Original Research Article

MDM Medical Decision Making

Medical Decision Making
1–14

DOI: 10.1177/0272989X211042894
journals.sagepub.com/home/mdm
SAGE

Evaluating Risk-Stratified HPV Catch-up Vaccination Strategies: Should We Go beyond Age 26?

Fan Wang, Kristen N. Jozkowski, and Shengfan Zhang

Background. Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States. HPV can cause genital warts and multiple types of cancers in females. HPV vaccination is recommended to youth age 11 or 12 years before sexual initiation to prevent onset of HPV-related diseases. For females who have not been vaccinated previously, catch-up vaccines are recommended through age 26. The extent to which catch-up vaccines are beneficial in terms of disease prevention and cost-effectiveness is questionable given that some women may have been exposed to HPV before receiving the catch-up vaccination. This study aims to examine whether the cutoff age of catch-up vaccination should be determined based on an individual woman's risk characteristic instead of a one-size-fits-all age 26. **Methods**. We developed a microsimulation model to evaluate multiple clinical outcomes of HPV vaccination for different women based on a number of personal attributes. We modeled the impact of HPV vaccination at different ages on every woman and tracked her course of life to estimate the clinical outcomes that resulted from receiving vaccines. As the simulation model is risk stratified, we used extreme gradient boosting to build an HPV risk model estimating every woman's dynamic HPV risk over time for the lifetime simulation model. **Results**. Our study shows that catch-up vaccines still benefit all women after age 26 from the perspective of clinical outcomes. Women facing high risk of HPV infection are expected to gain more health benefits compared with women with low HPV risk. **Conclusions**. From a cancer prevention perspective, this study suggests that the catch-up vaccine after age 26 should be deliberately considered.

Keywords
human papillomavirus, vaccination, medical decision making, microsimulation, machine learning in health care

Date received: June 2, 2018; accepted: August 6, 2021

Human papillomavirus (HPV) is the most common sexually transmitted disease in the United States. Over 14 million people are newly infected with HPV each year.[1] Although most HPV strains are asymptomatic and often regress spontaneously, some HPV subtypes can cause genital warts and multiple types of cancers for females, including cervical, anal, vaginal, vulvar, and oropharyngeal cancers. In particular, HPV types 16 and 18 result in 70% of cervical cancers, which account for the second largest cause of cancer deaths among women worldwide.[2,3]

There are 3 major types of HPV vaccinations: bivalent, quadrivalent, and 9-valent vaccines. The 9-valent vaccine, which was recently approved by the US Food

Walmart, Inc, Bentonville, AR, USA (FW); Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, USA (KNJ); Department of Industrial Engineering, University of Arkansas, Fayetteville, AR, USA (SZ). The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported in part by funds from the National Science Foundation (NSF: #1920920) and the Women's Giving Circle at the University of Arkansas.

Corresponding Author
Shengfan Zhang, Department of Industrial Engineering, University of Arkansas, 4207 Bell Engineering Center, Fayetteville, AR 72701, USA. (shengfan@uark.edu)

and Drug Administration (FDA), provides protection against HPV types of 6, 11, 16, 18, 31, 33, 45, 52, and 58. Since the end of 2016, only 9-valent vaccine has been available in the United States, while bivalent and quadrivalent vaccines continue to be available in other countries.[4] Although the HPV subtypes covered by the 3 vaccines vary, all of them provide strong protection against HPV 16 and 18, which account for most HPV-related cancers. The Advisory Committee on Immunization Practices (ACIP) recommends that children should receive HPV vaccination at ages 11 to 12.[4] For those who have not been vaccinated, catch-up HPV vaccinations are also recommended for both females and males through age 26. According to the latest estimates, 60% of US teens aged 13 to 17 years received at least 1 dose of the HPV vaccine.[5] Although the guidelines proposed do not recommend vaccination for women over age 26, recent studies indicated that these women still bear significant HPV risk.[6] Previous clinical trials also found that the peak antibody titers and 4-month follow-up plateaus of HPV vaccines for women over 26 were noninferior to those induced at ages 16 to 26.[7] In Australia, HPV vaccines are approved for use by females up to 45 years old.[8] Thus, being vaccinated after age 26 may be beneficial.[9–11] In 2018, the FDA approved a supplemental application for Gardasil 9 (9-valent vaccine Merck & Co., NJ, USA), expanding the use of the vaccine to include ages 27 through 45 in the United States.[12] The FDA's approval of the new age range suggests the potential necessity and possibility of extending HPV vaccination above age 26 years in the future national vaccination guidelines.

Previous studies evaluated HPV vaccination programs at the population level. Most of these studies assessed the cost-effectiveness and clinical outcomes of HPV vaccination programs based on simulation models.[13–17] In particular, a recent study by Brisson and Laprise[18] discussed the cost-effectiveness of extending the established HPV vaccine program in the United States to both women and men aged 27 to 45 years. The study found extending vaccination to middle-aged adults only produced small additional reductions in predicted cases of genital warts, cervical intraepithelial neoplasia, and cervical cancers compared with the current vaccination program. Van de Velde et al.[19] concluded that differences in the elements of model design, such as natural immunity, partnership duration, HPV types, and waning of vaccine protection, resulted in significant differences in the estimated effectiveness of the vaccine. It is reasonable to expect that people with different levels of HPV risk, based on individual characteristics, have different clinical postvaccination outcomes. To our knowledge, there are relatively few studies quantifying the effect of HPV vaccination at the individual level, especially from the cancer prevention perspective. However, a number of individual-level risk factors that determine the level of HPV risk on different women, including demographic attributes such as age, personal lifestyle, and sexual behavior, have been identified.[20–22] Thus, a risk-stratified evaluation model incorporating individual HPV risk could more precisely reveal the different impacts of HPV vaccination on different people.

This study aims to provide a risk-stratified evaluation model of HPV vaccination strategy and examine the clinical outcomes of HPV vaccination for women older than 26. We seek to investigate the cutoff age of an HPV catch-up vaccine based on every single woman's risk characteristics. We developed a microsimulation model to evaluate multiple clinical consequences after a woman receives the vaccines based on a number of personal attributes. We estimated the following patient-specific clinical consequences: 1) lifetime risk of developing HPV-related cancers, 2) life expectancy, and 3) life years saved by vaccines. While the prior cost-effectiveness study on HPV vaccination showed a relatively small population-level health benefit of extending the HPV vaccination up to age 45,[18] we expect those clinical outcomes to differentiate at the individual level obviously. As previous population-level studies of male HPV vaccination suggested that male vaccination was likely not cost-effective if the vaccine coverage of females was high ($>80\%$),[16] we focus this study on females' HPV vaccination. Specifically, we choose the clinical outcomes related to female cancers as the metrics evaluated in the simulation. The contribution of our study is 2-fold. First, this is the only patient-specific simulation model for the evaluation of HPV vaccination that provides more practical and accurate decision support for both individual women and care providers. Second, our study considers vaccination after age 26, which exceeds the recommended age of catch-up vaccines in the United States. The results of the study would be instrumental for medical decision makers to rationally determine the catch-up vaccination and potentially amplify the public health benefits of HPV vaccines.

Methods

The proposed HPV vaccination evaluation model consists of 2 submodels. The main body of the model is a microsimulation model that keeps track of every individual woman's course of life, which involves multiple deterministic events and probabilistic events. An HPV risk model dynamically estimates the patient-specific

Table 1 Main Events in the Simulation Model

Deterministic Events	Probabilistic Events
■ The woman will get vaccinated at a specific age. ■ She will die at age 100 if not dying at an earlier age.	■ The woman may be infected with the high-risk human papillomavirus (HPV). ■ After the woman is infected with the high-risk HPV, HPV may clear spontaneously with or without naturally acquired immunity. ■ After the woman is infected with the high-risk HPV, HPV may progress to cancer. ■ If the woman develops an HPV-related cancer, the cancer may result in death or be successfully cured. ■ The woman may die at any age.

high-risk HPV (i.e., HPV 16/18) risk of every simulated woman's life course so as to update the likelihood of probabilistic events in the simulation.

Estimation of Clinical Consequences Using Microsimulation

We model every woman's life course in different vaccination scenarios (i.e., varying age at vaccination). The model simulates every woman's life course with the given risk characteristic repeatedly over a planning horizon from a certain starting age to her death, which does not exceed age 100. The system clock is incremented by a fixed amount of time (1 year) at each step of the simulation. We use t to represent the current time epoch (i.e., a woman's age). Five main states are used to represent a woman's health status (Table 1):

- **Susceptible (S):** the woman has no immunity against HPV and therefore is susceptible.
- **Immune (I):** the woman has been vaccinated or naturally acquired immunity and therefore is immune to the specific high-risk HPV.
- **HPV infection (H):** the woman without immunity is infected with high-risk HPV.
- **Cancer (C):** the woman has developed an HPV-related cancer.
- **Death (D):** the woman dies from HPV-related cancers or other causes.

A woman's state at time t is denoted by s_t, where $s_t \in \{\mathrm{S}, \mathrm{I}, \mathrm{H}, \mathrm{C}, \mathrm{D}\}$. During the course of a woman's simulated life, her health status will switch between the 5 states until she enters death or reaches age 100. Table 1 summarizes the main deterministic events and probabilistic events in a woman's simulated life course. Figure 1 shows the 5 states and the specific events resulting in the transitions between these states. In particular, we assume all women complete 3-dose HPV vaccines and acquire

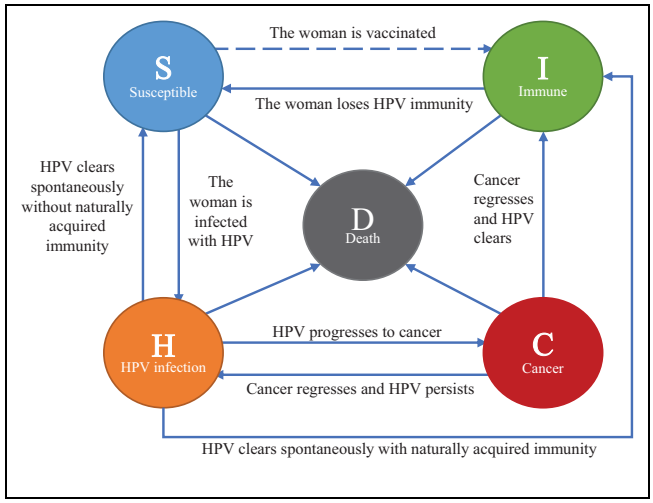

Figure 1 The main structure of the simulation model.

full immunity against HPV. As long-term clinical trials examining protection duration for HPV vaccines are still ongoing and have reported almost persistent efficacy during the whole follow-up period,[23] we also assume that the vaccines provide women with full lifetime immunity against HPV infection in the base case analysis. However, as the HPV attack rate and associated high-grade precancerous outcomes are relatively low after age of 25 years,[24] the long-term effectiveness of HPV vaccination in mid-adult women remains unclear. Previous trials showed that 96% of women had anti-HPV 16 seropositivity, while 84% of women expressed anti-HPV 18 seropositivity at 10 years after receiving bivalent vaccines.[7] In contrast, loss of seropositivity over years for anti-HPV 18 remains problematic regardless of age of receiving quadrivalent vaccines. There are no trials of 9-valent vaccines in mid-adult-aged women. Thus, we set multiple sensitivity analysis scenarios by considering different levels of vaccine efficacy against persistent HPV infections

Table 2 States Transitions in the Simulation Model

Probability	Transition	Value
$P_t(H\|S)$	A woman is infected with human papillomavirus (HPV).	Individual annual HPV incidence rate
$P_t(S\|H)$	HPV regresses spontaneously without naturally acquired immunity.	One-year regression rate of HPV adjusted for the possibility of no HPV immunity
$P_t(I\|H)$	HPV regresses spontaneously with naturally acquired immunity.	One-year regression rate of HPV adjusted for the possibility of acquiring HPV immunity
$P_t(D\|H)$	The infected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(C\|H)$	HPV progresses to an HPV-related cancer.	Cancer-specific 1-year progression rate of HPV
$P_t(D\|C)$	The woman dies from any reasons, including HPV-related cancer.	Cancer-specific 1-year death rate (including other reasons)
$P_t(D\|I)$	The uninfected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(S\|I)$	The woman loses the HPV immunity.	One-year waning rate of HPV immunity
$P_t(D\|S)$	The uninfected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(I\|S)$	The woman gets vaccinated.	100% at a specified age 0% at other ages
$P_t(H\|C)$	The HPV-related cancer regresses. HPV persists.	HPV persistent rate after cancer treatment or cancer regression
$P_t(I\|C)$	The HPV-related cancer regresses. HPV clears.	1−HPV persistent rate (excluding death in 1 year)

after age 26. In addition, when a woman's HPV clears spontaneously, she may obtain naturally acquired immunity, which is possibly waning over time.[25] Thus, a woman with protection against HPV may still enter "HPV infection" or "susceptible" states under the scenarios of nonlifetime vaccine protection or naturally acquired immunity. At each step of the simulation, the system determines the occurrences of the events based on the woman's current state and the corresponding likelihoods of probabilistic events. The woman stays in one of the states until an event occurs and changes her state.

Let $P_t(s_{t+1}|s_t)$ denote the transition probability from state s_t to state s_{t+1} for a woman at age t. Table 2 summarizes the major transition probabilities used in the simulation. The HPVs involved in the states all refer to high-risk HPV 16/18. We separate HPV 16 and 18 and simulate the 2 HPV subtypes' trajectories independently. The states denoted by H include HPV 16 infection, HPV 18 infection, and HPV 16/18 coinfection. Thus, some transitions in Table 2 correspond to multiple actual transitions in the simulation (e.g., $P_t(H|S)$).

Different from the other 4 states, "cancer" is a complex state that consists of several subsequent processes, which represent the development, diagnosis, and treatment of an HPV-related cancer. Figure 2 illustrates subsequent processes that may take place after HPV develops into cancer. We use multiple substates to differentiate detected and undetected HPV-related cancers as well as different stages of the cancers, as they have different death rates. The simulation clock is still incremented by 1 year at each step after a woman enters the substates of the "cancer" state. When a woman completes the transitions from "HPV infection" state to "cancer" state, she will be immediately assigned to the "precancer/in situ cancer" state. Then the woman's precancer or in situ cancer may progress to an invasive cancer or be diagnosed. The state transitions in the substructure are similar to the transitions in the main structure of the simulation model.

HPV has been identified as an important cause of at least 5 cancers experienced by women, including cervical cancer, anal cancer, vaginal cancer, vulvar cancer, and oropharyngeal cancer.[26] The simulation model takes these 5 HPV-related cancers into account. As an example, cervical cancer development and prognosis are explained in detail to demonstrate the simulation processes after a woman's HPV progresses to cancer. Every year, an HPV 16/18 infection may naturally progress to an HPV-related cancer with a certain probability. The first status after the progression is the precancer or in situ cancer stage. For cervical cancer, this status includes 3 states: cervical intraepithelial neoplasia (CIN) 1, CIN 2, and CIN 3, which are nonmalignant precancer stages of cervical cancer with the propensities of regression and progression.[27] These precancer stages may be detected by routine cervical cancer screening (i.e., Pap test) with a certain probability every year. Once detected, the CINs

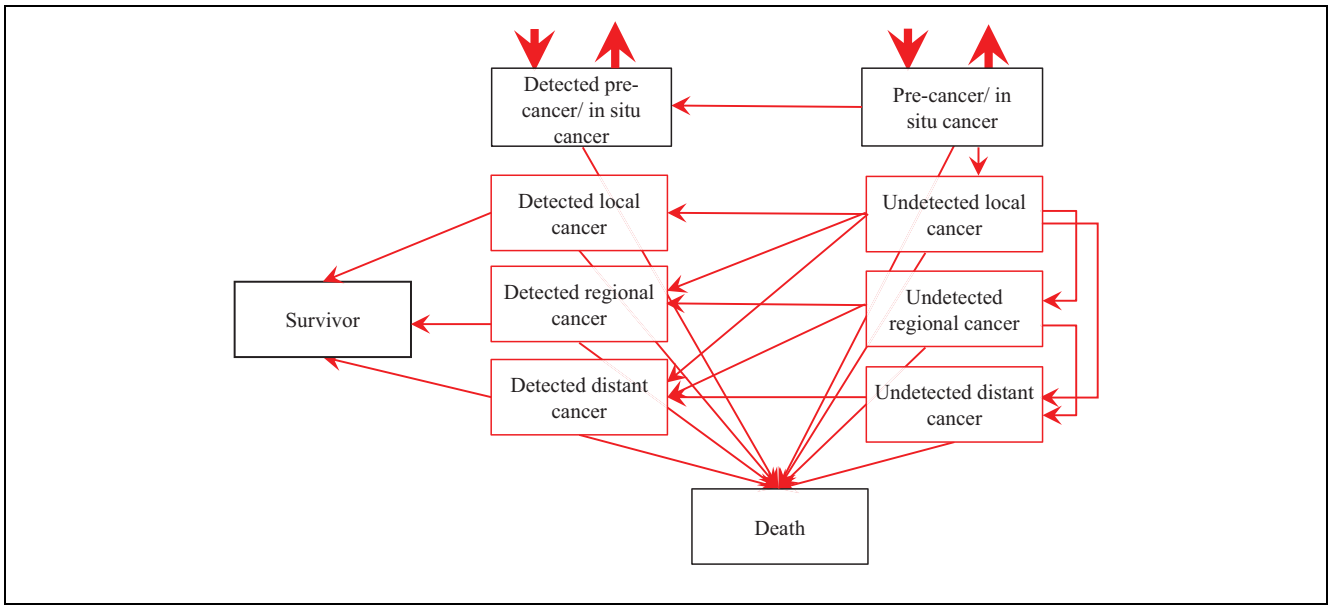

Figure 2 The sub-structure of the "cancer" state.

will be treated. Then the simulated woman will reenter the "HPV infection" state or the "immune" state in Figure 1, depending upon whether the HPV infection persists or not. During the year, the woman may also die from a cause other than cervical cancer. If the CINs are not detected, the woman's status may stay the same, regress to "HPV infection," regress to "immune," or progress to a local cervical cancer. When the woman enters the "undetected local cancer" state, similarly, the local cancer has the potential of evolving to a regional cervical cancer and being detected. Since the woman has cervical cancer, there is a certain possibility for her to develop cancer-related symptoms, which finally result in diagnosis and treatment of the cancer. The combination of the symptom development rate and cervical cancer screening rate is the transition probability from undetected local cancer to detected local or regional cancer. If the cancer is detected, the woman will enter the "detected local cancer" state and end up with "survivor." The "survivor" state makes the simulated woman quit the simulation and be assigned with a lump-sum life expectancy based on her specific cancer stage.

We assume that a woman will never develop multiple HPV-related cancers simultaneously, as the incidences of synchronous primary cancers of the female genital tract are very rare.[28] Once a woman develops an HPV-related cancer in the simulation, she is temporarily free from the other cancers.

The goal of the simulation is to estimate patient-specific clinical consequences in different vaccination scenarios (i.e., different ages at vaccination) with given individual HPV risk characteristics. A woman's life course is simulated multiple times to derive the average values of lifetime risk of developing HPV-related cancers and life expectancy. By changing the age at vaccination and HPV risk, we expect to observe dramatically different outcomes of these metrics.

HPV Risk Model

In the simulation model, individual HPV risk is embodied in the transition possibility $P_t(\mathrm{H}|\mathrm{S})$, which is the transition from "susceptible" to "HPV infection." Different women's possibilities of being infected with HPV at every age are estimated based on their personal risk characteristics. Previous studies have identified several behavioral and demographic risk factors of HPV.[20–22,29,30] We use a number of identified risk factors associated with HPV or HPV-related cancers to build a penalized regression model to estimate the personal HPV risk for every individual, including demographic attributes, personal lifestyle, and sexual behaviors.

Table 3 summarizes the 14 candidate risk factors considered in the HPV risk model. The risk factors are employed as predictors to estimate the HPV risk. Since HPV 16 and 18 are responsible for most HPV-related

Table 3 Candidate Human Papillomavirus Risk Factors

Demographic Attributes	Personal Lifestyle	Sexual Behaviors
Age	Alcohol use	Age at first sex
Marital status	Smoking	Lifetime number of sex partners
Education level		Number of recent sex partners
Ratio of family income to poverty		Sexual orientation
Race		Ever had sex
Age at first menarche		
Parity history		

cancers and preventable by the 3 approved vaccines, we only take these 2 HPV subtypes into account and treat the incidences of HPV 16 and 18 as binary variables. Two separate risk models are built for HPV 16 and 18, respectively. The response variable is whether a woman has HPV (16 or 18) at her current age.

There are numerous regression and classification approaches to build risk models for a problem with a binary response variable, such as logistic regression, support vector machine, and various decision tree–based models. Among all the decision tree–based models, the extreme gradient boosting (XGBoost) has received extensive attention in the machine learning community since it was proposed in 2014.[31] XGBoost is an ensemble algorithm for decision trees. The algorithm trains new decision trees, adding to the pool in a sequential way, which uses the optimization steps to improve the classification or the loss functions of regression models as every new tree is generated. XGBoost is well known for its superior prediction performance and has been a leading algorithm in machine learning competitions such as Kaggle and KDD Cup.[31,32] We choose this particular algorithm to build the HPV risk model, not only due to its generally excellent prediction performance for binary classification problems but also because of its significant advantages in some special problem settings. First, XGBoost is able to optimally handle missing values by finding the best direction to split, which is exactly the case we encounter in our HPV risk data. Second, XGBoost is fully compatible with L1 and L2 regularizations, making it efficient to perform feature selection based on a number of candidate predictors. Primarily, we want to build a parsimonious model—some risk factors are dynamic and therefore difficult to be tracked over time, such as number of sex partners. A simple model with as few predictors as possible would greatly reduce the effort put into data preparation. In addition, multicollinearity may exist in the model, as many variables are inherently correlated, such as marital status and number of sex partners. Additionally, incorporating too many variables may result in overfitting, which leads to the model's poor out-of-sample performance when being applied to the new data. As such, we use XGBoost with optimally tuned L1 regularization to build the risk model. The optimal tuning penalty parameter (lambda) is determined by a 5-fold cross-validation with the area under curve (AUC) as the performance metric, since the HPV risk data are highly imbalanced.

Once the HPV risk models are built, we use the estimated HPV risk as the input for the transition probability $P_t(\mathrm{H}|\mathrm{S})$, which represents the change from the "susceptible" state to the "HPV infection" state. Then the simulation is individualized for different women and generates patient-specific clinical outcomes.

Numerical Experiments

Since the incidences of the HPV-related cancers among teenagers are very low,[33] we assume women do not have HPV-related cancer before age 20. With the proposed simulation framework, we simulate the life courses of women with different HPV risk characteristics from age 20 to their deaths. The simulation estimates the average values of the gain in life expectancy, lifetime risk of developing cervical cancer, and lifetime risk of HPV-related cancers after 100,000 replications. The gain in life expectancy is defined as the difference of life expectancy between a woman receiving vaccines at a specific age and the same woman who never receives HPV vaccines.

Parameter Estimation for the Simulation Model

Table 4 summarizes the values or major data sources used to estimate the transition probabilities between the main states in the simulation model. It is worth mentioning that our HPV risk model actually estimates the prevalence of HPV associated with people having specific risk characteristics, which is equivalent to the initial state distribution at the beginning of the simulation. Each simulated woman is randomly assigned to either "susceptible" or "HPV infection" states according to the

Table 4 Sources of Input Data for Transition Probabilities

Transition Probability	Value or Source for Parameter Estimation
P_t(H|S)	Refer to the HPV Risk Model section
P_t(S|H)	Sanders and Taira,[13] Matthijsse et al.[25]
P_t(I|H)	Sanders and Taira,[13] Matthijsse et al.[25]
P_t(D|H)	Arias et al.[34]
P_t(C|H)	Elbasha et al.[14]
P_t(D|C)	Elbasha et al.[14]
P_t(D|I)	Arias et al.[34]
P_t(D|S)	Arias et al.[34]
P_t(I|S)	100% at specified vaccination age and 0% at other ages
P_t(S|I)	100% (80% or 50% in the scenarios of sensitivity analysis)
P_t(H|C)	Sanders and Taira,[13] Elbasha et al.,[14] Matthijsse et al.[25]
P_t(I|C)	Sanders and Taira,[13] Elbasha et al.,[14] Matthijsse et al.[25]

estimated prevalence. The epidemiologic equation Prevalence = Incidence × Duration[35] is used to calculate P_t(H|S), which is the incidence associated with a specific risk characteristic. The 1.2-year duration of oncogenic HPV 16/18 infection is adopted.[14]

As mentioned above, we assume that women will not develop multiple HPV-related cancers at the same time. An HPV carrier may develop 1 of the 5 HPV-related cancers with a certain incidence rate every year. The incidence rates of the 5 cancers on HPV-infected women are estimated through multiplying the percentage of a cancer attributable to HPV 16/18 by the corresponding cancer's population-level incidence rate.[15] For cervical cancer, the annual progression rates, annual regression rates, annual symptom development probabilities, annual screening rate, and stage-specific annual death rates from cervical cancer, we use the data reported by Sanders and Taira[13] as well as Elbasha et al.[14] The stage-specific lump-sum life expectancy of the absorbing state "survivor" is estimated by the DEALE method with the 5-year survival of invasive cervical cancer.[33,36] However, unlike cervical cancer, there are relatively limited epidemiological data on the natural history, diagnosis, and prognosis of anal cancer, vaginal cancer, vulvar cancer, and oropharyngeal cancer, which do not sufficiently support the complex substructures of "cancer" state. In addition, many prior studies on HPV vaccination evaluation only took cervical cancer into consideration due to the low incidences rates of these 4 cancers.[13,14] Hence, our study simplifies the impacts of the 4 HPV-related cancers on women's life expectancy by deducting the loss of life years associated with different cancers reported by Chesson et al.[15] rather than modeling their full substructures of "cancer" state to simulate these cancers.

Parameter Estimation for the HPV Risk Model

We use the National Health and Nutrition Examination Survey (NHANES) 2015–2016 data[37] to build the individual HPV risk model. These publicly available data sets include data for people (n = 15,327 from 30 different study locations across the United States) aged 18 to 59 years. The NHANES data report a large number of personal information of participants, including demographic, socioeconomic, dietary, and health-related information as well as medical laboratory exam results, which cover all the variables summarized in Table 3. Particularly, the NHANES 2015–2016 data include exam results of multiple HPV subtypes.

Although 15,327 people were included in the survey, only 1,872 observations were from females and reported HPV 16/18 data. Among the 1,872 observations, only 1,516 subjects were age 20 or older. As we set age 20 as the starting point of the simulation, we use the 1,516 observations to build the risk model. However, a critical issue of the NHANES data is its large number of missing data. Only 373 subjects have complete data in all the 15 candidate variables. As mentioned above, thanks to XGBoost's exceptional feature of finding the best direction to split for missing or unknown values, we are able to leverage both complete and incomplete data to train the risk models.

Ten of the 14 candidate risk factors are dynamic and change over time, including age, marital status, education level, ratio of family income to poverty, birth history, pregnancy history, alcohol use, smoking, lifetime number of sex partners, and number of recent sex partners. Age is simply increased by 1 with every increment of the system clock. For the remaining 9 dynamic variables, longitudinal data tracking their changes over time

Table 5 Model Selection Result

Candidate Variable	Original Data Type	Property	Selected
Age	Numerical	Dynamic	Yes
Marital status	Categorical	Dynamic	Yes
Education level	Categorical	Dynamic	
Ratio of family income to poverty	Numerical	Dynamic	Yes
Race	Categorical	Fixed	Yes
Age at first menarche	Numerical	Fixed	Yes
Parity history	Numerical	Dynamic	Yes
Alcohol use	Numerical	Dynamic	Yes
Smoking	Numerical	Dynamic	Yes
Age at first sex	Numerical	Fixed	Yes
Ever had sex	Categorical	Dynamic	
Lifetime number of sex partners	Numerical	Dynamic	Yes
Number of recent sex partners	Numerical	Dynamic	
Sexual orientation	Categorical	Fixed	

are needed for the lifetime simulation model. To simplify the tracking process, we convert all numerical dynamic variables to categorical variables.

As many of the risk factors, such as lifetime number of sex partners and number of recent sex partners, are inherently correlated, we fit the XGBoost model along with optimally tuned L1 regularization to control the number of selected risk factors. The L1 regularization makes the model as parsimonious as possible to save the effort in collecting data for dynamic risk factors. We perform the entire analysis in R 3.5.2. The final models for HPV 16 and 18 selected 10 risk factors in total (Table 5). Except for age, race, and age at first sex, the remaining 7 selected variables dynamically fluctuate over years and have to be modeled throughout every woman's whole life. During the simulation, if the values of these variables have any changes, the HPV risk is also updated accordingly. We use a variety of data sources to model these dynamic variables as follows.

Marital status. We model women's changes in marital status based on the method proposed by Ley-Chavez.[38] Five statuses are considered in the model, including first marriage, first divorce, second marriage, second divorce, and never married. The duration of each status is modeled by a discrete distribution depending on age. The simulation randomly assigns an age of first marriage to women based on the probability distribution estimated from the Survey of Income and Program Participation.[39] The durations of marriage and divorce last for a certain amount of time or a lifetime, based on the probability distributions estimated using the Survey of Income and Program Participation[36] and the National Survey of Family Growth.[37]

Ratio of family income to poverty. Ratio of family income to poverty is encoded as a binary variable representing ratio lower or not lower than 1. The variable's transitions are modeled as a distribution stratified by marital status using data from the US Census Bureau.[39]

Parity. The parity in this study is defined by a binary variable indicating if a woman has had a live birth, which is the probability of already having a first birth. The probability is stratified by 3 factors: age, marital status, and race. We assume the probability of already having a first birth is equally distributed in the 5 years of each 5-year age group to get age-specific probabilities.

Alcohol use. Women's alcohol use is defined by 3 types of drinking behavior: nondrinker, moderate drinker, and heavy drinker (i.e., ever have 4 or more drinks every day). The shifts between the 3 types of drinking behavior are modeled based on data reported in the longitudinal analysis by Molander et al.[40] We assume the probability of a transition from one behavior to another behavior is equally distributed in the 5 years to get yearly transition rates.

Smoking. For smoking behavior, women are classified into 2 groups: smokers or nonsmokers. The likelihood of a nonsmoker turning into a smoker is directly estimated using the NHANES data, which also report participants'

Table 6 Simulation Scenarios

Variable	High-Risk Woman	Low-Risk Woman
Age at vaccination	20, 26, 30, 35, 40, 45	20, 26, 30, 35, 40, 45
Age	20	20
Marital status	Unmarried	Married
Ratio of family income to poverty	<1	≥ 1
Race	African American	Asian American
Age at first menarche	7 to 9	13 to 15
Parity history	No	Yes
Alcohol use	Heavy drinker	Nondrinker
Smoking	Yes	No
Age at first sex	≤ 12	20–24
Lifetime number of sex partners	≥ 11	1

ages when they started smoking cigarettes regularly. We use the recent smoking cessation rate reported by the Centers for Disease Control and Prevention[41] to evaluate the yearly likelihood of a smoker to quit smoking.

Lifetime number of sex partners. Lifetime number of sex partners is encoded as a categorical variable with 6 levels: 0, 1, 2 to 4, 5 to 7, 8 to 10, and 11 or more. A 2-stage model is built to model the yearly new sex partners acquisition of women. As the NHANES 2015–2016 data also reported if every woman had sex with a new person, we fit an XGboost model to predict a woman's possibility of having sex with new partners in the current year using the same data for the HPV risk model. For those who are predicted to have new partners, the numbers of their new sex partners are then estimated based on the method proposed by Ley-Chavez.[38] The number of new sex partners a woman acquires annually is categorized into the following groups: 0, 1, 2, and 3 or more. The probability of women having 0 to 3 or more new partners each year is determined by an age-specific discrete distribution. The category "3 or more" actually assigns 3 new partners to women in the simulation. Thus, a woman with new sex partners is randomly assigned to 1 of the 4 categories according to the discrete distribution based on her age group.

The detailed distributions and parameter values are provided in the Appendix.

Design of Numerical Experiments

Based on the HPV risk models, 2 typical risk characteristics, a typical 20-year-old high-risk woman and a typical 20-year-old low-risk woman, are made up using the selected variables (Table 6). As we assume no woman has HPV-related cancer before age 20, the first vaccination scenario is that the 2 women are both vaccinated at age 20. We also examine the outcomes of vaccination at age 26, which is the current age limit of catch-up HPV vaccination. Then the 4 ages beyond the limit are investigated: ages 30, 35, 40, and 45. The simulation is run for the 2 women under each of the vaccination scenarios for 100,000 times. Then the average values of the 3 metrics are reported for each woman under different scenarios, including average gain in life expectancy, lifetime risk of developing cervical cancer, and lifetime risk of developing HPV-related cancers. The average gain in life expectancy is derived by calculating the difference of expected life years between an unvaccinated woman and a vaccinated woman, both of whom have the same risk characteristics.

As a validation, we also use the HPV incidences of the general female population to simulate average women's life courses under the same 6 scenarios. The simulation results of average women are expected to lie between high- and low-risk women's clinical outcomes. The 3 groups' HPV 16/18 incidences in the simulation are directly using the HPV risk models and NHANES data. Table 7 shows the incidences of 3 groups at the 6 vaccination ages (the actual incidence of high-risk and low-risk in the simulation may vary due to the dynamic changes in their risk characteristics).

Results

Table 8 presents the clinical outcomes of vaccinations at different ages. The clinical outcomes show dramatic differences among the women with different levels of HPV risk. The gain in life expectancy of the high-risk woman is over 14 times higher than that of the low-risk woman. However, no matter what the risk, receiving catch-up vaccines almost always benefits a woman. The high-risk

Table 7 Human Papillomavirus (HPV) Incidences of 3 Risk Groups[a]

Age	High-Risk Woman, %	Low-Risk Woman, %	Average-Risk Woman, %
20	4.0	0.2	1.3
26	8.2	0.4	5.0
30	7.6	0.3	4.2
35	6.9	0.7	3.5
40	6.3	0.7	3.6
45	9.4	0.3	2.6

[a]The annual incidences are combined incidence of HPV 16 and HPV 18.

woman has the highest gain in life expectancy when receiving vaccines at age 20, which amounts to 304 days (0.833 years) on average. The effect of vaccination on an average-risk woman falls in between a high-risk woman and a low-risk woman. When getting vaccinated at age 20, an average-risk US woman improves her life expectancy by 147 days (0.403 years), while the low-risk woman only gains 20 additional days (0.056 years) in her life expectancy. The current HPV vaccination policy only recommends catch-up vaccines for women up to age 26, but we still see significant improvements in life expectancy on women older than 26, especially those bearing a high HPV risk. At age 30, the high-risk woman is expected to gain 273 days (0.748 year) by getting vaccinated, while an average-risk woman also gets a 133-day (0.364 years) gain in life expectancy. Even at age 45, the vaccination still shows a considerable value on high-risk women—a 0.491 life-year gain, which is equivalent to 179 days. For an average woman, getting vaccinated at age 45 generates a 0.189 life-year gain on average. In addition, it would result in a lifetime cervical cancer risk of 0.41% for her, which is a slight but significant reduction from the current US female population's lifetime risk of developing cervical cancer (i.e., 0.62%).[42] These results suggest that catch-up vaccines after age 26 should be deliberately considered.

Sensitivity Analysis

The base case analysis assumes that the vaccines provide women with lifetime full immunity against HPV infection. However, as we mentioned above, the HPV attack rate and associated high-grade precancerous outcomes are relatively low after age 25.[24] The long-term effectiveness of HPV vaccination in mid-adult women remains debated. We perform a sensitivity analysis to examine the impacts of different vaccine protection rates on the outcomes (Table 9). Four different scenarios considering age-dependent and time-dependent partial protections against HPV are investigated. Figure 3 presents varying average gain in life expectancy as a metric to test under different scenarios. The results show that lowering vaccine protection after age 25 in the simulation has significant impacts on the high-risk and average-risk women. A partial protection of 50% after age 25 drastically lowers the high-risk and average-risk women's life-year gains from receiving catch-up vaccines to a level similar to that of the low-risk woman (scenario 2). In contrast, the difference of life-year gains between the full and partial protection assumptions is less obvious for low-risk women. In general, varying vaccine protection rates does not produce unexpected outcomes.

Discussion

In this study, we estimate the clinical outcomes that women from typical risk-stratified groups are expected to have under different vaccination scenarios. Although the average gain in life expectancy from receiving catch-up vaccines after age 26 is numerically modest even for the high-risk individual, the potentially averted HPV-related cancer cases can aggregate to substantially large numbers on a group of women with the same risk characteristics. For instance, the high-risk woman vaccinated at age 30 has a lifetime risk of developing cervical cancer equal to 0.24%. If the catch-up vaccine is delayed to age 45, the risk rises to 0.68%. The increased risk (i.e., 0.44%) is equivalent to a significant number of cancer cases when applying to the population with the same high-risk characteristics. Assuming there are 100,000 women having the high-risk profile, if we consider the vaccination at age 45 as a baseline, the catch-up vaccination at age 26 would prevent additional 577 cancer cases compared with the vaccination at age 45. However, the vaccination as late as age 30 could still prevent 480 cervical cancer cases compared with the baseline age 45. In contrast, the increased risk of cancer on low-risk woman is relatively low, implying the necessity of a risk-stratified catch-up vaccination policy.

Table 8 Clinical Outcomes of Human Papillomavirus (HPV) Vaccination at Different Ages

Age at Vaccination	High-Risk Woman	Low-Risk Woman	Average-Risk Woman
Average gain in life expectancy (95% confidence interval)			
20	0.833 (0.820, 0.845)	0.056 (0.044, 0.068)	0.403 (0.153, 0.155)
26	0.789 (0.776, 0.802)	0.056 (0.044, 0.068)	0.390 (0.080, 0.080)
30	0.748 (0.736, 0.761)	0.054 (0.042, 0.066)	0.364 (0.057, 0.057)
35	0.710 (0.698, 0.723)	0.051 (0.039, 0.063)	0.338 (0.326, 0.030)
40	0.608 (0.596, 0.621)	0.037 (0.002, 0.002)	0.277 (0.265, 0.290)
45	0.491 (0.478, 0.504)	0.014 (0.002, 0.027)	0.189 (0.176, 0.202)
Lifetime risk of developing cervical cancer, %			
20	0.04	0.00	0.01
26	0.11	0.00	0.03
30	0.20	0.01	0.09
35	0.30	0.01	0.16
40	0.49	0.04	0.27
45	0.68	0.07	0.41
Lifetime risk of developing HPV-related cancer, %			
20	0.05	0.00	0.01
26	0.14	0.00	0.04
30	0.24	0.01	0.10
35	0.34	0.01	0.17
40	0.57	0.05	0.30
45	0.81	0.09	0.48

Table 9 Vaccine Protection Scenarios in the Sensitivity Analysis

Scenario	Protection Scenario
1	Protection rate decreases to 80% for women over age 25.
2	Protection rate decreases to 50% for women over age 25.
3	Protection rate decreases to 80% after 10 years of vaccination for women.
4	Protection rate decreases to 50% after 10 years of vaccination for women.

In short, this study proposes a novel vaccination evaluation framework, which combines a microsimulation with a dynamically updated risk model. Specifically, we show how to address dynamic risk factors and model these factors over every woman's life course. The modeling process involves data from a variety of sources, but the results can be validated. The same approach can be applied to a clinical outcome evaluation or cost-effectiveness study on the vaccination of another disease such as herpes simplex virus.

Our findings are aligned with the well-established public health recommendations of HPV catch-up vaccination and confirm that the earlier a woman gets vaccinated, the less HPV-related cancer risk she would be exposed to. However, our study differs from most of the previous studies in that the model works at the individual level rather than the population level. We evaluate the impact of HPV vaccines on different women and reveal that the catch-up vaccines after age 26 are still very beneficial to US women, especially those with high HPV risk. We also confirm that Australia's HPV vaccination policy, which provides vaccines to women up to 45 years old, is also rational for some US women. Our results are consistent with many previous arguments that support providing catch-up HPV vaccines to older women.[8–10] Although this study demonstrates the significant health benefits of customizing the cutoff age for different women, whether the current one-size-fits-all HPV vaccination catch-up policy has to be individualized or risk stratified still relies on a comprehensive cost-effectiveness study. This study also creates a new perspective on the cost-effectiveness of HPV vaccination for the countries that still use bivalent or quadrivalent HPV vaccines. As the bivalent and quadrivalent vaccines provide a considerable level of protection against the HPV-related cancers at a much lower cost compared to the 9-valent vaccine, vaccinating older

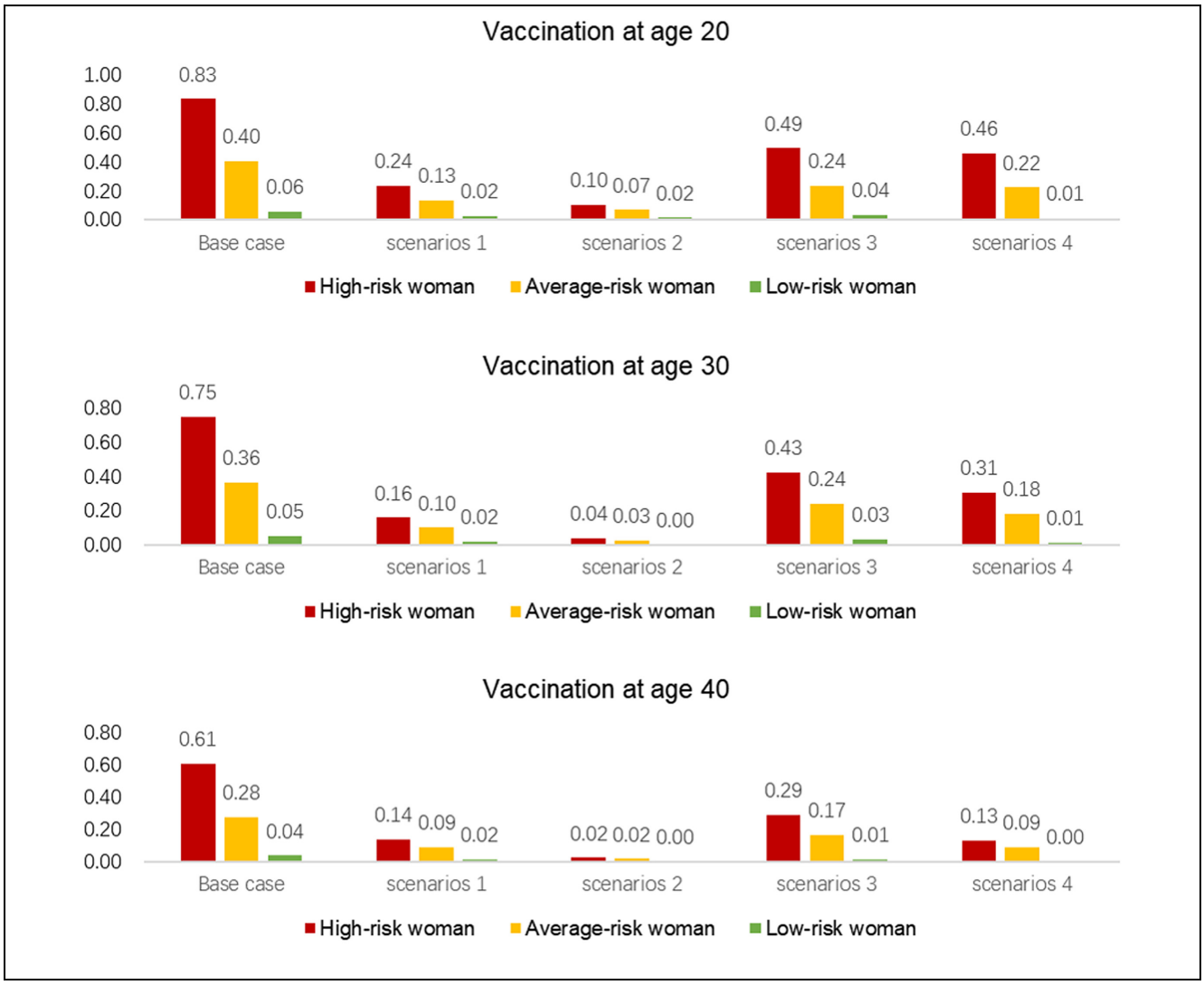

Figure 3 The average gains in life expectancy associated with different protection rates.

women with bivalent or quadrivalent vaccines may exempt some low-risk women at older ages from HPV-related cancer screening (e.g., the Pap test), which is eventually more cost-effective from a systematic point of view.

Our analysis also has some limitations. First, due to the data scarcity, the diagnosis and prognosis of the HPV-related cancers other than cervical cancer are not sufficiently modeled. The simplified method may result in inaccuracy in the final outcomes. Second, we use annual cycles in the simulation as most of the raw data for parameter estimations were reported as annual rates. The model is highly complicated and involves a wide range of parameter estimation, including the dynamic change of various risk-related personal behaviors, the natural history of HPV, and the development, diagnosis, and progression of HPV-related cancers in every cycle. It is difficult to accurately quantify most of these parameters at an interval of less than 1 year. However, the annual update is potentially inadequate to reflect the subtle changes in patients' personal conditions, such as some repeated cancer screenings at intervals of less than 1 year. Third, we make several assumptions in the parameter estimation. For instance, we assume the duration of protection against HPV is lifelong, as the current long-term clinical trials that follow up the protection duration from HPV vaccines are still ongoing and have only reported 10 years' effective duration.[23] Fourth, it is

difficult to find comparable data to validate the results produced by our simulation model. Most of the empirical data and previous studies were focused on the vaccination policy for the entire population and usually put a wide range of ages at vaccination (e.g., 11 to 26 years old) in the same pool to obtain the population-level estimates, while this study only investigates the catch-up vaccination for females at the individual level. Last, prior studies suggested that the 9-valent vaccine targeting HPV 31/33/45/52/58 may prevent an additional 4.2% to 18.3% of cancers on top of the cancers caused by HPV 16/18.[43] Our study focuses on the cancers attributable to HPV 16/18, which potentially underestimates the impact of 9-valent vaccines on cancer prevention. As a future research direction, we can incorporate all HPV-related cancers and noncancer diseases (e.g., genital warts) targeted by 9-valent HPV vaccines in a simulation study to evaluate catch-up vaccination policies.

ORCID iD

Fan Wang https://orcid.org/0000-0001-8860-2150

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at http://journals.sagepub.com/home/mdm.

References

1. Satterwhite CL, Torrone E, Meites E, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis*. 2013;40(3):187–93.
2. Arbyn M, Castellsagué X, de Sanjosé S, et al. Worldwide burden of cervical cancer in 2008. *Ann Oncol*. 2011;22(12):2675–86.
3. Lowy DR, Schiller JT. Reducing HPV-associated cancer globally. *Cancer Prev Res*. 2012;5(1):18–23.
4. Meites E, Szilagyi PG, Chesson HW, Unger ER, Romero JR, Markowitz LE. Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep*. 2019;68(32):698–702.
5. Walker TY, Elam-Evans LD, Singleton JA, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13–17 years—United States, 2016. *MMWR Morb Mortal Wkly Rep*. 2017;66(33):874.
6. Velentzis LS, Sitas F, O'Connell DL, et al. Human papillomavirus 16/18 seroprevalence in unvaccinated women over 30 years with normal cytology and with high grade cervical abnormalities in Australia: results from an observational study. *BMC Infect Dis*. 2014;14(1):3861.
7. Harper DM, DeMars LR. HPV vaccines: a review of the first decade. *Gynecol Oncol*. 2017;146(1):196–204.
8. Australian Government Department of Health. *The Australian Immunisation Handbook*. Department of Health and Ageing; 2013. Canberra.
9. Schwarz TF, Spaczynski M, Schneider A, et al. Immunogenicity and tolerability of an HPV-16/18 AS04-adjuvanted prophylactic cervical cancer vaccine in women aged 15–55 years. *Vaccine*. 2009;27(4):581–7.
10. Westra TA, Rozenbaum MH, Rogoza RM, et al. Until which age should women be vaccinated against HPV infection? Recommendation based on cost-effectiveness analyses. *J Infect Dis*. 2011;204(3):377–84.
11. Mazza D, Petrovic K, Grech C, Harris N. HPV vaccination in women aged 27 to 45 years: what do general practitioners think? *BMC Womens Health*. 2014;14(1):91.
12. U.S. Food and Drug Administration. FDA Approves Expanded Use of Gardasil 9 to Include Individuals 27 through 45 Years Old. 2018. Available from: https://www.fda.gov/news-events/press-announcements/fda-approves-expanded-use-gardasil-9-include-individuals-27-through-45-years-old
13. Sanders GD, Taira AV. Cost effectiveness of a potential vaccine for human papillomavirus. *Emerg Infect Dis*. 2003;9(1):37–48.
14. Elbasha EH, Dasbach EJ, Insinga RP. Model for assessing human papillomavirus vaccination strategies. *Emerg Infect Dis*. 2007;13(1):28.
15. Chesson HW, Ekwueme DU, Saraiya M, Markowitz LE. Cost-effectiveness of human papillomavirus vaccination in the United States. *Emerg Infect Dis*. 2008;14(2):244–51.
16. Chesson HW, Ekwueme DU, Saraiya M, Dunne EF, Markowitz LE. The cost-effectiveness of male HPV vaccination in the United States. *Vaccine*. 2011;29(46):8443–50.
17. Guzzetta G, Faustini L, Panatto D, Gasparini R, Manfredi P. The impact of HPV female immunization in Italy: model based predictions. *PLoS One*. 2014;9(3):e91698.
18. Brisson M, Laprise J. Cost-effectiveness of extending HPV vaccination above age 26 years in the U.S. 2019. Available from: https://stacks.cdc.gov/view/cdc/78081/cdc_78081_DS1.pdf
19. Van de Velde N, Brisson M, Boily MC. Understanding differences in predictions of HPV vaccine effectiveness: a comparative model-based analysis. *Vaccine*. 2010;28(33):5473–84.
20. Winer RL, Lee SK, Hughes JP, Adam DE, Kiviat NB, Koutsky LA. Genital human papillomavirus infection: incidence and risk factors in a cohort of female university students. *Am J Epidemiol*. 2003;157(3):218–26.
21. Dempsey AF. Human papillomavirus: the usefulness of risk factors in determining who should get vaccinated. *Rev Obstet Gynecol*. 2008;1(3):122.
22. Shi R, Devarakonda S, Liu L, Taylor H, Mills G. Factors associated with genital human papillomavirus infection among adult females in the United States, NHANES 2007–2010. *BMC Res Notes*. 2014;7(1):544.

23. Romanowski B, Schwarz TF, Ferguson L, et al. Sustained immunogenicity of the HPV-16/18 AS04-adjuvanted vaccine administered as a two-dose schedule in adolescent girls: five-year clinical data and modeling predictions from a randomized study. *Hum Vaccines Immunother*. 2016; 12(1):20–9.
24. Wheeler CM, Skinner SR, Del Rosario-Raymundo MR, et al. Efficacy, safety, and immunogenicity of the human papillomavirus 16/18 AS04-adjuvanted vaccine in women older than 25 years: 7-year follow-up of the phase 3, double-blind, randomised controlled VIVIANE study. *Lancet Infect Dis*. 2016;16(10):1154–68.
25. Matthijsse SM, van Rosmalen J, Hontelez JA, et al. The role of acquired immunity in the spread of human papillomavirus (HPV): explorations with a microsimulation model. *PLoS One*. 2015;10(2):e0116618.
26. Chaturvedi AK. Beyond cervical cancer: burden of other HPV-related cancers among men and women. *J Adolesc Health*. 2010;46(4):S20–6.
27. Schiffman MH, Bauer HM, Hoover RN, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst*. 1993;85(12):958–64.
28. Tong SY, Lee YS, Park JS, Bae SN, Lee JM, Namkoong SE. Clinical analysis of synchronous primary neoplasms of the female reproductive tract. *Eur J Obstet Gynecol Reprod Biol*. 2008;136(1):78–82.
29. Moscicki A-B, Palefsky J, Gonzales J, Schoolnik GK. Human papillomavirus infection in sexually active adolescent females: prevalence and risk factors. *Pediatr Res*. 1990;28(5):507–13.
30. Reiter PL, McRee AL. HPV infection among sexual minority women: does it matter how sexual orientation is measured? *Cancer Epidemiol Prevent Biomarkers*. 2016;25(3): 559–60.
31. Chen T, Guestrin C. 2016. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by Balaji Krishnapuram, 785–94. San Francisco: Association for Computing Machinery.
32. Nielsen D. Tree boosting with XGBoost: why does XGBoost win "every" machine learning competition? Master's thesis, NTNU, 2016.
33. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2010*. Bethesda, MD: National Cancer Institute; 2013.
34. Arias E, Heron M, Xu J. United States life tables, 2012. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics Report. 2016;65(8):1.
35. Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*. Burlington, MA: Jones & Bartlett; 2013.
36. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"): II. Use in medical decision-making. *Am J Med*. 1982;73(6):889–97.
37. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data 2011–2012*. Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2016. Available from: https://www.cdc.gov/nchs/nhanes/nhanes2011-2012/overview_g.htm.
38. Ley-Chavez A. Quantitative models to design and evaluate risk-specific screening strategies for cervical cancer prevention. Doctoral dissertation, The Ohio State University, 2012.
39. U.S. Census Bureau: Survey of income and program participation. Washington, DC: U.S. Department of Commerce; 2014 Available from: https://www.census.gov/programs-surveys/sipp/data/datasets.html
40. Molander RC, Yonker JA, Krahn DD. Age-related changes in drinking patterns from mid-to older age: results from the Wisconsin longitudinal study. *Alcoholism*. 2010;34(7):1182–92.
41. Centers for Disease Control and Prevention. Quitting smoking among adults—United States, 2001–2010. *MMWR Morb Mortal Wkly Rep*. 2011;60:1513–9.
42. Dunne EF, Markowitz LE, Saraiya M, et al. CDC grand rounds: reducing the burden of HPV-associated cancer and disease. *MMWR Morb Mortal Wkly Rep*. 2014;63(4): 69–72.
43. Saraiya M, Unger ER, Thompson TD, et al. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. *J Natl Cancer Inst*. 2015;107(6): djv086.