Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

A COMPARTMENTALIZED SIMULATION MODEL FOR EVALUATION OF HPV VACCINATION POLICIES IN COLOMBIA

Daniela Angulo Díaz
Raha Akhavan-Tabatabaei
Ivan Mura

Department of Industrial Engineering
Universidad de los Andes
Carrera Primera n. 18A-12
Bogotá, COLOMBIA

ABSTRACT

Cervical cancer (CC) is the second leading cause of cancer-related deaths among Colombian women, caused most commonly by Human Papillomavirus (HPV) infection. Screening programs, vaccination against HPV and improved socio-economic conditions have significantly reduced CC mortality rate over the last 40 years. Understanding the transmission dynamics of HPV infection is essential to the definition of cost-effective disease control strategies. We propose a compartmentalized epidemiological simulation model based on differential equations, which represents HPV transmission within the population, likelihood of infection clearance, virus induced appearance of precancerous lesions and eventually of CC. Time-dependent birth and natural mortality rates inferred from census are used to calibrate model population dynamics. Literature data and 5-years medical records of 3,428 Colombian women are used to estimate the infection dynamics and cancerous stages. The model allows evaluating the predicted effects of vaccination strategies against HPV, providing valuable support to healthcare decision-makers.

1 INTRODUCTION

Cervical cancer is the fourth most common cancer in women worldwide, with an estimated 530,000 new cases diagnosed and 270,000 deaths per year, of which more than 85% occur in low and middle income countries (World Health Organization 2015). In Colombia, approximately 4,462 new detected cases and 1,861 deaths are reported per year (Pardo and Cendales 2015). Although these figures remain alarming, Colombia has made significant progress in reducing them over the last 40 years. Through improvements in socio-economic conditions, implementing more effective screening programs and teaching self-care procedures the mortality rates have been reduced from 14 deaths per 100,000 women in 1987 to 7.08 deaths per 100,000 in 2013 (Ministerio de Salud 2013).

Cervical cancer is defined as the condition where there is an unregulated growth of malignant cells in the lower part of the uterus (American Cancer Society 2014a). This abnormal growth is almost always caused by the Human Papillomavirus (HPV)(World Health Organization 2015). In most cases, HPV is sexually transmitted and it is almost certain that any sexually active individual will be infected at some point in their lives (World Health Organization 2015). The majority of HPV infections resolve spontaneously within less than 2 years. However, persistent infection with specific types of HPV, most frequently types 16 and 18, may lead to precancerous lesions (World Health Organization 2015). *Cervical Intraepithelial Neoplasia (CIN)* is one of most common terms used to refer to precancerous lesions. CIN is graded on a scale of 1 to 3 based on how much of the cervical tissue looks abnormal when viewed under the microscope (American Cancer Society 2014b).

There are two vaccines that protect against HPV types 16 and 18, which are the most common causes of CC among the HPV strands. It has been proven that both vaccines work best if they are administrated prior to exposure to HPV, therefore the ideal time to get vaccinated is before the first sexual activity (World Health Organization 2015). Since the vaccine is preventive rather than therapeutic, once the infection is developed or the person acquieres some kind of pre-cancerous lesion the vaccine has no effect. That is why, although the vaccine can prevent the infection, it is not enough as the sole means of CC prevention (Profamilia 2015). Cytology and HPV-DNA tests are two of the most common screenings for cervical cancer detection in women worldwide. The importance of an effective screening and treatment lies in the fact that pre-cancerous cervical lesions can regress. Therefore, cervical cancer can be avoided.

Understanding the transmission dynamics of HPV infection is essential to the definition of cost-effective disease control strategies. Consequently, the objective of this paper is to create a simulation model that is able to reproduce Colombian population dynamics and the transmission of HPV within the population, which allows evaluating various vaccination polices. In order to accomplish our purpose, we create a compartmentalized epidemiological simulation model based on differential equations. Our model divides the population into three different categories; gender, age group and biological stage of HPV infection and cancerous lesions. This stratification allows us to observe the transmission of infection among a population that is homogeneous within each age compartment. Furthermore, our model also considers the likelihood of infection clearance, the virus induced appearance of precancerous lesions and eventually of CC. In addition, as a prevention policy, our model considers the options of vaccinating male and female members of the population. In order to calibrate the model we used data from the literature and five-year medical records of 3,428 Colombian women.

The rest of the paper is organized as follows: In Section 2 we review the existing literature on epidemiological simulation models of HPV infection. Section 3 covers the general methodology we used to build the model. In Section 4 we present the methods for data processing, as well as the validation and results of our model. Finally, Section 5 presents our concluding remarks and suggested future work.

2 LITERATURE REVIEW

Several simulation models have been developed to analyze the progression of infectious diseases. Dimitrov and Mayers (2010) proposed a mathematical approach for modeling the spread of infectious diseases such as Tuberculosis or HPV infection. They evaluate a compartmental Susceptible Infectious Recovered model with differential equations to reproduce disease dynamics and a contact network model as an analytical framework that captures the interactions of the population for disease transmission. Kim and Goldie (2008) compared the cost-effectiveness of vaccinating pre-adolescent girls against HPV with the cost-effectiveness of vaccinating older girls and women in catch-up programs in the United States. They developed an epidemiological model to simulate the transmission of HPV infection (specifically types 16 and 18) among the population, considering the dynamics of HPV transmission and the duration and effectiveness of the vaccine. They concluded that the effectiveness of the vaccination program will depend on the duration of the vaccine, and it will be optimized with higher vaccination coverage in pre-adolescent girls.

Orrego (2013) proposed a stochastic discrete compartmental model that simulates the transmission of HPV infection within the Colombian population. Their model evaluates different vaccination policies such as vaccinating girls, girls and boys or none, taking into consideration the economic and public health implications of each vaccination program. They conclude that the incidence and prevalence of the infection is significantly lower in vaccination programs compared to no vaccination. However, their study showed a significant reduction in deaths due to CC is only achievable if both women and men receive vaccination.

No previous model has been proposed to simulate HPV infection in Colombia as a deterministic continuous simulation model, with parameters taken from Colombian literature or estimated from Colombian medical records, taking into account time dependent birth and deaths rates and the possible regression from cancer to pre-cancerous lesions.

3 METHODOLOGY AND DATA PROCESSING

Since our goal is to create an epidemiological compartmentalized model, we first create a population model that replicates Colombian population dynamics, using differential equations. After validating this model, we add the dynamics of HPV transmission, creating a model that describes the dynamics of the infection within the population. Finally, we include the vaccination policy, thereby creating a simulation model that represents the infection transmission dynamics in Colombian population under a vaccination policy.

To model the population dynamics, we compartmentalize the population according to gender and age group, considering 15 age ranges: [0,4] [5,9] ... [65,69] and [70 or older]. We will refer to them as 1,2,3,..15 respectively. We assume that, within each of the 30 compartments, every individuals follows the same dynamics. The differential equations only consider births, deaths and population aging (moves through age compartments). Table 1 presents the variables and parameters used in the differential equations. The values of all parameters are given in the Appendix.

Table 1: Variables and parameters used in the population model.

Variable	Description
$F_i(t)$	Number of women of age range i at time t
$M_i(t)$	Number of men of age range i at time t
Parameter	
p_f	Probability that a newborn is a woman
$\pi_i(t)$	Reproductive rate for women in age range i
$\mu_i^F(t)$	Death rate for women in age range i
$\mu_i^M(t)$	Death rate for men in age range i

Equations (1)-(6) represent the population dynamics in Colombia, where θ is a constant representing the aging rate, i.e. the rate with which people move from age range i to age range $i+1$, $i = 1,2,$ $ldots,14$. For the sake of conciseness, we do not write the parameter t in the equations.

$$d/dt(F_1) = p_f \sum_j \pi_j F_j - \theta F_1 - \mu_1^F F_1 \tag{1}$$

$$d/dt(M_1) = (1-p_f) \sum_j \pi_j F_j - \theta M_1 - \mu_1^M M_1 \tag{2}$$

$$d/dt(F_i) = \theta(F_{i-1} - F_i) - \mu_i^F F_i \quad i = 2,3,\ldots,14 \tag{3}$$

$$d/dt(M_i) = \theta(M_{i-1} - M_i) - \mu_i^M M_i \quad i = 2,3,\ldots,14 \tag{4}$$

$$d/dt(F_{15}) = \theta F_{14} - \mu_{15}^F F_{15} \tag{5}$$

$$d/dt(M_{15}) = \theta M_{14} - \mu_{15}^M M_{15} \tag{6}$$

Equation (1) represents the change in the number of girls in age range 1 (0-4 years) at time t. This is equal to newborn girls, less those who pass to the next age compartment or die. This same logic applies to Equation (2) which is for number of boys in age range 1 at time t. Equations (3) and (4) are for the individuals, women and men respectively, in age range $i \leq 14$ at time t. This is equal to the individuals who come from an earlier age compartment, less the individuals who pass to the next age compartment or die. Equation (5) and (6) represents the changes in the number of individuals, women and men respectively, in age range [70 or older] at time t, given by the individuals who come from an earlier age compartment, less those who die.

We extend the population model by adding HPV transmission dynamics. We divide each previous compartment in multiple compartments, representing the biological state of HPV infection. Any individual at any age can be classified into one of the following categories:

- Susceptible: the individual has never had the infection and is susceptible to become infected.
- Infected: the individual is HPV infected, therefore can infect other people.

- Partially immune: the individual has already recovered from an HPV infection, and can become infected again with a reduced probability.

Additionally, any woman at any age, can also be classified as being in stage CIN1, when the VPH infection has evolved to CIN1, stage CIN23, when the infection has evolved to CIN2 or CIN3, and finally Cervical Cancer, when the infection has already led to advanced cancerous lesions. Figure 1 shows the possible transitions between biological compartments and defines the abbreviated names of biological compartments we shall be using in the models.

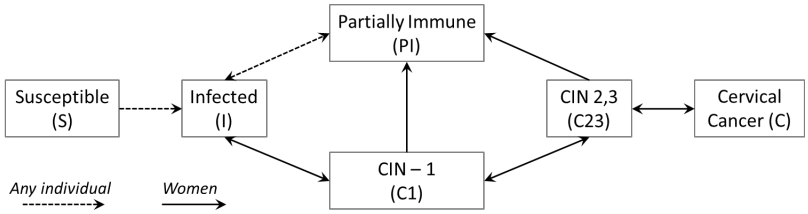

Figure 1: HPV transition diagram.

The infection model adds the HPV transmission within the population. Tables 2 provides the description of variables and parameters used in the model. When describing the infection model, we first provide the

Table 2: Variables and parameters in the infection model.

Variable	Description
$S_i^F(t)$	Number of susceptible women of age range i at time t
$S_i^M(t)$	Number of susceptible men of age range i at time t
$I_i^F(t)$	Number of infected women of age range i at time t
$I_i^M(t)$	Number of infected men of age range i at time t
$P_i^F(t)$	Number of women with partial immunity of age range i at time t
$P_i^M(t)$	Number of men with partial immunity of age range i at time t
$C1_i(t)$	Number of women with C1 of age range i at time t
$C23_i(t)$	Number of women with C23 of age range i at time t
$C_i(t)$	Number of women with cervical cancer of age range i at time t
Parameter	
$\mathbb{P}^F$	Transition probability matrix between biological stages for women
$\mathbb{P}^M$	Transition probability matrix between biological stages for men
μ_{ca}	Death rate due to CC

equations for men and and then for women. Differential equations for men are as follows:

$$d/dt(S_1^M) = (1-p_f)\textstyle\sum_j \pi_i F_i - \theta \mathbb{P}_{S,S}^M S_1^M - \mu_1^M S_1^M \tag{7}$$

$$d/dt(S_i^M) = \theta[\mathbb{P}_{S,S} S_{i-1}^M - (\mathbb{P}_{S,I} + \mathbb{P}_{S,S}) S_i^M] - \mu_i^M S_i^M \quad i = 2,3,\ldots,14 \tag{8}$$

$$d/dt(S_{15}^M) = \theta(\mathbb{P}_{S,S} S_{14}^M - \mathbb{P}_{S,I} S_{15}^M) - \mu_{15}^M S_{15}^M \tag{9}$$

$$d/dt(I_i^M) = \theta[\mathbb{P}_{I,I} I_{i-1}^M + \mathbb{P}_{S,I} S_{i-1}^M + \mathbb{P}_{PI,I} P_{i-1}^M - (\mathbb{P}_{I,PI} + \mathbb{P}_{I,I}) I_i^M] - \mu_i^M I_i^M \quad i = 1,2,\ldots,14 \tag{10}$$

$$d/dt(I_{15}^M) = \theta[\mathbb{P}_{I,I} I_{14}^M + \mathbb{P}_{S,I}(S_{14}^M + S_{15}^M) + \mathbb{P}_{PI,I}(P_{14}^M + P_{15}^M) - \mathbb{P}_{I,PI} I_{15}^M] - \mu_{15}^M I_{15}^M \tag{11}$$

$$d/dt(P_i^M) = \theta(\mathbb{P}_{PI,PI} P_{i-1}^M + \mathbb{P}_{I,PI} I_{i-1}^M - \mathbb{P}_{PI,PI} P_i^M) - \mu_i^M P_i^M \quad i = 1,2,\ldots,14 \tag{12}$$

$$d/dt(P_{15}^M) = \theta[\mathbb{P}_{PI,PI} P_{14}^M + \mathbb{P}_{I,PI}(I_{14}^M + I_{15}^M)] - \mu_{15}^M P_{15}^M \tag{13}$$

Differential equations for women:

$$d/dt(S_1^F) = p_f \sum_j \pi_i S_i^F - \theta \mathbb{P}_{S,S}^F S_1^F - \mu_1^F S_1^F \tag{14}$$

$$d/dt(S_i^F) = \theta[\mathbb{P}_{S,S}^F S_{i-1}^F - (\mathbb{P}_{S,I}^F + \mathbb{P}_{S,S}^F) S_i^F] - \mu_i^F S_i^F, \quad i = 2,3,\ldots,14 \tag{15}$$

$$d/dt(S_{15}^F) = \theta(\mathbb{P}_{S,S}^F S_{14}^F - \mathbb{P}_{S,I}^F S_{15}^F) - \mu_{15}^F S_{15}^F \tag{16}$$

$$\begin{aligned} d/dt(I_i^F) = {} & \theta[\mathbb{P}_{I,I}^F I_{i-1}^F + \mathbb{P}_{S,I}^F S_{i-1}^F + \mathbb{P}_{PI,I}^F P_{i-1}^F + \mathbb{P}_{C1,I}^F C1_{i-1}] \\ & -\theta(\mathbb{P}_{I,C1}^F + \mathbb{P}_{I,PI}^F + \mathbb{P}_{I,I}^F) I_i^F - \mu_i^F I_i^F, \quad i = 1,2,\ldots,14 \end{aligned} \tag{17}$$

$$\begin{aligned} d/dt(I_{15}^F) = {} & \theta[\mathbb{P}_{I,I}^F I_{14}^F + \mathbb{P}_{S,I}^F (S_{14}^F + S_{15}^F) + \mathbb{P}_{PI,I}^F (P_{14}^F + P_{15}^F) + \mathbb{P}_{C1,I}^F (C1_{14} + C1_{15}) \\ & -\theta([\mathbb{P}_{I,C1}^F + \mathbb{P}_{I,PI}^F) I_{15}^F - \mu_{15}^F I_{15}^F \end{aligned} \tag{18}$$

$$\begin{aligned} d/dt(C1_i) = {} & \theta(\mathbb{P}_{C1,C1}^F C1_{i-1} + \mathbb{P}_{I,C1}^F I_{i-1}^F + \mathbb{P}_{C23,C1}^F C23_{i-1}) \\ & -\theta(\mathbb{P}_{C1,PI}^F + \mathbb{P}_{C1,C23}^F + \mathbb{P}_{C1,C1}^F) C1_i - \mu_i^F C1_i, \quad i = 1,2,\ldots,14 \end{aligned} \tag{19}$$

$$\begin{aligned} d/dt(C1_{15}) = {} & \theta[\mathbb{P}_{C1,C1}^F C1_{14} + \mathbb{P}_{I,C1}^F (I_{14}^F + I_{15}^F) + \mathbb{P}_{C23,C1}^F (C23_{14} + C23_{15})] \\ & -\theta(\mathbb{P}_{C1,PI}^F + \mathbb{P}_{C1,C23}^F) C1_{15} - \mu_{15}^F CIN1_{15} \end{aligned} \tag{20}$$

$$\begin{aligned} d/dt(C23_i) = {} & \theta(\mathbb{P}_{C23,C23}^F C23_{i-1} + \mathbb{P}_{C1,C23}^F C1_{i-1} + \mathbb{P}_{C,C23}^F C_{i-1}) \\ & -\theta(\mathbb{P}_{C23,PI}^F + \mathbb{P}_{C23,C1}^F + \mathbb{P}_{C23,C}^F + \mathbb{P}_{C23,C23}^F) C23_i - \mu_i^F C23_i, \quad i = 1,2,\ldots,14 \end{aligned} \tag{21}$$

$$\begin{aligned} d/dt(C23_{15}) = {} & \theta[\mathbb{P}_{C23,C23}^F C23_{14} + \mathbb{P}_{C1,C23}^F (C1_{14} + C1_{15}) + \mathbb{P}_{C,C23}^F (C_{14} + C_{15}) \\ & -\theta(\mathbb{P}_{C23,PI}^F + \mathbb{P}_{C23,C1}^F + \mathbb{P}_{C23,C}^F) C23_{15} - \mu_{15}^F C23_{15} \end{aligned} \tag{22}$$

$$d/dt(C_i) = \theta[\mathbb{P}_{C,C}^F C_{i-1} + \mathbb{P}_{C23,C}^F C23_{i-1} - (\mathbb{P}_{C,C23}^F + \mathbb{P}_{C,C}^F) C_i - (\mu_i^F + \mu_{ca}) C_i, \quad i = 1,2,\ldots,14 \tag{23}$$

$$d/dt(C_{15}) = \theta[\mathbb{P}_{C,C}^F C_{14} + \mathbb{P}_{C23,C}^F (C23_{14} + C23_{15}) - \mathbb{P}_{C,C23}^F C_{15}] - (\mu_{15}^F + \mu_{ca}) C_{15} \tag{24}$$

$$\begin{aligned} d/dt(P_i^F) = {} & \theta(\mathbb{P}_{PI,PI}^F P_{i-1}^F + \mathbb{P}_{I,PI}^F I_{i-1}^F + \mathbb{P}_{C1,PI}^F C1_{i-1} + \mathbb{P}_{C23,PI}^F C23_{i-1}) \\ & -\theta(\mathbb{P}_{PI,PI}^F + \mathbb{P}_{PI,I}^F) P_i^F - \mu_i^F P_i^F, \quad i = 1,2,\ldots,14 \end{aligned} \tag{25}$$

$$\begin{aligned} d/dt(P_{15}^F) = {} & \theta[\mathbb{P}_{PI,PI}^F P_{14}^F + \mathbb{P}_{I,PI}^F (I_{14}^F + I_{15}^F) + \mathbb{P}_{C1,PI}^F C1_{14} + \mathbb{P}_{C23,PI}^F (C23_{14} + C23_{15})] \\ & -\theta \mathbb{P}_{PI,I}^F P_{15}^F - \mu_{15}^F P_{15}^F \end{aligned} \tag{26}$$

Equation (7) represents the change in the number of susceptible boys between 0 and 4 years, given by the total of births, less those who die or move to the next age compartment. The same logic applies to equation (14), for the number of susceptible girls in the first age range. The change in number of susceptible men and women in age range i, $2 \leq i \leq 14$ at time t, is provided by equations (8) and (15) respectively, and it is equal to the the number of individuals who come from an earlier age compartment, less those who die, get infected with the virus or move to the next age compartment. Equations (9) and (16) represent the change in the number of susceptible men and women, respectively, in the age range [70 or older] at time t, equal to the number of individuals that came from and earlier age compartment, less those who die or get infected. These two equations have a slightly different form because individuals in the last age range cannot move to other age ranges and can only change biological compartment.

The equations (10)-(11) and equations (17)-(18) provides the change in the number of infected men and women. These equations are analogous to the equations for susceptible individuals. They all have in their right-hand side positive terms which account for the individuals who move into the compartment from the immediately preceding age range, according to the possible biological transitions, and negative terms that account for the transition to the next age range and for deaths. Again, in the case of the last age range, transitions are only possible towards other biological compartments.

The equations for the remaining biological stages and age ranges are analogous.

Finally, we implement the prevention model, which takes into account vaccination as a prevention policy and hysterectomy, as a radical form of risk reduction strategy applied to women diagnosed with CC. We assume that only susceptible individuals can receive vaccination, and that a vaccinated individual will acquire a permanent immunity. The prevention mode considers one additional biological compartment, to account for HPV immune (i.e vaccinated) individuals. Table 3 describes the new variables and parameters used.

Table 3: Description of the parameters and variables used the prevention model.

Variable	Description
$R_t^F(i)$	Number of vaccinated women of age range i at time t
$R_t^M(i)$	Number of vaccinated men of age range i at time t
Parameter	
$\mathbb{P}^F$	Transition probability between biological stages for women
$\mathbb{P}^M$	Transition probability between biological stages for men

Since only susceptible individuals can be vaccinated, and only women diagnosed with CC undergo hysterectomy, the differential equations that represents the change in the number of infected and partially immune individuals, women in C1, C23 do not change with respect to the infection model. The modified differential equations of the prevention model for men are the following ones.

$$d/dt(S_1^M) = (1-p_f)\sum_j \pi_i S_i^F - \theta(\mathbb{P}_{S,R}^M + \mathbb{P}_{S,S}^M)S_1^M - \mu_1^M S_1^M \tag{27}$$

$$d/dt(S)_i^M = \theta[\mathbb{P}_{S,S}S_{i-1}^M - (\mathbb{P}_{S,R} + \mathbb{P}_{S,I} + \mathbb{P}_{S,S})S_i^M(t)] - \mu_i^M S_i^M, \quad i = 2,3,\ldots,14 \tag{28}$$

$$d/dt(S)_{15}^M = \theta[\mathbb{P}_{S,S}S_{14}^M - (\mathbb{P}_{S,R} + \mathbb{P}_{S,I})S_{15}^M(t)] - \mu_{15}^M S_{15}^M \tag{29}$$

$$d/dt(R_i^M) = \theta(R_{i-1}^M + \mathbb{P}_{S,R}S_{i-1}^M - R_i^M) - \mu_i^M R_i^M, \quad i = 1,2\ldots,14 \tag{30}$$

$$d/dt(R_{15}^M) = \theta(R_{14}^M + \mathbb{P}_{S,R}S_{14}^M) - \mu_{15}^M R_{15}^M \tag{31}$$

The modified differential equations for women are as follows.

$$d/dt(S_1^F) = p_f\sum_j \pi_i S_i^F - \theta(\mathbb{P}_{S,R}^F + \mathbb{P}_{S,S}^F)S_1^F - \mu_1^F S_1^F \tag{32}$$

$$d/dt(S_i^F) = \theta[\mathbb{P}_{S,S}^F S_{i-1}^F - (\mathbb{P}_{S,R}^F + \mathbb{P}_{S,I}^F + \mathbb{P}_{S,S}^F)S_i^F - \mu_i^F S_i^F, \quad i = 2,3,\ldots,14 \tag{33}$$

$$d/dt(S_{15}^F) = \theta[\mathbb{P}_{S,S}^F S_{14}^F - (\mathbb{P}_{S,R}^F + \mathbb{P}_{S,I}^F)S_{15}^F - \mu_{15}^F S_{15}^F \tag{34}$$

$$d/dt(C_i) = \theta(\mathbb{P}_{C,C}^F C_{i-1} + \mathbb{P}_{C23,C}^F C23_{i-1} - \mathbb{P}_{C,C23}^F C_{i-1} - \mathbb{P}_{C,C}^F C_i) - (\mu_i^F + \mu_{ca})C_i, \quad i = 1,2,\ldots,14 \tag{35}$$

$$d/dt(C_{15}) = \theta(\mathbb{P}_{C,C}^F C_{14} + \mathbb{P}_{C23,C}^F C23_{14} - \mathbb{P}_{C,23}^F C_{15}) - (\mu_{15}^F + \mu_{ca})C_{15} \tag{36}$$

$$d/dt(R_i^F) = \theta(R_{i-1}^F + \mathbb{P}_{S,R}^F S_{i-1}^F + \mathbb{P}_{C,R}^F C_{i-1}(t) - R_i^F) - \mu_i^F R_i^F, \quad i = 1,2,\ldots,14 \tag{37}$$

$$d/dt(R_{15}^F) = \theta(R_{14}^F + \mathbb{P}_{S,R}^F S_{14}^F + \mathbb{P}_{C,R}^F C_{14}) - \mu_{15}^F R_{15}^F \tag{38}$$

In the equations of the prevention model, the possibility that individuals get immunized is reflected into a transition to the immune biological stage. As we assume the immunity is perpetual, once an individual moves to the corresponding compartment, he/she will only change the age range (until the last age range is reached) but not the biological stage.

4 MODEL VALIDATION AND RESULTS

To fully instanciate the models for the Colombian population, we must estimate the corresponding parameters. We have mainly estimated these parameters through published literature and five-year medical records of a cohort of 3,428 Colombian women. Annually, the Colombian National Administrative Statistics Department

(DANE) publishes the number of births and deaths, discriminated by age range and gender, that occur in Colombia (de Estadística 2007). Based on these data, we estimated the birth and death rates, for every age range and gender, since 1995. We fitted to every combination of age range and gender, exponential, polynomial and potential functions, as these three functions provide a wide range of shapes. In most cases, potential functions fit the data best.

To determine the transition probabilities between biological states, we used the probability matrix estimated by Wilches (2015) from the medical records of a cohort of 3,428 Colombian women. This matrix provides, for each pair of biological state we considered, the probability that a woman experiences a transition between states in a year. As our age ranges span over 5 years, we estimated the 5-step transition probability matrix from these source. Finally, to find the existing population in each of the compartments.

To determine the initial state of the model, we use population registers published by DANE for 2016 and medical studies that could indicate the probability of being in each compartment. Camargo (2011) conducted a study with 2110 Colombian women, from different regions and with different socioeconomic levels. Each woman took the HPV-DNA test to determine whether at that time they had the virus, the results were taken as the initial frequencies in the biological compartments.

After instantiating the models, we ran deterministic continuous simulations. We chose this type of simulation for efficiency reasons: the number of entities in each compartment of the model is very large, and a stochastic simulation approach would require significantly higher computation time to simulate a single trajectory, whereas the numerical integration of the differential equations is very fast (simulation time is less than 1 sec on a standard Intel i5 machine). Moreover, as the model is deterministic, we only need to simulate a single trajectory. On the other hand, the large number of entities is ensuring that a continuous approximation is adequate for the system. Although we gain in computational efficiency, we acknowledge that this form of simulation does not allow incorporating variability. We ran the simulation models for 10 years, starting from 2016 until 2026.

The results of the population model for 2020 are shown in Figure 2, where we can see the predicted percentage of women and men discriminated by age range in Colombia. Although we ran the model for 10 years we can only compare the results obtained until 2020, as DANE population projections only reach that date. We can see that a large percentage of the population is concentrated between 0 and 4 years and 70 years or older. In addition, given that the size of the base is approximately equal to the size of the next step, it is possible to deduce that the death rate in children between 0 and 4 is small. Similarly, given that changes in size between adjacent steps are smaller for women than for men, it is valid to conclude that death rates are less variant in women than in men.

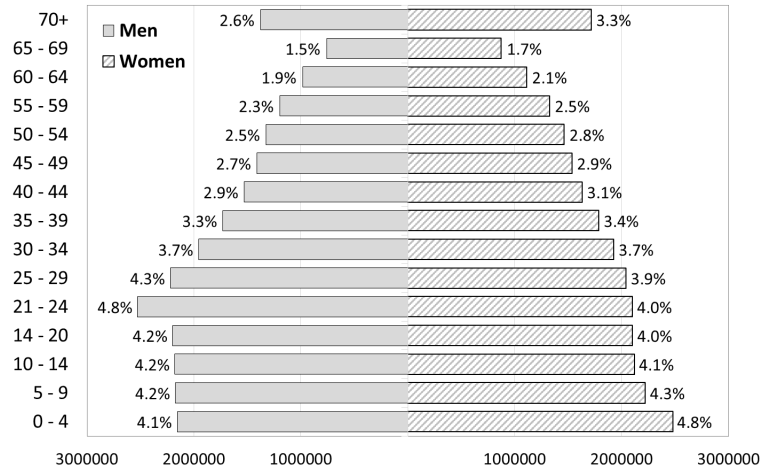

Figure 2: Population pyramid by 2020.

We validated our population model by estimating the parameters for the period 2010-2015 based on historical data from 2000 to 2009, and obtained an average error of 7% compared to actual data provided by DANE. Then, we compared the population predictions provided by our model with DANE population projections. Figure 3 (left panel) graphically illustrates the comparison of the two predictions, showing that the differences in each of the age compartments are really small, indicating that our model closely follows the same trends as DANE projections. Finally, as an accuracy measure we calculate the correlations between DANE and our model predictions. As it can be seen in the right panel in Figure 3, the correlations are very high, in all cases greater than 97%, giving an indication of successful parameter estimation and model implementation.

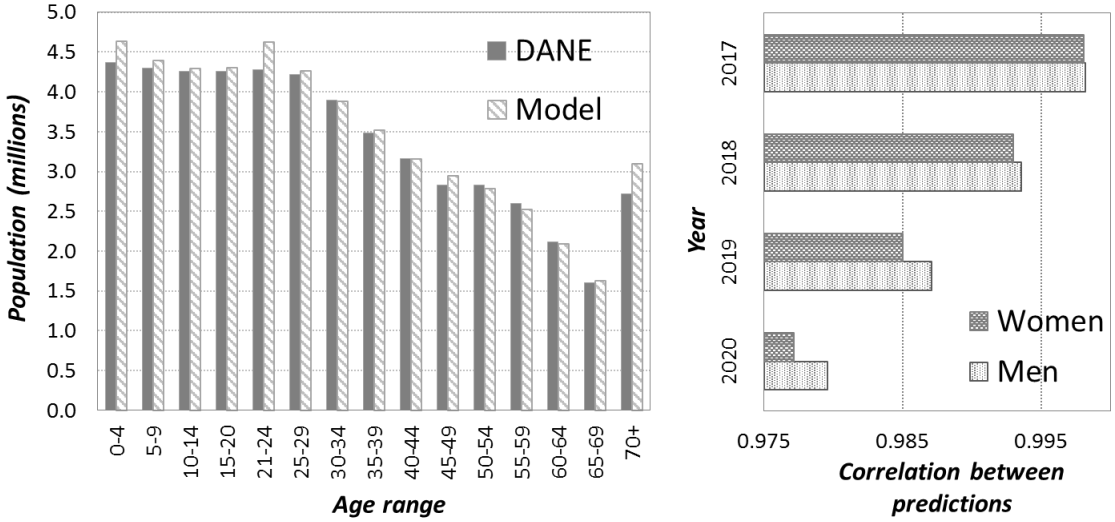

Figure 3: Comparison between our model predictions and DANE prediction (left panel) for year 2020, and correlations between DANE projections and our model (right panel).

We simulate the infection model to obtain predictions about the incidence and prevalence of cancer over the next 10 years. In Figure 4 we report the CC prevalence predictions, which are increasing over time, reflecting the expected growth of the Colombian population and the absence of prevention mechanisms included in the model.

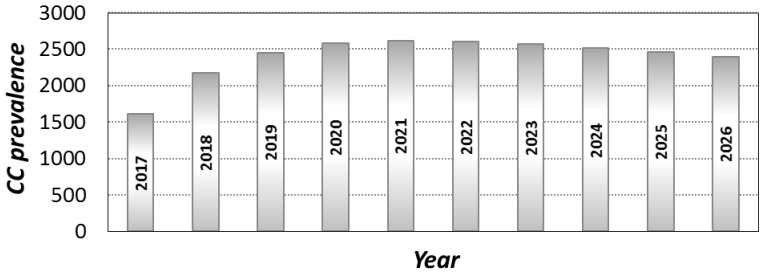

Figure 4: Predicted CC prevalence in Colombia in the period 2017-2026.

Last, we simulate the prevention model to estimate the number of CC deaths over a 30 year-horizon simulation. We evaluate different vaccination scenarios in order to evaluate their effectiveness in terms of deaths due to CC.

In Figure 5 we show the trends of predicted death rate per 100.000 women in a scenario were a partial vaccination strategy would be used. We compare a no vaccination policy (which we call *NoVaccin*), with

the prediction for a scenario in which 40% of girls in the age group 2 are immunized (*Vaccin1* policy), 40% of girls in age ranges 2 and 3 are vaccinated (*Vaccin2* policy) and 40% of girls in age ranges 2, 3 and 4 receive the vaccination (*Vaccin3* policy). As it is shown in the graph, when vaccination is applied, the death rate decreases over time, proportional to the number of vaccinated girls. It is important to note that the predicted death rate shows an increasing trend for all policies, because since the vaccine is being applied to young girls, a delay will be experienced before its effects on CC will be measurable.

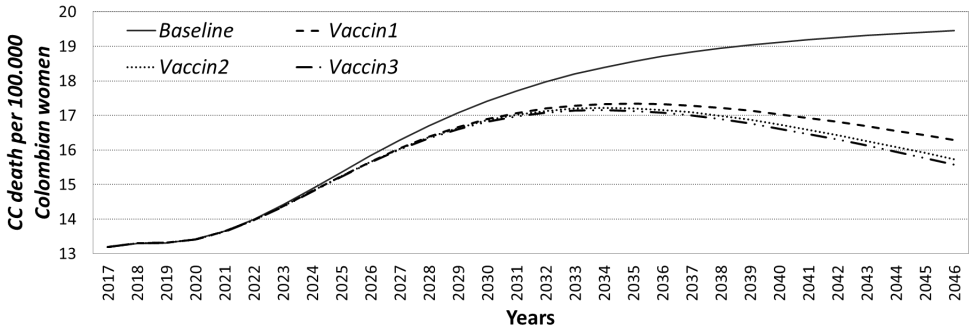

Figure 5: Number of CC deaths, per 100.000 Colombian women - Partial vaccination scenario.

Figure 6 compares the predicted number of deaths, per 100.000 women, in a scenario when 100% of girls in the age group 2 were vaccinated (*Vaccin4* policy), 100% of girls in age ranges 2 and 3 are vaccinated (*Vaccin5* policy), and 100% of girls in age ranges 2, 3 and 4 are vaccinated (*Vaccin6* policy). As it can be observed, the death rate decreases over time once the effect of vaccination kicks in. As in the first scenario, the higher the number of vaccinated girls, the faster decreases the death rate. In this second scenario has 100% coverage of vaccination, the decline in death rate is faster over time.

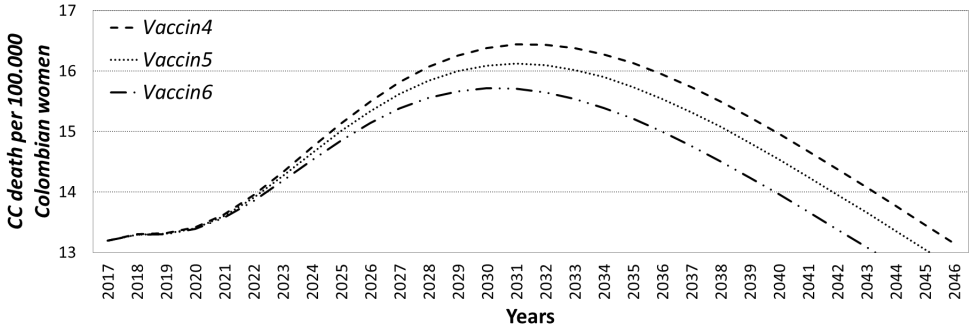

Figure 6: Number of CC deaths, per 100.000 Colombian women - Total vaccination scenario.

Finally, we performed a sensitivity analysis to evaluate the effect of uncertainties that affect the probabilities of transition between biological stages. To this end, we varied each probability by $+/-$ 5%. We took as the reference scenario one in which 20% of girls in the age range 2 receive the vaccine. The model does not show significant sensitivity to any of these parameters.

5 CONCLUSIONS AND FUTURE WORK

We proposed a compartmentalized epidemiological simulation model based on differential equations, that reproduces the Colombian population dynamics and the transmission of HPV infection within the population, with the possibility to evaluate different prevention policies, such as vaccination. In order to achieve this, we first created a model that reproduces the Colombian population dynamics. After validating this model,

we added the virus transmission dynamics, and finally the vaccination policy. The parameters used in our models were taken from Colombian literature (birth and death rates and initial frequencies in the biological compartments) or estimated from Colombian medical records (transition probability matrix). The model predicts that the prevalence of CC would increases over time if no vaccination policies are applied, while cancer incidence would decrease under several vaccination schemes. However, deaths due to CC will continue to increase reaching the incidence values, indicating that vaccination policies decrease the infection transmission rate, but will not prevent CC deaths. Future work will focus on two aspects: adding a new socio-economic compartment to the model to allow studying the implications of the infection transmission in different strata of the Colombian society, and evaluating the impact of additional preventive policies beyond vaccination, such as educational programs.

A APPENDIX: VALUES OF PARAMETERS USED IN THE MODELS

A.1 Population Model

Probability that a newborn is a woman:

$$p_f = 0.5$$

Birth rate per woman:

$$p_f = 0.0665t^{-0.16}/year$$

Death rates:

$$\begin{array}{ll}
\mu_1^F(t) = 0.0035e^{-0.091t}/year & \mu_1^M(t) = 0.0043e^{-0.092t}/year \\
\mu_2^F(t) = 0.0003e^{-0.077t}/year & \mu_2^M(t) = 0.0004e^{-0.086t}/year \\
\mu_3^F(t) = 0.0003e^{-0.054t}/year & \mu_3^M(t) = 0.0005e^{-0.087t}/year \\
\mu_4^F(t) = 0.0007e^{-0.086t}/year & \mu_4^M(t) = 0.0016e^{-0.016t}/year \\
\mu_5^F(t) = 0.0008e^{-0.081t}/year & \mu_5^M(t) = 0.0044e^{-0.128t}/year \\
\mu_6^F(t) = 0.0009e^{-0.056t}/year & \mu_6^M(t) = 0.0051e^{-0.114t}/year \\
\mu_7^F(t) = 0.001e^{-0.043t}/year & \mu_7^M(t) = 0.0045e^{-0.092t}/year \\
\mu_8^F(t) = 0.0014e^{-0.059t}/year & \mu_8^M(t) = 0.0047e^{-0.108t}/year \\
\mu_9^F(t) = 0.0015e^{-0.035t}/year & \mu_9^M(t) = 0.0055e^{-0.126t}/year \\
\mu_{10}^F(t) = 0.0028e^{-0.078t}/year & \mu_{10}^M(t) = 0.0055e^{-0.094t}/year \\
\mu_{11}^F(t) = 0.0044e^{-0.079t}/year & \mu_{11}^M(t) = 0.0071e^{-0.08t}/year \\
\mu_{12}^F(t) = 0.0063e^{-0.066t}/year & \mu_{12}^M(t) = 0.0092e^{-0.052t}/year \\
\mu_{13}^F(t) = 0.0105e^{-0.081t}/year & \mu_{13}^M(t) = 0.0146e^{-0.061t}/year \\
\mu_{14}^F(t) = 0.0189e^{-0.143t}/year & \mu_{14}^M(t) = 0.0248e^{-0.121t}/year \\
\mu_{15}^F(t) = 0.07e^{-0.002t}/year & \mu_{15}^M(t) = 0.0533e^{-0.001t}/year
\end{array}$$

A.2 Infection Model

Probabilities of transition between biological stages:

$$\mathrm{P}^M = \begin{array}{c} \\ S \\ I \\ PI \end{array} \begin{array}{c} \begin{array}{ccc} S & I & PI \end{array} \\ \begin{bmatrix} 0.45 & 0.55 & 0.0 \\ 0 & 0.54 & 0.46 \\ 0 & 0.19 & 0.81 \end{bmatrix} \end{array}$$

$$\mathrm{P}^F = \begin{array}{c} \\ S \\ I \\ CIN1 \\ CIN23 \\ C \\ PI \end{array} \begin{array}{c} \begin{array}{cccccc} S & I & CIN1 & CIN23 & C & PI \end{array} \\ \begin{bmatrix} 0.45 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0.54 & 0.07 & 0 & 0 & 0.39 \\ 0 & 0.76 & 0.03 & 0.15 & 0 & 0.06 \\ 0 & 0 & 0.59 & 0.07 & 0.28 & 0.05 \\ 0 & 0 & 0 & 0.03 & 0.97 & 0 \\ 0 & 0.19 & 0 & 0 & 0 & 0.81 \end{bmatrix} \end{array}$$

Death rate for CC:

$$\mu_{ca} = 0.5/year$$

A.3 Prevention model

Probability that a woman having CC would undergo an hysterectomy:

$$Hys = 0.0056$$

REFERENCES

American Cancer Society 2014a. " What is Cervical Cancer?". http://www.cancer.org/cancer/cervicalcancer/detailedguide/cervical-cancer-what-is-cervical-cancer. Online; accessed 12 April 2016.

American Cancer Society 2014b. "How is Cervical Cancer Diagnosed?". http://www.cancer.org/cancer/cervicalcancer/detailedguide/cervical-cancer-diagnosis. Online; accessed 12 April 2016.

Camargo, S. 2011. "Determinación de la Prevalencia de Infección y Coinfección por Virus del Papiloma Humano (VPH) y Asociación con Diferentes Factores de Riesgo". Master's Thesis, Universidad del Rosario, Bogotá, Colombia.

Departamento Administrativo Nacional de Estadística 2007. "Boletín Censo General 2005: Datos Desagregados por Sexo". https://www.dane.gov.co/files/censos/libroCenso2005nacional.pdf. Online; accessed 12 April 2016.

Dimitrov, N. B., and L. A. Meyers. 2010. "Mathematical Approaches to Infectious Disease Prediction and Control". *JJ Hasenbein, ed. INFORMS Tutorials in Operations Research* 7:1–25.

Hernando Navarro-Newball, M. 2005. "Histerectomía Laparoscópica: Sistematización del Procedimiento y Propuesta para Programas de Capacitación". *Colombia Médica* 36 (2): 115–119.

Kim, J., and S. Goldie. 2008. "Health and Economic Implications of HPV Vaccination in the United States". *The New England Journal of Medicine* 8 (359): 821–832.

Ministerio de Salud 2013. " Bajan Tasas de Incidencia y Mortalidad de Cáncer de Cuello Uterino en Colombia". https://www.minsalud.gov.co/Paginas/Bajan-tasas-incidencia-mortalidad-cancer-cuello.aspx. Online; accessed 12 April 2016.

Namen, M. I. 2014. "A POMDP Approach to Age-dependent Primary Screening Policies for Cervical Cancer in Colombia". Master's Thesis, Universidad de los Andes, Bogotá, Colombia.

Orrego, J. 2013. "Evaluating HPV Vaccination Policies in Colombia". Undergraduate Thesis, Universidad de los Andes, Bogotá, Colombia.

Pardo, C., and R. Cendales. 2015. *Incidencia, Mortalidad y Prevalencia de Cáncer en Colombia 2007-2011*. Instituto Nacional de Cancerología, Bogotá, Colombia.

Profamilia 2015. "¿Qué Hace la Vacuna para Prevenir el Cáncer de Cuello Uterino?". http://profamilia.org.co/inicio/mujer-2/preguntas-y-respuestas-mujer/cancer-de-cuello-uterino/?id=29. Online; accessed 12 April 2016.

Ruiz, E. G. V., R. Sánchez, and J. M. Lozano. 2013. "Análisis Epidemiológico sobre la Incidencia y Resolución de las Infecciones por el Virus del Papiloma Humano en Mujeres Sometidas a un Estudio de Seguimiento Retrospectivo". *Revista Colombiana de Ciencias Químico Farmacéuticas* 42 (2): 284–297.

Sanitas Internacional Organization 2014. " Virus de Papiloma Humano, Primera Causa de Cáncer de Cuello Uterino". http://prestadores.colsanitas.com/portal/documents/233493/2438755/Boletin_Epidemiologico_Semanal_No_7.pdf. Online; accessed 12 April 2016.

Wilches, M. 2015. "Estimating the Parameters of a POMDP Model for Cervical Cancer Primary Screening with a Colombian Data Base". Undergraduate Thesis, Universidad de los Andes, Bogotá, Colombia.

World Health Organization 2015. "Human Papillomavirus (HPV) and Cervical Cancer". http://www.who.int/mediacentre/factsheets/fs380/en/. Online; accessed 12 April 2016.

AUTHOR BIOGRAPHIES

DANIELA ANGULO DÍAZ is an undergraduate student of industrial engineering at Universidad de los Andes, Bogotá, Colombia with emphasis in Operation Research. E-mail: kd.angulo2295@uniandes.edu.co

RAHA AKHAVAN-TABATABAEI is an associate professor at the Department of Industrial Engineering, Universidad de los Andes in Bogotá, Colombia. Prior to this position she was a senior industrial engineer at Intel Corporation, Chandler, AZ, USA. She received her Ph.D. in Industrial Engineering with minor in Operations Research from the Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, USA. Her research is focused on stochastic modeling of complex systems applied to healthcare, transportation and logistics, revenue management and production and manufacturing. E-mail: r.akhavan@uniandes.edu.co

IVAN MURA received his first degree in Computer Science and a PhD in Computer Science Engineering from the University of Pisa, Italy, and a Master of Science in Information Technology Project Management from the George Washington University School of Business. He is currently a visiting professor at the Department of Industrial Engineering, Universidad de los Andes, Bogotá, Colombia. His research interests include the mathematical modeling of artificial and living systems, with continuous-deterministic and discrete-stochastic state-based techniques. E-mail: i.mura@uniandes.edu.co