Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011)
P. Qian, Y. Zhou, C. Rudin, eds.

AN MDP APPROACH TO CERVICAL CANCER SCREENING

Diana Sánchez
Centro para la Optimización y Probabilidad Aplicada (COPA)
Departamento de Ingeniería Industrial
Universidad de los Andes
Bogotá D.C., Colombia

Raha Akhavan-Tabatabaei
Centro para la Optimización y Probabilidad Aplicada (COPA)
Departamento de Ingeniería Industrial
Universidad de los Andes
Bogotá D.C., Colombia
r.akhavan@uniandes.edu.co

Thomas G. Yeung
Department of Industrial Engineering & Automatic Control
Ecole des Mines de Nantes / IRCCyN
Nantes, France 44307
thomas.yeung@emn.fr

Cristina López Garzón
Centro para la Optimización y Probabilidad Aplicada (COPA)
Departamento de Ingeniería Industrial
Universidad de los Andes
Bogotá D.C., Colombia

Abstract

Cervical cancer, caused by the human Papillomavirus (HPV), is the second most common cancer in women after breast cancer. Proper screening through the Pap test can reduce cancer incidences by 60%. Abnormal Pap tests are followed by a colposcopy procedure in order to obtain further information. We propose a finite-horizon Markov Decision Process (MDP) to find the optimal screening policy for cervical cancer prevention. The model decides whether an additional screening should be performed beyond the annual Pap frequency and whether the Pap test should be bypassed in favor of a direct colposcopy to eliminate an additional procedure and trip to the doctor's office. Decisions are made based on the women's current diagnosis, age, HPV contraction risk, and screening results. The cost function considers the tradeoffs between prevention and treatment procedures and the risk of taking no action while taking into account the life-quality of patients. We apply the model to data collected from a representative sample of 1,141 Colombian women. The results suggest that in order to effectively track disease progress and avoid higher cancer rates and future costs, more frequent screenings should be carried out on patients with higher risk profiles. However, the annual guideline is sufficient for healthy women.

Keywords: Cervical cancer screening, PAP test, Markov Decision Processes.

1. Introduction

Cervical cancer in women is the second most common cancer around the world, and the first cause of cancer related deaths in Colombia [2]. Finding effective prevention policies can reduce the mortality by 69% to 81% depending on the strategy [1], and also can decrease the incidence rates of cervical cancer, which is of high importance for Latin American health systems. In this paper, we propose a finite-horizon Markov Decision Process model to find an optimal policy for prevention of cervical cancer. We develop the model based on information provided by The Colombian National Cancer Institute (INC) and Colsubsidio, a Colombian health management company.

Several mathematical models have been proposed to study cervical cancer through analyzing the natural history of the disease, the implication of the Human Papillomavirus (HPV) types on the disease, screening and prevention programs among others. Nevertheless, in these types of medical models the behavior of the natural disease varies across countries and regions.

Gamboa *et. Al.* [1] asses the cost-effectiveness of cervical cancer screening in Colombia, and find that doing an HPV DNA test every 5 years is cost-effective only if the test does not exceed USD $31, otherwise doing PAP tests (with 1-1-3 scheme) is the most cost-effective strategy. Murillo [6] points out that it is not sufficient to only have high coverage of population, it is also important to do a follow-up for women with abnormal results in the Pap test.

McLay and Merrick [5] develop a simulation-based optimization model to find the optimal age to begin performing screening and the results suggest that doing less screening per year reduces the cost and offers the same health benefits. Schaefer *et al.* [7] describe MDP modeling as an appropriate technique for stochastic and dynamic medical decision-making.

2. Problem Statement

Currently, the advice given by the Ministry of Health in Colombia is that all women above the age of 25 and those below 25 who have already started their sexual activity, should repeat the Pap test in a 1-1-3 scheme. This scheme implies performing the test in two consecutive years and repeat in three years if the results of the first two are negative, regardless of the individual risk factors of women.

The principal screening procedures implemented in Colombia are the Pap test, colposcopy and biopsy. The Pap test diagnoses whether or not a particular patient has any abnormal cells in her cervix. Currently in Colsubsidio, if the result of the test is normal, the patient is recommended to perform a Pap test every year and the patient is classified as healthy. However, in the case of abnormal results, the patient must receive a colposcopy. Finally, according to the colposcopy results, the patient must receive a biopsy for determining the stage of the abnormality.

Figure 1 shows the existing screening and treatment practice performed by Colsubsidio in cervical cancer patients. In this figure, circles represent the possible diagnosed states of a patient and rectangles the possible treatment that a patient can receive for each diagnosed state and her individual status (age and HPV risk). The diagnosis of atypical cells corresponds to Atypical Squamous Cells of Undetermined Significance (ASCUS). The Squamous Intraepithelial Lesion (SIL) can be classified as high (HSIL) or low (LSIL). Along with this, “Caut” refers to cauterization, “Con” refers to conization, “Hyst” refers to hysterectomy, “CH-R” to chemotherapy and radiotherapy, and “Pal” to palliative care.

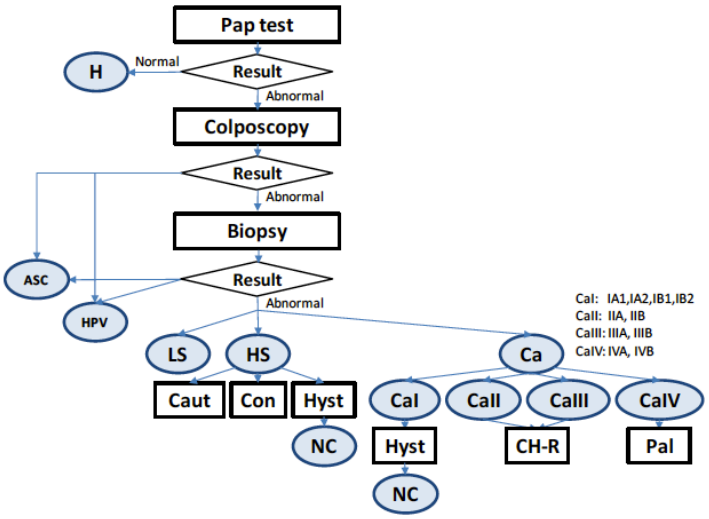

Figure 1 - Current practice of Colsubsidio in screening and treatment of cervical cancer patients

As discussed in the previous section, in our study we focus on the preventive part and not on the treatment decisions after the patient has been diagnosed with cancer. More specifically, we do not consider treatment decisions related to a cancer patient such as hysterectomy, chemotherapy, radiotherapy or palliative care.

3. Model Formulation

In this research our objective is to find the optimal prevention treatment policy for the current practice in Colsubsidio based on their personal information such as age, the type of virus acquired, the level of abnormality, and the record of the last Pap test performed. We formulate the problem as an MDP, minimizing the expected costs of screening and treatment over a finite horizon (expected normal lifetime of a patient) and propose optimal preventive policies for different categories of patients.

3.1 Components of the MPD Model

The model is designed such that it guarantees one Pap test per year for any given patient to accomplish the government regulation and Colsubsidio's minimum requirements. Then after six months we evaluate the action of either repeating it, or performing a colposcopy without doing a prior Pap test. Otherwise we can decide to take no action. We choose the decision epoch of every 6 months and a finite decision horizon. We choose to model a finite horizon MDP since human patients have finite lifetime expectancy.

We define the state space using three variables that describe the principal aspects regarding the status of a patient. We denote these variables by s, a and z. The first variable s corresponds to the current diagnosis of the patient including information about the risk related to the HPV type, if applicable. The second variable a represents the age of the patient that starts from 15 and goes beyond 65 years. Finally, the binary variable z tracks the record of the last screening test. A value of 1 indicates that the patient received a screening test in the last decision epoch and 0 if she did not. Table 1 describes the notation and all possible values that each variable can take on.

Table 1 - Feasible observations in the state elements

STATE →X : { s, a, z}

s	Diagnosis of the patient	Healthy	ASCUS	HPV Low Risk	HPV High Risk	LSIL Low Risk	LSIL High Risk	HSIL	Cancer	NO Cervix	Dead	NO info.
	Notation	H	ASC	HPV_LR	HPV_HR	LS_LR	LS_HR	HS	CA	NC	DEAD	NOI
a	Age	15 to 65+										
z	Record of last screening test	1	If the patient received a screening test in the last decision epoch.									
		0	If the patient DID NOT received a screening test in the last decision epoch.									

The set of actions that can be taken include "taking no action" denoted as do nothing (DN), "performing a Pap test" referred to as Pap test (P), "performing colposcopy without doing Pap test" expressed as colposcopy (C). We do not take into account other procedures since they are mandatory as they follow directly from the state diagnosis and the individual condition of the patient.

3.2. Data for Transition Probabilities

The transition probabilities are calculated taking into account the possible transitions that a patient can make from a specific state as shown in Figure 2. The following assumptions regarding the progression and regression of the disease are considered: (1) The HPV risk of a patient cannot change in one decision epoch, because they are related with two different types of virus and a patient needs to get rid of one type to acquire the other. (2) All screening tests are 100% accurate, with complete sensitivity and specificity. (3) Once the cancer develops it never regresses without the treatment intervention.

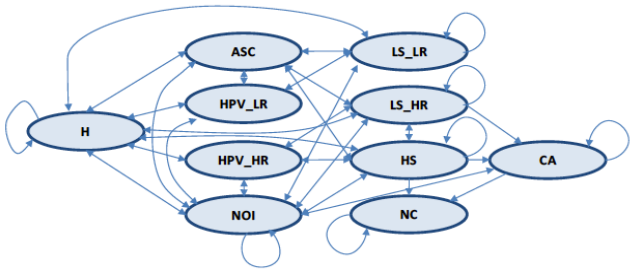

Figure 2 - Transition diagram for the MDP

In order to find the transition probabilities we acquired information from different Colombian patient sources. For a population of Colombian women, Gamboa *et al.* [1] gives the transition probabilities between states for the natural behavior of cervical cancer. For the other two decisions, namely P and C, Colsubsidio provides information from their patients.

3.3. Value Function

The value function integrates the immediate and the future expected costs of taking an action. The immediate cost includes cost of the treatments performed in each state, and all costs related with the action taken, as well as an intangible cost related to the life-quality of a patient in each state. The future cost involves the optimal value function of the states that are reachable starting at a specific state.

The cost of treatments in our model only takes values greater than 0 for HSIL and CA states as they are the only diagnosed states where the patients receive a treatment, as demonstrated in Figure 1. The second cost component depends on the possible actions namely, Pap test and Colposcopy, and considers the cost of a Biopsy if the patient has a second visit. The above mentioned cost components are listed in Table 2 along with their current values in Colombia (in US dollars).

Table 2 Costs of treatment and Cost that depend on the action

State	Description	Treatment Cost (US dolars)
HS	HSIL	$ 360.00
CA	Cancer	$ 3,086.00

Notation	Description	Value (US dolars)
gp	Pap test	$ 6.34
gc	Colposcopy	$ 45.61
gb	Biopsy	$ 90.34
g2	2° Visit to the doctor	$ 10.00

Table 3 Life-quality associated with each state and range age

State \ Age	15	20	25	30	35	40	45	50	55	60	65
H	0,93000	0,91000	0,91000	0,91000	0,89000	0,89000	0,86000	0,86000	0,80000	0,80000	0,74000
ASC	0,85560	0,83720	0,83720	0,83720	0,81880	0,81880	0,79120	0,79120	0,73600	0,73600	0,68080
HPV_LR	0,69750	0,68250	0,68250	0,68250	0,66750	0,66750	0,64500	0,64500	0,60000	0,60000	0,55500
HPV_HR	0,69750	0,68250	0,68250	0,68250	0,66750	0,66750	0,64500	0,64500	0,60000	0,60000	0,55500
LSIL_LR	0,76260	0,74620	0,74620	0,74620	0,72980	0,72980	0,70520	0,70520	0,65600	0,65600	0,60680
LSIL_HR	0,76260	0,74620	0,74620	0,74620	0,72980	0,72980	0,70520	0,70520	0,65600	0,65600	0,60680
HSIL	0,74400	0,72800	0,72800	0,72800	0,71200	0,71200	0,68800	0,68800	0,64000	0,64000	0,59200
NC	0,65100	0,63700	0,63700	0,63700	0,62300	0,62300	0,60200	0,60200	0,56000	0,56000	0,51800
NOI	0,58590	0,57330	0,57330	0,57330	0,56070	0,56070	0,54180	0,54180	0,50400	0,50400	0,46620
CA	0,51026	0,49929	0,49929	0,49929	0,48832	0,48832	0,47186	0,47186	0,43894	0,43894	0,40602
Dead	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000

The last component of the value function is the intangible cost. We include this cost through the value of life-quality related to each healthy state. We take the quality of life associated with all symptoms based on the HALex (Health and Activity Limitation Index) [4]. We consider the quality associated with each state including not only the health state but also the age. In Table 3 we show the life-quality associated with each state and age range, where the maximum value is 1, which represents a perfectly healthy state value, and the minimum value is 0, which means the worst state, death.

3.4. Minimizing The Cost Function

The MDP formulation shown in Equation (1) attempts to find the set of policies that minimize the costs across a finite horizon, where t is the decision moment, x_t denotes the current state at time t, from the set of feasible states X_t, and d denotes the action taken from the set of feasible actions. Also, $x_t{}'$ denotes the transitioning state from the set of feasible actions, $P_t(x_t{}'|x_t, d)$ the probability of transitioning from x_t to $x_t{}'$ given action d, $r_t(x_t, d)$ represents the immediate cost at decision moment t of being in state x_t and taking decision d, and $V_t^*(x_t)$ is the minimal value at decision moment t for state x_t. The immediate cost or $r_t(x_t, d)$ is indicated Equation (2) and can be: $DN(x_t)$ if the decision is Do Nothing, $P(x_t)$ if the decision is Pap test and $C(x_t)$ if the decision is Colposcopy.

$$V_t^*(x_t) = \min_{d \in D_{x_t}} \left\{ r_t(x_t, d) + \sum_{x_t \in X_t} P_t(x_t{}'|x_t, d) * V_{t+1}^*(x_t) \right\} \quad \forall\, t, x_t \qquad (1)$$

$$r_t(x_t, d) = \min\{DN(x_t), P(x_t), C(x_t)\} \qquad \forall\, t, x_t \qquad (2)$$

3.5. Numerical Results

The following results are calculated assuming that the outcome in screening tests follows a decreasing function probability according to a health status. For example, the probability that the Pap test has normal results for a patient that starts in the system as Healthy should be higher than for a patient that at the beginning is in a LSIL. Table 4 shows the results for optimal preventive policy given the data from Colsubsidio. We consider that the diagnosis of the patient deteriorates the normal outcomes probabilities

by decreasing in 10 % the initial probability, the one given by Colsubsidio and the one that characterizes the healthy state. We denote by 1 the Pap test, by 2 the Colposcopy and by 3 the Do Nothing decisions. For example according to these results, when a patient is Healthy receives a Pap test every year and a extra Pap test par year should not be perform until the patient reaches 65 years. If a patient has ASCUS and is older than 25 the optimal policy is to perform two Pap tests per year until the atypical cells are identified as dangerous or benign. Finally, if a patient is younger than 20, and has a lesion more severe that HPV she should always undergo a colposcopy.

Table 4 Policy assuming probabilities of having normal screening outcomes reduced by 10% as the lifetime quality of health state decreases

	Optimal Actions to take for each state and each decision moment																																																		
AGE	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	Policy cost
H, 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.583,3482
H, 1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	$ 12.573,0136
ASC, 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.586,4058
ASC, 1	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.585,9348
HPV_L, 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.589,2953
HPV_L, 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.589,2953
HPV_H, 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.592,0084
HPV_H, 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.592,0084
LS_H, 0	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.594,4694
LS_H, 1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.594,4694
LS_L, 0	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.596,6553
LS_L, 1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.596,6553
HS, 0	2	2	2	2	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.598,6541
HS, 1	2	2	2	2	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.598,6541
NC, 1	2	2	2	2	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.600,4970
NOI, 0	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.602,1117
CA, 1	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$ 12.603,4820
	Policy Cost																																																			$ 226.690,8384

4. Conclusions

We propose a finite-horizon MDP model to find optimal policies for cervical cancer prevention in Colombian women that consist not only of which screening test to perform but also the frequency of performing it, according to different conditions of the patient such as life-quality, age and her risk of acquiring Human Papillomavirus (HPV).

An important follow up to this work is to obtain more specific data not only for the transition probabilities related to the HPV risk, but also for the intangible costs of each action in terms of adapting to analyze QUALY´s against just quality. In this work we take into account that the exams are 100 % accurate, so for future work it is of interest to consider the error margin that this type of exams has.

5. References

1. Gamboa, O.-A., Chicaíza, L., Garcia-Molina, M., Diaz, J., Gonzalez, M., Murillo, R., et al. (2008). Cost-effectiveness of conventional cytology and HPV DNA testing for cervical cancer screening in Colombia. *Salud Pública de México Vol.50 No.4* .
2. GLOBOCAN 2008. *International Agency for Research on Cancer.* Cancer Incidence and Mortality Worldwide in 2008: http://globocan.iarc.fr/
3. International Agency for Reaserch on Cancer - *World Health Organization*. FIGO staging of cervical carcinomas: http://screening.iarc.fr/viaviliappendix1.php
4. Gold, M.R., Franks, P., McCoy, K. I., & Fryback, D. G. (1998) Toward Consistency in Cost -Utility. Analyses Using Natianl Measures to Create Condition-. *Medical Care Vol. 36 N.6, 778-792.*
5. McLay, L. A., & Merrick, J. (2010). Using simulation-optimization to construct screening strategies for cervical cancer. *Health Care Manag Sci* .
6. Murilo, R. (2007). Programas de Tamización de Cáncer de cuello uterino basados en citología en América Latina. *HPV TODAY Newsletter on Human Papillomavirus No.12*.
7. Schaefer, A., Bailey, M., Shechter, S, & Roberts, M. Modeling Medical Treatment Using Markov Decision Processes. *Operations Research and Health Care* , 598-616.