant p-value of 0.219. The sensitivity of VIA test is also significant higher in the positive category of hybrid capture2 than negative category with p-value < 0.001. The results of the student t-test from the other covariate measured in the cervical cancer study show that the significant effect of smoking and hybrid capture2 in the detection rate of VIA test measured by physician.

The specificity of VIA test measured by physician was significantly different between two categories of smoking, marital and hybrid capture2 based on the student t-test p-value < 0.001. These significant p-values show the effect of the covariates on the specificity of VIA test measured by physician.

4.4 Latent Class Modeling

The results of the classical LCM with conditional independence fitted for the two screening tests, VIA and VILI tests both measured by two raters, nurse and physician during cervical cancer screening in Kinshasa are presented in Table 8.

In Table 8, A denotes the VIA test measured by nurse, B denotes the VIA test measured by physician, C denotes the VILI test measured by nurse and D denotes the VILI test measured by physician. The G^2 is the likelihood ratio Chi-square statistic to assess the extent to which maximum likelihood estimates for the expected frequencies differ from the corresponding observed frequencies. As shown in Table 8, the G^2 value is high (324.80) with 6 degree of freedom. The 6 degree of freedom for G^2 is due to the four tests (A, B, C, D) outcomes that were grouped into 16 categories minus 8 parameters of the four tests (since each test has two levels), and the prevalence to be estimated with one constraint for the grand total. Based on the G^2 value and the p-value, it shows that the fit of the model is poor. Therefore, there is a need for diagnostic test so as to know which interaction terms to be included in the log-linear model. The conditional positive response probabilities for the LCM with two latent classes with latent variables X1 and X2 are shown in Table 9.

Table 8: Results of classical LCM for VIA and VILI test outcomes measured by nurse and physician

Test Outcomes A	B	C	D	Observed Freq	Estimated Freq	Standardized residual
1	1	1	1	846	794.83	1.82
1	1	1	2	7	6.35	0.26
1	1	2	1	15	61.08	-5.90
1	1	2	2	1	2.45	-0.92
1	2	1	1	26	30.82	-0.87
1	2	1	2	51	16.36	8.57
1	2	2	1	2	9.94	-2.52
1	2	2	2	18	44.18	-3.94
2	1	1	1	112	165.14	-4.14
2	1	1	2	2	4.80	-1.28
2	1	2	1	69	14.33	14.45
2	1	2	2	7	10.03	-0.96
2	2	1	1	40	19.89	4.51
2	2	1	2	34	79.82	-5.13
2	2	2	1	25	38.98	-2.24
2	2	2	2	262	218.02	2.98
G^2					324.80	
DF					6	
P-value					< 0.001	
AIC (log-likelihood)					5089.64	
BIC (log-likelihood)					5137.56	

1 = negative result; 2 = positive result

Based on the results in Table 9, the sensitivity and the specificity of the latent variable for the VIA test measured by nurse is 83% respectively. The false positive and false negative error rates for the sensitivity and specificity of VIA test measured by nurse is 17%.

Table 9: Estimated sensitivity and specificity of VIA and VILI test outcomes measured by nurse and physician with false positive and false negative error rates using classical LCM

Test	Latent class 1 (X1)	Latent class 2 (X2)
A 1	0.83	0.17
A 2	0.17	0.83
B 1	0.97	0.04
B 2	0.03	0.96
C 1	0.93	0.27
C 2	0.07	0.73
D 1	0.99	0.15
D 2	0.01	0.85
Latent class prevalence	0.71	0.29

For the VILI test measured by nurse, the sensitivity and specificity of the latent variable is also 73% and 93% respectively with false positive error rate of 27% and 7% for the false negative error rate respectively. The sensitivity and specificity of the latent variable for the VIA test measured by physician is 96% and 97% respectively with false positive error rate of 4% and 3% for the false negative error rate respectively. For the VILI test measured by physician, the sensitivity and specificity for the latent variable is 85% and

99% respectively while the false positive error rate is 0.7% and 15% for the false negative error rate respectively.

Table 10: Results of the diagnostic test for conditional dependence using classical LCM

Tests	Expected log odds ratio	Standard. Error	Observed log odds ratio	Z- value	G^2
A B	2.75	0.14	2.83	0.60	0.35
A -- C	2.02	0.13	3.91	14.53 *	135.00
A D	2.83	0.15	2.66	-1.12	1.32
B C	3.15	0.15	3.06	-0.61	0.37
B D	5.06	0.24	5.48	1.78	2.68
C -- D	3.04	0.15	3.34	2.07 *	4.02

* $p < .05$, two-tailed; A=VIA test measured by nurse; B = VIA test measured by physician; C = VILI test measured by nurse; D = VILI test measured by physician.

As shown in Table 10, VIA and VILI tests measured by nurse appear to be conditionally dependent as indicated by the large Z-value likewise VILI test measured by nurse and VILI test measured by physician. The G^2 shown in last column of Table 10 is the likelihood-ratio Chi-squared statistic comparing the observed and expected two-way table for each pair of test. For each of the pair of tests that are locally dependent, G^2 value is also large. These indicate that two way interaction terms between the pair of tests that were dependent need to be included in the loglinear model.

Table 11: Results of loglinear LCM for VIA and VILI test outcomes measured by nurse and physician

Test Outcomes A	B	C	D	Observed Freq	Estimated Freq	Standardized residual
1	1	1	1	846	845.83	0.01
1	1	1	2	7	7.20	-0.07
1	1	2	1	15	15.09	-0.02
1	1	2	2	1	0.84	0.18
1	2	1	1	26	33.61	-1.31
1	2	1	2	51	43.36	1.16
1	2	2	1	2	1.77	0.18
1	2	2	2	18	18.30	-0.07
2	1	1	1	112	112.17	-0.02
2	1	1	2	2	1.66	0.27
2	1	2	1	69	68.91	0.01
2	1	2	2	7	7.31	-0.11
2	2	1	1	40	32.39	1.34
2	2	1	2	34	41.78	-1.20
2	2	2	1	25	25.24	-0.05
2	2	2	2	262	261.55	0.03
G^2					6.51	
DF					2	
P-value					0.039	
AIC (log-likelihood)					4779.35	
BIC (log-likehood)					4848.57	

1 = negative result; 2 = positive result

The results of the loglinear model with two way interaction terms are shown in Table 11. From Table 11, it could be observed that there is an improvement on the model compared to the results in Table 9. The expected frequencies are very close to the observed

frequencies except few. The G^2 value has reduced from 324.80 to 6.51 with reduction from 6 degrees of freedom to 2 degrees of freedom. The BIC value also reduced from 5137.56 to 4848.57. These indices show that the fit of the log-linear model is better than classical LCM though the p-value is less than 5%.

To be sure that the interaction terms included into the log-linear model have taking the dependence between the pair of the tests into account, there is a need to conduct another diagnostic test on the output from the log-linear model. The conditional positive response probabilities for the log-linear model with two latent classes with latent variables X1 and X2 are shown in Table 12.

Table 12: Estimated sensitivity and specificity of VIA and VILI test outcomes measured by nurse and physician with false positive and false negative error rates using log-linear LCM

Test	Latent class 1 (X1)	Latent class 2 (X2)
A 1	0.83	0.21
A 2	0.17	0.79
B 1	1.00	0.02
B 2	0.00	0.98
C 1	0.92	0.33
C 2	0.08	0.67
D 1	0.99	0.20
D 2	0.01	0.80
Latent class prevalence	0.69	0.31

From Table 12, the sensitivity and the specificity of the latent variable for the VIA test measured by nurse is 79% and the 83% respectively. The false positive error rate for the sensitivity is 17% and the false negative error rate for the specificity of VIA test measured by nurse is 21%. For the VILI test measured by nurse, the sensitivity and specificity of the latent variable is also 67% and 92% respectively with false positive error rate of 8% and false negative error rate of 33.0% respectively. The sensitivity of the latent variable for the VIA test measured by physician is 98% and 100% respectively with false positive error rate of 0% and false negative error rate of 2% respectively. For the VILI test measured by physician, the sensitivity and specificity for the latent variable is 80% and 99% respectively while the false positive error rate is 1% and false negative error rate is 20% respectively.

Table 13: Results of the diagnostic test for conditional dependence using log-linear LCM

Tests	Expected log odds ratio	Standard. Error	Observed log odds ratio	Z- value	G^2
A B	2.83	0.14	2.83	0.01	0.00
A -- C	3.91	0.19	3.91	-0.00	0.00
A D	2.82	0.15	2.66	-1.07	1.20
B C	3.06	0.15	3.06	0.02	0.00
B D	5.48	0.27	5.48	-0.00	0.00
C -- D	3.34	0.16	3.34	-0.00	0.00

* $p < .05$, two-tailed; A=VIA test measured by nurse; B = VIA test measured by physician; C = VILI test measured by nurse; D = VILI test measured by physician.

Meanwhile, the diagnostic test conducted through CONDEP shown in Table 13 shows that the conditional dependence earlier seen in the standard LCM model is no more

present. This shows that the log-linear model with the interaction terms have resolved the problem of the local dependence. It also shows that the log-linear LCM model is better than the classical LCM model.

4.5 Measure of agreement between the screening tests raters: Kappa Statistic

Looking at the estimated Kappa values from Table 14 and Table 15, it is very difficult to ascertain if the values follow particular pattern or trend across age and parity groups. To confirm this, weighted linear regression model was fitted using the Kappa values as continuous response variable and the mid-value of age and parity groups was used as continuous predictor variables. The inverse of the square of the estimated standard error for Kappa values was then used as weight. The estimated slopes for age and parity in Table 14 and Table 15 were tested if they are different from zero using Z-test. For marital status, smoking, and hybrid capture2, Z-test for independent samples was conducted to compare the Kappa value between the two categories of each covariate.

Table 14: Estimated Kappa statistic for the agreement between nurse and physician on VIA test outcomes by covariates

Variable	Categories	Kappa statistic (95% CI)	‡ P-value	% Expected Agreement	Agreement % (95% CI)	† P-value
Age group	30-34	0.46 (0.35, 056)	0.003	49.95	73.05 (67.47, 78.14)	< 0.001
	35-39	0.54 (0.42, 0.64)	-	55.32	79.27 (73.66, 84.16)	< 0.001
	40-44	0.58 (0.47, 0.67)	-	53.29	80.53 (75.21, 85.15)	< 0.001
	45-49	0.58 (0.47, 0.68)	-	60.74	83.57 (78.75, 87.67)	< 0.001
	50-54	0.52 (0.36, 0.64)	-	61.10	81.42 (75.02, 86.78)	< 0.001
	55-59	0.71 (0.52, 0.83)	-	62.71	88.99 (81.56, 94.18)	< 0.001
	60-64	0.55 (0.29, 0.75)	-	67.61	85.51 (74.96, 92.83)	< 0.001
	65-69	0.94 (0.69, 0.99)	-	50.69	96.97 (84.24, 99.92)	< 0.001
	70+	0.76 (0.50, 0.90)	-	55.23	89.36 (76.90, 96.45)	< 0.001
Smoking	No	0.54 (0.48, 0.58)	0.009	54.46	78.89 (76.34, 81.29)	< 0.001
	Yes	0.68 (0.60, 0.75)	-	58.10	86.77 (83.21, 89.83)	< 0.001
Marital status	Married/ cohabitating	0.56 (0.51, 0.61)	0.319	55.17	80.33 (77.90, 82.61)	< 0.001
	Otherwise	0.62 (0.53, 0.70)	-	56.13	83.33 (79.19, 86.95)	< 0.001
Parity	0-1	0.60 (0.42, 0.73)	0.574	64.46	85.61 (78.66, 90.98)	< 0.001
	2-5	0.55 (0.47, 0.62)	-	51.75	78.28 (74.35, 81.86)	< 0.001
	6-9	0.58 (0.51, 0.64)	-	57.15	81.88 (78.83, 84.67)	< 0.001
	10+	0.59 (0.45, 0.70)	-	56.23	82.01 (75.78, 87.21)	< 0.001
Hybrid Capture2	Negative	0.56 (0.50, 0.60)	0.841	56.55	80.75 (78.38, 82.97)	< 0.001
	Positive	0.57 (0.44, 0.68)	-	50.08	78.70 (71.75, 84.61)	< 0.001

‡ P-value testing for Kappa trend across age and parity groups and for difference in Kappa value between two groups; † P-value for testing if Kappa value is due to chance

In Table 14 and Table 15, the p-value obtained from Kappa value trend test and Z-test for independent samples are shown in the fourth column. Meanwhile, the p-value in the seventh column of Table 14 and Table 15 was obtained from testing if the estimated Kappa value is due to chance. Based on the results from the trend test in Table 14, there is significant difference in the Kappa values across age group but not in parity group.

However in Table 15, there is no significant difference in Kappa value across both age and parity groups. No significant difference in the Kappa values between the two categories of other covariates, marital and hybrid capture2 except smoking in Table 14 and Table 15.

The estimated value for the agreement in the fifth column of Table 14 and Table 15 suggested substantial and almost perfect agreement. But, the estimated value for the expected agreement for the measurement of nurse and physician shows that there is either moderate or substantial agreement except in age group 65-69 years old in Table 14. The significant p-values on the seventh column of Table 14 and Table 15 indicate that the estimated values for the Kappa statistics were not due to chance. These significant Kappa values were driven by the large sample size of the data set.

Table 15: Estimated Kappa statistic for the agreement between nurse and physician on VILI test outcomes by covariates

Variable	Categories	Kappa statistic (95% CI)	‡ P-value	% Expected Agreement	Agreement % (95% CI)	† P-value
Age group	30-34	0.65 (0.55, 0.74)	0.803	55.90	84.75 (80.02, 88.74)	< 0.001
	35-39	0.56 (0.43, 0.66)	-	61.49	82.93 (77.63, 87.41)	< 0.001
	40-44	0.68 (0.57, 0.76)	-	56.99	86.26 (81.49, 90.19)	< 0.001
	45-49	0.66 (0.54, 0.75)	-	67.24	88.81 (84.57, 92.22)	< 0.001
	50-54	0.70 (0.54, 0.81)	-	69.33	90.71 (85.54, 94.49)	< 0.001
	55-59	0.57 (0.37, 0.73)	-	65.64	85.32 (77.26, 91.37)	< 0.001
	60-64	0.53 (0.25, 0.74)	-	72.21	86.96 (76.68, 93.86)	< 0.001
	65-69	0.59 (0.25, 0.81)	-	55.37	81.82 (64.54, 93.02)	< 0.001
	70+	0.78 (0.50, 0.91)	-	61.97	91.49 (79.62, 97.63)	< 0.001
Smoking	No	0.61 (0.56, 0.66)	0.030	61.34	85.07 (82.81, 87.14)	< 0.001
	Yes	0.74 (0.65, 0.80)	-	62.91	90.26 (87.06, 92.89)	< 0.001
Marital status	Married/ cohabitating	0.65 (0.59, 0.70)	0.939	62.12	86.65 (84.54, 88.58)	< 0.001
	Otherwise	0.64 (0.55, 0.72)	-	60.71	85.98 (82.06, 89.32)	< 0.001
Parity	0-1	0.55 (0.35, 0.71)	0.958	72.80	87.77 (81.14, 92.71)	< 0.001
	2-5	0.67 (0.60, 0.74)	-	57.95	86.27 (82.89, 89.20)	< 0.001
	6-9	0.65 (0.58, 0.71)	-	62.75	87.02 (84.30, 89.42)	< 0.001
	10+	0.58 (0.44, 0.70)	-	62.02	84.13 (78.12, 89.03)	< 0.001
Hybrid Capture2	Negative	0.63 (0.57, 0.68)	0.849	63.59	86.46 (84.37, 88.36)	< 0.001
	Positive	0.64 (0.51, 0.74)	-	51.87	82.84 (76.29, 88.20)	< 0.001

‡ P-value testing for Kappa trend across age and parity groups and for difference in Kappa value between two groups; † P-value for testing if Kappa value is due to chance

CHAPTER 5

Discussion

Estimation of the sensitivity and specificity of diagnostic or screening tests through logistic regression model has been described in the literature [9]. The primary objective of this thesis is to investigate how the covariates affect the performance of the screening tests. The secondary objectives are to adjust for false positive and false negative error rates in the screening tests through latent class modeling, and to evaluate the agreement between the ratings of the screening test outcomes by nurse and physician. In this thesis, I have generalized the method applied in Coughlin et al, 1992 [9] by including the disease-by-covariates interaction terms and the quadratic terms of age and parity into the model. This approach is to know if there is difference in the effect of age and parity between the diseased and non-diseased women screened for cervical cancer. The data set considered for thesis was measured by a nurse and a physician during cervical cancer screening through VIA and VLI tests.

Different statistical methods have been applied to achieve the objectives of this thesis. These methods could be applied to other screening data set and settings where there are more than 4 ratings. From empirical analysis, no significant trend of both the sensitivity and the specificity of VIA and VILI tests measured by nurse and by physician could be observed across age and parity groups. The plots of both the sensitivity of VIA and VILI tests have wider confidence intervals as measured by both nurse and physician. These wider confidence intervals could be attributed to small number of women with positive test results in some of the categories of age and parity. The plots of the specificity of VIA and VILI tests also have narrower confidence intervals. These could also due to much data available for those women screened with negative results.

By comparing the empirically estimated sensitivity and specificity of VIA and VILI tests measured by nurse between the two categories of other covariates, the sensitivity of VIA test is significantly different but no difference in specificity between "married/cohabiting" and "other" group of screened women. For VILI test, there is no significant difference in the sensitivity and specificity between "married/cohabiting" and "other" group of screened women. There is significant difference in the sensitivity but no difference in the specificity of VIA and VILI tests measured by nurse for hybrid capture2. However, the sensitivity and specificity of VIA and VILI test measured by nurse are significantly different between smokers and non-smokers among the screened women. Similar results were obtained when the empirically estimated sensitivity and specificity of VIA and VILI tests measured by physician were compared between "married/cohabiting" and "other" group of screened women. The sensitivity of VIA test measured by physician is significantly different between smokers and non-smokers but there is no significant difference in specificity between the two groups. For VILI test measured by physician, both the sensitivity and the specificity are not significantly different between smokers and non-smokers. The sensitivity and specificity of VIA test are significantly different between group of women with positive hybrid capture2 and

those with negative hybrid capture2. For the VILI test measured by physician, only the sensitivity is significantly different between group of women with positive hybrid capture2 and those with negative hybrid capture2.

From logistic regression analysis, age, parity and their respective quadratic terms have shown significant effect on the probability of VIA test measured by nurse to detect cervical cancer. For VILI test measured by nurse and physician, age, parity and the quadratic term of age have shown significant effects. The plots of the estimated sensitivity and specificity for both VIA and VILI tests measured by nurse and by physician are smoother than the plots in the empirical analysis. The smoothness in the plots could be due to the inclusion of the quadratic terms of the continuous age and parity as well as the shape of the logistic models. The effect of the variability, fewer and much data across the different groups of age and parity still reflected in the estimated confidence intervals. Among the covariates considered for the analysis, our results shows that there are significant effect of age and parity on the sensitivity and specificity of VIA and VILI tests measured by the physician and those measured by nurse. From other covariates in the model, there is also an effect of hybrid capture2 on the sensitivity and specificity of VIA and VILI tests measured by physician alone. The Hosmer and Lemeshow test for the goodness of fit test shows that all the final 4 models fitted for the VIA and VILI test outcomes measured by nurse and by physician have good fit.

On the latent class modeling analysis, both the classical and the log-linear LCM have been modeled. The results of the estimated sensitivity and specificity of the latent variable likewise the false negative and false positive from these two models were very similar. However, the results from the log-linear LCM were considered to be better than classical model. The reason for chosen the log-linear model was due to its ability to correct for the association between the locally dependent manifest variables.
Based on the results from log-linear LCM, the false positive error rate and the false negative error rate for VIA test measured by nurse is 17% and 21% respectively. The false positive error rate and false negative error rate for the VILI test measured by nurse is 8% and 33% respectively. For the VIA test outcomes measured by physician, false positive error rate and false negative error rate is 0% and 2% respectively. Also for the VILI test outcomes measured by physician, the false positive error rate and false negative error rate is 1% and 20% respectively. These results shows that the false negative and false positive error rate in the sensitivity and specificity of VIA and VILI tests outcomes measured by nurse is quite substantial. For the VIA and VILI tests outcomes measured by physician, the false negative and false positive error rate is small except for the specificity of VILI test.

The results of the measure of agreement for both VIA and VILI test outcomes measured by nurse and by physician shows that all the estimated Kappa statistic values are not due to chance. In addition, the estimated percentage of agreement falls between the substantial and almost perfect levels of agreement. The parametric trend test for the Kappa values between nurse and physician ratings for VIA test shows that there is difference in their level of agreement across age and parity groups. However, this is not the case for the Kappa values between nurse and physician ratings for VILI test across

age and parity groups. The level of agreement for the nurse and the physician ratings for VIA and VILI test between the two categories of marital status and hybrid capture2 is not different but significantly different between the two categories of smoking.

However, there are still some limitations to this study. These limitations arose from the measurements by nurse compared to that of physician, latent class model assumption and Kappa statistic. For the data itself, the limitation is the sample size. Looking at the trend of the sensitivity and the specificity of VIA and VILI tests measured by physician and that by nurse across the age and parity groups, it could be observed that there is difference in the trend. The sensitivity and specificity measured by physician is higher than those measured by the nurse. In another case, the observed categories measured by both the nurse and physician are different for both VIA and VILI tests outcomes. These differences between the categories of the two tests outcomes lead to dichotomizing the tests outcomes which is loss of information.

On the latent class modeling assumption for local independent, only the inclusion of the interaction terms into the log-linear latent models have been applied. Yet, the p-value for the likelihood ratio test statistic is significant but other indices for measuring the goodness of fit of the models suggest that the model is better than classical LCM model. This is one of the limitations of this study because other methods such as inclusion of other latent class have not been investigated. The limitations of the Kappa statistic on the sensitiveness of the confidence intervals of the agreement and the p-value for the Kappa statistic value as seen in the results suggest using other methods such as model-based approach.

CHAPTER 6

Conclusion

With all the statistical methods applied to achieve the objectives of this thesis, the findings show that there is significant effects of age, parity and the quadratic term of age on the performance of VIA and VILI tests outcomes measured by both nurse and physician in a low resource setting like Kinshasa. Apart from age and parity, only hybrid capture2 has also shown significant effect on the performance of VIA and VILI tests outcomes measured by physician alone.

For future cervical screening study in a low resource setting like Kinshasa, it would be good if more training is provided for nurses on the case report and definition of outcomes. This would eliminate or reduce the discrepancies in the observed categories of VIA and VILI tests outcomes between the nurse and the physician.

References

1. Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncology* 2001; **2**: 533-43.
2. Denny L, Quinn M, Sankaranarayanan R. Chapter 8: Screening for cervical cancer in developing countries. *Vaccine* 2006; S3/71-S3/77.
3. Sankaranarayanan R, Basu P, Wesley RS, et al. Accuracy of visual screening for cervical neoplasia: Results from an IARC multicentre study in India and Africa. *International Journal of Cancer* 2004; **110**: 907-913.
4. Arbyn M, Raifu AO, Autier P, Ferlay J. Burden of cervical cancer in Europe: estimates for 2004. *Annal of Oncol*ogy 2007; **18**(10):1708-15.
5. Bhatla N, Mukhopadhyay A, Joshi S, et al. Visual inspection for cervical cancer screening: evaluation by doctor versus paramedical worker. *India Journal of Cancer* 2004; **41**(1): 32-36.
6. Rothman KJ 2002. *Epidemiology*. Oxford University Press Inc., New York.
7. Walter SD. Estimation of Test Sensitivity and Specificity when Disease Confirmation is Limited to Positive Results. *Epidemiology* 1999; **10**(1): 67-72.
8. Sangwa-Lugoma G, Mahmud S, Nasr SH, et al. Visual inspection as a cervical cancer screening method in a primary health care setting in Africa. *International Journal of Cancer* 2006; **119**:1389-1395.
9. Coughlin SS, Trock B, Criqui MH, et al. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *Journal of Clinical Epidemiology* 1992; **45**(1):1-7.
10. The frequency distribution in Table 4 was generated by SAS software Version 9 of the SAS System for Windows. Copyright © 2002 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks of SAS Institute Inc., Cary, NC, USA.
11. StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
12. Vermunt, JK. 1997. LEM: A General Program for the Analysis of Categorical Data. Department of Methodology and Statistics, Tilburg University.
13. Uebersax, JS. 1999. Probit latent class analysis: conditional independence and conditional dependence models. *Applied Psychological Measurement*, 23(4), 283-297, 1999.
14. Cornfield J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. *Federation Proceedings* 1962; 21: 58-61.
15. Hosmer DW and Lemeshow S. 2000. *Applied Logistic Regression*, Second Edition. John Wiley & Sons, Inc., New York.
16. Agresti A. 2002. *Categorical Data Analysis*, Second Edition. John Wiley & Sons, Inc., New York.
17. Neter J, Kutner MH, Nachtsheim, Wasserman W (1996). *Applied Linear Statistical Models*, Fourth Edition. The McGraw-Hill Companies, Inc., New York.

18. Hosmer DW and Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics* 1980; **A10**: 1043-1069
19. Vermunt JK and Magidson J. 2000a. "Latent Class Cluster Analysis", chapter 3 in J.A. Hagenaars and A.L. McCutcheon (eds.), *Advances in Latent Class Analysis.* Cambridge University Press.
20. Vermunt JK and Magidson J. 2000b. *Latent GOLD 2.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.
21. Walter SD, Franco EL. Use of latent class models to accommodate inter-laboratory variation in assessing genetic polymorphism associated with disease risk. *BMC Genetics* 2008; **9**: 51
22. Yang I and Becker MP. Latent Variable Modeling of Diagnostic Accuracy. *Biometrics* 1997; **53**: 948-958.
23. Lazarsfeld PF and Henry NW. 1968. *Latent structure analysis*. Boston: Houghton Mill.
24. Goodman LA. Exploratory latent structure analysis using both indentifiable and unidentifiable models. *Biometrika* 1974; **61**: 215-231.
25. Clogg CC. Some latent structure models for the analysis of Likert-type data. *Social Sciences Research* 1979; **8**: 287-301.
26. Clogg CC and Goodman, LA. Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association* 1984; **79**: 762-771.
27. Uebersax JS and Groove WM. Latent class analysis of diagnostic agreement. *Statistics in Medicine* 1990; **9**: 559-572.
28. Formann AK. Measurement errors in caries diagnosis: Some further latent class models. *Biometrics* 1994; **50**: 865-871.
29. Hagenaars JA. Latent structure models with direct effects between indicators: local dependence models. *Sociological Methods and Research* 1988; **16**:379-405.
30. Garrett ES and Zeger SL. Latent class model diagnosis. *Biometrics* 2000; **56** (4): 1055-1067.
31. Espeland MA and Handelman SA. Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics* 1989; **45**: 587-599.
32. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; **20**: 37-46.
33. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968; **70**: 213-220.
34. Reichenheim ME. Confidence intervals for the kappa statistic. *The Stata Journal* 2004; **4** (4): 421-428.
35. Fleiss JL, Cohen J., and Everitt BS. Large sample standard errors for kappa and weighted kappa. *Psychological Bulletin* 1969; **72**: 323-327.
36. Cantor AB. Sample size calculations for Cohen's k. *Psychological Methods* 1996; **1**: 150-153.
37. Viera AJ and Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine* 2005; **37**(5): 360-363.
38. Cuzick, J. A Wilcoxon-type test for trend. *Statistics in Medicine* 1985; **4**: 87–90.
39. Altman, DG. 1991. *Practical Statistics for Medical Research*. London: Chapman and Hall, 215–217.

40. Stepniewska, KA. and Altman DG. snp4: Non-parametric test for trend across ordered groups. *Stata Technical Bulletin* 1992; **9**: 21–22.
41. Hui SL and Walter SD. Estimating the Error Rates of Diagnostic Tests. *Biometrics* 1980; **36**: 167-171.
42. Chu H, Zhou Y, Cole SR, and Ibrahim JG. On the estimation of disease prevalence by latent class models for screening studies using two screening tests with categorical disease status verified in test positives only. *Statistics in Medicine* 2010; **29**: 1206-1218.
43. Hui SL and Zhou XH. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**: 354-370.
44. Engel B, Backer J, and Buist W. Evaluation of the Accuracy of Diagnostic Tests from Repeated Measurements without a Gold Standard. *Journal of Agricultural, Biological, and Environmental Statistics* 2009; **15** (1): 83-100.
45. Pepe MS and Janes H. Insights into Latent Class Analysis of Diagnostic Test Performance. *Biostatistics* 2006; **1** (1): 1-18.
46. Nyankomo WM. 2006. Impact of imperfect reference tests on the evaluation of the accuracy of new cervical cancer screening tests applied in developing countries. Adjustment for gold standard misclassification using latent class modelling. University of Hasselt, Belgium. Unpublished MSc Thesis.
47. Raifu AO. and Arbyn M. Description of cervical cancer mortality in Belgium using Bayesian age-period-cohort models. *Archives of Public Health* 2009; **67** (3): 100-115.

Appendix

Table 16: Results of the univariate logistic regression for model selection

Variable	Categories	VIA test (by nurse)		VILI test (by nurse)	
		OR (95% CI)	P-value	OR (95% CI)	P-value
Age group	1 = 30-34	1		1	
	2 = 35-39	0.545 (0.384, 0.774)	0.001	0.796 (0.535, 1.183)	0.258
	3 = 40-44	0.672 (0.479, 0.942)	0.021	0.954 (0.650, 1.400)	0.808
	4 = 45-49	0.386 (0.273, 0.545)	< 0.001	0.570 (0.383, 0.850)	0.006
	5 = 50-54	0.470 (0.319, 0.691)	< 0.001	0.548 (0.343, 0.875)	0.012
	6 = 55-59	0.338 (0.208, 0.552)	< 0.001	0.787 (0.456, 1.359)	0.390
	7 = 60-64	0.227 (0.119, 0.434)	< 0.001	0.458 (0.227, 0.924)	0.029
	8 = 65-69	0.816 (0.396, 1.682)	0.582	1.248 (0.548, 2.846)	0.598
	9 = 70+	0.608 (0.323, 1.144)	0.123	0.934 (0.438, 1.989)	0.859
Province	1 = BAND	1		1	
	2 = Others	1.106 (0.840, 1.457)	0.438	0.945 (0.698, 1.282)	0.718
Religion	1 = Catholic	1		1	
	2 = Protestants	1.085 (0.838, 1.405)	0.534	1.061 (0.801, 1.406)	0.679
	3 = Others	1.050 (0.784, 1.496)	0.743	1.048 (0.763, 1.440)	0.771
Kinshasa	Continuous	0.999 (0.991, 1.007)	0.787	0.999 (0.990, 1.008)	0.845
Education	1 = None	1		1	
	2 = Elementary	1.044 (0.783, 1.393)	0.769	1.089 (0.779, 1.524)	0.618
	3 = Secondary	1.322 (0.976, 1.790)	0.071	1.147 (0.804, 1.637)	0.450
	4 = High school	1.576 (0.603, 4.119)	0.354	0.692 (0.191, 2.508)	0.576
	5 = University	Dropped (3 obs)		1.5 (0.134, 16.824)	0.742
Schoolyrs	Continuous	1.022 (0.999, 1.045)	0.060	1.02 (0.994, 1.041)	0.157
Profession	0 = None	1		1	
	1 = Manual	0.842 (0.673, 1.053)	0.131	0.880 (0.80, 1.140)	0.335
	2 = Skilled worker	0.635 (0.419, 0.961)	0.032	0.967 (0.616, 1.520)	0.886
	3 = Professional	0.768 (0.259, 2.278)	0.635	0.254 (0.032, 2.002)	0.193
SLI	1 = 0-18 (Low)	1		1	
	2 = 19-24 (Medium)	0.828 (0.639, 1.074)	0.155	0.800 (0.593, 1.0790	0.143
	3 = 25-45 (High)	0.905 (0.700, 1.171)	0.448	0.875 (0.650, 1.178)	0.379
Smoking	1 = No	1		1	
	2 = Yes	0.735 (0.578, 0.933)	0.011	0.953 (0.725, 1.252)	0.728
Marital status	1 = Married/ cohabitating	1		1	
	2 = Otherwise	0.887 (0.693, 1.134)	0.339	1.180 (0.893, 1.559)	0.245
Co-wives	1 = None	1		1	
	2 = 1+	0.918 (0.722, 1.167)	0.483	0.989 (0.752, 1.301)	0.936
Parity	1 = 0-1	1		1	
	2 = 2-5	2.275 (1.490, 3.473)	< 0.001	2.773 (1.596, 4.816)	< 0.001
	3 = 6-9	1.548 (1.023, 2.342)	0.039	2.069 (1.120, 3.566)	0.009
	4 = 10+	1.546 (0.949, 2.518)	0.080	2.373 (1.284, 4.385)	0.006
Contragrp	0 =None	1		1	
	1 = Natural	0.948 (0.763, 1.178)	0.631	0.912 (0.709, 1.172)	0.472
	2 = Medical	1.459 (0.849, 2.504)	0.171	1.336 (0.717, 2.489)	0.362
Hybrid Capture 2	0 = Negative	1		1	
	1 = Positive	1.581 (1.140, 2.193)	0.006	1.600 (1.135, 2.256)	0.007
Hybrid Capture 2a	0 = ≤ 0.75	1		1	
	1 = 0.76 – 1.25	1.107 (0.536, 2.288)	0.784	1.144 (0.523, 2.501)	0.736
	2 = 1.26 – 79.99	1.348 (0.872, 2.084)	0.179	1.321 (0.829, 2.104)	0.242
	3 = 80+	1.960 (1.192, 3.225)	0.008	2.024 (1.217, 3.366)	0.007

Table 17: Results of the univariate logistic regression for model selection

Variable	Categories	VIA test (by physician)		VILI test (by physcian)	
		OR (95% CI)	P-value	OR (95% CI)	P-value
Age group	1 = 30-34	1		1	
	2 = 35-39	0.613 (0.421, 0.892)	0.011	0.668 (0.449, 0.993)	0.046
	3 = 40-44	0.566 (0.389, 0.823)	0.003	0.823 (0.562, 1.207)	0.319
	4 = 45-49	0.413 (0.283, 0.903)	< 0.001	0.454 (0.303, 0.680)	< 0.001
	5 = 50-54	0.319 (0.200, 0.507)	< 0.001	0.400 (0.245, 0.651)	< 0.001
	6 = 55-59	0.506 (0.297, 0.862)	0.012	0.587 (0.334, 1.032)	0.064
	7 = 60-64	0.375 (0.197, 0.714)	0.003	0.367 (0.178, 0.757)	0.007
	8 = 65-69	0.884 (0.395, 1.980)	0.764	0.668 (0.272, 1.641)	0.379
	9 = 70+	0.566 (0.267, 1.199)	0.137	0.734 (0.339, 1.592)	0.434
Province	1 = BAND	1		1	
	2 = Others	0.958 (00.716, 1.281)	0.770	0.841 (0.614, 1.153)	0.283
Religion	1 = Catholic	1		1	
	2 = Protestants	1.143 (0.873, 1.498)	0.330	1.057 (0.794, 1.407)	0.705
	3 = Others	1.127 (0.832, 1.528)	0.440	1.087 (0.788, 1.499)	0.611
Kinshasa	Continuous	0.992 (0.984, 1.001)	0.088	0.995 (0.986, 1.004)	0.275
Education	1 = None	1		1	
	2 = Elementary	1.220 (0.876, 1.700)	0.240	1.285 (0.901, 1.832)	0.166
	3 = Secondary	1.614 (1.141, 2.83)	0.007	1.541 (1.063, 2.233)	0.022
	4 = High school	0.979 (0.305, 3.144)	0.972	0.888 (0.244, 3.231)	0.856
	5 = University	1.469 (0.131, 16.4700	0.755	1.923 (0.171,21.622)	0.596
Schoolyrs	Continuous	1.029 (1.004, 1.055)	0.023	1.037 (1.010, 1.065)	0.008
Profession	0 = None	1		1	
	1 = Manual	0.889 (0.693, 1.141)	0.357	0.824 (0.634, 1.072)	0.150
	2 = Skilled worker	1.040 (0.675, 1.601)	0.860	1.007 (0.641, 1.584)	0.975
	3 = Professional	0.800 (0.209, 3.057)	0.744	0.588 (0.125, 2.755)	0.500
SLI	1 = 0-18 (Low)	1		1	
	2 = 19-24 (Medium)	0.854 (0.641, 1.138)	0.281	0.869 (0.641, 1.177)	0.364
	3 = 25-45 (High)	0.882 (0.662, 1.175)	0.392	0.908 (0.671, 1.230)	0.533
Smoking	1 = No	1		1	
	2 = Yes	0.932 (0.717, 1.211)	0.597	1.045 (0.794, 1.375)	0.755
Marital status	1 = Married/ cohabitating	1		1	
	2 = Otherwise	1.088 (0.830, 1.426)	0.540	1.140 (0.858, 1.515)	0.365
Co-wives	1 = None	1		1	
	2 = 1+	1.044 (0.803, 1.356)	0.749	0.980 (0.742, 1.296)	0.889
Parity	1 = 0-1	1		1	
	2 = 2-5	2.252 (1.391, 3.646)	0.001	1.805 (1.087, 2.998)	0.02
	3 = 6-9	1.448 (0.900, 2.328)	0.127	1.482 (0.901, 2.437)	0.121
	4 = 10+	1.767 (1.021, 3.056)	0.042	1.357 (0.756, 2.436)	0.306
Contragrp	0 =None	1		1	
	1 = Natural	0.857 (0.673, 1.092)	0.212	0.799 (0.618, 1.034)	0.088
	2 = Medical	1.411 (0.775, 2.570)	0.260	1.355 (0.727, 2.525)	0.339
Hybrid Capture 2	0 = Negative	1		1	
	1 = Positive	2.557 (1.840, 3.555)	< 0.001	2.711 (1.939, 3.789)	< 0.001
Hybrid Capture 2a	0 = ≤ 0.75	1		1	
	1 = 0.76 – 1.25	1.177 (0.551, 2.514)	0.673	0.783 (0.319, 1.923)	0.594
	2 = 1.26 – 79.99	1.648 (1.058, 2.567)	0.027	1.783 (2.130, 2.814)	0.013
	3 = 80+	5.180 (3.055, 8.782)	< 0.001	5.566 (3.321, 9.327)	< 0.001

Table 18: Estimated sensitivity and specificity of VIA test measured by the Nurse with 95% confidence interval using empirical method

Variable		TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Total		57	21	498	952	1528	73.1 (61.8, 82.5)	65.7 (63.1, 68.1)
Age (Year)	30-34	10	5	133	135	283	66.7 (38.4, 88.2)	50.4 (44.2, 56.5)
	35-39	11	5	77	153	246	68.8 (41.3, 89.0)	66.5 (60.0, 72.6)
	40-44	15	1	92	155	263	93.8 (69.8, 99.8)	62.8 (56.4, 68.8)
	45-49	6	1	76	207	290	85.7 (42.1, 99.6)	73.1 (67.6, 78.2)
	50-54	5	2	55	123	185	71.4 (29.0, 96.3)	69.1 (61.8, 75.8)
	55-59	5	2	23	79	109	71.4 (29.0, 96.3)	77. 5 (68.1, 85.1)
	60-64	1	3	12	53	69	25.0 (0.6, 80.6)	81.5 (70.0, 90.1)
	65-69	1	0	14	18	33	100.0 (2.5, 100.0)	56.3 (37.7, 73.6)
	70+	3	2	15	27	47	60.0 (14.7, 94.7)	64.3 (48.0, 78.4)
Marital	Married, Cohabiting	42	11	383	710	1146	79.2 (65.9, 89.2)	65.0 (62.0, 67.8)
	Others	15	10	115	242	382	60.0 (38.7, 78.9)	67.8 (62.7, 72.6)
Parity	0-1	1	2	34	103	140	33.3 (0.8, 90.6)	75.2 (67.1, 82.2)
	2-5	22	8	190	269	489	73.3 (54.1, 87.7)	58.6 (53.9, 63.2)
	6-9	28	8	213	453	702	77.8 (60.8, 89.9)	68.0 (64.3, 71.5)
	10+	6	2	58	123	189	75.0 (34.9, 96.8)	68.0 (60.6, 74.7)
Smoking	No	33	14	385	660	1092	70.2 (55.1, 82.7)	63.2 (60.2, 66.1)
	Yes	24	7	112	291	434	77.4 (58.9, 90.4)	72.2 (67.6, 76.5)
Hybrid Capture2	Negative	20	11	396	756	1183	64.5 (45.4, 80.8)	65.6 (62.8, 68.4)
	Positive	31	8	46	84	169	79.5 (63.5, 90.7)	64.6 (55.8, 72.8)

TP: True Positive; FN: False Negative; FP: False Positive; TN: True Negative

Table 19: Estimated sensitivity and specificity of VILI test measured by the Nurse with 95% confidence interval using empirical method

Variable		TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Total		55	23	346	1104	1528	70.5 (59.1, 80.3)	76.1 (73.9, 78.3)
Age (Year)	30-34	8	7	83	185	283	53.3 (26.6, 78.7)	69.0 (63.1, 74.5)
	35-39	11	5	55	175	246	68.8 (41.3, 89.0)	76.1 (70.0, 81.4)
	40-44	15	1	68	179	263	93.8 (69.8, 99.8)	72.5 (66.4, 77.9)
	45-49	5	2	56	227	290	71.4 (29.0, 96.3)	80.2 (75.1, 84.7)
	50-54	5	2	34	144	185	71.4 (29.0, 96.3)	80.9 (74.3, 86.4)
	55-59	5	2	20	82	109	71.4 (29.0, 96.3)	80.4 (71.4, 87.6)
	60-64	2	2	10	55	69	50.0 (6.8, 93.2)	84.6 (73.5, 92.4)
	65-69	1	0	11	21	33	100.0 (2.5, 100.0)	65.6 (46.8, 81.4)
	70+	3	2	9	33	47	60.0 (14.7, 94.7)	78.6 (63.2, 89.7)
Marital	Married, Cohabiting	38	15	258	835	1146	71.7 (57.7, 83.2)	76.4 (73.8, 78.9)
	Others	17	8	88	269	382	68.0 (46.5, 85.1)	75.4 (70.5, 79.7)
Parity	0-1	1	2	19	118	140	33.3 (0.8, 90.6)	86.1 (79.2, 91.4)
	2-5	23	7	130	329	489	76.7 (57.7, 90.1)	71.7 (67.3, 75.8)
	6-9	25	11	149	517	702	69.4 (51.9, 83.7)	77.6 (74.3, 80.7)
	10+	6	2	46	135	189	75.0 (34.9, 96.8)	74.6 (67.6, 80.8)
Smoking	No	30	17	263	782	1092	63.8 (48.5, 77.3)	74.8 (72.1, 77.4)
	Yes	25	6	82	321	434	80.6 (62.5, 92.5)	79.7 (75.4, 83.5)
Hybrid Capture2	Negative	18	13	279	873	1183	58.1 (39.1, 75.5)	75.8 (73.2, 78.2)
	Positive	31	8	28	102	169	79.5 (63.5, 90.7)	78.5 (70.4, 85.2)

TP: True Positive; FN: False Negative; FP: False Positive; TN: True Negative

Table 20: Estimated sensitivity and specificity of VIA test measured by the Physician with 95% confidence interval using empirical method

Variable		TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Total		68	10	394	1056	1528	87.2 (77.7, 93.7)	72.8 (70.5, 75.1)
Age (Year)	30-34	13	2	110	158	283	86.7 (59.5, 98.3)	59.0 (52.8, 64.9)
	35-39	16	0	61	169	246	100.0 (79.4, 100.0)	73.5 (67.3, 79.1)
	40-44	16	0	69	178	263	100.0 (79.4, 100.0)	72.1 (66.0, 77.6)
	45-49	6	1	65	218	290	85.7 (42.1, 99.6)	77.0 (71.7, 81.8)
	50-54	5	2	32	146	185	71.4 (29.0, 96.3)	82.0 (75.6, 87.4)
	55-59	5	2	21	81	109	71.4 (29.0, 96.3)	79.4 (70.3, 86.8)
	60-64	3	1	12	53	69	75.0 (19.4, 99.4)	81.5 (70.0, 90.1)
	65-69	1	0	13	19	33	100.0 (2.5, 100.0)	59.4 (40.6, 76.3)
	70+	3	2	10	32	47	60.0 (14.7, 94.7)	76.2 (60.5, 87.9)
Marital	Married, Cohabiting	47	6	298	795	1146	88.7 (77.0, 95.7)	72.7 (70.0, 75.4)
	Others	21	4	96	261	382	84.0 (63.9, 95.5)	73.1 (68.2, 77.6)
Parity	0-1	2	1	27	110	140	66.7 (9.4, 99.2)	80.3 (72.6, 86.6)
	2-5	27	3	155	304	489	90.0 (73.5, 97.9)	66.2 (61.7, 70.5)
	6-9	31	5	159	507	702	86.1 (70.5, 95.3)	76.1 (72.7, 79.3)
	10+	7	1	51	130	189	87.5 (47.3, 99.7)	71.8 (64.7, 78.2)
Smoking	No	42	5	296	749	1092	89.4 (76.9, 96.5)	71.7 (68.8, 74.4)
	Yes	26	5	96	307	434	83.9 (66.3, 94.5)	76.2 (71.7, 80.3)
Hybrid Capture2	Negative	24	7	305	847	1183	77.4 (58.9, 90.4)	73.5 (70.9, 76.1)
	Positive	38	1	45	85	169	97.4 (86.5, 99.9)	65.4 (56.5, 73.5)

TP: True Positive; FN: False Negative; FP: False Positive; TN: True Negative

Table 21: Estimated sensitivity and specificity of VILI test measured by the Physician with 95% confidence interval using empirical method

Variable		TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Total		67	11	318	1132	1528	85.9 (76.2, 92.7)	78.1 (75.8, 80.2)
Age (Year)	30-34	13	2	83	185	283	86.7 (59.5, 98.3)	69.0 (63.1, 74.5)
	35-39	16	0	46	184	246	100.0 (79.4, 100.0)	80.0 (74.2, 85.0)
	40-44	15	1	67	180	263	93.8 (69.8, 99.8)	72.9 (66.9, 78.3)
	45-49	6	1	51	232	290	85.7 (42.1, 99.6)	82.0 (77.0, 86.3)
	50-54	5	2	27	151	185	71.4 (29.0, 96.3)	84.8 (78.7, 89.8)
	55-59	5	2	18	84	109	71.4 (29.0, 96.3)	82.4 (73.6, 89.2)
	60-64	3	1	8	57	69	75.0 (19.4, 99.4)	87.7 (77.2, 94.5)
	65-69	1	0	9	23	33	100.0 (2.5, 100.0)	71.9 (53.3, 86.3)
	70+	3	2	9	33	47	60.0 (14.7, 94.7)	78.6 (63.2, 89.7)
Marital	Married, Cohabiting	46	7	239	854	1146	86.8 (74.7, 94.5)	78.1 (75.6, 80.6)
	Others	21	4	79	278	382	84.0 (63.9, 95.5)	77.9 (73.2, 82.1)
Parity	0-1	2	1	23	114	140	66.7 (9.4, 99.2)	83.2 (75.9, 89.0)
	2-5	26	4	114	345	489	86.7 (69.3, 96.2)	75.2 (70.9, 79.1)
	6-9	31	5	142	524	702	86.1 (70.5, 95.3)	78.7 (75.4, 81.7)
	10+	7	1	37	144	189	87.5 (47.3, 99.7)	79.6 (72.9, 85.2)
Smoking	No	41	6	236	809	1092	87.2 (74.3, 95.2)	77.4 (74.8, 79.9)
	Yes	26	5	80	323	434	83.9 (66.3, 94.5)	80.1 (75.9, 83.9)
Hybrid Capture2	Negative	23	8	243	909	1183	74.2 (55.4, 88.1)	78.9 (76.4, 81.2)
	Positive	38	1	36	94	169	97.4 (86.5, 99.9)	72.3 (63.8, 79.8)

TP: True Positive; FN: False Negative; FP: False Positive; TN: True Negative

Table 22: Estimated sensitivity and specificity from logistic regression model using VIA and VILI test outcomes measured by the Nurse with their 95% confidence interval

Variable	Category	VIA		VILI	
		Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Age (Year)	30-34	69.8 (49.6, 89.9)	52.5 (46.0, 59.0)	59.6 38.0, 81.1)	69.3 (64.0, 74.7)
	35-39	77.8 (66.6, 89.0)	60.5 (55.6, 65.4)	71.9 (59.7, 84.1)	73.7 (70.2, 77.3)
	40-44	79.7 (68.5, 91.0)	66.8 (62.0, 71.5)	77.9 (66.5, 89.4)	76.6 (73.3, 79.8)
	45-49	80.6 (68.0, 93.2)	71.8 (66.8, 76.8)	80.2 (67.5, 92.9)	78.9 (75.5, 82.5)
	50-54	78.8 (64.2, 93.4)	73.4 (68.5, 78.2)	81.9 (69.0, 94.8)	80.0 (76.6, 83.4)
	55-59	75.3 (59.2, 91.4)	74.6 (69.5, 79.7)	77.5 (62.3, 92.6)	80.6 (76.9, 84.2)
	60-64	71.3 (52.9, 89.7)	72.5 (66.5, 78.5)	73.7 (56.5, 90.9)	79.7 (75.2, 84.1)
	65-69	59.7 (33.5, 86.0)	69.0 (60.7, 77.2)	64.8 (39.8, 89.7)	78.1 (71.5, 84.7)
	70+	45.0 (5.6, 84.4)	63.6 (51.4, 75.8)	50.0 (10.4, 89.6)	76.4 (66.2, 86.5)
Marital	Married, Cohabiting	75.8 (60.8, 90.8)	64.9 (59.5, 70.3)	72.3 (56.9, 87.6)	75.6 (71.6, 79.6)
	Others	70.2 (51.2, 89.1)	68.0 (62.1, 73.9)	69.6 (50.7, 88.5)	77.6 (73.2, 82.0)
Parity	0-1	61.1 (37.5, 84.6)	71.4 (63.8, 78.9)	55.1 (32.5, 77.6)	78.5 (73.3, 83.7)
	2-5	72.8 (56.2, 89.4)	62.5 (57.4, 67.6)	67.4 (49.7, 85.1)	75.8 (71.8, 79.8)
	6-9	76.5 (61.5, 91.5)	65.2 (60.2, 70.2)	74.8 (59.6, 90.0)	76.2 (72.5, 79.8)
	10+	72.3 (54.5, 90.1)	71.0 (63.9, 78.2)	77.8 (62.5, 93.0)	74.9 (69.6, 80.1)
Smoking	No	75.6 (60.6, 90.7)	63.3 (57.9, 68.7)	69.6 (53.5, 85.8)	75.5 (71.4, 79.7)
	Yes	71.6 (53.6, 89.7)	71.8 (65.8, 77.8)	74.1 (57.1, 91.0)	77.6 (73.6, 81.7)
Hybrid Capture2	Negative	75.4 (59.6, 91.2)	65.6 (60.1, 71.2)	73.3 (57.3, 89.3)	76.1 (72.0, 80.2)
	Positive	72.9 (56.3, 89.5)	64.7 (59.2, 70.1)	69.5 (52.6, 86.4)	75.4 (71.4, 81.0)

Table 23: Estimated sensitivity and specificity from logistic regression model using VIA and VILI test outcomes measured by the Physician with their 95% confidence interval

Variable	Category	VIA		VILI	
		Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Age (Year)	30-34	93.5 (88.6, 98.4)	60.3 (53.6, 67.0)	91.8 (86.0, 97.6)	69.5 (63.2, 75.8)
	35-39	91.6 (85.6, 97.6)	69.1 (64.6, 73.7)	90.3 (83.8, 96.9)	75.4 (71.1, 79.6)
	40-44	87.1 (78.4, 95.8)	74.2 (70.2, 78.4)	86.1 (77.2, 95.0)	78.8 (74.9, 82.6)
	45-49	83.3 (72.5, 94.1)	78.1 (73.9, 82.2)	82.5 (71.7, 93.2)	81.6 (77.7, 85.5)
	50-54	83.3 (72.6, 94.1)	79.5 (75.4, 83.5)	82.5 (71.7, 93.2)	82.8 (79.0, 86.6)
	55-59	82.1 (70.7, 93.6)	80.2 (76.1, 84.3)	81.0 (69.4, 92.6)	83.9 (80.1, 87.6)
	60-64	84.7 (74.4, 95.0)	78.6 (73.5, 83.6)	82.6 (71.6, 93.6)	83.4 (78.8, 88.0)
	65-69	89.4 (81.4, 97.4)	75.2 (67.6, 82.8)	86.9 (77.4, 96.3)	82.0 (75.3, 88.7)
	70+	89.4 (80.5, 98.2)	71.5 (59.7, 83.3)	85.4 (73.7, 97.0)	81.0 (71.0, 91.0)
Marital	Married, Cohabiting	88.9 (81.3, 96.5)	72.1 (67.2, 76.9)	87.6 (79.5, 95.7)	77.5 (73.1, 82.2)
	Others	87.3 (78.5, 96.2)	74.6 (69.3, 80.0)	85.6 (76.0, 95.2)	79.9 (75.0, 84.8)
Parity	0-1	93.1 (87.8, 98.4)	73.6 (67.3, 80.0)	91.5 (85.4, 97.7)	79.3 (73.4, 85.1)
	2-5	88.3 (80.2, 96.4)	71.0 (65.9, 76.0)	86.6 (77.8, 95.4)	77.0 (72.4, 81.7)
	6-9	87.8 (79.5, 96.1)	73.7 (69.3, 78.0)	86.5 (77.8, 95.2)	78.8 (74.7, 82.8)
	10+	89.9 (82.5, 97.3)	72.8 (66.7, 78.9)	88.5 (80.4, 96.6)	78.1 (72.5, 83.8)
Smoking	No	89.4 (82.1, 96.8)	71.7 (66.7, 76.8)	88.0 (80.0, 95.9)	77.5 (72.8, 82.2)
	Yes	86.7 (77.6, 95.8)	75.2 (70.3, 80.1)	85.3 (75.6, 94.9)	80.0 (75.4, 84.5)
Hybrid Capture2	Negative	84.5 (74.3, 94.7)	73.7 (69.2, 78.3)	82.9 (72.1, 93.6)	80.0 (74.9, 83.3)
	Positive	91.4 (85.1, 97.7)	63.6 (54.6, 72.5)	90.1 (83.2, 97.0)	70.1 (61.5, 78.7)