2010
2011

FACULTY OF SCIENCES

Master of Statistics: Epidemiology & Public Health Methodology

Masterproef

Diagnostic accuracy evaluation fo screening tests for detecting high grade cervical leasions in Flanders

Promotor :
Prof. dr. Marc AERTS

Promotor :
Dr. GAËLLE BOULET

Raoul Ouandji Djiokep

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Epidemiology & Public Health Methodology

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

2010
2011

FACULTY OF SCIENCES

Master of Statistics: Epidemiology & Public Health Methodology

Masterproef

Diagnostic accuracy evaluation fo screening tests for detecting high grade cervical leasions in Flanders

Promotor :
Prof. dr. Marc AERTS

Promotor :
Dr. GAËLLE BOULET

Raoul Ouandji Djiokep

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Epidemiology & Public Health Methodology

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Interuniversity Institute for Biostatistics and Statistical Bioinformatics

Universiteit Hasselt

DIAGNOSTIC ACCURACY EVALUATION OF SCREENING TESTS FOR DETECTING CIN2+ LESIONS IN FLANDERS

By: *Ouandji Djiokep Raoul*

Supervisors: *Prof. Dr. Marc Aerts*
Prof. Dr. Johannes Bogers
Dr. Gaëlle Boulet

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Statistics: Epidemiology and Public Health Methodology

September 2011.

CERTIFICATION

This is to certify that this report was written by *Ouandji Djiokep Raoul* under our Supervision.

Ouandji Djiokep Raoul

Student

……..

Signature

…….

Date

Prof. Dr. Marc Aerts

Supervisor

……

Signature

……..

Date

Prof. Dr. Johannes Bogers

Supervisor

……

Signature

……..

Date

Dr. Boulet Gaëlle

Supervisor

……

Signature

……..

Date

ACKNOWLEDGEMENT

First and foremost I am grateful to the Almighty God for the health and intercessions bestowed upon me to complete this master programme. I wish to express my sincere gratitude to my supervisors Dr. Boulet Gaëlle for her thorough advises and guidance, Prof. Dr. Marc Aerts and Prof. Dr. Bogers Johannes for their insights, suggestions and pertinent questions during discussions.

My gratitude goes to my friends Nico Ntellah, Elvis Ndah, Abatih Emmanuel, John Ediebah alongside all the CamUHasselt fellows and UB coursemates (Leon, Sylvia, Mirande, Alino, Amega, Parfait) for their advises and moral support.

My very special thanks go to my parents (Mr. & Mrs. Djiokep), brothers (Rodrigue, Hippolyte, Malcom), sisters (Viviane, Clemence, Arlette, Amelie) for their constant support, encouragement, love and without whom it would not be feasible to complete this master of statistics programme.

SUMMARY

Cervical cancer is the second most common cancer among women worldwide and about 54% of the disease is caused by the most common high-risk human papillomavirus (HR-HPV) infection, HPV16. Screening methods such as cytology and HR-HPV testing present some important shortcomings and require additional testing to distinguish transforming from productive HPV infections. It is apparent that future screening should combine multiple diagnostic tests to identify women at risk for cervical cancer and reduce unnecessary follow-up of transient HPV infections. It emerged from this thesis that via the aid of a logistic regression model and the AUC as discriminatory tool of classification accuracy, markers combinations such as the viral load and the amount of HPV E2 genes per cell on the one hand, and the viral integration state and cytological outcome on the other hand could enhance the sensitivity of cervical cancer screening programmes in Flanders. More specifically, it resulted that the combination of cytological evaluation with the viral integration state has a better discriminatory ability for CIN2+ lesions compared to the combination of viral load with the amount of HPV E2 genes per cell.

Keywords: discriminatory ability; cytology; integration state; risk model; ROC curve; viral load.

Contents

List of Tables

List of Figures

1. INTRODUCTION

Biomarker development is a major focus of cancer research. Biomarkers such as clinical signs, symptoms of disease and laboratory tests results are used to detect subclinical cancer, make precise diagnoses, predict disease outcome, and monitor disease progression or response to treatment (Pepe *et al.*, 2004). The efficacy of secondary prevention through screening depends on the natural history of the disease, the sensitivity of the screening test and the screening policy. It is expected that multiple biomarkers will be needed for detecting subclinical cancer with adequate sensitivity and specificity (Pepe *et al.*, 2001). In this respect, statistical modelling offers a valuable approach to identify statistically and/or clinically useful biomarkers and enable risk stratification.

It is firmly established that high-risk human papillomavirus (HR-HPV) infections are responsible for approximately 5% of all cancers worldwide, causing cervical cancer, other anogenital tumors and head-and-neck cancers (Boulet, 2010). The focus of this thesis is on cervical cancer, which is the second most common cancer among women worldwide, and more specifically on the most common high-risk human papillomavirus (HR-HPV) infection, HPV16, that causes about 54% of all cervical cancers (Clifford *et al.*, 2006).

As stated by Pepe *et al.* (2001), multiple biomarkers and clinical data are needed to optimize cancer prediction/classification and to detect subclinical cancer with adequate sensitivity and specificity. The purpose of this thesis is to evaluate the performance of viral load and integration state as biomarkers for cervical disease and to assess performance improvements when combined with cytological evaluation.

The data considered in this thesis consist of patient characteristics, the cytological outcome and polymerase chain reaction (PCR) results of cervical samples retrieved retrospectively from women classified as case or control based on their histopathological diagnosis. Boulet *et al.* (2009) primarily analyzed the data using a contingency table analysis and the area under the receiver-operating characteristic curve (ROC) approach to compare the distributions of the viral load and

integration state between cases and controls, and to evaluate the diagnostic accuracy of the markers in distinguishing cases from controls. This thesis extends Boulet *et al.*'s work and adopts more classy statistical modelling tools that account for the heterogeneity induced by longitudinal measurements from some women in the data.

Two statistical approaches have basically been implemented for biomarkers evaluation. The first assesses the biomarker(s) by its effects on the risk of disease (or disease outcome) conditional on other predictors via a risk model, such as the logistic regression model. Moons & Harrell (2003) argued in favour of this approach since it enables one to know subjects' risk of disease given biomarker(s) measurements and identifies markers that are strongly related to risk. Pepe *et al.* (2004) however documented that this approach is adequate for etiologic purposes, but a marker strongly related to risk may be a poorly performing classifier. They rather advocated to summarize the marker performance with accuracy measures such as the area under the ROC curve when the purpose is classification. In this thesis, risk modelling supplemented by ROC analysis was performed to avoid being misled by summary measures, such as the misclassification rate or the magnitude of the odds ratio, in assessing the discriminatory ability of the resulting classifier(s). More specifically, a multiple logistic regression approach was primarily adopted since the aim was to classify subjects as cases/controls based on multiple factors. Then, analysis improvements such as the generalized estimating equations (GEE) and generalized linear mixed models (GLMM) were implemented to account for the underlying longitudinal nature of the data. The results were summarized via the traditional measures of accuracy used in medicine, the true- and false-positive fractions (TPF and FPF, see figure 1), also known as sensitivity and $1 - specificity$ respectively, in order to obtain the ROC curves.

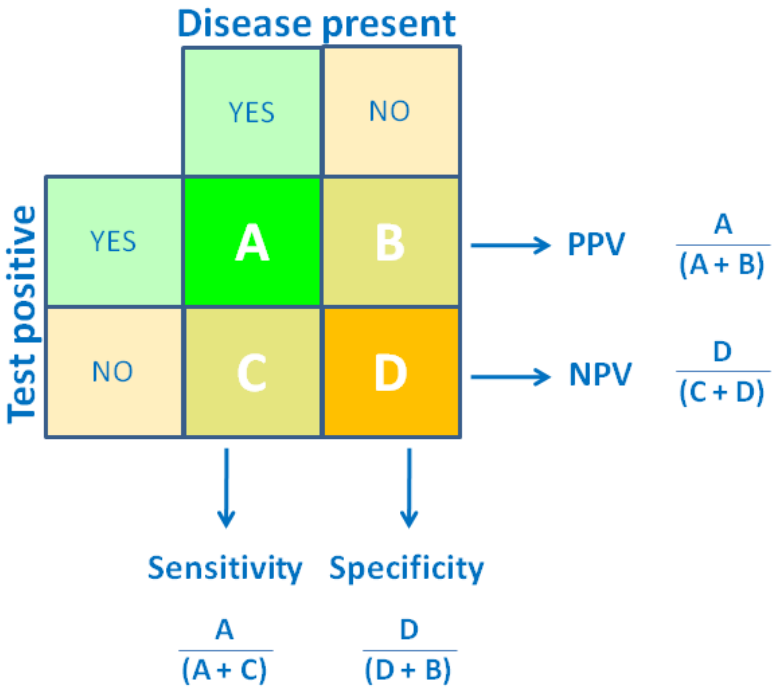

Figure 1: Measures of accuracy of diagnostic tests: definition.

Sensitivity: proportion of disease-positives that are test-positive. *Specificity*: proportion of disease-negatives that are test-negative; *Positive Predictive Value (PPV)*: proportion of test-positive subjects that are correctly identified as disease; *Negative Predictive Value (NPV)*: proportion of test-negative subjects that are correctly identified as healthy (Adapted from Boulet, 2010).

The remainder of this thesis is organised as follows. Section 2 presents some background on the HPV-induced carcinogenesis and the screening methodologies. A brief overview of the data then follows in section 3. Description of the screening approach, sample collection and the statistical analysis methods are given in section 4. Section 5 is devoted to the implemented analyses results and section 6 paraphrases the report with some remarks, discussions and ideas for further works.

2. BACKGROUND

HPV is a small non-enveloped virus with a 55-nm diameter icosahedral capsid that contains a double-stranded DNA genome of approximately 8000 base pairs (bp). Eight open reading frames (ORFs) are encoded, which are transcribed as polycistronic messenger RNAs (mRNAs). The viral proteins are classified as `early´ (E1-E2, E4-E7) or `late´ (L1, L2) depending on the time of expression during the viral life cycle (Fehrmann & Laimins, 2003). E6, E7 and E2 are the most important proteins in the HPV-induced cervical carcinogenesis.

2.1. Role of HPV in the cervical carcinogenesis

In a nutshell, the HPV E2 gene is disrupted during integration of the viral DNA into the host genome. Loss of the E2 gene, which functions as a transcriptional repressor of the E6/E7 expression, results in overexpression of E6 and E7 oncoproteins that inactivate the tumor suppressor genes p53 and pRb, respectively. As a consequence, constant activity of HPV E6 and E7 leads to increasing genomic instability, accumulation of oncogenes mutations, further loss of cell-growth control and ultimately cancer (Muñoz *et al.*, 2006; Boulet, 2010).

HPV is one of the most common sexually transmitted infections. The lifetime incidence of cervical infection is estimated to be as high as 80%. Most HPV infections are transient and spontaneously cleared by host's immune responses or suppressed into long-term latency. The remainder produces dysplastic squamous intraepithelial lesions (SILs), histopathologically described as CIN, which may also regress (see figure 2). Mild dysplasia, low-grade SIL (LSIL) or CIN1, represents an established HPV infection in which HPV DNA is episomal with the occurrence of intact virion production and shedding. Compared with clearance, viral persistence is uncommon (only 10 to 20% of genital HPV infections). Clonal progression of the persistently-infected epithelium leads to severe cervical intraepithelial neoplasia, CIN3 or high-grade SIL (HSIL), which represents the precursor lesion of cervical cancer. This carcinogenic progression is associated with integration of the viral DNA, loss

of E2, deregulated viral oncogene expression and genomic instability. The timing from infection to CIN3 varies from 2 to 15 years and further progression to cancer typically takes an additional 10 or more year (Moscicki *et al.*, 2006). These well-defined stages from non-invasive HPV infection to cervical cancer via carcinogenic progression make screening for the disease very amenable.

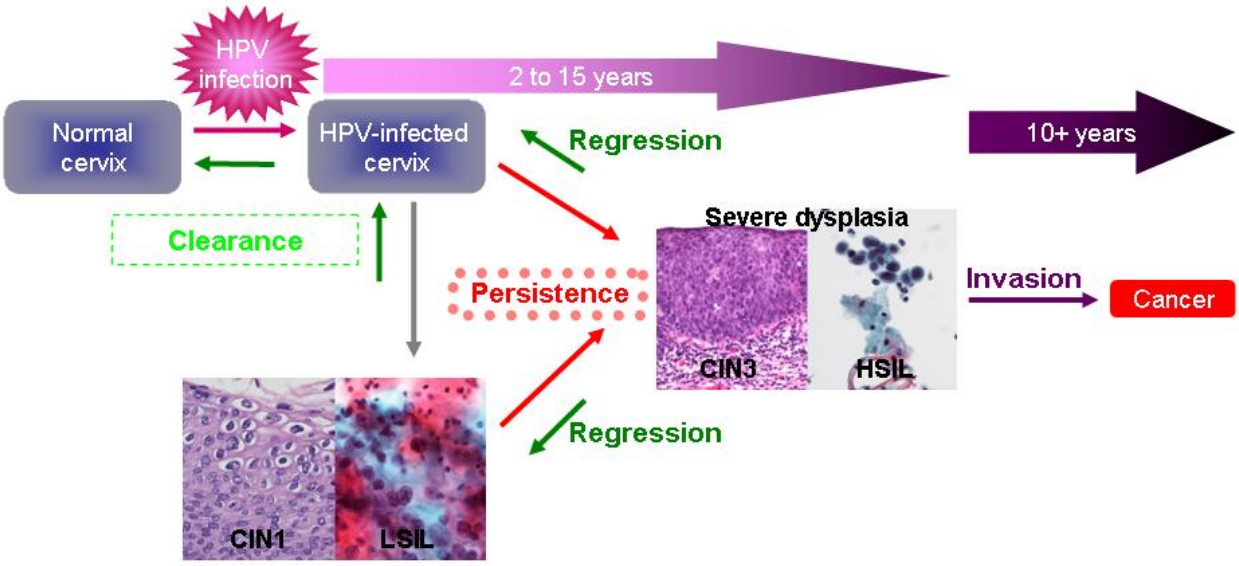

Figure 2: Major steps in the cervical carcinogenesis: HPV infection, viral persistence rather than clearance, progression to CIN3 and invasion (Adapted from Boulet, 2010).

2.2. Screening Methods

Several methodologies are available for cervical cancer screening. Cervical cytology is the standard screening approach and involves the morphological evaluation of cervico-vaginal (or Pap) smears to identify cancer (precursor) cells, followed by histological confirmation of the diagnosis on excised tissue. This approach presents some shortcomings such as a low sensitivity of high-grade cervical lesions (CIN2+) and a poor reproducibility (Boulet, 2010, chapter 2; European Commission, 2008, chapter 3).

The recognition of the strong causal relationship between persistent infection with HR-HPV and cervical cancer has resulted in the introduction of HPV DNA testing in clinical practice. HR-HPV

DNA detection has proven to be more sensitive than cytology in identifying CIN2+. However, due to the high prevalence of transient infections, its specificity and positive predictive value (PPV) are low. Therefore, the introduction of HR-HPV testing as a screening tool requires additional testing to distinguish transforming from productive HPV infections. Cytological triage has proven to be a feasible strategy, but biomarkers associated with lesion progression could also be useful. HPV type, load, integration status, E6/E7 mRNA and cell-cycle regulating proteins are candidate markers under investigation to increase the sensitivity of cervical screening with a reduction in unnecessary follow-up of transient infections. Overall, it becomes more apparent that future screening should combine multiple diagnostic tests to identify women at risk for cancer. In the conventional multi-step algorithms, the most sensitive test, i.e. HPV DNA detection, would be performed first, followed by the most specific triage assay. However, the diagnostic accuracy might be improved by using all diagnostic and clinical information available through the application of appropriate statistical models.

3. DATA

The data were collected from the Laboratory for Clinical Pathology (Labo Lokeren, Campus RIATOL, Antwerp, Belgium), which receives cervical samples for cytological evaluation. Women with known histological outcome and an HPV16 infection in their clinical history were selected and classified into cases or controls based on their histopathological diagnosis. More specifically, women with normal to mild CIN (i.e. normal histology or inflammation, CIN1) were classified as controls, and those with moderate to severe CIN (i.e. CIN2, CIN3 or invasive carcinoma) were classified as cases. For each woman, all archival DNA samples dated prior to the histological outcome were retrieved and the quantitative polymerase chain reaction (qPCR) analysis was performed as described subsequently in section 4.1. Table 1 gives a brief description of the variables involved.

Table 1*: Brief description of the data set variables*

Variables	Description
Response Variables	
Groups	Groups of subjects (cases coded as 1 & controls coded as 0) included in the study and classified according to histological endpoint.
Histology	Histological outcome of the biopsy classified as Normal (Inflammatory or not), CIN1, CIN2, CIN3 (or CIS) and Invasive Carcinoma (or SPINO).
Continuous Variables	
Age	Age of the subject at the time of the biopsy.
Time	Time lag (in days) between biopsy and cervical sample measurements from a given subject.
DNA_Count	Standardized DNA concentration of a given subject.
E6/Cell	Number of E6 copies per cell or Viral Load.
E2/Cell	Number of E2 copies per cell.
Ratio	Ratio of E2 by E6 copies per cell.
Categorical Variables	
Integration	Determines the HPV physical state as Episomal (2), Mixed (1) and Integrated (0) based on the E2/E6 ratio.
Cytology	Diagnosis of cytology sample prior to biopsy and classified as Normal, LSIL, HSIL and ASC.
Ratio Code	Code for classification according to the detection of DNA, E6, E2. The code was used to exclude samples for which PCR technical problems occurred (e.g. too little cells present in sample). Only samples with reliable PCR results were considered in the analysis (code 1 OR 4).
DNA Code	DNA concentration classified as Low(<1) and High (≥ 1).
Patient Count	Determines the number of measurements taken per subject.
Identification Variables	
Patient ID	Identification numbers of subjects.
Biopt_Date	Date at which the biopsy was performed.

4. MATERIALS AND STATISTICAL METHODOLOGY

This section elaborates on the screening approach used for data collection and on the methods used for data exploration and analysis.

4.1. Screening Approach and Sample Collection

As mentioned above, women with a known histological outcome were classified as case or control and cervical samples preceding biopsy were retrieved and analyzed.

The idea to evaluate the HPV integration state and viral load, surrogated by the E2/E6 ratio and the amount of HPV E6 oncogenes respectively, as biomarkers for HPV infection was motivated as follows: firstly, the E6, E7 and E2 proteins are fundamental for the HPV-induced carcinogenesis and the carcinogenic process is initiated when E2 is disrupted and E6/E7 is overexpressed; secondly, the available screening approaches such as cytology and HPV DNA testing, present some important shortcomings such as low cost-effectiveness, poor reproducibility, low sensitivity and/or specificity as highlighted in section 2.2.

As such, a real-time qPCR assay was developed to simultaneously measure E2 and E6 copy numbers and assess the amount of HPV present in the cervical sample. The PCR results determined the viral load and the viral integration rate. Other factors were further recorded as illustrated in section 3.

PCR is a method that allows exponential amplification of short DNA sequences within a longer double stranded DNA molecule. It's worth mentioning that the approach included sample cellularity normalization by means of β-globin to minimize inaccuracies in the quantification of DNA concentration (see Boulet, 2010 chapter 4 for more details).

4.2. Exploratory Data Analysis (EDA)

This was imperative to get acquainted with the data and detect underlying patterns necessary for the statistical modelling. To this effect, descriptive statistics, mean profiles and bivariate tests for comparison were used.

4.3. Multiple Logistic regression model

Multiple logistic regression (MLR) is an extension of contingency table analysis when several exploratory variables of continuous and/or categorical types are involved.

MLR is specified via three components: a random component identifying the binary response variable $\boldsymbol{Y}$ with a binomial probability distribution; a systematic component specifying a linear combination of (possibly transformed or categorized) p exploratory variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$; and a link function $g()$ specified to adequately map the systematic component onto the probability range. The most general form of the MLR model illustrated by Agresti (2002) is given by:

$$logit\left(P\left(Y_i = 1|X_{ij}\right)\right) = P\left(\frac{P\left(Y_i = 1|X_{ij}\right)}{1 - P\left(Y_i = 1|X_{ij}\right)}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \qquad (1)$$

where $P()$ denotes the probability function, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, X_p)$ is a $p+1$ vector of parameters of interest, $i = 1, \ldots, sample\ size$, and $j = 1, \ldots, p$.

The *logistic* function used in (1) has the core advantage over any other monotone function (such as the *probit* or *complementary log-log*) mapping the real line onto the [0; 1] interval since it is valid even for case-control or retrospective studies though yielding in this case an underestimated value of the intercept by a constant proportional to the sampling rate between cases and controls (Agresti, 2002 p. 171).

The parameters $\boldsymbol{\beta}$ and their corresponding standard errors are estimated via maximum likelihood (ML). Test of significance is done here using the Wald or likelihood ratio test statistics. Assessing the logistic model fit prior to any inference is done using the Pearson or deviance statistics for grouped data; or the Hosmer-Lemeshow statistic when the systematic component includes at least one continuous variable.

4.4. Generalized Estimating Equations (GEE)

This approach was proposed by Liang & Zeger (1986) for the marginal modelling of data that are clustered or repeated but also caters for data exhibiting dispersion.

GEE is a very appealing approach since it basically does not assume any particular distribution for the GLM's random component. More specifically, it solely requires: a model for the mean $E(Y_{ij}|X_{ij}) = \mu_{ij}$ assumed to depend on the systematic component through a given link function $g()$ i.e. $g(\mu_{ij}) = X'_{ij}\beta$ with notation consistent as in equation (1); a variance function describing the dependence of the variance on the mean; and a working guess for the correlation structure amongst values of the repeated response variable Y_{ij}'s (Agresti, 2002; Molenberghs & Verbeke, 2006). Several choices for the correlation structure are available and choosing an `adequate´ correlation structure primarily depends on the number of measurements per subject and on the need for parsimony.

GEE parameter estimates are obtained as solutions of the estimating equations i.e. the score equations of the underlying basic GLM with an updated version of the variance function to incorporate the working correlation. Significance testing and inference on parameters are asymptotically based on the Wald statistic.

GEE is deemed robust in the sense that it yields consistent effect estimates under misspecification of the correlation structure though with incorrect model-based standard errors. Nevertheless, robust standard errors are obtained from the empirical or `sandwich´ estimates (Liang & Zeger, 1986). It's

worth mentioning that severe misspecifications of the correlation structure do however affect the efficiency of the estimators. The choice of the most adequate correlation structure thus however boils down to that yielding the smallest discrepancy between the model-based or `naïve´ standard errors and their empirical counterparts.

4.5. Generalized Linear Mixed Models (GLMM)

In contrast to the GEE marginal model that models the correlation between pair of repeated measurements over the same subject via a working correlation assumption, the GLMM is an extension of GLM that induces the underlying correlation by inclusion of random effects $\boldsymbol{b}_i$ for each subject/cluster to capture heterogeneity.

For a binary response Y_{ij} taking values 0/1, the logistic GLMM could be written as:

$$logit\left(P\left(Y_{ij}=1|\boldsymbol{b_i}, X_{ij}\right)\right) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b_i} \qquad (2)$$

where $\boldsymbol{b_i} \sim N(\mathbf{0}, \boldsymbol{D})$ denotes the vector of random effects values for cluster *i* with **D** having unknown variance components and possibly correlation parameters; $\boldsymbol{X}_i$ denotes a vector of explanatory variables values for fixed effect model parameters $\boldsymbol{\beta}$; and $\boldsymbol{Z}_i$ denotes a vector of the random effects' explanatory variables. It's worth mentioning that conditioned on $\boldsymbol{b_i}$'s, GLMM (2) treats the Y_{ij}'s as independent over *i* and *j*.

Estimates of $\boldsymbol{\beta}$ and $\boldsymbol{D}$ are obtained via ML or restricted maximum likelihood with the likelihood function obtained by marginalizing/integrating over the random effects.

Guerin & Stroup (2000) documented that inference from a GLMM is severely compromised with a blatantly poor choice of the covariance model. Choice of an adequate covariance structure is often done by comparing models with various structures via information criteria such as the AIC and/or BIC.

It's worth mentioning that one should test for the need of random-effects prior to inference on the parameters. This is done using a likelihood ratio test with a mixture of chi-square as asymptotic

distribution due to boundary conditions on the variance components (Agresti, 2002; Molenberghs & Verbeke, 2006).

4.6. The Area under the ROC Curve Approach

This approach presents a framework for assessing the classification ability of a given classifier. The ROC curve is a two dimensional measure for evaluating the discriminatory accuracy of diagnostic tests. More specifically, it is the graph obtained by plotting the true positive fraction (sensitivity) against their corresponding false positive fraction $(1 - specificity)$ for a set of specified discrimination threshold values (Zhou, Obuchowski & McClish, 2002; Pepe, 2003). The area under the ROC curve (AUC) is a scalar measure representing expected performance of the ROC curve; the closer to one it is, the more accurate the test is. The AUC is appealing since it provides an appropriate common scale for comparing predictors (or their linear combination) even if they are not measured in the same measurement units (Pepe, 2003).

For a given classifier $L_\beta(\boldsymbol{X}) = P(Y = 1|\boldsymbol{X}) = X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$, the AUC of the ROC curve is equivalent to the Mann-Whitney U statistics (Pepe, 2003; Pepe *et al.*, 2006) and is given by:

$$\widehat{AUC}(\beta) = \frac{\sum_{i=1}^{n_Y}\sum_{j=1}^{n_{\bar{Y}}}\left\{I\left[L_\beta(\boldsymbol{X}_{Yi}) > L_\beta\left(\boldsymbol{X}_{\bar{Y}j}\right)\right] + 0.5 \times I\left[L_\beta(\boldsymbol{X}_{Yi}) = L_\beta\left(\boldsymbol{X}_{\bar{Y}j}\right)\right]\right\}}{n_Y n_{\bar{Y}}} \qquad (3)$$

where $\left\{L_\beta(\boldsymbol{X}_{Yi}),\ i = 1, \dots, n_Y;\ L_\beta\left(\boldsymbol{X}_{\bar{Y}j}\right)\ i = 1, \dots, n_{\bar{Y}}\right\}$ are the values of $L_\beta(\boldsymbol{X})$ for cases and controls respectively with corresponding sample size n_Y and $n_{\bar{Y}}$.

This approach was adopted here to primarily compare the classification performances of the fitted models. It was also used to optimize parameter estimates in order to obtain those yielding the `best´ performance accuracy i.e. $\widehat{\boldsymbol{\beta}}_{AUC} = argmax\left(\widehat{AUC}(\beta)\right)$. This was motivated by the fact that estimates with high discriminatory accuracy are obtained –over all possible linear combinations– even when the GLM (1) doesn't hold in the sense that its link function, random and/or systematic components might be incorrect (Pepe & Thompson, 2000; Pepe *et al.*, 2006). More importantly, it

has been proven that uniqueness, consistency and asymptotical normality of $\widehat{\boldsymbol{\beta}}_{AUC}$ are guaranteed under some mild assumptions (Sherman, 1993; Pepe *et al.*, 2006; Huang *et al.*, 2010).

Huang *et al.* (2010) advocated the cross-validation approach for estimating the AUC associated with estimated parameters to adjust for the upward bias induced when using (3) to optimize the parameters and re-substituting these latter to calculate the AUC.

4.7. Model Selection and Diagnostics

Model selection was done with respect to the MLR as the `starting point´. Variables' main effects were primarily included and the less significant ones were systematically removed based on likelihood ratio tests. Interactions of the remaining significant variables were further included and likewise systematically taken off. The GEE model was selected by comparing the discrepancy between the empirical and model-based standard errors of the auto-regressive, exchangeable and independence working correlations. The structure yielding the smallest difference was then considered more appropriate. The LRT was used to select the `best´ systematic component of the GLMM and the AIC was further used for selecting the correlation structure. The structure with the smallest AIC was thus considered more plausible relative to the others.

The MLR was diagnosed using the Hosmer-Lemeshow statistics displayed via the predictiveness curve. It is worth mentioning that the classifier performance was assessed via the AUC of the ROC curve and the classifier yielding the AUC closer to one was deemed more relevant.

All models were implemented using SAS version 9.2 and R 2.10.1. Tests of significance were all performed at a 5% level of significance.

5. RESULTS

This section presents EDA's findings, results of the various fitted models and implemented approaches.

5.1. Exploratory Data Analysis

The data basically consists of a total of 510 observations on 326 distinct women. Measurements vary from 1 to 6 with about 62% of the women having a single measurement and 2% having more than 4 measurements. It is worth mentioning that 5 (4 recorded as cases and one as a control) observations reported missing values for age and cytology. These were discarded from subsequent analyses without loss of generality. Of the resulting 323 women remaining in the study sample, 154 (48%) were classified as controls and 169 (52%) as cases.

A glance at the variables TIME, E6_count and E2_count revealed that they all reported very large values as shown by the plots of their histograms which are highly skewed to the right (Figures 1A, 2A and 3A in the appendix). The traditional log transformation was applied on these variables in order to recalibrate their variability for a better fit of the subsequent models.

Inspecting the evolution of the proportion of cases within time intervals of length 1 (on the *log* scale) revealed that the number of cases seems to be high at the time prior to biopsy and decreases back in time (Figure 3, left panel). This insinuates that controls seem to have more records compared to cases. One also indeed noticed that older women show a tendency to be classified as cases (Figure 3, right panel).

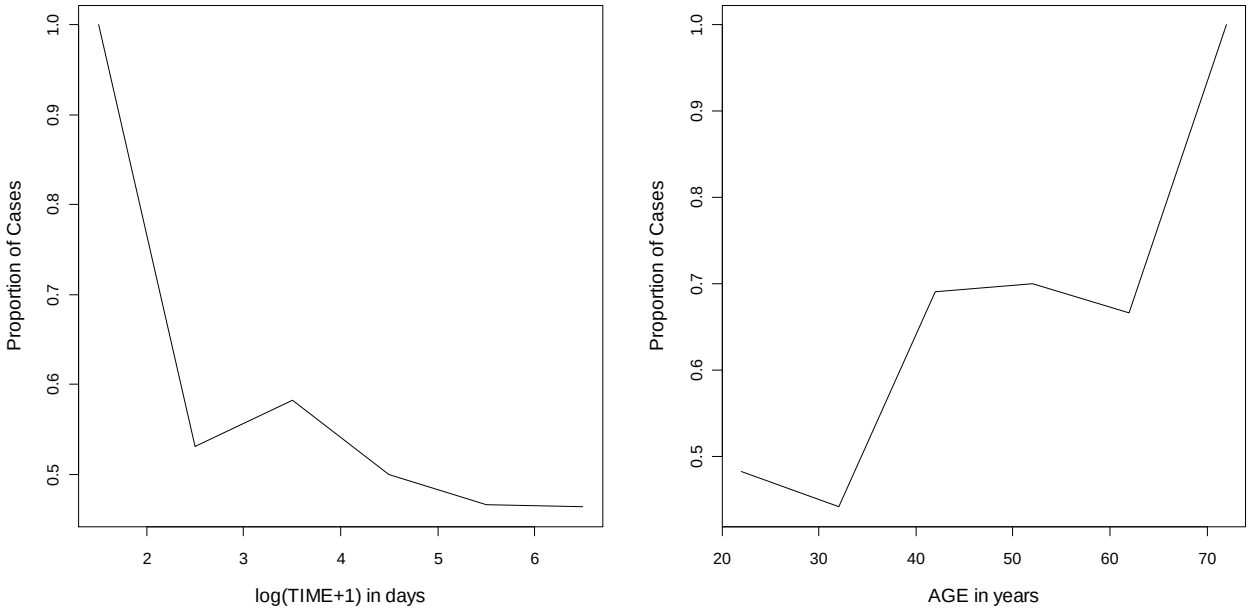

Figure 3: Evolution of the proportion of cases/controls in Time (left) and with Age (right panel).

Given the retrospective nature of the study, one was curious to assess eventual systematic differences between cases and controls at baseline or time of biopsy for each of the covariates. To this effect, Fisher chi-square, Fisher exact and the t-test were performed depending on the nature of the covariate and the sparseness of the contingency table for categorical variables.

The results of these analyses are shown in Table 2 and it emerged that the distribution of the variables DNA_Count, E2_count, Ratio_code were the same between cases and controls. Also, controls (Mean age=31, s.e.=7.71) were 5 years younger than cases (Mean age=36, s.e.=12.39) on average. The E2/E6 ratio seemed higher in controls (Mean=1.02, s.e.=0.34) than in cases (Mean=0.86, s.e.=0.32) indicating that the episomal HPV fraction is higher in controls compared to cases. One also noticed that cases were on average biopsied earlier than controls.

Table 2*: Characteristics of the study subjects at time T_0 (or baseline) of biopsy*

	Controls			*Cases*			*P-value*
	Frequency(%)	**Mean**	**SD**	**Frequency(%)**	**Mean**	**SD**	
Cytology							<.0001
Normal (1)	14 (4.44)			1 (0.32)			
ASCUS (2)	23 (7.30)			15 (4.76)			
ASC-H (3)	18 (5.71)			17 (5.40)			
LSIL (4)	24 (7.62)			121 (38.41)			
HSIL (5)	69 (21.90)			13 (4.13)			
Integration							0.0006[$]
0	1 (0.32)			7 (2.22)			
1	45 (14.29)			78 (24.76)			
2	102 (32.38)			82 (26.03)			
Ratio Code							**0.0708[$]**
1	147 (46.67)			160 (50.79)			
4	1 (0.32)			7 (2.22)			
Age		31.11	7.71		36.09	12.42	0.0001[*]
***log*(Time+1)**		3.20	1.94		2.56	1.83	0.0028[*]
DNA_Count		12.49	10.73		13.60	9.45	**0.3325[*]**
***log*(E6/Cell+1)**		3.97	2.36		4.46	2.00	0.0503
***log*(E2/Cell+1)**		3.97	2.39		4.24	2.16	**0.2851[*]**
E2/E6 Ratio		1.02	0.34		0.86	0.32	0.0001[*]

P-value for: [§]chi-square test; [$]Fisher exact test; [*] t-test between cases and controls.

5.2. The Risk Models

This subsection entails the risk models resulting from the MLR, GEE and GLMM.

Recall that the core aim of the thesis is to assess the impact of ratio and/or the amount of E6/E2 HPV DNA copies per cell as primary markers for cervical screening, and then assess how much screening is enhanced when combining them with cytology. To this effect, two models (with and without cytology) were sought for in the model selection process.

The logistic regression models were primarily fitted and a model selection was done with the variables (Age, Age^2, *log*(time+1), log^2(time+1), *log*(E6_count+1), *log*(E2_count+1), Ratio, Integration) included with their interactions of order 1. It emerged that the model with variables Age, *log*(time+1), *log*(E6_count+1), *log*(E2_count+1) and the interaction of the latter two was deemed `best´ (AIC= 643.671) for the model without cytology (model A); whereas the systematic component of that with cytology (model B) was made of variables Age, *log*(time+1), ratio, cytology, and interactions of the latter two with time (AIC= 540.937). One could indeed notice a considerable improvement when using cytology as an additional marker to classify women as cases or control for further follow-up/diagnostic tests.

As highlighted in figure 1, most infections do not persist into carcinogenesis since they are cleared by immunological intervention. This motivated the interest to examine whether the risk to be classified as a case/control was the same for women older/younger than 30 years. The age variable was thus dichotomized to this effect and it emerged that for a given amount of E6 and E2 copies per cell at a given point in time, the risk to be classified as case/control was the same (OR=1.32, CI=[0.9; 1.92]) for women older/younger than 30 years. Model B also reported a non-significantly different risk between women younger/older than 30 years (OR= 1.18, CI=[0.76; 1.82]). An `optimum´ age cutoff value was thus sought that would yield significantly different risks for the age groups. This was done by specifying a handful of cutoff values and finding that which maximizes the AIC. For model A and B, it resulted that the risk to be classified as a case was significantly higher for women older than 41 years compared to those younger than 41 years i.e. women older than 41 years were more likely to be classified as cases relative to those younger than 41 years. Moreover, with this dichotomized version of age, one noticed an improvement in AIC for both models (AIC= 634.614 and 531.260 respectively). The latter models were thus preferred over their counterparts with a continuous version of Age. Since there is `no free lunch´ in statistics, this was done at a price of information loss in terms of continuity, but nevertheless with easy interpretation

and use of the risk model by clinicians for future classification. Table 3 illustrates the parameter estimates of the resulting models.

Prior to using the models for inference, it was crucial to assess their goodness-of-fit. This was graphically done for the MLR via the predictiveness curve combined with observed proportions (Pepe *et al.*, 2007) as shown in figure 4. One clearly sees that the risk models slightly overestimate the observed data. These slight discrepancies nevertheless appeared to be statistically non-significant as revealed by the Hosmer & Lemeshow tests ($\chi^2 = 9.81$ and $\chi^2 = 7.10$ with $df = 8$ for models A and B respectively).

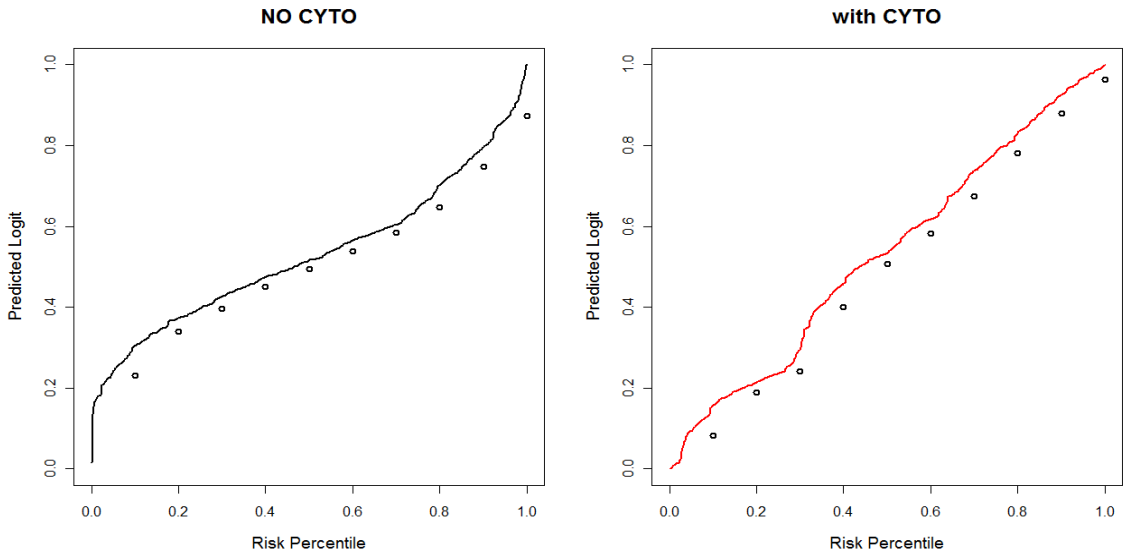

Figure 4: Lack-of-fit assessment of models A and B by calibration.

A glance at the parameter estimates (MLR) in Table 3 reveals that for a given amount of E6/E2 DNA copies at a given point in time, the odds that HPV infection persists is about 3.9 (C.I.=[2.22, 6.79]) times on average for women older than 41 years relative to those younger than 41 years. Taking the cytology outcome into account slightly reduces this odds to about 3.7 (C.I.=[1.93; 7.26]) with respect to a given cytological outcome and HPV DNA integration level at a given point in time. One however notices the wider confidence interval for the OR estimates with model B that could result from the lack of unanimity amongst clinicians in assessing cytological outcomes. It is

worth noticing that the effect of the amount of E2 copies and viral load on the probability of HPV persistence depends on each other for model A. Indeed, as one could expect, an increase in viral load (for a given amount of at least 1.5×10^{12} E2 HPV copies) will lead to transforming infection whereas, an increase in the amount of E2 HPV copies (for any given viral load of at least 1 E6 copy) will result into a productive infection that will be cleared by immunological response. Again, for women of age more than 41, the odds for clearance is smaller and that of carcinogenesis is higher on average compared to those younger than 41 years.

Model B revealed that the odds to be classified as a case based on cytological outcome or the integration level depends on the time of biopsy. Nevertheless, women with a given cytological outcome and known age group at time 0 are very likely on average to be classified in the control group for a unit increase in the integration level. Still at time 0, for a given integration level and known age group, women with ASCH, ASCUS, LSIL and HSIL are more likely to be classified as cases compared to those with Normal cytology. These odds nevertheless decrease retrospectively in time and women with LSIL became more likely (relative to ASCH, ASCUS and HSIL) to report viral clearance compared to those with Normal cytology.

The GEE was implemented to account for dispersion and/or the longitudinal nature of the data. The interaction term emerged to be non-significant for model A and was dropped. For both model A and B, the independence working assumption was preferred over the exchangeability and auto-regressive of order 1 correlation assumptions since the latter both reported very unstable parameter estimates with standard errors very close to 0. As can be seen in Table 3 under the GEE (Inde.) column, similar conclusions could be drawn as in the MLR since they indeed both report similar parameter estimates. One particularly observed that the empirical standard errors of model B are smaller than their model-based counterparts from the MLR. This could be explained by the fact that a total of 323 patients involved in the sample seems not to be large enough (with respect to the amount of covariates in the systematic component) to meet the asymptotic requirements for GEE inference so that empirical standard errors could have been underestimated. This however should

not jeopardize the trust in the `robust´ standard errors since there is no rule of thumb affirming that sandwich standard errors should always exceed model-based counterparts.

In order to explicitly account for longitudinal measurements and investigate woman-specific odds of viral persistence, a random intercept GLMM model was implemented and the result also presented in table 3. Testing for the need of random-intercept appeared to be highly significant (p-value<0.001). GLMM for Model A reported that for a given woman in the study, the effect of the viral load and the amount of HPV E2 DNA copies on the odds of transforming infection is positive and negative respectively. Nevertheless, one interestingly notices that their estimates are roughly equal so that for a unit increase in both the viral load and E2 copies, the woman has a 50% chance of viral clearance and carcinogenesis. Model B on the other side highlighted that a given woman with cytological outcome is very likely to observe viral clearance for a unit increase in her integration level. Further, given women with ASCH, ASCUS and HSIL cytological outcome are more likely to encounter viral persistence compared to Normal cytology, whereas women with LSIL cytology are more likely to encounter clearance compared to Normal cytology.

Both model A and B also revealed a very high intra-woman correlation derived as:

$$\rho = \frac{\sigma^2}{\frac{\pi^2}{3} + \sigma^2} \approx 0.999$$

indicating a very low within woman variability and consequently a high between woman variability.

Table 3*: Estimates of the logistic models fitted.*

*Model without Cytology (**MODEL A**)*					*Model with Cytology (**MODEL B**)*				
	Type of Logit Models[a]					Type of Logit Models[a]			
	MLR	**GEE (Inde.)**	**GLMM**	**Optim.**		**MLR**	**GEE (Inde.)**	**GLMM**	**Optim.**
Intercept	-1.21 (0.40)*	-1.21 (0.43)*	-12.03 (3.74)*		**Intercept**	-2.06 (2.13)	-2.06 (1.91)	0.42 (5.89)	
Age (≥ 41 years)	1.36 (0.29)*	1.36 (0.38)*	26.69 (7.05)*	1.36 (0.25)	**Age (≥ 41 years)**	1.32 (0.34)*	1.32 (0.43)*		1.31 (0.55)
Log(Time+1)	-0.10 (0.05)*	-0.10 (0.05)	-0.31 (0.50)	-0.10 (0.9)	**Log(Time+1)**	0.43 (0.38)	0.43 (0.34)		0.42 (0.94)
Log(E6+1)	1.96 (0.37)*	1.96 (0.42)*	10.75 (2.67)*	1.96 (0.9)	**Cyto** *ASCUS*	4.83 (2.08)*	4.83 (1.90)*	1.54 (6.06)	4.84 (0.82)
Log(E2+1)	-1.26 (0.35)*	-1.26 (0.39)*	-10.55 (2.60)*	-1.27 (0.9)	*ASC-H*	5.87 (2.05)*	5.87 (1.78)*	1.12 (6.26)	5.86 (0.66)
E6 x E2	-0.07 (0.02)*	-0.07 (0.02)*		-0.07 (1.7)	*LSIL*	3.75 (2.00)	3.75 (1.79)*	-20.59 (7.28)*	3.75 (0.38)
				s.e. x 10^{-3}	*HSIL*	8.05 (2.00)*	8.05 (1.76)*	23.68 (5.73)*	8.06 (0.61)
					Ratio	-3.66 (1.14)*	-3.66 (1.09)*	-2.10 (3.73)	-3.66 (0.55)
					Time x*ASCUS*	-0.74 (0.36)*	-0.74 (0.33)*		-0.77 (0.85)
					Time x*ASC-H*	-1.01 (0.36)*	-1.01 (0.31)*		-1.02 (0.53)
					Time x *LSIL*	-0.83 (0.34)*	-0.83 (0.31)*		-0.83 (0.97)
					Time x *HSIL*	-1.29 (0.35)*	-1.29 (0.31)*		-1.30 (1.45)
					Time x Ratio	0.51 (0.24)*	0.51 (0.22)*		0.52 (1.34)
									s.e. x 10^{-2}
			σ=62.26 (10.42)*					σ =113.33 (23.24)*	
	AIC=634.6	*QIC=640.05*	*AIC=429.9*			*AIC=531.26*	*QIC=535.41*	*AIC=412.8*	

*: *significant at a 5% level.*

[a]: *MLR* is the multiple logistic regression; *GEE(Inde.)* is the generalized estimating equations model under the independence working correlation; *GLMM* is the generalized linear mixed-effect model with a random-intercept; *Optim.* is the approach that searches parameter estimates which optimize the AUC with respect to the data. Values in parentheses are standard errors.

5.3. The AUC framework for assessing the discriminatory accuracy

Having obtained the parameter estimates via the risk models, their discriminatory abilities were then assessed. To this effect, the ROC curves of the MLR/GEE model were evaluated and plotted in figure 5 (left panel). The respective AUCs were obtained via cross-validation as advocated by Huang *et al*. (2010). More specifically, a 10-fold cross-validation (Hastie *et al.*, 2009 p.242) was adopted where at each iteration, 90% of cases/controls were used to learn the models and the remaining 10% as validation set. It emerged that model A reported an AUC of about 0.71 (s.e.=0.07) indicating that the expected true positive fraction taken uniformly over every possible false positive fractions in [0; 1] is about 71%, whereas that of model B was about 83% (s.e.=0.08). Similar results were obtained when optimizing the parameters to maximize the AUC (column optim. in Table 3) rather than the likelihood function as done with the MLR. The AUCs thus obtained were indeed higher (AUC=0.72 (s.e.=0.08), AUC=0.84 (s.e.=0.08) for model A and B respectively) compared to those resulting from the MLR likelihood. In addition to the Hosmer & Lemeshow goodness-of-fit results previously obtained, this insinuates that the MLR model could be deemed appropriate for the data. As expected, one could see that classification performance is indeed enhanced when cytology supplements the viral DNA integration state compared to when the viral load is combined with the amount of E2 DNA copies. Moreover, figure 5 (right panel) shows that the curves resulting from model A and B seem significantly different from each other since their confidence bands do not overlap.

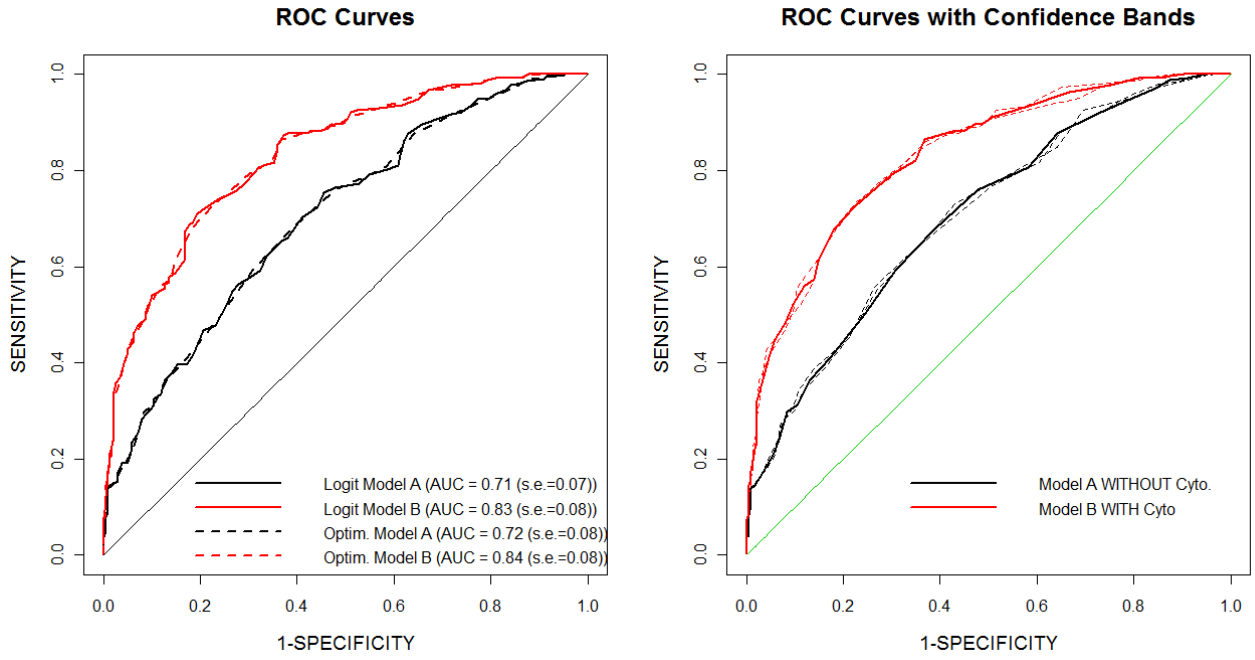

Figure 5: ROC Curves for parameters obtained via MLR and Maximizing the AUC.

The AUC of the GLMM was obtained by marginalizing over the random intercepts via numerical averaging. To this effect, 20 values of the random intercept were arbitrarily drawn from $N(0;\ 10.42^2)$ and $N(0;\ 23.24^2)$ for model A and B respectively. The expected probabilities were then obtained from:

$$P(Cases = 1) = \frac{1}{20}\sum_{i=1}^{20} expit(\beta_1 + \beta_2 Age\_strat + \beta_3 ltime + \beta_4 \log\_E6 + \beta_5 \log\_E2 + + \beta_6 E6 \times E2 + \mathrm{b_{i1}})$$

$$P(Cases = 1) = \frac{1}{20}\sum_{i=1}^{20} expit(\beta_1 + \beta_2 ASCUS + \beta_3 ASCH + \beta_4 \mathrm{LSIL} + \beta_5 \mathrm{HSIL} + \beta_6 \mathrm{Ratio} + \mathrm{b_{i2}})$$

where the βs are obtained from table 3. It unfortunately emerged that marginalizing over the random effects reported very poor performances (AUC=0.52 and 0.51 for model A and B respectively). This could be due to the very diffuse distribution of the random intercepts so that large values are very likely to be drawn when sampling. Setting the random effects to zero however yielded AUCs of 0.68 and 0.79 for model 1 and 2 respectively. As such, one could indeed rely on

the MLR for future patients' classification. One should nevertheless be cautious when using the model to distinguish between persistent and productive infections since the women involved in the study do not represent a cross-section of the population of all HPV infected women in Flanders in the sense that some cases existing in the population, but not yet diagnosed in the Laboratory for Clinical Pathology have not been included in the study.

6. DISCUSSION AND CONCLUSION

HPV is one of the most common sexually transmitted infections and it is nowadays firmly established that high-risk human papillomavirus (HR-HPV) infections are responsible for approximately 5% of all cancers worldwide, causing cervical cancer, other anogenital tumors and head-and-neck cancers (Boulet, 2010). In particular, HPV16 infection that causes about 54% of all cervical cancers was the focus of this thesis (Clifford *et al.*, 2006).

About 80% of HPV16 infections are cleared by the host's immune system and the remainder could persist into severe cervical intraepithelial neoplasia, the precursor lesion of cervical cancer. Secondary prevention of the disease through screening is amenable since the disease evolves via well-defined stadia from HPV infection through invasive cancer. The standard screening approach often used in practice, cervical cytology, has proven to be feasible but is hampered by high cost-effectiveness, poor reproducibility due to high interobserver variability, and very low true positive fraction (Boulet, 2010, chapter 2; European Commission, 2008, chapter 3). The recognition of the strong causal relationship between persistent infection with HR-HPV and cervical cancer has resulted in the introduction of HPV DNA testing in clinical practice. HR-HPV DNA detection has proven to be more sensitive than cytology in identifying CIN2+. However, due to the high prevalence of transient infections, its specificity and positive predictive value (PPV) are low. Therefore, the introduction of HR-HPV testing as a screening tool requires additional testing to distinguish transforming from productive HPV infections. Proteins E6, E7 and E2 are known to

play fundamental roles in the HPV-induced carcinogenesis that is associated with integration of the viral DNA, loss of E2, deregulated viral oncogene expression and genomic instability. The purpose of this thesis was to evaluate biomarkers associated with lesion progression such as the viral integration state, viral load and the amount of E2 HPV DNA copies per cell in order to enhance the sensitivity and specificity of cervical screening in distinguishing transforming from productive HPV infections.

Statistical models served for this purpose since multiple biomarkers and clinical data are needed to optimize cancer prediction/classification in order to detect subclinical cancer with adequate sensitivity and specificity (Pepe *et al.*, 2001). More specifically, risk models such as the logistic regression, GEE and GLMM were implemented to model women's risks on transforming HPV infections. As advocated by Pepe *et al.* (2004), AUC (area under the ROC curve) supplemented risk modelling in the assessment of discriminatory accuracy of resulting classifiers in order to avoid being misled by commonly used measures, such as the overall misclassification error or the magnitude of the odds ratio estimates.

Hesselink *et al*. (2008) reported that in a cervical screening setting viral load assessment of HPV16 has no additive value to stratify high-risk HPV positive women for risk of CIN2 or CIN3. It emerged herein that combining the viral load with the amount of E2 HPV DNA copies per cell might result in an increased sensitivity for CIN2+, but nevertheless at the cost of a relatively low specificity (black curve on figure 5). Risk model A also revealed that at a given point in time, for a given age group and amount of E2 DNA copies, increased viral loads are associated with an increased risk of CIN2+ and reduced amounts of viral DNA to the absence of CIN2 and viral clearance (Monnier-Benoit *et al.*, 2006; Hesselink *et al*. 2008). Moreover, implementing more classy statistical analyses such as the MLR and/or GLMM, as an improvement to Boulet *et al.* (2009), resulted into much better discriminatory accuracies when combining the amount of E2 HPV DNA copies per cell with viral load on the one hand and the integration state with cytology on the other hand.

In addition to the relatively good fit of the MLR models, their discriminatory accuracies were very similar to those under the AUC approach indicating that the MLR could be deemed appropriate for the data despite the underlying longitudinal nature. This is indeed not surprising given the considerable amount of single measurements per woman in the data and/or the relatively high between woman variability as revealed by the random-intercept model.

In conclusion, one could utter that irrespective of whether combining the amount of E2 DNA copies is with the viral load on the one hand, or the integration state and cytological evaluation on the other hand as biomarkers for cervical cancer screening in Flanders, women older than 41 years are more likely to develop CIN2+ lesions compared to those younger than 41 years. The effects of the viral load and amount of E2 HPV DNA copies depend on each other on average, with women more likely to report mild or normal displasia with an increase in the amount of E2 HPV DNA copies for a given amount of viral load. Also, women with increasing integration state were on average very likely to report mild or normal histology for a given cytology outcome at a given point in time. It also emerged that the effect of cytology on the histological outcome further depends (on average) on the time at which the cervical sample was analyzed. Overall, assuming that enough resources are available for cervical screening programs in Flanders, one could use the integration state alongside cytological evaluation as biomarkers to stratify HR-HPV women for risk of CIN2+.

One should however avoid being too optimistic in using the resulting model B as an eventual screening routine since the entailed study was retrospective in nature and the study sample does not represent a cross-section of the total population of cervical cancer patients in Flanders. These results could however serve as clues or be confirmed by cohort studies.

7. Recommendation

Future studies could also attempt to incorporate self-collected vaginal specimens since these have shown promising sensitivity compared to cytology (Monnier-Benoit *et al.*, 2006). Moreover, this would improve the screening population coverage since women with low socio-economic status and other hard-to-reach populations that are highly at risk or consequently reluctant to attend to clinician-based screening, would very likely be incorporated in the study sample.

As further recommendations, one could have incorporated the cross-validation in the variable selection process so that the model's systematic component will be made of relevant variables that optimize the AUC (Huang *et al.*, 2010). Also, it's worth mentioning that rather than simply calculating the AUC for given covariate parameters obtained either by maximizing the likelihood or the Mann-Whitney U statistic, one could have employed a ROC curve analysis via a binormal ROC curve model say, where accounting for clustering in the data can be done via random-effects, and covariates effects on the discriminatory capacity of the classifier could easily be assessed (Pepe, 2003 chapter 5 and 6). Moreover, as documented by Pepe *et al.*'s (2007), the predictiveness curve should be assessed in conjunction to performance measures such as the ROC curve since the latter does not display risk thresholds that are useful for risk stratification.

BIBLIOGRAPHY

Agresti, A. (2002). Categorical Data Analysis, Second Edition. United States of America: WILEYINTERSCIENCE.

Boulet, G., Benoy, I., Depuydt, C., Horvath, C., Aerts, M., Hens, N., Vereecken, A., and Bogers, J. (2009). HPV16 load and E2/E6 ratio in HPV16-positive women: biomarkers for CIN2+ in a liquid-based cytology setting? *Cancer Epidemiol Biomarkers Prev* **18(11)**, 2992-9.

Boulet, G. (2010). *Molecular Markers in Cervical Cancer Screening*. Faculty of Medecine: University of Antwerpen.

Clifford, G., Franceschi, S., Diaz, M., Muñoz, N. and Villa, L. L. (2006). Chapter 3: HPV Type-Distribution in Women with and without Cervical Neoplastic Diseases. *Vaccine* **24**, S3/26–S3/34.

European Commission (2008). *European Guidelines for Quality Assurance in Cervical Cancer Screening. 2nd ed*. Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, Wiener H, Daniel J, von Karsa L, editors. Luxembourg: Office for Official Publications of the European Communities.

Fehrmann, F. and Laimins, L., A. (2003). Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene* **22**, 5201-7.

Guerin, L., and Stroup, W. (2000). "A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data." In *Proceedings of the Twelfth Annual Conference on Applied Statistics in Agriculture*. Manhattan: Kansas State University.

Hastie, T.; Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd edition. New York:Springer.

Hesselink, A., T., Berkhof, J., Heideman, D., A., Bulkmans, N., W., van Tellingen, J., E., Meijer, C., J., and Snijders, P., J. (2008). High-risk human papillomavirus DNA load in a population-based cervical screening cohort in relation to the detection of high-grade cervical intraepithelial neoplasia and cervical cancer. *Int. J. Cancer* **124**, 381-86.

Huang, X., Qin, G. and Fang, Y. (2010). Optimal combinations of diagnostic tests based on AUC. *Biometrics* **67**, 568-76.

Liang, K. Y., and Zeger, S., L. (1986). Longitudinal data analysis using generalized linear models. *Biometrics* **73**, 13-22.

Molenberghs, G. and Verbeke, G. (2006). Models for Discrete Longitudinal Data. New York: Springer.

Moons, K., G., M. and Harrell, F., E. (2003). Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. *Acad Radiol* **10**, 670-2.

Moscicki, A.B., Schiffman, M., Kjaer, S. and Villa, L.L. (2006). Chapter 5: updating the natural history of HPV and anogenital cancer. *Vaccine* **24 (Supplement 3)**, S3/42-S3/51.

Monnier-Benoit, S., Dalstein V., Riethmuller, D., Lalaoui, N., Mougin, C. and Pretet, J., L. (2006). Dynamics of HPV16 DNA load reflect the natural history of cervical HPV-associated lesions. *J Clin Virol* **35**, 270–7.

Muñoz, N., Castellsagué, X., de González, A. B. and Gissmann, L (2006). Chapter 1: HPV in the etiology of human cancer. *Vaccine* **24 (Supplement 3)**, S3/1- S3/10.

Pepe, M. S. and Thompson, M. L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics* **1**, 123-40.

Pepe, M. S., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M.L., Thornquist, M., Winget, M., and Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W. and Newcomb, P. (2004). Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology* **159**, 882-90.

Pepe, M. S., Cai, T. and Longton, G. (2006). Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* **62**, 221-29.

Pepe, M. S., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. M., Zheng, Y. (2007). Integrating the Predictiveness of a Marker with Its Performance as a Classifier. *American Journal of Epidemiology* **167 (3)**, 362–68.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrics* **61**, 123-37.

Zhou, X. H., Obuchowski, N.A. and McClish, D. K. (2002). *Statistical Methods in Diagnostic Medecine*. New York: John Wiley & Sons.

APPENDIX

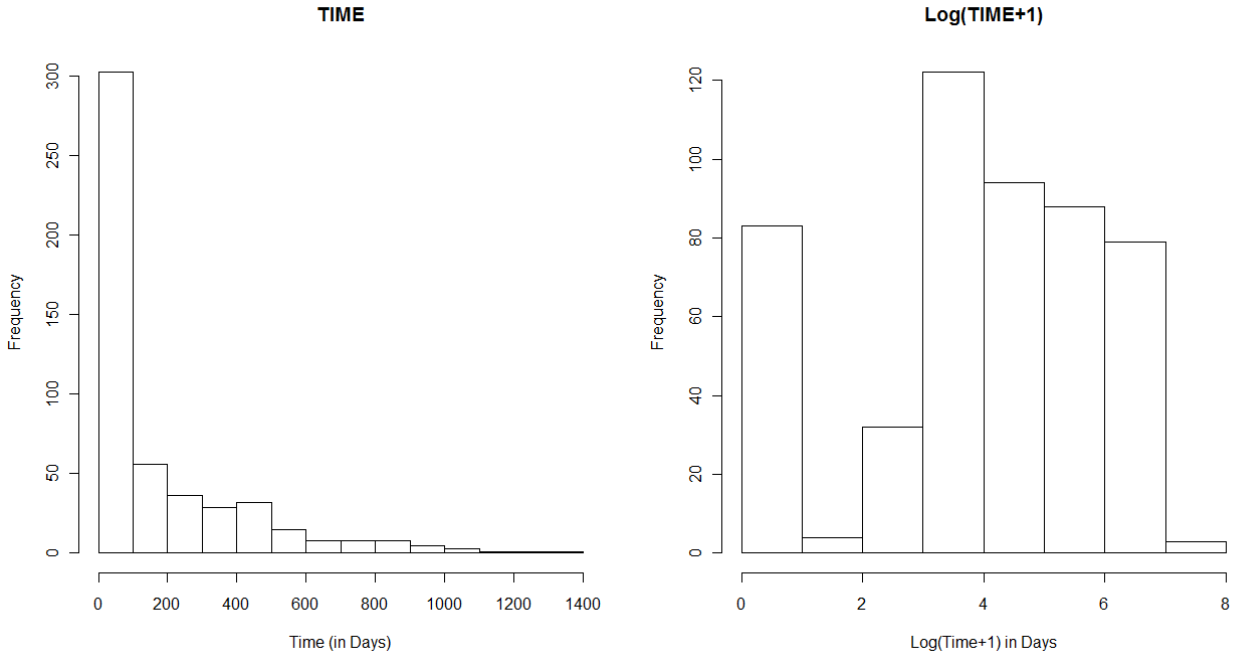

Figure 1A*: Histograms for the TIME variable and its log-transformed version*

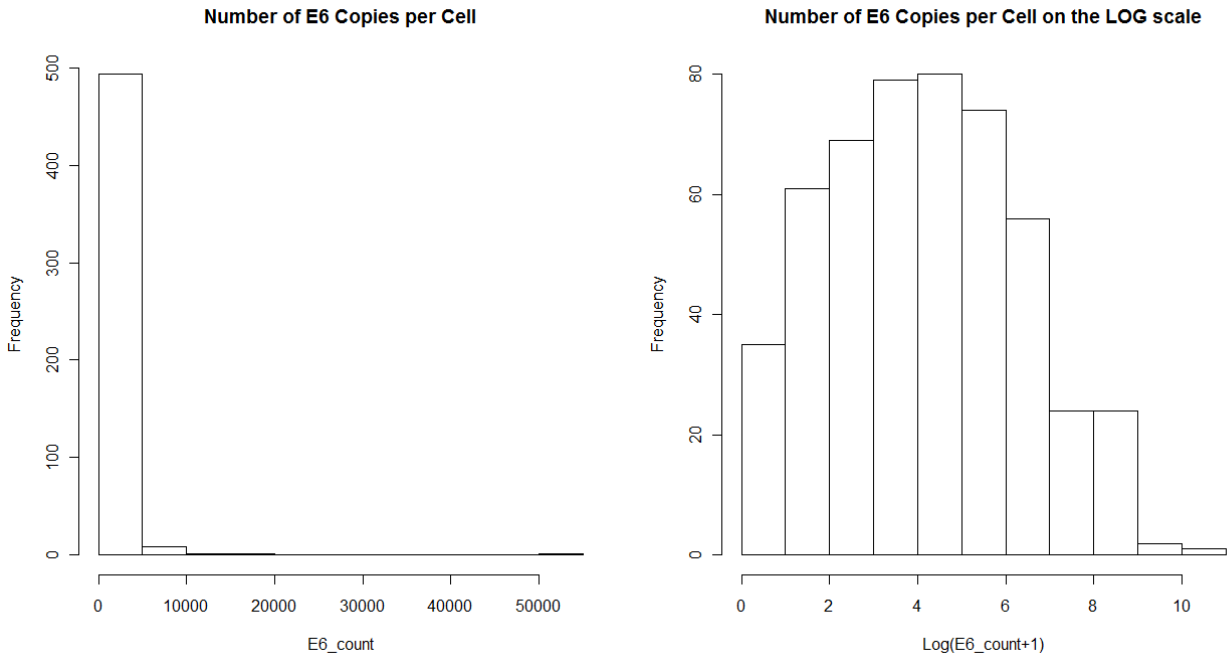

Figure 2A*: Histograms for the amount of E6 copies per cell and its log-transformed version*

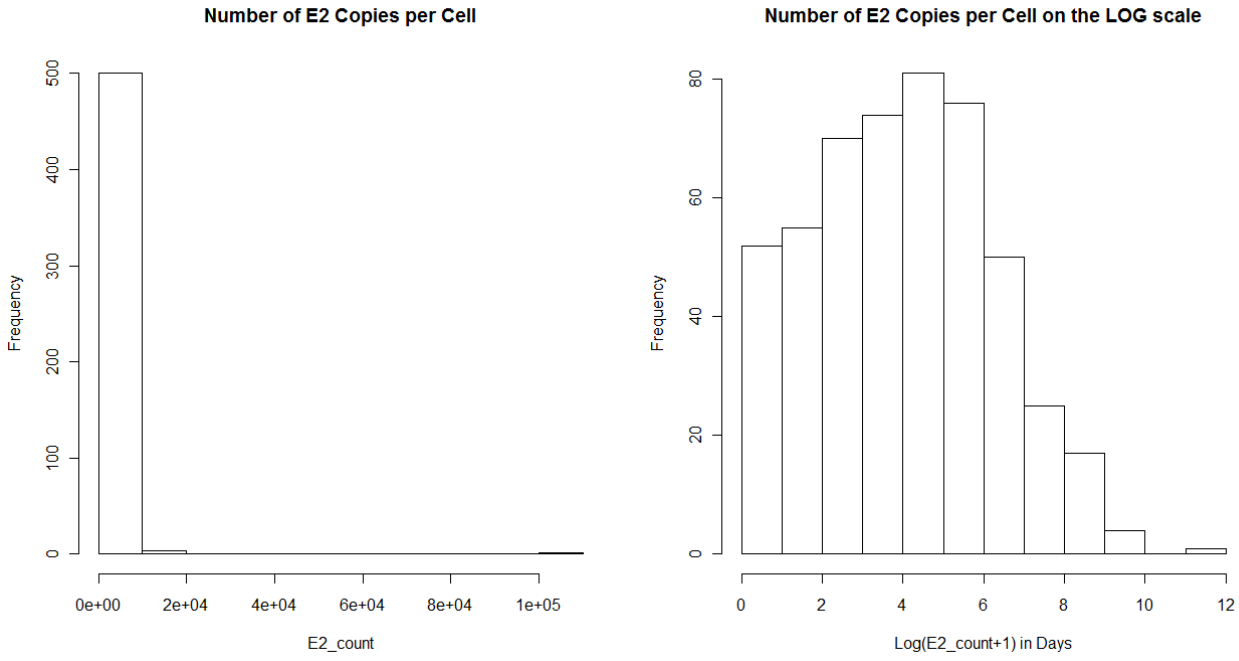

Figure 3A*: Histograms for the amount of E2 copies per cell and its log-transformed version*

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Diagnostic accuracy evaluation fo screening tests for detecting high grade cervical leasions in Flanders

Richting: **master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2011**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Ouandji Djiokep, Raoul

Datum: **12/09/2011**