Bayesian Estimation of Performance Measures of Cervical Cancer Screening Tests in the Presence of Covariates and Absence of a Gold Standard

Edson Zangiacomi Martinez[1], Francisco Louzada-Neto[2], Sophie Françoise Mauricette Derchain[3], Jorge Alberto Achcar[1,2], Renata Clementino Gontijo[3], Luis Otávio Zanatta Sarian[3] and Kari Juhani Syrjänen[4]

[1]Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil. [2]Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, SP, Brazil. [3]Departamento de Tocoginecologia, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brazil. [4]Department of Oncology and Radiotherapy, Turku University Hospital, Turku, Finland.

Abstract: In this paper we develop a Bayesian analysis to estimate the disease prevalence, the sensitivity and specificity of three cervical cancer screening tests (cervical cytology, visual inspection with acetic acid and Hybrid Capture II) in the presence of a covariate and in the absence of a gold standard. We use Metropolis-Hastings algorithm to obtain the posterior summaries of interest. The estimated prevalence of cervical lesions was 6.4% (a 95% credible interval [95% CI] was 3.9, 9.3). The sensitivity of cervical cytology (with a result of $\geq$ ASC-US) was 53.6% (95% CI: 42.1, 65.0) compared with 52.9% (95% CI: 43.5, 62.5) for visual inspection with acetic acid and 90.3% (95% CI: 76.2, 98.7) for Hybrid Capture II (with result of $>$1 relative light units). The specificity of cervical cytology was 97.0% (95% CI: 95.5, 98.4) and the specificities for visual inspection with acetic acid and Hybrid Capture II were 93.0% (95% CI: 91.0, 94.7) and 88.7% (95% CI: 85.9, 91.4), respectively. The Bayesian model with covariates suggests that the sensitivity and the specificity of the visual inspection with acetic acid tend to increase as the age of the women increases.

The Bayesian method proposed here is an useful alternative to estimate measures of performance of diagnostic tests in the presence of covariates and when a gold standard is not available. An advantage of the method is the fact that the number of parameters to be estimated is not limited by the number of observations, as it happens with several frequentist approaches. However, it is important to point out that the Bayesian analysis requires informative priors in order for the parameters to be identifiable. The method can be easily extended for the analysis of other medical data sets.

Keywords: Bayesian analysis, diagnostic tests, latent variables, cervical cytology, visual inspection with acetic acid, Hybrid Capture II

1. Introduction

The sensitivity (S_e) and the specificity (S_p) are the two most common measures of the performance of a diagnostic test, where S_e is the probability of a diseased individual to be correctly identified by the test while S_p is the probability of an individual without the disease (or condition) of interest to be correctly identified by the same test. When the outcomes of the diagnostic test are represented in a continuous scale, a cut-off value should be chosen in order to determine when an individual is classified as positive or negative. Generally, individuals with test outcome larger or at least equal to this fixed cut-off are classified as positive while individuals with test outcomes inferior to this fixed cut-off are classified as negative.

Although the real disease status of the individual could be verified by a procedure generically denominated gold standard, it is common to find situations where a proportion of the sampled individuals cannot be verified on their real disease status. The problem can occur especially when the gold standard is an invasive and/or risky procedure and the definitive verification for apparently healthy individuals is thus neither practical nor ethical. In order to overcome this problem, many studies on the evaluation of the diagnostic test are carried out by considering only verified individuals. However, this approach can lead to measures that are usually biased, leading to studies denominated verification bias or workup bias.

Correspondence: Edson Zangiacomi Martinez, Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP, 14049-900, Brazil, Email: edson@fmrp.usp.br

Unbiased estimators for S_e and S_p are introduced by Begg and Greenes [1] and Zhou [2].

Another problem appears when all individuals can not be verified by a gold standard. This occurs when there is not a definitive test for detection of the disease or the verification by a gold standard is an impracticable procedure according to its cost, accessibility or risks. In this situation, maximum likelihood estimators are proposed by Hui and Walter [3]. However, these estimators are reasonable only in situations where the number of observations is larger or equal to the number of parameters, which is not our case, as we will see later. Free of this limitation, a Bayesian approach was introduced by Joseph et al. [4]. However, the method of Joseph et al. [4] do not consider the presence of covariates, which are very common on data from diagnostic test studies.

The objective of the present study is to verify the performance measures of cervical cytology, Hybrid Capture II (HC II) and visual inspection with acetic acid (VIA) in the detection of cervical precursor lesions, using a Bayesian statistical method that allows for the estimation of these measures, although part of the sampled women was not verified by a gold standard. We also consider the presence of covariates in our study. Since Bayesian methods are based on incorporation of historical information and expert opinion into the modelling strategy (the called prior information), these elements could be too subjective, with source for other bias. In other words, inadequate prior information can imply in a biased estimator. However, a careful verification of the prior information and a subsequent analysis of its changes in the outcomes can result in reasonable estimates for the tests performance measures.

Thus, the new methodological contribution of the present paper is an extension of the Bayesian method proposed by Joseph et al. [4] for estimating the performance measures of screening tests introducing a vector of covariates. The rest of the paper is organized as follows. In Section 2, we discuss the definition of a gold standard in accuracy studies of cervical cancer screening tests. In Section 3, there is a description of the method of Joseph et al. [4] for estimating S_e and S_p related to two diagnostic tests in the absence of a gold standard. We also introduce in this section the notation used in the paper. In the following, we introduce the methodology for estimating S_e and S_p in the presence of a covariate. The cervical cancer screening data set is described in Section 4. The application of the proposed methodology on the analysis of the data set is presented in Section 5. Concluding remarks are given in Section 6.

2. Accuracy of Cervical Cancer Screening Tests

South and Central America have some of the highest incidence rates for cervical carcinoma in the world, ranging from 30/100,000 women to 40/100,000 women, or three to four times the incidence in developed countries [5]. In Brazil, crude estimates of incidence and mortality are given by 19,82/100,000 and 4,49/100,000 women, respectively [6]. Thus, it is strongly justified to analyze the accuracy of different diagnostic tools for cervical carcinoma and their efficacy in screening programs.

In assessing the accuracy of cervical cancer screening tests, it is not straightforward to define an ideal gold standard. In many studies, the gold standard for evaluating the accuracy of screening tests in detecting true positive lesions is histopathology. If biopsies are not obtained, colposcopy is accepted as the final diagnosis. However, colposcopy can give many false negative results when used to discriminate between normal and abnormal tissues (see, for example, Mitchell et al. 1998 and Hopman et al. 1998) [7,8].

The reference test used in these studies, defined by the results of histology or colposcopy, is thus subject to errors and its estimates for sensitivity and specificity can be biased. Another type of bias is evident when only a part of the sampled individuals will have their real disease status confirmed by the biopsy and the remainders are not included in the calculations of the sensitivity and specificity. This occurs principally when only the women with positive result for one or more diagnostic tests (or with positive clinical signals) are submitted to the gold standard and this selection results in an overestimated sensitivity and an underestimated specificity [9].

Despite of the appearance of new methods developed to estimate the sensitivity and specificity of screening tests without a gold standard [10] or in the presence of the verification bias [11], many studies on the accuracy of cervical cancer screening tools present biased results due to the limitations of the proposed gold standard. For example, in a recent meta-analysis of the studies on performance of conventional cervical cytology, McCrory et al. (1999) evaluated 939 studies, where 84 took care of the standards established by the authors to guarantee the quality of the results. Of these 84 studies, only three did not have their results affected by the verification bias [12].

Many studies are introduced in the literature in the absence of a gold standard. For instance, Hui and Walter (1980) derived equations that compute estimates and standard errors of sensitivity, specificity and prevalence, without considering a reference test [3]. Joseph et al. (1995) introduced a Bayesian model using latent variables [4], and Dendukuri and Joseph (2001) extended this method to account for conditional dependence between the diagnostic tests [13]. Other important statistical contributions were provided by Faraone and Tsuang (1994) [14], Qu et al. (1996) [15] and Hadgu and Qu (1998) [16].

3. The Bayesian Framework

Considering k diagnostic tests, let $T_m = 1$ if the result of test m is positive and $T_m = 0$ if the result of test m is negative, for $m = 1, ..., k$. Let S_{e_m} and S_{p_m} be the sensitivity and the specificity of the test m, respectively and let g be an observation of a binary latent variable G, introduced in the model aiming to simulate a non-observable gold standard [17]. Denoting the set of the observations and this latent variable for the i-th individual by $x_i^T = \{t_{1_i}, t_{2_i}, t_{3_i}, g_i\}$, where T_m is an observation of T_m, we have the joint density function

$$f(x_i) = p^{g_i}(1-p)^{1-g_i} \prod_{m=1}^{k} S_{e_m}^{t_{m_i} g_i}$$

$$(1-S_{e_m})^{(1-t_{m_i})g_i} S_{p_m}^{(1-t_{m_i})(1-g_i)} (1-S_{p_m})^{t_{m_i}(1-g_i)}$$

assuming that the test outcomes are independents. We have $2k + 1$ parameter to be estimated, or say, k pairs (S_{e_m}, S_{p_m}), and the prevalence p. The likelihood function $L(\theta)$, where $\theta = (S_{e_m}, S_{p_m}, p : m = 1, ..., k)$ is the vector of parameters, is given by

$$L(\theta) = p^{\sum_{i=1}^{n} g_i}(1-p)^{n-\sum_{i=1}^{n} g_i} \prod_{m=1}^{k} S_{e_m}^{\sum_{i=1}^{n} t_{m_i} g_i}$$

$$(1-S_{e_m})^{\sum_{i=1}^{n}(1-t_{m_i})g_i} S_{p_m}^{\sum_{i=1}^{n}(1-t_{m_i})(1-g_i)} \quad (1)$$

$$(1-S_{p_m})^{\sum_{i=1}^{n} t_{m_i}(1-g_i)}.$$

The latent variable G, following the Bayes equation, has a Bernoulli distribution, that is,

Considering beta prior densities $Beta(\alpha_\theta, \beta_\theta)$ for all parameters in θ, where α_θ and β_θ generically denotes fixed hyperparameters and combining the likelihood function for $\theta(1)$ with the prior densities, we use the Gibbs sampling algorithm [18,19] to simulate samples for the posterior distribution for θ. These samples are simulated from the full conditional posterior distributions for p, S_{e_m} and S_{p_m}.

Following Equations (1) and (2) and considering k diagnostic tests, the conditional posterior distributions for the components of θ needed for the Gibbs sampling algorithm are given by

$$P | \mathbf{X}, \alpha_p, \beta_p \sim Beta\left(\sum_{i=1}^{n} g_i + \alpha_p; \right.$$

$$\left. n - \sum_{i=1}^{n} g_i + \beta_p \right),$$

$$S_{e_m} | \mathbf{X}, \alpha_{S_{e_m}}, \beta_{S_{e_m}} \sim Beta\left(\sum_{i=1}^{n} t_{m_i} g_i + \alpha_{S_{e_m}}; \right.$$

$$\left. \sum_{i=1}^{n} (1-t_{m_i}) g_i + \beta_{S_{e_m}} \right), \text{and}$$

$$S_{p_m} | \mathbf{X}, \alpha_{S_{p_m}}, \beta_{S_{p_m}}$$

$$\sim Beta\left(\sum_{i=1}^{n} (1-t_{m_i})(1-g_i) + \alpha_{S_{p_m}}; \right.$$

$$\left. \sum_{i=1}^{n} t_{m_i}(1-g_i) + \beta_{S_{p_m}} \right),$$

for $m = 1, ..., k$. This model is analogous to model developed by Joseph et al. (1995) [4].

Let $\mathbf{w}_i$ be a sample observation of $\mathbf{W}_i$, a vector of L covariates. For the sake of simplicity and without lack of generality, we assume that T_i is a random variable (with observation t_i) related to result of only one diagnostic test, with Bernoulli distribution with success probability $p_i S_{e_i} + (1-p_i)(1-S_{p_i})$, $i = 1, ... n$. In the presence of a vector of covariates, let us assume the logit links for S_{e_i}, S_{p_i} and p_i, given by $\theta_{li} = \exp\left(\sum_{j=0}^{L} \beta_{lj} w_{ji}\right)\left[1+\exp\left(\sum_{j=0}^{L} \beta_{lj} w_{ji}\right)\right]^{-1}$,

$$G_i \sim Bernoulli\left(\frac{p \prod_{m=1}^{k} S_{e_m}^{t_{m_i}} (1-S_{e_m})^{1-t_{m_i}}}{p \prod_{m=1}^{k} S_{e_m}^{t_{m_i}} (1-S_{e_m})^{1-t_{m_i}} + (1-p) \prod_{m=1}^{k} S_{p_m}^{1-t_{m_i}} (1-S_{p_m})^{t_{m_i}}} \right). \quad (2)$$

where $l = 1, 2, 3$, $W_{0i} = 1$, $\theta_{1i} = S_{ei}$, $\theta_{2i} = S_{pi}$, $\theta_{3i} = p_i$, for $i = 1, ..., n$. In this way, we have a vector of parameters given by $\beta = (\beta_1, \beta_2, \beta_3)$, where $\beta_1 = (\beta_{l0}, \beta_{l1}, ..., \beta_{lL})$, $l = 1, 2, 3$. Assuming prior independence among the parameters, we consider the prior densities for β_{lj} with normal distribution with fixed hyperparameters a_{lj} (means) and b_{lj}^2 (variances), $l = 1, 2, 3, j = 0, 1, ..., L$. The likelihood function for β is given by

$$L(\beta) = \frac{\exp\left[\sum_{j=0}^{L} \beta_{1_j} \sum_{i=1}^{n} w_{ji} t_i g_i + \sum_{j=0}^{L} \beta_{2_j} \sum_{i=1}^{n} w_{ji}(1-t_i)(1-g_i) + \sum_{j=0}^{L} \beta_{3_j} \sum_{i=1}^{n} w_{ji} g_i\right]}{\prod_{i=1}^{n}\left\{\left[1+\exp\left(\sum_{j=0}^{L} \beta_{1_j} w_{ji}\right)\right]^{g_i}\left[1+\exp\left(\sum_{j=0}^{L} \beta_{2_j} w_{ji}\right)\right]^{1-g_i}\left[1+\exp\left(\sum_{j=0}^{L} \beta_{3_j} w_{ji}\right)\right]\right\}},$$

where g is an observation of the latent variable G, given by (2). Combining the prior distributions with $L(\beta)$, we have the conditional posterior distributions for β given by

sample. In each cycle of the algorithm is generated a new value for the latent variable G as (2).

In studies of the performance of two or more independent diagnostic tests applied to a selected group of individuals, where none of these tests can be considered the gold standard, a straightforward extension of this model can be used. Considering the three diagnostic tests, cervical cytology, VIA and HC II, the vector of unknown parameters is now given by $\beta = (\beta_1, ..., \beta_7)$, where $\beta_l = (\beta_{l0}, \beta_{l1}, ..., \beta_{lL})$, $l = 1, ..., 7$, are vectors of parameters related to the sensitivity and the specificity of each test and the prevalence of

$$\pi\left(\beta_{1_j} \mid {}_{(\beta_{1_j})}\beta, \mathbf{X}, \mathbf{W}\right) \propto N\left(a_{1_j}; b_{1_j}^2\right) \times \exp\left\{\beta_{1_j} \sum_{i=1}^{n} w_{ji} t_i g_i - \sum_{i=1}^{n} g_i \ln\left[1+\exp\left(\sum_{k=0}^{L} \beta_{1_k} w_{ki}\right)\right]\right\},$$

$$\pi\left(\beta_{2_j} \mid {}_{(\beta_{2_j})}\beta, \mathbf{X}, \mathbf{W}\right) \propto N\left(a_{2_j}; b_{2_j}^2\right) \times \exp\left\{\beta_{2_j} \sum_{i=1}^{n} w_{ji}(1-t_i)(1-g_i)\right.$$

$$\left. -\sum_{i=1}^{n}(1-g_i)\ln\left[1+\exp\left(\sum_{k=0}^{L} \beta_{2_k} w_{ki}\right)\right]\right\}, \text{ and}$$

$$\pi\left(\beta_{3_j} \mid {}_{(\beta_{3_j})}\beta, \mathbf{X}, \mathbf{W}\right) \propto N\left(a_{3_j}; b_{3_j}^2\right) \times \exp\left\{\beta_{3_j} \sum_{i=1}^{n} w_{ji} g_i - \sum_{i=1}^{n} \ln\left[1+\exp\left(\sum_{k=0}^{L} \beta_{3_k} w_{ki}\right)\right]\right\}.$$

where $j = 0, 1, ..., L$ and $\beta_{(\beta_{10})}$ is the vector of all parameters except β_{10} (for example). Observe that we should simulate samples for all parameters considering the Metropolis-Hastings algorithm [47] since their conditional distributions are difficult to

cervical lesions. Let T_{m_i} be a random variable with observation t_{m_i} related to test m, $m = 1, 2, 3$. Using logit link function to relate the vector $\mathbf{W}_i$ of L covariates to the screening performance measures, $i = 1, ..., n$, the likelihood function for β is now given by

$$L(\beta) = \frac{\exp\left[\sum_{i=1}^{3}\sum_{j=0}^{L} \beta_{lj} \sum_{i=1}^{n} w_{ji} t_{l_i} g_i + \sum_{l=4}^{6}\sum_{j=0}^{L} \beta_{lj} \sum_{i=1}^{n} w_{ji}(1-t_{(l-3)_i})(1-g_i) + \sum_{j=0}^{L} \beta_{7j} \sum_{i=1}^{n} w_{ji} g_i\right]}{\prod_{i=1}^{n}\left\{\prod_{l=1}^{3}\left[1+\exp\left(\sum_{j=0}^{L} \beta_{lj} w_{ji}\right)\right]^{g_i}\prod_{l=4}^{6}\left[1+\exp\left(\sum_{j=0}^{L} \beta_{lj} w_{ji}\right)\right]^{1-g_i}\left[1+\exp\left(\sum_{j=0}^{L} \beta_{7j} w_{ji}\right)\right]\right\}}.$$

In this expression, the vectors of parameters β_1, β_2 and β_3 are related to the sensitivities of the cervical cytology, VIA and HC II, respectively; β_4, β_5 and β_6 are related to the specificities of the cervical cytology, VIA and HC II, respectively; and the vector β_7 is related to the prevalence of cervical lesions. We consider the prior densities for β_{lj} with normal distribution with fixed hyperparameters a_{lj} (means) and b_{lj}^2 (variances), $l = 1, .., 7, j = 0, 1, ..., L$. Combining the prior distributions with $L(\beta)$, we have the conditional posterior distributions for β and the Metropolis-Hastings algorithm is used to generate samples from the each parameter.

4. Data Set

The data set is from a European Commission funded ongoing study known as the LAMS (Latin American Screening) study, where PAP smear/ liquid-based cytology and screening colposcopy were compared with three optional screening tools (visual inspection with acetic acid or Lugol's iodine and cervicography) and with Hybrid Capture II from conventional samples and from self-samples, in women at different risk for cervical cancer in three Brazilian arms (São Paulo, Campinas and Porto Alegre) and one Argentine arm (Buenos Aires). The study design and baseline data of the LAMS study were presented by Syrjanen et al. (2005) [20]. Partial results from the LAMS study were provided by Sarian et al. (2005) [21].

In the present study, we considered the data from Campinas, one of the three Brazilian arms of the LAMS study. From February to December 2002, 1,195 women were recruited at a basic health unit and from July to December 2002, 221 women were recruited at the University Hospital (Centro de Atenção Integral à Saúde da Mulher-CAISM). Both services are situated in Campinas, a 969,396 inhabitants city in Brazil's southeast region. Among these 1.416 women, 809 women were eligible for the study related to the sensitivity and specificity of three cervical cancer screening tests (cervical cytology, visual inspection with acetic acid and Hybrid Capture II) in the presence of covariates and in the absence of a gold standard and were willing to participate. Women were eligible if they were between 18 and 60 years of age, if they had been submitted at all three diagnostic methods and if they had intact uterus. Patients previously subjected to treatment for condylomas or with history of current abnormal cytology were excluded. Women who presented with confirmed immunossupression, immunodeficiency or HIV infection, who had sexual intercourse or vaginal medication in the last three days were not included. Informed consent was obtained from all participant women. The study protocol was reviewed and approved by the Committee of Ethics in Research of the Medical Science School of the State University of Campinas.

Cervical cytology was collected after evaluation and treatment for possible infectious processes. Ayre spatulas and cervical brushes were used for these samplings. The samples were stained according to the Papanicolaou method and evaluated using the Bethesda System [21]. Cytology was considered positive if showing cellular atypia, irrespective of their severity. Ecto- and endo-cervical samples were collected for second generation Hybrid Capture (HC II) using sterile endo-cervical brushes supplied by Digene Diagnostics and processed following the instructions of the manufacturer (Digene Diagnostics Inc.). The HC-II is a molecular biological method that tests the presence of the HPV-DNA, through a chemoluminescent reaction. HC-II is commercialized as standard kits and it is based on a reaction of hybridization realized into several sorts of solutions with non-radioactive probes of known ribonucleic acids. Viral load was measured in relative light units (RLU/CO) and HC II results were categorized as negative if <1 RLU/CO and positive otherwise. After the collection of the cervical cytology and HC II, dilute 5 percent acetic acid was applied to the cervix. One minute afterwards, the cervix was illuminated with adapted spotlights (100 Watts) and naked-eye examined for acetowhite areas. The visual appearance was classified according to the Atlas for Unaided Visual Inspection of the Cervix [22] using the categories: normal, atypical, intra-epithelial neoplasia or suggestive of cervical cancer. Normal or atypical results were classified as negative and intra-epithelial neoplasia or suggestive of cervical cancer were classified as positive. More details on the study protocol may be found in Syrjanen et al. (2005) [20].

5. Results

First of all, the sensitivity and specificity of cervical cytology (T_1), VIA (T_2) and HC II (T_3) were estimated by a Bayesian approach proposed by Joseph et al. (1995) [4]. This method was developed

for the situation where a reference test is not available and it has the assumption that the tests are conditionally independents. Seven parameters were estimated, including the prevalence of preneoplasic or neoplasic lesions and the sensitivity and specificity pairs relative to the three diagnostic methods under evaluation. An important feature of the Bayesian approach is the combination of the data obtained by the current sampling scheme with prior information about the parameters of interest. This prior information is quantitatively introduced in the statistical analysis and it can represent the pooled subjective opinions of the experts, or information derived from the published literature. In the present study, we initially defined the prior information from the medical literature, using beta probability distributions.

The prior information about the sensitivity and the specificity of cervical cytology was based on the systematic review of Nanda et al. (2000), who presented sensitivity for atypical squamous cells of undetermined significance (ASC-US) or worse being ranged from 29 percent to 56 percent and specificity from 97 percent from 100 percent [23]. The studies of Belinson et al. (2001) and of the University of Zimbabwe and JHPIEGO Cervical Cancer Project (1999) were used as references for the choice of the prior information about the sensitivity and the specificity for the VIA [25–26]. In these studies, the sensitivity of VIA for at least CIN II was estimated in 55 percent and 64 percent, respectively and the specificity was estimated as 76 percent and 67 percent, respectively. The prior information of the accuracy measures of HC II test was based on the studies of Schiffman et al. (2000) and Wright et al. (2000), who estimated sensitivities by 88.4 percent and 81.3 percent (at 1 RLU cut-off), respectively and specificities by 89.0 percent and 84.5 percent, respectively [27–28].

However, the choice of informative prior distributions based only in a summary of previous studies can be a complex task, since each study has elements of subjectivity, error-proneness and possible potential for bias. Thus, a panel of experts on cervical cancer was asked to provide their best estimate for the sensitivities and specificities of the tests and the prior distributions that summarise the information provided by the literature review corrected by the experts were derived. The assessment of beta distribution priors for each test parameter considered the method presented by Joseph et al. (1995) [4], where the hyperparameters are defined by matching the center of a range of plausible values of sensitivity and sensitivity with the mean of the beta distribution and matching the standard deviation of the beta distribution with one quarter of the total range. We considered a vague prior distribution for the prevalence of precursor cervical lesions (a Beta distribution with hyperparameters 0.5 and 0.5, see [29]) motivated by a little background knowledge about this parameter.

The median age of the 809 women who participated of the study was 34 years. Approximately three quarters of these women lived with a partner (73.0 percent) and one-third had 8 or more years of education (33.3 percent). The majority self-reported to be white (67.2 percent) and 64.3 percent reported not to be a smoker. Half of the women (50.3 percent) reported to have had only one lifetime sexual partner and almost three quarters (72.9 percent) had initiated the sexual life in the teenage. Only 1.5 percent of the women entered in the study with less than one year since her first sexual intercourse and the majority (86.1 percent) reported to have had only one sexual partner during the last 12 months. The percentage of women who are pregnant at the time of the study was around 7.7 percent.

Based on the cases for which cervical cytology was available, 758 (93.7 percent) had normal results, 12 (1.5 percent) had low-grade squamous intraepithelial lesion (LSIL), 4 (0.5 percent) had high-grade squamous intraepithelial lesion (HSIL) and 35 (4.3 percent) were ASC-US. Table 1 shows the results of the three tests for the 809 available cases. For a Bayesian data analysis, the Gibbs sampler algorithm was run for 100,000 cycles, where the first 20,000 were used to assess convergence and the last 80,000 were used for inferences. For each parameter of interest, the arithmetic mean of these 80,000 Gibbs samples is a natural Bayesian estimator. These arithmetic means are showed in Table 2, with the respective 95 percent credible intervals. Table 2 also shows positive and negative predictive values for each diagnostic test, calculated in each cycle of the Gibbs algorithm from the estimated sensitivities and specificities and the prevalence figures (a mathematical approach is presented by Altman and Bland [30]).

The results suggest a low sensitivity for cervical cytology to detect ASC-US or worse (53.6 percent) as well as for VIA (52.9 percent), but indicate a high sensitivity for the HC II (90.3 percent). All screening methods presented relatively high

Table 1. Results of cervical cytology, VIA and HC II in 809 women who underwent all three tests.

	Cervical cytology +		Cervical cytology –		Total
	HC II +	HC II –	HC II +	HC II –	
VIA+	9	2	15	35	61
VIA–	21	19	87	621	748
Total	30	21	102	656	809

Abbreviations: VIA: visual inspection with acetic acid; HC II: Hybrid Capture II.

specificities, 97.0 percent for cervical cytology, 93.0 percent for VIA and 88.7 percent for the HC II. As is evident in Table 2, all methods presented very low positive predictive values (PPV) due to the low prevalence of cervical lesions [30]. Although with high sensitivity and specificity, HC II did not present a high PPV (35.3 percent), which is similar to that estimated for VIA. On the other hand, all methods presented high negative predictive values (NPV) (Table 2). The prevalence of precursor lesion was estimated as 6.4 percent and this low prevalence naturally imply few diseased individuals and consequently low PPVs.

Another important result from the Bayesian model is related to the estimates for the expected value of true positives for each combination of the three screening methods, as shown in Table 3. This estimator is a numerical approximation and in this way it may accept decimals. The proportion of the predicted number of positives from the total of sampled individuals is the estimate of PPV. In Table 3, we notice that for 9 women, all three tests

reported positive outcomes and the expected PPV is 98.6 percent. On the other hand, when all the tests are negative, it is expected that only 0.2 percent of the women with this outcome will have cervical lesions.

Tables 4 and 5 shows as secondary results the sensitivities and specificities for each combination of the screening tests. Table 4 summarizes the results when the tests are evaluated in serial combination (positive when both tests were positive and negative otherwise) and Table 5 summarizes the results when the tests are evaluated in parallel combination (positive when at least one of the tests was positive and negative otherwise). When the association between two tests is considered, the results of the third test are not considered. In Table 4, the serial combination between cervical cytology and visual inspection has a sensitivity of 78.0 percent, which is higher than the sensitivity of each test individually (53.6 and 52.9 percent respectively, see Table 2). This result suggests an apparent improvement in sensitivity, but at a cost

Table 2. Bayesian estimates for sensitivities, specificities, positive and negative predictive values for each screening test, and for the prevalence of cervical lesions.

Test		%	95% CI
Cervical cytology	sensitivity	53.6	42.1–65.0
	specificity	97.0	95.5–98.4
	PPV	55.3	37.4–73.6
	NPV	96.8	94.8–98.3
Visual inspection with acetic-acid			
	sensitivity	52.9	43.5–62.5
	specificity	93.0	91.0–94.7
	PPV	34.0	22.4–46.2
	NPV	96.6	94.6–98.1
Hybrid Capture II	sensitivity	90.3	76.2–98.7
	specificity	88.7	85.9–91.4
	PPV	35.3	22.8–48.8
	NPV	99.2	97.8–99.9
prevalence		6.4	3.9–9.3

Abbreviations: CI: credible interval; PPV: positive predictive value; NPV: negative predictive value.

Table 3. Estimates for the number of true positives for each combination of the screening methods and their expected positive predictive values.

Cervical cytology	Test VIA	HC II	Total of sampled individuals	Predicted number of positives	Expected positive predictive value (%)
+	+	+	9	8.87	98.6
+	+	−	2	0.93	46.7
+	−	+	21	17.33	82.5
+	−	−	19	1.37	7.2
−	+	+	15	9.85	65.6
−	+	−	35	0.97	2.8
−	−	+	87	10.39	11.9
−	−	−	621	1.19	0.2

Abbreviations: VIA: visual inspection with acetic acid; HC II: Hybrid Capture II.

in specificity, when these two tests are used jointly. However, this increase in sensitivity must be interpreted with caution. Franco [31] advised that nominal increase in sensitivity always occurs by chance whenever an adjunct test, as HC II, is used in combination with a conventional test, as Pap cytology, even if the complementary test was totally random with respect to the disease being evaluated.

In a second instance, we introduced in the model the age of the women (X) as a continuous covariate. The covariate W_1 is given by $(X - \bar{x})/10$, where $\bar{x}$ is the sample mean of X. The quotient 10 is only considered for avoiding numerical instability related to large values in exponential functions present in the conditional posterior densities of interest. We also introduced in the model the variable W_2, a dichotomous variable that denotes whether or not the woman is actually pregnant (1 if pregnant and zero otherwise). Firstly, we considered 13 the interaction between W_1 and W_2 in the model. However, all interaction parameters were estimated to be close to zero (ranged from

−0.031 to 0.009) and it were excluded from the final model.

From the conditional densities for the parameters in β, we generated 100,000 Gibbs samples. From this chain, we discarded the first 20,000 (regarded as burn-in samples). The convergence of the Gibbs samples was monitored by standard existing methods [32] and the trace plots obtained are shown in Figure 1. The convergence was observed for all parameters. Prior distributions for the intercept parameters β_{10} to β_{70} were assumed with fixed hyperparameters based in estimates obtained in the previous analysis without covariates. For example, we noted that the estimated sensitivity of the cervical cytology was 0.536 (see Table 2), and considering the inverse of the logit function, the hyperparameter a_{10} is thus given by log $(0.536/(1 - 0.536))$. All the other hyperparameter values were chosen to have noninformative priors. Thus, we used an empirical Bayesian modelling approach [33]. For each parameter, we considered every 50th draw, which totalizes a sample of size 1, 600. Considering

Table 4. Bayesian estimates of the sensitivities and specificities for the combinations between screening tests in serial mode*.

Combination between tests		%	95% CI
Cervical cytology + and VIA +	sensitivity	78.0	70.3–84.8
	specificity	90.2	87.9–92.4
Cervical cytology + and CH II +	sensitivity	95.5	88.7–99.4
	specificity	86.1	82.9–89.1
VIA + and CH II +	sensitivity	95.4	88.5–99.4
	specificity	82.5	79.3–85.6
All tests positive	sensitivity	97.9	94.5–99.7
	specificity	80.1	76.6–83.4

*Positive when all tests were positive and negative otherwise.
Abbreviations: VIA: visual inspection of the cervix with acetic-acid; HC II: Hybrid Capture II; CI: credible interval.

Table 5. Bayesian estimates of the sensitivities and specificities for the combinations between screening tests in parallel mode*.

Combination between tests		%	95%CI
Cervical cytology + or VIA +	sensitivity	28.3	20.5–36.9
	specificity	99.8	99.7–99.9
Cervical cytology + or CH II +	sensitivity	48.3	36.4–60.3
	specificity	99.7	99.4–99.8
VIA + or CH II +	sensitivity	47.7	36.8–58.2
	specificity	99.2	98.8–99.4
At least one test positive	sensitivity	25.6	17.8–34.2
	specificity	99.98	99.96–99.99

*Positive when at least one of the tests was positive and negative otherwise.
Abbreviations: VIA: visual inspection of the cervix with acetic-acid; HC II: Hybrid Capture II; CI: credible interval.

that a logit link was used, the regression coeffcients in β are interpreted as being the logarithm of the odds ratios (OR). These odds ratios represent an association measure between the variables W_1 and W_2 and the operating characteristics of the screening tests.

In Table 6, we have the posterior summaries for the exponential funcion of the parameters of interest in β, interpreted as odds ratios. We observe that the 95 percent credible intervals for the parameters $e^{\beta_{11}}$ to $e^{\beta_{71}}$ included the value 1, suggesting that there is no evidence for the effect of pregnancy in $e^{\beta_{12}}$ and $e^{\beta_{52}}$ S_e and S_p measures for all tests. The parameters $e^{\beta_{12}}$ and $e^{\beta_{52}}$ were estimated in 2.033 and 1.615, respectively and its credibility interval does not include the value 1. This result suggests that the sensitivity and the specificity of VIA increases as the age of the women increases. In fact, in the medical literature several authors have described that methods for detection of precursor lesions of cervical cancer have different performances according to the age of women [26, 34, 35]. The prevalence of cervical lesions, as expected, tends to increase as the age of the women increases (*OR* estimated in 0.483 and the respective 95% credible interval do not included the value 1). This effect of age on the prevalence is well-known in the medical literature since the disease is more incident in sexually active women.

6. Concluding Remarks
In this article, we introduced a Bayesian approach based on a Markov chain Monte Carlo (MCMC) algorithm that allows the performance measures estimations of diagnostic tests in the presence of covariates and when a gold standard is not available. An advantage of the proposed methodology is the fact that the number of parameters to be estimated is not limited by the number of observations as it happens when we use the method introduced by Hui and Walter [3]. We used the logit link function to relate the covariates linearly to the screening performance measures, but it is possible to use other link functions than the logit function, in according to the nature of the data. For comparison, we also adjusted Bayesian models to estimating the sensitivity and the specificity of the cervical cancer screening tests here presented based on other link functions, as the log-log complementary function, but we do not observe significant changes in the parameter estimates and its inferences (results not shown). However, a misspecified model could arise from an incorrect link function and the use of model comparison measures, as the Deviance Information Criterion (DIC) of Spiegelhalter et al. [36], allow us to decide which of these functions give us the most appropriate model. An advantage of the logit link function over other functions is that it provides estimates of odds ratios, a meaningful and well-known measure of association.

An important consideration in the use of the proposed model is its dependence on the prior information. In a sensitivity analysis, we noted, for example, that the prevalence of cervical lesions is increased when we used all non-informative prior distributions and other substantial changes in the sensitivity and specificity. In the presented model, the lack of a gold standard is counterbalanced by the introduction of a latent variable G (2) that best describe the data simulating a reference test. This latent variable has a Bernoulli distribution with success probability given in function of the
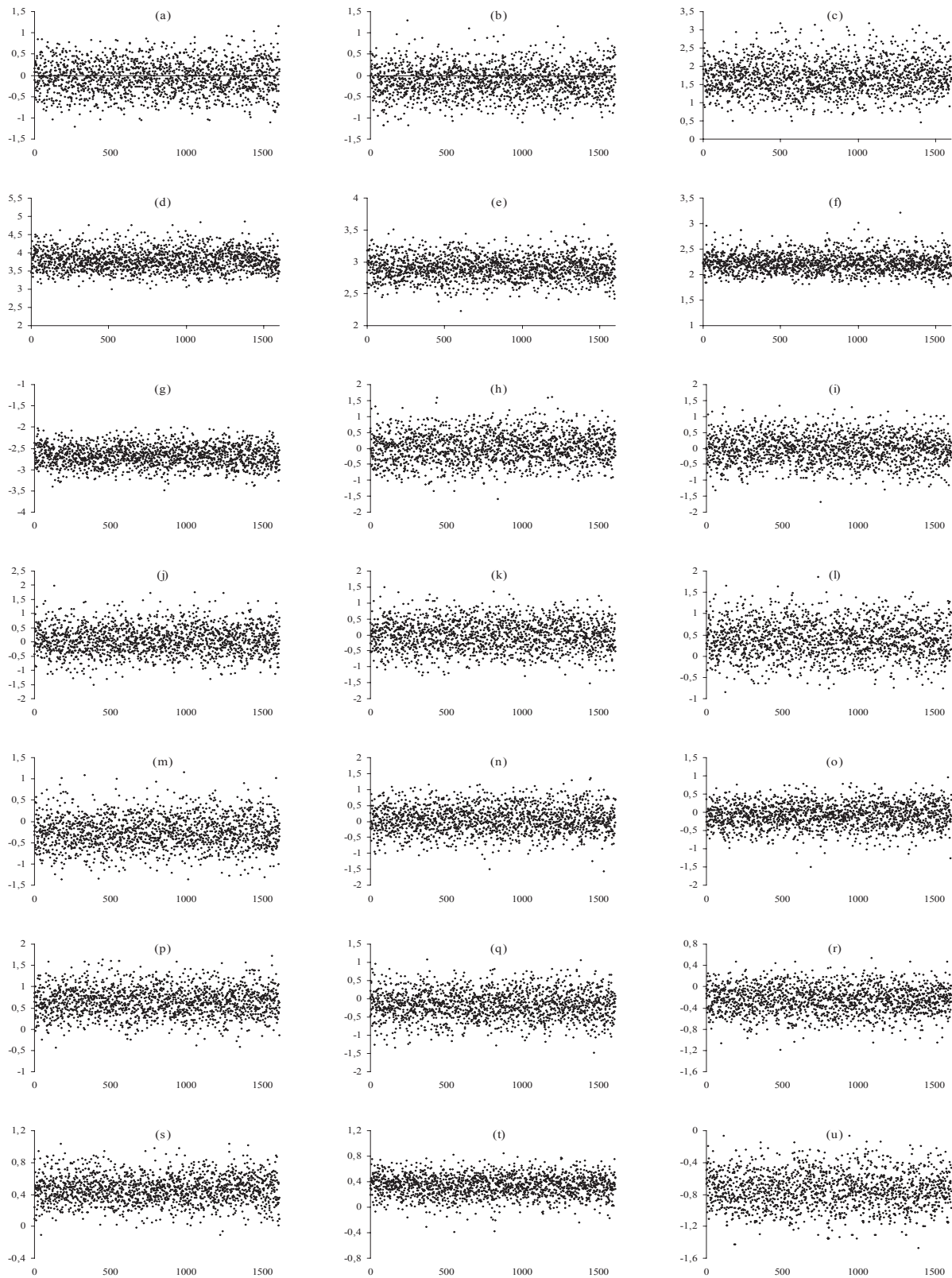
Figure 1. Trace plots of the sample values versus iteration for the parameters β_{10} (a), β_{20} (b), β_{30} (c), β_{40} (d), β_{50} (e), β_{60} (f), β_{70} (g), β_{11} (h), β_{21} (i), β_{31} (j), β_{41} (k), β_{51} (l), β_{61} (m), β_{71} (n), β_{12} (o), β_{22} (p), β_{32} (q), β_{42} (r), β_{52} (s), β_{62} (t) and β_{72} (u).

Table 6. Posterior odds ratios as association measures between pregnancy and age and the performance measures of the cervical cytology, VIA and HC II.

parameter	measure	mean	SD	95% CI	
$e^{\beta 11}$	effect of pregnancy on S_{e1}	1.142	0.566	0.420	2.496
$e^{\beta 21}$	effect of pregnancy on S_{e2}	1.079	0.502	0.400	2.291
$e^{\beta 31}$	effect of pregnancy on S_{e3}	1.214	0.667	0.413	2.864
$e^{\beta 41}$	effect of pregnancy on S_{p1}	1.105	0.528	0.416	2.384
$e^{\beta 51}$	effect of pregnancy on S_{p2}	1.626	0.710	0.670	3.425
$e^{\beta 61}$	effect of pregnancy on S_{p3}	0.840	0.343	0.381	1.692
$e^{\beta 71}$	effect of pregnancy on P	1.168	0.509	0.479	2.451
$e^{\beta 12}$	effect of age on S_{e1}	0.973	0.318	0.481	1.694
$e^{\beta 22}$	effect of age on S_{e2}	2.033	0.699	1.013	3.737
$e^{\beta 32}$	effect of age on S_{e3}	0.914	0.364	0.393	1.856
$e^{\beta 42}$	effect of age on S_{p1}	0.804	0.202	0.469	1.239
$e^{\beta 52}$	effect of age on S_{p2}	1.615	0.282	1.131	2.272
$e^{\beta 62}$	effect of age on S_{p3}	1.422	0.238	0.996	1.925
$e^{\beta 72}$	effect of age on P	0.483	0.109	0.313	0.738

Abbreviations: SD: standard deviation; CI: credibility interval.

performance screening measures and its subjectiveness from the respective prior distributions. Therefore, more accurate results would be given if we incorporate reasonable prior distributions based on prior knowledgement of clinical experts.

Although the proposed model is able to estimate useful performance measures of serial and parallel combinations of the screening tests (see Tables 4 and 5), the relative gains in sensitivity and losses in specificity can be misleading (see details in Franco e Ferenczy [37]). Macaskill et al. [38] argued that the expected number of additional true positive and false positive results (or true negative and false negative results) can be used as the basis for deciding whether to use tests in combination when neither the combined nor a component test shows superior test performance based on their likelihood ratios. Thus, the comparison between the likelihood ratios for two competing tests can be used to assess the incremental gain from an adjunct test. An extension of the presented Bayesian model considering the inclusion of parameters that describe the comparison between likelihood ratios as proposed by Macaskill et al. [38] should be also considered in future studies.

The major shortcoming of the Bayesian estimating method resides in the necessary presumption that the diagnostic tests are statistically and conditionally independent. This presupposition might not be invariably true [39] and alternative methods were proposed by Espeland and Handelman (1989); Yang and Becker (1997) and Dendukuri

and Joseph (2001) to address those situations [40, 41, 13]. However, all of these approaches address situations in which the correlation between two screening tests is considered and extensions for three or more tests are not found in the literature. Bayesian models that include the conditional dependence between multiple screening tests should be considered in future studies.

It is also important to point out that the diagnostic tests evaluated in this study have some inherent flaws. The lack of accuracy and reproducibility of cervical cytology is explained by the biological variability, sample quality, subjective interpretation of morphological abnormalities and examiners fatigue derived from repetitive procedures [42]. VIA requires training, although an obvious trend towards examiner subjectivity is always present. HC II results are more reproducible than those of VIA and cervical cytology.

Thus, VIA suggested by Belinson et al. (2001) as a screening method is likely to assume a central role in the prevention of cervical cancer in many countries. This simple and inexpensive method does not require complex technical supplies and it allows diagnosis and treatment at a single visit [25]. Coste et al. (2003) evaluated the performances of conventional cytology, liquid-based cytology and HC II in detecting cervical lesions with a sample of 1,757 women, with a combination of colposcopy and biopsy as the gold standard [43]. This impressive number of colposcopies and biopsies certainly reduced the verification bias, but our results still substantiate the view that this type of statistical modelling could provide reliable results

using fewer patients either subjected to the gold standard or simply with no gold standard. This is because of the natural influence of the disease prevalence on the values of PPV and NPV. The prevalence of histologically confirmed cervical abnormalities is necessarily low in a healthy population and screening tests usually have low PPV and high NPV, because true positives are rare and true negatives are abundant [30].

Three Indian studies in the late 1990s provided evidence supporting VIA as a viable alternative to cytology as a primary screening test [44–47]. In one of these studies, Londhe et al. [45] evaluated 12,372 women that underwent VIA, Pap smear and colposcopy in a gynecology outpatient clinic. VIA identified 78 percent of high-grade cervical lesions diagnosed with colposcopy, 3.5 percent more than were identified by cytology. In a 1998 Indian study [46] involving 3,000 women, VIA and cytology (done only by cytotechnicians) performed very similarly (sensitivity ratio of 1.05) in terms of detecting moderate/severe dysplasia. The approximate specificity of VIA in this study was 92.2 percent compared with 91.3 percent for cytology. In a third Indian study published in 1999, Sankaranarayan et al. found that VIA detected significantly more moderate/severe lesions than cytology but its specificity was significantly lower [47]. A large-scale study (over 10,000 women) in Zimbabwe compared the performance of VIA and the Pap smear in the hands of nurse midwives in primary health clinics. Phase II of this study was the first to provide direct estimates of sensitivity/specificity because all women testing negative or positive were offered the reference standard (colposcopy and biopsy, if indicated). In this study, the sensitivity of VIA (for high-grade positivity) was 1.75 times higher than that of cytology (76.7 vs. 44.3 percent respectively) whereas the specificity was 1.4 times lower [26].

The studies mentioned above yielded valuable information regarding the performances of VIA and cytology. Our Bayesian estimates provided performance values that can be compared with the results obtained in the standard manner, that is, with the use of a gold standard. In our estimates, VIA and cytology had similar sensitivity and specificity, as well as PPV and NPV. These figures contradict the previously published direct estimates that reported a superiority of VIA in detecting cervical lesions [46,47]. These differences may be attributable to methodological incompatibilities in

sample collection or processing and, as mentioned before, to the inherent difficulties derived from the variability of cytology and VIA interpretation.

The performance of HPV test in screening settings has been extensively studied, but to a lesser extent in comparison with VIA and cytology. Denny et al. (2000) published their data on 2,944 women subjected to VIA, cytology and HPV testing. In this study, VIA and HPV ($>$1 RLU) were similar to cytology in their performance of detecting high-grade lesions, but VIA yielded the largest number of false-positives among the three testing modalities [48]. More recently, these same authors [49] published a methodologically similar study testing cervicography, VIA, HPV test and cytology. In this study, 2,754 previously unscreened South African women were subjected to the four exams and VIA detected significantly more high-grade lesions as compared to the other screening tests.

The shortcomings of the tests (as reproducibility, subjective interpretation of results and required training of professional) did not hamper the Bayesian estimates but, in contrast, would enhance the search for realistic estimating equations. Thus, the inclusion of covariates (or control variables) should be encouraged in studies designed for the evaluation of estimating methodologies. According to Parmigiani (2002), prediction models used to support the clinical and health policy decision making need to consider the course of the disease over an extended period of time and draw evidence from a broad knowledge base, including epidemiological cohort and case control studies, randomized clinical trials, expert opinions and more [50]. In these cases, Bayesian decision theory and the tools typically used to describe the uncertainties involved could be extremely useful. The age of the patient is an important covariate in the study of cervical carcinoma precursor lesions. Koss [51] and Schiffman et al. [27,44] did not recommend the use of HPV testing in young women because the prevalence of the virus is exceptionally high in this group and the majority of such infections will spontaneously regress in the short term. Cervical lesions only develop in the presence of persistent HPV infections, thus HPV testing in young women will reveal an excessive number of HPV positive subjects that will never develop HPV-related cervical precancer lesions.

We can conclude that the estimated performances of VIA, HC II and cytology clearly show

that the Bayesian method is a remarkable tool for validating diagnostic tests when a gold standard is available for a very limited number of cases or not available at all.

Acknowledgements

Sources of Funding

References

[1] Begg, C.B. and Greenes, R.A. 1983. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, 39:207–215.

[2] Zhou, X. 1983. Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Commun Statis Theory Meth.*, 22:3177–3198.

[3] Hui, S.L. and Walter, S.D. 1980. Estimating the error rates of diagnostic tests. *Biometrics*, 36:167–171.

[4] Joseph, L., Gyorkos, T.W. and Coupal, L. 1985. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*, 141:263–272.

[5] Hitt, E. 2003. Cancer in the Americas. *Lancet Oncol*, 4:9.

[6] Brasil. Ministério da Saúde. Secretaria Nacional de Assistência à Saúde. Instituto Nacional do Câncer. Estimativas da incidˆencia e mortalidade por cˆancer no Brasil. Rio de Janeiro: INCA; 2002. Available on website: <http://www.inca.org.br/cancer/epide miologia/estimativa2002/estimativas.html>.

[7] Mitchell, M.F., Schottenfeld, D., Tortolero-Luna, G. and Cantor, S.B. 1998. Richards-Kortum R. Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol*, 91:626–631.

[8] Hopman, E.H., Kenemans, P. and Helmerhorst, T.J. 1998. Positive predictive rate of colposcopic examination of the cervix uteri: an overview of literature. *Obstet Gynecol Surv*, 53:97–106.

[9] Begg, C.B. 1987. Biases in the assessment of diagnostic tests. *Stat Med*, 6:411–423.

[10] Hui, S.L. and Zhou, X.H. 1998. Evaluation of diagnostic tests without gold standards. *Stat Methods Med Res*, 7:354–370.

[11] Zhou, X.H. 1998. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res*, 7:337–353.

[12] McCrory, D.C., Matchar, D.B., Bastian, L. et al. 1999. Evaluation of cervical cytology. Evidence report/technology assessment n.5. (Prepared by Duke University under Contract n. 290-97-0014). AHCPR publication n. 99-E010. Rockville: Agency for Health Care Policy and Research.

[13] Dendukuri, N. and Joseph, L. 2001. Bayesian approaches to modelling the conditional dependence between multiple diagnostic tests. *Biometrics*, 57:208–217.

[14] Faraone, S.V. and Tsuang, M.T. 1994. Measuring diagnostic accuracy in the absence of a "gold standard". *Am J Psychiatry*, 151:650–657.

[15] Qu, Y., Tan, M. and Kutner, M.H. 1996. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52:797–810.

[16] Hadgu, A. and Qu, Y. 1998. A biomedical application of latent models with random effects. *Appl Statist*, 47:603–616.

[17] Tanner, M. and Wong, W. 1987. The calculation of posterior distributions by data augmentation. *J Am Statist Ass*, 82:528–550.

[18] Gelfand, A.E. and Smith, AFM. 1990. Sampling Based Approaches to Calculating Marginal Densities. *J Am Stat Assoc*, 85:398–409.

[19] Gelfand, A.E. 2000. Gibbs sampling. *J Am Stat Assoc*, 95:1300–1304.

[20] Syrjanen, K., Naud, P., Derchain, S. et al. 2005. Comparing PAP smear cytology, aided visual inspection, screening colposcopy, cervicography and HPV testing as optional screening tools in Latin America. Study design and baseline data of the LAMS study. *Anticancer Res*, 25:3469–3480.

[21] Sarian, L.O., Derchain, S.F., Naud, P. et al. 2005. Evaluation of visual inspection with acetic acid (VIA), Lugol's iodine (VILI), cervical cytology and HPV testing as cervical screening tools in Latin America. This report refers to partial results from the LAMS (Latin AMerican Screening) study. *J Med Screen*, 12:142–149.

[22] Solomon, D., Davey, D., Kurman, R. et al. 2002. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA*, 287:2114–2119.

[23] Blumenthal, P. and Sanghvi, H. 1997. Atlas for unaided visual inspection of the cervix. Baltimore and Harare: JHPIEGO Corporation and University of Zimbabwe Medical School.

[24] Nanda, K., McCrory, D.C., Myers, E.R. et al. 2000. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*, 132:810–819.

[25] Belinson, J.L., Pretorius, R.G., Zhang, W.H., Wu, L.Y., Qiao, Y.L. and Elson, P. 2001. Cervical cancer screening by simple visual inspection after acetic acid. *Obstet Gynecol*, 98:441–444.

[26] University of Zimbabwe/JHPIEGO Cervical Cancer Project. Visual inspection with acetic acid for cervical-cancer screening: test qualities in a primary care setting. *Lancet* 1999; 353:869–73.

[27] Schiffman, M., Herrero, R., Hildensheim, A. et al. 2000. HPV DNA testing in cervical cancer screening: results from women in a high-risk province of Costa Rica. *JAMA*, 283:87–93.

[28] Wright, T.C., Jr, Lynette, D., Kuhn, L., Pollack, A. and Lorincz, A. 2000. HPV DNA testing of self-collected vaginal samples compared with cytologic screening to detect cervical cancer. *JAMA*, 283:81–86.

[29] Box, GEP. and Tiao, G.C. 1992. Bayesian Inference in Statistical Analysis. Reprint edition. New York: Wiley-Interscience.

[30] Altman, D.G. and Bland, J.M. 1994. Diagnostic tests 2: predictive values. *BMJ*, 309:102.

[31] Franco, E.L. 2003. Primary screening of cervical cancer with human papillomavirus tests. *J Natl Cancer Inst Monogr*, 31:89–96.

[32] Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics 4*; Bernardo, J.M.; Berger, J.O.; Dawid, A.P and.Smith, A.F.M., Eds.; Clarendom Press: Oxford, 169–94.

[33] Carlin, B.P. and Louis, T.A. 2000. Bayes and empirical Bayes methods for data analysis. *2nd ed. London: Chapman and Hall/CRC*,

[34] Koss, L.G. 2000. Human papillomavirus testing as a screening tool for cervical cancer. *JAMA*, 283:2525.

[35] Shlay, J.C., Dunn, T., Byers, T., Barón, A.E. and Douglas, J.M. 2000. Prediction of cervical intraepithelial neoplasia grade 2–3 using risk assessment and human papillomavirus testing in women with atypia on Papanicolaou Smears. *Obstet Gynecol*, 96:410–416.

[36] Spiegelhalter, D.J., Best, N.G. and Carlin, B.P. 2002. van der Linde A Bayesian measures of model complexity and fit (with discussion). *J Roy Statist Soc B*, 64:583–640.

[37] Franco, E.L. and Ferenczy, A. 1999. Assessing gains in diagnostic utility when human papillomavirus testing is used as an adjunct to papanicolaou smear in the triage of women with cervical cytologic abnormalities. *Am J Obstet Gynecol*, 181:382–386.

[38] Macaskill, P., Walter, S.D., Irwig, L. and Franco, E.L. 2002. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Statis Med*, 21:2527–2546.

[39] Brenner, H. How independent are multiple "independent" diagnostic classifications?. *Stat Med*,

[40] Espeland, M.A. and Handelman, S.L. 1989. Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics*, 45:587–599

[41] Yang, I. and Becker, M.P. 1997. Latent variable modeling of diagnostic accuracy. *Biometrics*, 53:948–958.

[42] Ratman, S., Franco, E.L. and Ferenczy, A. 2000. Human papillomavirus testing for primary screening of cervical cancer precursors. *Cancer Epidemiol Biomarkers Prev*, 9:945–951.

[43] Coste, J., Cochand-Priollet, B., de Cremoux, P. et al. 2003. Cross sectional study of conventional cervical smear, monolayer cytology and human papillomavirus DNA testing for cervical cancer screening. *BMJ*, 326:733.

[44] Schiffman, M., Hildesheim, A., Herrero, R. and Bratti, M. 2000. In reply: human papillomavirus testing as a screening tool for cervical cancer. *JAMA*, 283:2525–2526.

[45] Londhe, M., George, S.S. and Seshadri, L. 1997. Detection of CIN by naked eye visualization after application of acetic acid. *Indian J Cancer*, 34:88–91.

[46] Sankaranarayanan, R., Wesley, R., Somanathan, T. et al. 1998. Visual inspection of the uterine cervix after the application of acetic acid in the detection of cervical carcinoma and its precursors. *Cancer*, 83:2150–2156.

[47] Sankaranarayanan, R., Shyamalayumary, B., Wesley, R., Sreedevi Amma, N., Parkin, D.M. and Nair, M.K. 1999. Visual inspection with acetic acid in the early detection of cervical cancer and precursors. *Int J Cancer*, 80:161–163.

[48] Denny, L., Kuhn, L., Pollack, A., Wainwright, H. and Wright, TC Jr. 2000. Evaluation of alternative methods of cervical cancer screening in resource-poor settings. *Cancer*, 89:826–833.

[49] Denny, L., Kuhn, L., Pollack, A. and Wright, TC Jr. 2002. Direct visual inspection for cervical cancer screening: an analysis of factors influencing test performance. *Cancer*, 94:1699–1707.

[50] Parmigiani, G. 2002. Measuring uncertainty in complex decision analysis models. *Stat Methods Med Res*, 11:513–537.

[51] Koss, L.G. 2000. Human papillomavirus testing as a screening tool for cervical cancer. *JAMA*, 283:2525–2526.

[52] Smith, AFM. and Roberts, G.O. 1993. Bayesian Computation via the Gibbs Sampler and Related MCMC Methods. *J R Stat SocB* , 55:3–24.