MODELING THE IMPACT AND INTERVENTION OF A SEXUALLY TRANSMITTED
DISEASE: HUMAN PAPILLOMA VIRUS

Courtney D. Corley

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2006

APPROVED:

Armin Mikler, Major Professor
Rada Mihalcea, Committee Member
Joseph Oppong, Committee Member
Diane Cook, Committee Member
Karan Singh, Committee Member
Krishna Kavi, Chair of the Department of
  Computer Sciences
Oscar Garcia, Dean of the College of
  Engineering
Sandra L. Terrell, Dean of the Robert B.
  Toulouse School of Graduate Studies
Corley, Courtney D. *Modeling the Impact and Intervention of a Sexually Transmitted Disease: Human Papilloma Virus*. Master of Science (Computer Science), May 2006, 58 pp., 7 tables, 10 figures, references, 53 titles.

Many human papilloma virus (HPV) types are sexually transmitted and HPV DNA types 16, 18, 31, and 45 account for more than 75% if all cervical dysplasia. Candidate vaccines are successfully completing US Federal Drug Agency (FDA) phase III testing and several drug companies are in licensing arbitration. Once this vaccine become available it is unlikely that 100% vaccination coverage will be probable; hence, the need for vaccination strategies that will have the greatest reduction on the endemic prevalence of HPV. This thesis introduces two discrete-time models for evaluating the effect of demographic-biased vaccination strategies: one model incorporates temporal demographics (i.e., age) in population compartments; the other non-temporal demographics (i.e., race, ethnicity). Also presented is an intuitive Web-based interface that was developed to allow the user to evaluate the effects on prevalence of a demographic-biased intervention by tailoring the model parameters to specific demographics and geographical region.
ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Armin Mikler, for his unwaivering commitment to see me succeed in this field. Thank you, Dr. Rada Mihalcea – your support, guidance, and advice on how to survive graduate school has been instrumental in my career. I would not have been able to complete this work without the motivation, watchful eye and love of my Higher Power.

I would also like to thank Maciej Boni at Stanford University for his many phone and email conversations about mathematical model sensitivity analyses. I would also like to acknowledge Donna Rickert at the Center for Disease Control and Prevention for her introduction to the development of a human papilloma virus vaccine.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
</tbody>
</table>

### Chapters

1. **INTRODUCTION** .................................................................................................................. 1
   1.1 Motivation ...................................................................................................................... 1
   1.2 Thesis Objectives and Outline ..................................................................................... 2

2. **REVIEW OF CLASSIC AND CURRENT METHODS** .......................................................... 4
   2.1 Background .................................................................................................................... 4
   2.2 Towards Computational Epidemiology .......................................................................... 4
   2.3 Methods in Computational Epidemiology ..................................................................... 6
       2.3.1 Mathematical Models ............................................................................................. 6
       2.3.2 Modeling with Cellular Automata .......................................................................... 9
       2.3.3 Agent-Based Models .............................................................................................. 9
       2.3.4 Related Methods in Modeling the Impact of Human Papilloma Virus ................ 10
   2.4 Limitations of Current Models ..................................................................................... 11

3. **HUGHES, GARNETT, ANDERSON AND KOUTSKY MODEL** ........................................ 12
   3.1 Population-level Impact of a HPV Vaccine ................................................................. 12
       3.1.1 Model Outline ......................................................................................................... 13
   3.2 Parameter Space ............................................................................................................. 14
       3.2.1 Population ................................................................................................................ 14
       3.2.2 Disease .................................................................................................................... 17
       3.2.3 Vaccine .................................................................................................................... 17
   3.3 Mathematical Description ............................................................................................... 18
       3.3.1 Population ................................................................................................................ 18
       3.3.2 Infection Dynamics ................................................................................................. 19
       3.3.3 Compartments ......................................................................................................... 21
3.4 Non-temporal Demographics .......................................................... 23

4. TWO DISCRETE-TIME EPIDEMIC MODELS .................................. 25
4.1 Introduction to Discrete-Time Models .......................................... 25
4.2 Temporal Model ........................................................................ 26
  4.2.1 Temporal Model Description – First Age Stratum ............... 27
  4.2.2 Temporal Model Description – Remaining Age Strata ...... 30
4.3 Non-Temporal Model ................................................................ 32
4.4 Sensitivity Analysis .................................................................. 35

5. POPULATION-LEVEL IMPACT: MODELING PREVALENCE THROUGH
   DEMOGRAPHICS ................................................................................ 40
5.1 Integrating Demographics ........................................................... 40
  5.1.1 Sexual Mixing Patterns ....................................................... 41
5.2 Tool Interface ............................................................................ 42
5.3 Results ....................................................................................... 47
5.4 Discussion .................................................................................. 48
  5.4.1 Analysis of Denton County, TX ......................................... 48

6. CONCLUSION .................................................................................. 53
6.1 Discussion .................................................................................. 53
6.2 Limitations and Future Work ..................................................... 53

BIBLIOGRAPHY .................................................................................... 55
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Model parameters</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>Temporal model sensitivity analysis</td>
<td>39</td>
</tr>
<tr>
<td>5.1</td>
<td>US county abbreviations</td>
<td>40</td>
</tr>
<tr>
<td>5.2</td>
<td>US Census data for select US counties</td>
<td>41</td>
</tr>
<tr>
<td>5.3</td>
<td>Sexual activity group contact rates per demographic stratum</td>
<td>45</td>
</tr>
<tr>
<td>5.4</td>
<td>Temporal model: Endemic prevalence and relative reduction in prevalence per vaccine policy</td>
<td>50</td>
</tr>
<tr>
<td>5.5</td>
<td>Non-temporal model: Endemic prevalence and relative reduction in prevalence per vaccine policy</td>
<td>51</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Infection timeline</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>SIR state diagram</td>
<td>6</td>
</tr>
<tr>
<td>2.3</td>
<td>SIR graph</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>Sexually active population subset</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>HGAK model state diagram</td>
<td>14</td>
</tr>
<tr>
<td>3.3</td>
<td>Example interaction matrix with 3 activity classes</td>
<td>20</td>
</tr>
<tr>
<td>4.1</td>
<td>Temporal model state diagrams</td>
<td>28</td>
</tr>
<tr>
<td>5.1</td>
<td>HPV tool interface</td>
<td>43</td>
</tr>
<tr>
<td>5.2</td>
<td>HPV tool interface</td>
<td>44</td>
</tr>
<tr>
<td>5.3</td>
<td>Temporal model output graphs</td>
<td>46</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. Motivation

Human papilloma virus (HPV) DNA is found in 99.7% of all cervical cancers with types 16, 18, 31 and 45 accounting for 75% of cervical dysplasia (cancer) [1]. In the United States (US), 13,000 women are diagnosed with cervical dysplasia and 5,000 die annually. By the age 50, 80% of women will have acquired genital HPV infection [2]. Currently, 20 million people are infected with HPV in the U.S. with 5.5 million new cases annually [3]. In 2004, U.S. health care system spent over $1.6 billion treating HPV symptoms and an additional $5-6 billion on screening tests, including pap smears.

An effective HPV vaccine would have significant impact on HPV infections and cervical disease. Candidate vaccines finished phase II testing and phase III trials have begun with the Food and Drug Administration [4]. Due to the health care and human costs associated with this virus, it is vital to have an effective vaccination strategy in place when this vaccine becomes available.

Predictive models are important tools in determining disease transmission dynamics and effective vaccination solutions. Previously, Hughes, Garnett and Anderson have developed a model that predicts HPV prevalence [5]; however, this model does not allow for demographic stratification via time-dependent traits, such as age. With this in mind, we have developed a tool that facilitates predicting HPV prevalence in a variety of demographic settings and allows for various vaccination solutions. These predictive models stratify a population into different subgroups based on sexual mixing patterns. Population demographics and census data are analyzed to extract demographic parameters and youth and adult risk behavior surveys are studied to determine the sexual partner exchange rates for a population [6, 7].
1.2. Thesis Objectives and Outline

Disease intervention is critical to maintaining the general health of the population. In this thesis, we address the issue of providing a demographically-biased intervention on a sexually transmitted disease; specifically, human papilloma virus. Turner et al [8] address demographic disparity in the incidence of sexually transmitted disease; however, there is limited research in how a demographically-biased intervention would effect the prevalence of a disease in a population. This thesis introduces two discrete-time epidemic models capable of evaluating the effect of temporally (i.e. age) dependent demographic and non-temporally (i.e. ethnicity) dependent demographic interventions on the endemic prevalence of a sexually transmitted disease. This thesis also presents a user-centric web-application that interfaces to our models.

Computational epidemiology has been introduced as a new domain in computer science research and the motivation for a predictive epidemic model to evaluate the population-level impact of an intervention has been introduced in first two sections of this thesis. Chapter two defines classic and current methods in modeling epidemics including mathematical models, modeling with cellular automata, agent-based simulations, recent attempts at modeling HPV endemics and the limitations of the current models. Next, we provide an in-depth analysis of previous related work (Hughes et al) using a system or ordinary-differential equations to model human papilloma virus intervention. The discrete-time models defined in this thesis are constructed with similar concepts and dynamics as the work by Hughes et al.

Chapter four includes our contribution to the field of computational epidemiology. Two discrete-time models are defined that evaluate population-level impact of a demographically-biased intervention on HPV prevalence. The first model evaluates a population subset stratified by temporally dependent demographics (i.e.age) and the second model evaluates a non-temporally dependent demographic population subset (i.e. ethnicity). Chapter four also presents a numerical and sensitivity analysis of the temporal models.
Chapter five presents our analysis of the models in six select counties in the United States, chosen for their variance in demographic composition. Next, the web-tool that interfaces with the two models is introduced and the chapter concludes with results and discussions, including results as they related to Denton County, TX. Finally, in chapter 6 we present a concluding discussion and suggest possible future work.
CHAPTER 2

REVIEW OF CLASSIC AND CURRENT METHODS

2.1. Background

During the last century, medical science has made tremendous progress in identifying, treating, or even eradicating many infectious diseases. This can primarily be attributed to the increased understanding of the etiology and pathogenesis of such diseases. Newly emerging or re-emerging infectious diseases continue to occur regularly [9]. Some diseases have changed their appearance, some have become resistant to drug treatment, while others are so new that no previous outbreaks have ever been studied. It is ironic that epidemiologists have to take advantage of a disease outbreak in order to collect data necessary to formulate public health policy. Medical research has enhanced the understanding of disease characteristics in an individual. For example, the epidemiologic stages of Influenza as described by latent period, infectious period, and recovery period [10] are well known [10, 11]. So are the symptomatic stages of influenza (i.e., incubation period until symptoms occur) as shown in Fig. 2.2. The susctibles-infectes-removals (SIR) state diagram illustrates the course of a disease in an individual. The manifestation and spread of many infectious diseases in the population remain elusive and are dependent on socio-behavioral interaction patterns and population dynamics.

2.2. Towards Computational Epidemiology

With new and re-emerging local disease outbreaks and the increased threat of bio-terrorism, disease monitoring cannot continue to be fragmentary and inadequate, focusing on small special domains [12]. Computational models for simulation of disease dynamics will accelerate epidemiological research, disease tracking and surveillance. Study of infectious diseases needs models pertinent to spatially delineated environments, such as a tuberculosis
(TB) outbreak in a homeless shelter or factory setting, as well as non-delineated models for a geographic region, such as the progression of Influenza in specific regions of the United States. The study focusses on the progression of infectious diseases in non-delineated environments in diverse demographic and geographic settings.

The significance of computational epidemiology as a new field has been underscored by a special program at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) [13], funded as an NSF technology tenter. A 5-year program, consisting of working groups and short-courses focusing on computational and mathematical epidemiology began in summer of 2002. It emphasizes the development and strengthening of collaborations and partnerships between mathematicians, computer scientists, biologists, sociologists, biostatisticians, and epidemiologists. The computational epidemiology research laboratory [14] at the University of North Texas involves both graduate and under-graduate students in inter-disciplinary research studies of disease dynamics including influenza, tuberculosis, HPV and HIV\textsuperscript{1}.

\textsuperscript{1}This section also appears in NSF Proposal MGSGS 2005
2.3. Methods in Computational Epidemiology

To gain insight into the intricacies of disease dynamics in a specific population, statistical and mathematical models of infectious disease epidemics have been developed [15, 16, 17]. Recently, some computational disease models have emerged, which facilitate the simulation and investigation of different disease characteristics. These include models that exploit the *susceptibles-infectives-removals (SIR)* paradigm, *cellular automata (CA)* methodology, *agent-based modeling* and *bayesian reasoning* [18, 19, 17].

2.3.1. Mathematical Models

Newly emerging or re-emerging infectious diseases continue to occur regularly [9]. Identification, treatment and eradication of different infectious diseases can be attributed to the increased understanding of their etiology and pathogenesis. It is ironic that epidemiologists have to take advantage of a disease outbreak in order to collect requisite data to formulate public health policies. While medical research has enhanced the understanding of disease characteristics in an individual, manifestation and spread of infectious diseases in the population remains elusive.

![SIR state diagram](image)

**Figure 2.2. SIR state diagram**

2.3.1.1. *Susceptibles-Infectives-Removals Model.* Now, we introduce a family of SIR models. Most SIR mathematical models are based on the principle of interaction between groups of *susceptibles (S), exposed (E) / infectives (I)*, and *recovered/removed (R) individuals*, i.e.,
the \textit{SIR/SEIR} model. \textit{Susceptibles} are those individuals in a population who can be infected by the disease. \textit{Infectives} are those individuals who have been infected and are infectious. \textit{Removals} include all individuals that are incapable of transmitting the infection, and are either recovering, fully recovered or expired from the disease.

The SIR/SIRS state diagram (Figure 2.2) illustrates the course of a disease in an individual. A susceptible individual may be exposed to a disease pathogen and continue to be in the susceptible state. A susceptible becomes an infective, first when the susceptible individual becomes infected and then once the susceptible is able to transmit the pathogen onto others. The recovery state begins once the ability to infect ceases. The individual continues the state of recovery from the disease, or may expire. On full recovery, the individual may acquire full immunity from disease and hence is no longer susceptible (SIR model). The individual reverts to a susceptible on full recovery when lacking disease immunity (SIRS model).

The Kermack-McKendrick threshold theorem [20] is the basis for the \textit{SIR} model. A continuous influx of susceptibles is a requisite for sustained infection in a population. This is the case of endemic diseases, including HPV, that prevail in a community at all times. The model is based on the presumption of a closed population, assuming that the epidemic spreads rapidly enough that the changes brought in by births, deaths, migration and demographic changes are negligible [21].

During the start of a disease epidemic, the total population is susceptible, excluding those that have inherent immunity to the disease. The \textit{index case} is the first infected individual and the infection source. During the infectious period, the infection is transmitted to some susceptibles, who interact with the index case at close proximity to contract the infection. This triggers the cycle of infections progressing through the population. Infectious individuals become members of the removals category once they cease to be infectious. For the classic \textit{SIR} model, the total number of susceptibles ($S$), infectives ($I$), and removals ($R$) is constant. New infections occur until the rate of new disease cases, the disease incidence, reaches a
peak [20, 22]. Thereafter, the incidence starts to recede due to the decrease in the number of susceptibles, and diminishes eventually. Figure 2.3 illustrates the temporal flow of the population in each state for the classic SIR model.

![Figure 2.3. SIR graph](image)

The random mixing of susceptibles and infectives [21] is given by $S \ast I$. $\beta$ defines the transmission coefficient based on contact rate and disease infectivity [23]. $\gamma$ defines the rate of infectives ($I$) becoming non-infectious. Hence, the average duration of infectivity is $1/\gamma$ [21]. The set of differential equations used in classic SIR model for a closed population are summarized in Eq. 1. The transfer rates of individuals from $S \rightarrow I$ and $I \rightarrow R$ are given by $dS/dt$ and $dR/dt$ respectively. The rate of change of infectives is given by $dI/dt$.

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= +\beta SI - \gamma I \\
\frac{dR}{dt} &= +\gamma I
\end{align*}
\]

The SIR model provides a simple framework to represent disease prevalence. However, it does not provide sufficient insight into an epidemic composition in order to be used as a policy and planning tool for the allocation of public health resources. The SIR model does not take into account the geography or the demographics of a region.
2.3.2. *Modeling with Cellular Automata*

Cellular automata have been used for several decades [24] for computational models. A two dimensional automaton is used in epidemic models utilizing cellular automata [25, 24, 26, 27]. Each cell may represent an individual or a sub-population and is characterized with state and likelihood risks for exposure and contracting the disease. The disease progression is studied through its diffusion across the neighboring cells.

The earliest example of use of cellular automata is Bailey’s lattice model [22] for the spread of diseases from micro-level interactions. Di Stefano et al [27] have developed a lattice gas cellular automata model to analyze the spread of epidemics of infectious diseases. The model is based on individuals who can change their state independent of others and can move from one cell to another. However, this approach does not consider the critical factor of the infection time-line such as work by Venkatachalam et al. Fu has used stochastic cellular automata to model epidemic outbreaks that take into account the spatial heterogeneities [28]. The computational epidemiology research laboratory at the University of North Texas (UNT) has introduced the global stochastic field simulation (GSFS) paradigm, and use it in studies of infectious diseases’ epidemics and vaccination strategies [29, 19, 17].

2.3.3. *Agent-based Models*

Spatially delineated regions with a small or large populations can be constructed, using an agent-based approach, in which each individual is represented by an autonomous agent. The interaction parameters are pre-determined and population mixing patterns are studied to understand the progression of diseases [30]. Agent-based models have been used to analyze HIV/AIDS spread in the population and individual immune levels following the infection [31]. A survey of agent-based epidemic simulation models is available [15]. BioWar is an agent-based system that analyzes the disease spread, treatment, and recovery, by porting principles of interactions from social, knowledge and work networks [32]. Agent-based models have also
been applied to analyze real world outbreaks of tuberculosis in factory and homeless shelter settings [33].

2.3.3.1. Bayesian Models. The Bayesian paradigm incorporates the capabilities of probabilistic reasoning and reasoning under uncertainty. Probabilistic and stochastic analysis are integral to Bayesian methodologies and give a closer view of the real world compared with rule-based systems. Bayesian learning has been successfully applied in the areas of medical diagnosis, weather forecasting, gaming, and fault diagnosis. Nevertheless, in the field of modeling epidemics and their analysis, the Bayesian paradigm has been rarely utilized to its full potential.

In the Amazon region, where onchocerciasis (river blindness) is endemic, Bayesian reasoning has been used to identify communities that needed priority ivermectin treatment [34]. Spiegelhalter et al [35] investigated the utility of Bayesian analysis for health technology assessment and highlighted its practical advantages in handling complex inter-related problems. Bayesian classifiers are used in the real-time outbreak and disease surveillance (RODS) system [36], a computer based public health surveillance system that detects disease outbreaks. RODS was used in the 2002 Winter Olympics. Pennsylvania and Utah currently use RODS for public health surveillance. Bayesian monitoring of critical factors in cancer related clinical trials, such as toxicity and quality of life measures, led to higher accuracy [37]. An epidemiological model using Bayesian analysis has been developed for malaria in Ndiop, Senegal [38]. Bayesian learning has also been used to infer the dependency of disease incidence on the demographics in different geographic regions [39].

2.3.4. Related Methods in Modeling the Impact of Human Papilloma Virus

Markov models have been developed that are capable of simulating the natural history of HPV and type specific stages of cervical carcinogenesis [1, 40]. Improved Markov models simulate high-and-low risk HPV infection. They are capable of simulating non-persistent and persistent HPV infections that leads to cervical carcinogenesis [4]. Cost-effectiveness analysis
has been performed on the benefits of a HPV vaccine implementing decision and Markov models [3]. Additionally, models have been developed to investigate ethnic inequalities in the incidence of sexually transmitted infections [8]. However, many differential equation and Markov models ignore specific demographics and approach modeling at population-level [41, 5]. We provide a detailed review of one of the ODE models in Chapter 3.

2.4. Limitations of Current Models

The SIR model provides a simple representation of disease progression in a homogeneous population. However, it does not provide accurate insight into the composition of an epidemic to be used as a policy and planning tool for the allocation of public health resources. Naive cellular automata are impeded by a limited neighborhood, and the social interactions based on demographics are not readily incorporated. Bayesian principles have been applied to study disease outbreaks, but not as a predictive model to quantify the risk levels ahead of outbreaks. Agent-based models are implemented using rigid rules of interaction that differ from the real world social dynamics. The current models can be potentially extended to include geography, demographics, and social dynamics, but the associated drastic increase in intrinsic complexity renders the model computationally expensive and possibly infeasible. The restrictions and scalability limitations of the current SIR models will be addressed by research in field of computational epidemiology. Research in computational epidemiology will complement the current existing methodologies with studies of heterogeneous population, including interactions based on geography, demography, environment and migration patterns. Also, high performance computational platforms can provide the requisite cyber infrastructure to model and analyze disease progression and prevention strategies at higher levels of fidelity.
CHAPTER 3

HUGHES, GARNETT, ANDERSON AND KOUTSKY MODEL

3.1. Population-level Impact of a HPV Vaccine

The two-discrete time models defined in Chapter 4 of this thesis are constructed with the dynamics implemented in previous sexually transmitted disease models. In this section we describe one of these models, developed by Hughes, Garnett, Anderson and Koutsky (HGAK) [41, 42, 5], which evaluates the population-level impact of a pre-exposure human papilloma virus vaccine. This compartment model, based on the SIR paradigm, depicts endemic disease prevalence in a given population. Transmission and infection dynamics in this model occur within a sexually active population and this population is a subset, determined by age, of the total population. Given that individuals generally do not have the same contact rate, the sexually active population is sub-divided into disparate groups classified by their average contact rate with the population remaining constant throughout the execution of the model; however, a temporal flow is present with a population proportion aging-out and the same proportion aging-in continuously. The sexually active population is a subset of the total population, with age-related movement and constant population size (Fig. 3.1) and the model dynamics incorporate the concept of homogeneous sexual mixing, in a population, with varying contact rates between individuals. Effective disease transmission is described by varying contact rates between individuals, disease characteristics, and sexual mixing; the homogeneous sexual mixing that occurs is further described by the proportion of contacts that take place within group and conversely group-to-group. Now, endemic disease prevalence is determined when the infection rate does not change, this is called the infection equilibrium.
3.1.1. Model Outline

This model contains five compartments that represent an individual’s current disease state and each of these compartments is governed by an ordinary-differential equation. The model’s compartments are similar to the SIR model (susceptible, infectious, removed/immune) with two added states to include vaccination effect (vaccinated, vaccinated & infectious) and the dynamics of each compartment determine the level of infection in the population. Population size remains constant throughout the execution of the model through equal temporal age-related movement. Individuals age-into the model and are either susceptible or vaccinated upon entry. Members of the susceptible compartment either leave the population due to aging — out of the model, stay susceptible, become vaccinated, or become infectious. Members of the infectious compartment either leave the population due to aging out of the model, stay infectious or recover with no detectable HPV virus, viral load, in the body (viral clearance). Individuals in the vaccinated compartment either leave the population due to aging out of the model, stay vaccinated, become infectious or become susceptible due to vaccine duration. Members of the vaccinated infectious compartment either leave the population due to ageing out of the model, stay vaccinated infectious, or recover due to viral clearance. Individuals in the removed compartment either leave the population due to aging
out of the model, or stay removed. Figure 3.2 displays the population flow in this model and the list below describes the possible transitions in each state.

![Figure 3.2. HGAK model state diagram.](image)

Ultimately, the endemic prevalence of HPV is calculated at infection equilibrium. Infection equilibrium occurs when every infectious individual causes one new infection In an ordinary-differential equation model, infection equilibrium is calculated by solving the infectious ODE at zero.

3.2. Parameter Space

The HGAK model parameter space is categorized by one of the following characteristics: biological, vaccine, or population. The following section describes each category of parameters, defines possible value ranges, and explores the effect of on HPV endemic prevalence. A list of parameters and their default values can be found in Table 3.1 [5].

3.2.1. Population

The population and demographics in the HGAK model make-up the interaction base. This model segregates population into sexual activity groups, with partner change determining the
Table 3.1. Model parameters.

<table>
<thead>
<tr>
<th>Description</th>
<th>Variable</th>
<th>Type</th>
<th>Initial Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average duration in sexually active population</td>
<td>$1/\mu$</td>
<td>D</td>
<td>15</td>
</tr>
<tr>
<td>Mixing Parameter</td>
<td>$\epsilon$</td>
<td>D</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Demographic Sexual Activity specific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion in group $I$</td>
<td>$\omega_I$</td>
<td>D</td>
<td>†</td>
</tr>
<tr>
<td>Partner change rate for group $I$</td>
<td>$c_I$</td>
<td>D</td>
<td>‡</td>
</tr>
<tr>
<td><strong>Gender-specific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmission rate from infectious of gender $k'$ to susceptible of gender $k$</td>
<td>$\beta_k$</td>
<td>B</td>
<td>0.7, 0.8 ‡</td>
</tr>
<tr>
<td>Relative risk of transmission of a vaccinated individual compared with an</td>
<td>$r_k$</td>
<td>V</td>
<td>1.0</td>
</tr>
<tr>
<td>unvaccinated individual of gender $k$ to gender $k'$ with natural infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative susceptibility to infection of vaccinated of gender $k$ compared</td>
<td>$\psi_k$</td>
<td>V</td>
<td>0.25</td>
</tr>
<tr>
<td>with unvaccinated individuals of gender $k$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean duration of vaccine protection( yr.)</td>
<td>$1/\sigma_k$</td>
<td>V</td>
<td>10</td>
</tr>
<tr>
<td>Mean duration of infectiousness for individuals of gender $k$</td>
<td>$1/\gamma_k$</td>
<td>B</td>
<td>1.5</td>
</tr>
<tr>
<td>Relative rate of recovery from (breakthrough) infection in vaccinated</td>
<td>$\alpha_k$</td>
<td>V</td>
<td>1.0</td>
</tr>
<tr>
<td>individuals of gender $k$ compared with unvaccinated individuals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender and activity group specific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of individuals of gender $k$ and demographic activity group $I$</td>
<td>$\phi_{kI}$</td>
<td>V</td>
<td>0.9 ‡</td>
</tr>
<tr>
<td>who are vaccinated</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

source: Hughes et al Epidemiology, vol 13, no.6 pp. 631–639

*These parameters are the variables used in the equations described in [5].
† The type of each parameter is either D, B, or V corresponding to demographic, biological/natural history of HPV and vaccine
‡ Proportions are derived from population demographics with 0.03, 0.15, and 0.82 corresponding to high, moderate and low activity classes in a demographic group
§ Contact rate (partners/yr) correspond to high, moderate, and low activity groups in the demographic group
¶ Female-to-male and male-to-female, respectively
|| Effective vaccination coverage for all groups

activity group contact rate. Although, the number of sexual encounters is an important factor in disease transmission, the contact rate here is based on the number of partner changes in one time unit. A qualitative assumption can be made that disease transmission has a higher correlation to partner change than the number of sexual encounters.
The population compartments are as follows: $S_{ki}$ (susceptibles), $I_{ki}$ (infectious), $V_{ki}$ (vaccinated), $VI_{ki}$ (vaccinated and infectious), $R_{ki}$ (recovered or immune). Where $k$ is the gender, male ($k = 0$) or female ($k = 1$). Each of the compartments is further divided into sexual activity classes, indexed by $l$. The HGAK model considers three classes corresponding to low, moderate and high contact rates. These groups are classified by their average contact rate ($c_i$), where $c$ is 9, 3, 1.4 corresponding to high, moderate and low activity classes. The proportion of the population in each specific group is determined by $\omega_i$. The default proportions for each group are 0.03, 0.15, 0.82, corresponding to high, moderate and low activity classes.

Only heterosexual partnerships are evaluated and mixing is uniform inside each compartment’s sexual activity group. Since contact rate is measured via partnership change, this model assumes that before the age of 15 and after the age of 30, the partnership change variance is not significant. Therefore, the sexually active population modeled is between the ages of 15 and 30. There may be cases where demographics change the sexually active age range; hence, this model has the capability to model any age range ($1/\mu$). Although, mixing in each subgroup is homogeneous, the interaction and partnership formations between these subgroups may vary. It is not known if individuals with a high contact rate prefer to form partnerships with others with a high contact rate or with those with a low or moderate rate; hence, the parameter, $\epsilon$, is introduced. The parameter $\epsilon$ has a value between zero and one, ($0 \leq \epsilon \leq 1$). The closer the value to zero, partnerships occur within group, or are assortative (like with like); conversely, the closer the value to one, partnerships are distributed equally among all activity classes (like with unlike). So, $\epsilon$ determines the population proportion that interacts with individuals of the opposite sex in disparate demographic and sexual activity subgroups. The population parameters discussed above are gender ($k$), activity class ($l$), time in sexually active population ($1/\mu$), mixing parameter ($\epsilon$), and the compartments ($S_{ki}, I_{ki}, V_{ki}, VI_{ki}, R_{ki}$).
3.2.2. Disease

Several disease characteristics are vital for accurate portrayal of infection transmission, including infectivity and infectious period. The infection parameters in the HGAK model are obtained from viral and biological data. When a partnership is formed with one individual either being infectious or vaccinated and infectious, $\beta_k$ determines the proportion of the partnerships where infection is transmitted to a susceptible or vaccinated individual; specifically, $\beta_0$ is male-to-female and $\beta_1$ is female-to-male. HPV transmitted easily between hosts with more than 30 strains that are sexually transmittable. The transmission risk from an infectious individual to a susceptible of the opposite gender differs by gender and the effective transmission rate from male to female is 80% and for female to male is 70%. The effective transmission rate is derived from a binomial distribution, $1 - (1 - \zeta)^k$, over the average number of sex acts with a partner ($k$) and $\zeta$ is the risk of being infected in one sex act [43]. Here, the infectious period of HPV is approximately 1.5 years [40]. The infection parameters discussed above are partnership infectivity ($\beta$) and duration of infection ($1/\gamma$)

3.2.3. Vaccine

The goal of this model is to evaluate population-level prevalence after a vaccine is introduced. The vaccine can be arbitrarily placed in a segment of the population. The proportion of effective vaccination of gender $k$ and sexual activity class $l$ is $\phi_{kl}$, this is often referred to as the vaccine take. The vaccine can be targeted at high risk sub-groups or spread across portions or the entire population. The relative susceptibility to infection of vaccinated individuals of gender $k$ compared with unvaccinated individuals of gender $k$ is $\psi_k$, often referred to as the degree of infection or vaccine efficacy. Although, the vaccine is yet to be publicly available, the preliminary results from large human-trials show an average degree of 0.25 meaning that the vaccine takes and is efficacious in 0.75 of the population. Also, the average duration of the vaccine is assigned 10 years ($1/\sigma$). Other vaccine characteristics are considered as well. Since the exact effects of the vaccine are not known, it may be possible to increase or
decrease transmission risk and increase or decrease the recovery rate. The relative transmission risk of a vaccinated individual with breakthrough infection compared to an unvaccinated individual with natural infection is \( r_k \), with a value of 1.0 in our experiments meaning there is no change in risk of transmission due to the type of infection. The relative rate of recovery from (breakthrough) infection in vaccinated individuals compared with unvaccinated is \( \alpha_k \), with a default value of 1.0 meaning there is no difference in the relative rate of recovery between types of infection. The vaccine parameters discussed above are take \( (\phi_{ki}) \), degree \( (\psi_k) \), duration \( (1/\sigma) \), relative transmission risk \( (r_k) \), and relative recovery rate \( (\alpha_k) \). Thus, given population demographics, vaccine and HPV characteristics, endemic prevalence of HPV can be modeled for a region.

3.3. Mathematical Description

A set of differential equations define the HGAK model. Each compartment’s continuous change is evaluated with one of these equations. The model dynamics are similar to many other infectious disease compartment models. This section first details two equations that represent supersets of the disease state compartments. Next, infection dynamics are introduced and finally the compartment dynamics are described.

3.3.1. Population

The value of \( N_{kl} \) (Eq. 2) is the absolute population size of gender \( k \) and activity class \( l \), a superset of the compartments. The population in all five compartments of gender \( k \) and activity class \( l \) are summed to equal the value of \( N_{kl} \). This value will be used to define the total contacts generated in the model.

\[
N_{kl} = S_{kl} + I_{kl} + R_{kl} + V_{kl} + V'I_{kl}
\]

While the population remains constant, there is a temporal process in effect here. This process embeds moving population from the sexually non-active to the active population
modeled, and moving sexually active population to the non-active population, in each compartment and specifically each gender and activity class, to maintain constant population size. This proportion is exactly $\mu$ ($1/\mu$ is the number of years in the sexually active population). When individuals exit the active population from each compartment, the same proportion of individuals enter the total sexually active population ($\eta$), as either susceptible or vaccinated. The total sexually active population is the sum of male and females, in sexual activity classes. Equation 3 defines $\eta$, the total sexually active population.

\begin{equation}
\eta = \sum_k \sum_l N_{kl}
\end{equation}

3.3.2. Infection Dynamics

The deterministic infection dynamics occur due to homogeneous mixing within sub-groups and are parameterized (by $\epsilon$) in group-to-group mixing. The parameter $\rho$ generates contacts occurring within an individuals activity class ($l$) and the contacts ($N_m c_m$) made outside of the individuals activity group ($m$). The value of $m$ is equivalent to $l$ and introduced to differentiate between the current activity class being evaluated and the class with which interactions are made. Mixing occurs on a scale from fully assortative (like-with-like) to fully random (like-with-unlike). The range from assortative to random is parameterized by the value of $\epsilon$ ($0 \leq \epsilon \leq 1$). Equation 4 illustrates the proportion of contacts that occur between groups $l$ and $m$.

\begin{equation}
\rho_{lm} = (1 - \epsilon) \delta_{lm} + \epsilon \left[ \frac{N_m c_m}{\sum_s N_s c_s} \right]
\end{equation}

The following describes $\rho_{lm}$ in its two extremes. First, fully assortative mixing ($\epsilon = 0$) considers only interactions within sub-group. The second portion of the equation results in a zero value and can be ignored. Now, the proportion of population that mixes from group $l$ to $l$ is 1.0, conversely the proportion of population that mixes to all other groups is 0.0.
Figure 3.3. Example interaction matrix with 3 activity classes

$$\rho = \begin{bmatrix} 0.572 & 0.120 & 0.307 \\ 0.072 & 0.620 & 0.307 \\ 0.072 & 0.120 & 0.807 \end{bmatrix}$$

So, this can be represented as a mixing matrix equal to the identity matrix. Hence, $\delta$ is an identity matrix of size $m$ by $m$, where $m$ is the number of population sub-groups.

Second, fully random mixing ($\epsilon = 1$) accounts for all like-with-unlike combinations of subgroup-to-subgroup mixing. The first portion of the equation can be ignored since it evaluates to zero. The second portion of $\rho_{im}$ determines the proportion of interactions of group $m$. The product of the number of contacts for an activity class ($c_m$) and the population in the same group ($N_m$) are the total contacts of subgroup $m$. Now, all possible contacts is summed from the contacts of each subgroup, where $s$ is the number of activity classes. The quotient of the contacts from group $m$ and the total contacts is the proportion of contacts that occur like-with-unlike (random). This results in the interactions being spread uniformly across each disparate subgroup.

The force of infection in infectious disease epidemiology[44], denoted by $\lambda$, is defined as the rate at which susceptible individuals become infected by an infectious disease. $\lambda$ is often calculated in endemic states with homogeneous mixing and a rectangular population distribution. The force of infection in the HGAH model is defined in Eq. 13.

$$\lambda_{k,l} = c_i \beta_k \sum_m \left[ \rho_{im} \left( \frac{I_{km} + r V_{km}}{N_{km}} \right) \right]$$

Now, the force of infection is a logistic equation with $c_i$ as the number of contacts (partnerships) formed by activity class $l$, $\beta_k$ is the effective transmission rate from gender $k$
to gender $k'$, and $\sum_m$ is the population capable of transmitting the infection. The proportion of the population capable of transmitting infection is determined by the summation over all activity groups ($m$) of the opposite genders infectious population size, in activity class $m$, that mixes with every activity group according to $\rho$.

3.3.3. Compartments

The following section describes the ordinary differential equations (ODE) associated with each compartment. These equations represent continuous change in each sub-group population size.

**Susceptibles**

The susceptibles ODE is found in Eq. 6.

$$\frac{dS_{ki}}{dt} = .5\mu \omega_i(1 - \phi_{ki})\eta - (\lambda_{ki} + \mu)S_{ki} + \sigma V_{ki}$$

Equation 6 reads, the derivative of $S_{ki}$ with respect to time is the new susceptibles who age-in, the individuals whom become infectious or age-out, and those in which the vaccine has ceased to be effective. The new susceptibles are one half (approximate proportion of each gender) the product of individuals that age-in, at rate $\mu$, are not vaccinated (at rate $1 - \phi_{ki}$) and belong to activity class $l$ in the total sexually active population ($\omega_i\eta$). The negative change in this compartment occurs when susceptibles become infectious, with force of infection $\lambda_{ki}$, and when individuals age-out of the sexually active population, at rate $\mu$. The final change that occurs is due to the influx of individuals in which the vaccine has ceased to be effective, at a rate $\sigma$.

**Infectious**

The infectious ODE is found in Eq. 7.

$$\frac{dl_{ki}}{dt} = \lambda_{ki}S_{ki} - (\gamma + \mu)l_{ki}$$
Equation 7 reads, the derivative of $I_{kl}$ with respect to time is the new individuals who have become infectious and those whom are no longer infectious or age-out of the sexually active population. The new infectious are those susceptibles who acquire infection and become infectious at rate (force) $\lambda_{kl}$. The negative change occurs when infectious individuals clear the virus (at rate $\gamma$) and the population ages-out of this compartment, at rate $\mu$.

**Recovered or Immune**

The recovered/immune ODE is found in Eq. 8.

$$\frac{dR_{kl}}{dt} = \gamma I_{kl} + \alpha \gamma V I_{kl} - \mu R_{kl}$$  (8)

Equation 8 reads, the derivative of $R_{kl}$ with respect to time is the individuals who clear the infection, and those who age-out of the sexually active population. The new recovered individuals are those infectious individuals who clear the infection (at rate $\gamma$) and those vaccinated and infectious individuals who clear the infection at rate $\alpha \gamma$ where $\alpha$ is the relative recovery rate from breakthrough infection compared to natural infection.

**Vaccinated**

The vaccinated ODE is found in Eq. 9.

$$\frac{dV_{kl}}{dt} = .5 \mu \phi_{kl} \eta - (\mu + \sigma + \psi \lambda_{kl}) V_{kl}$$  (9)

Equation 9 reads, the derivative of $V_{kl}$ with respect to time is the new vaccinated who age-in, the individuals whom age-out, those in whom the vaccine ceases to be effective and the individuals who become infectious while vaccinated. The new vaccinated are one half (approximate proportion of each gender) the product of individuals that age-in, at rate $\mu$, are vaccinated (at rate $\phi_{kl}$) and belong to activity class $l$ in the total sexually active population ($\omega/\eta$). The negative population flow occurs when vaccinated individuals age-out (at rate $\mu$) and when the efficacious vaccine period expires (at rate $\sigma$); also, vaccinated individuals
become infectious at a rate $\psi \lambda$, where $\psi$ is the relative susceptibility of breakthrough infection compared to natural infection or degree and $\lambda$ is the rate of infection.

**Vaccinated and Infectious**

The vaccinated and infectious ODE is found in Eq. 10.

\[
\frac{dV_{I_{kl}}}{dt} = \psi \lambda_{ki} V_{ki} - (\mu + \alpha \gamma) V_{I_{kl}}
\]

Equation 10 reads, the derivative of $V_{I_{kl}}$ with respect to time is the new vaccinated and infectious individuals whom become infectious while vaccinated and those who exit the sexually active population or clear the infection. Vaccinated individuals become infectious at a rate $\psi \lambda$, where $\psi$ is the relative susceptibility of breakthrough infectious compared to natural infection. A proportion of the population exits this group by aging-out at a rate ($\mu$) and by clearing the infection at rate $\alpha \gamma$ where $\gamma$ is the recovery (clearance) rate and $\alpha$ is the relative recovery rate of breakthrough infection compared with natural infection.

3.4. Non-temporal Demographics

Initially, only partnership changes were used to classify the sub-groups; however, with slight modification the HGAK model can use other classifiers to sub-divide the population. The HGAK model allows for population stratification by demographic traits that are time-independent (e.g., race, ethnicity or income). Given the number of sexual activity classes ($x$) and the number of demographic sub-divisions ($y$), we can determine the total number of subgroups by the product of sexual activity classes and demographic sub-divisions ($l = x \cdot y$). Hence, the Hughes et al system of ODEs implements this new feature. The specific activity class can be determined by $l \mod x$ and similarly, the specific demographic group is found by $l \mod y$. This leads to a method for evaluating the effectiveness of vaccination policies on time-independent sub-populations. Next, to develop an efficient web-based interface to this
model we transpose the continuous HGAK model to a discrete-time model and also introduce a temporal model, that is capable of stratifying the population by time-dependent traits.
CHAPTER 4

TWO DISCRETE-TIME EPIDEMIC MODELS

4.1. Introduction to Discrete-Time Models

Discrete-time models or difference equations often describe various SI, SIR, and SIS models [45]. Ordinary differential equation (ODE) or continuous models are implemented more frequently due to their mathematical tractability. Consider the following two simple logistic equations (eqns. 11, 12), studied by R. May and also L. Olsen and W. Schaffer [46, 47].

\( x_{n+1} = (1 + r)x_n(1 - x_n) \)  \hspace{1cm} (11)

\( x_{n+1} = x_n \exp(r(1 - x_n)) \)  \hspace{1cm} (12)

In both these examples, the period doubles and the logistics eventually behave chaotically when the value of \( r \) increases above 2. Several discrete-time epidemic type models contain non-linear difference equations that are similar to the above logistic equations and are capable of producing chaotic behavior [45]. The infectives difference equations in the standard discrete-time SI and SIS models for sufficiently large time steps can behave chaotically. However, it has been shown that continuous (ODE) SIR models do not exhibit periodic behavior [48]. Similarly, the SIR difference equation form does not exhibit chaotic behavior and is there for tractable in most forms.

We extend the mathematical model by Hughes et al. to support multiple-demographic subgroups. This is critical for determining effective vaccination strategies. Surveys conducted by the Centers for Disease Control and Prevention Youth Risk Behavior Surveillance [6, 7] are incorporated in our demographically stratified model. These surveys state that reported
sexual activity varies by age and demographics. Our model illustrates temporal population flow and is described by a set of linear difference equations.

Our goal has been to provide public health officials a tool capable of qualitative-differential analysis of vaccination strategies. Providing a real-time application by solving for differential equations is computationally prohibitive; thus we introduce two discrete-time epidemic models that do not exhibit periodic or chaotic behavior. These two models also are computationally less intense and capable of real-time analysis.

4.2. Temporal Model

We have developed a discrete-time temporal model that is time-dependent and stratiﬁes the population via age. The temporal model predicts endemic prevalence of HPV by continually aging-in new susceptibles and aging-out a proportion \( v \) of each sexual activity class iteratively. Each class maintains a constant population size and the age range in each activity class is uniform. In our temporal model individuals \( \mu \) age-in to the model as either susceptible \( (1 - \phi) \) or vaccinated \( \phi \) individuals; however, individuals now will age-out of their current age stratum and age-in to the next contiguous age-stratum from their disease compartment at rate \( v \). Additionally a key concept in disease dynamics in the discrete-time models is the infectivity or force of infection (Equation 13) of HPV \( (\lambda_{ki}) \) of an infectious individual of gender \( k \) in activity group \( l \) to a susceptible of the opposite sex. We define this force of infection (infectivity) (Equation 13) as a binomial probability over the transmission risk in each demographic subgroup \( (\sum_{m}) \) as determined by the mixing parameter \( (\rho_{lm}) \).

The basis for interaction in the models is uniform mixing of heterosexual contacts. The age range of the sexually active population is modeled from 15 to 30; however, this range \( (1/\mu) \) can be varied depending on the demographics of a region. These models assume uniform mixing; however, the parameter, \( \epsilon \), determines the population proportion that interacts with individuals of the opposite sex in disparate demographic and sexual activity subgroups with the population interaction ranging from assortative to random \( (0 \leq \epsilon \leq 1) \). The parameter
\( \rho_{i,m} \) generates contacts occurring within an individuals activity group \((l)\) and the contacts \((N_{m}c_{m})\) made outside of the individuals activity group \((m)\).

\[
\lambda_{k,l} = 1 - \left( 1 - \beta_{k} \sum_{m} \left[ \rho_{i,m} \frac{l_{k,m} + rV_{l_{k,m}}}{N_{k,m}} \right] \right)^{c_{l}}
\]

Our temporal model combines demographic groups of different age ranges with sexual activity classes. Due to stratification of the population by age, individuals that \emph{age-out} of each demographic subgroup must \emph{age-in} to the next contiguous age group. Individuals in the last age group exit the sexually active population. Next, the following notation is used in our models: let \( d \) be the number of demographic groups, let \( n \) be the number of sexual activity classes in each demographic group, and let \( l \) be the total number of subgroups modeled \((d \cdot n)\). Figure 2(a) graphically describes population movement for the first demographic group \((n\) sub-groups); similarly, fig. 2(b) describes the population movement for the remaining demographic subgroups \((l - n)\) sub-groups.

The models allows arbitrary vaccination in segments of the population \((\phi_{k,l} \text{ with gender } k \text{ and sexual activity class } l)\) and vaccine coverage can be targeted in high-risk sub-groups or spread across the entire population, where high-risk is determined to be the sexual activity class with the highest average contact rate. The HPV vaccine efficacy in the general population is yet unknown; however, the model has encapsulated the needed functionality to vary the vaccine efficacy by the degree of vaccination \((\psi)\). The model also incorporates other characteristics, such as time-line until the vaccine ceases to be effective \((1/\sigma)\), relative risk of transmission from a vaccinated individual compared with an unvaccinated individual \((r)\) and infectious period of non-vaccinated \((1/\gamma)\) and vaccinated \((1/\alpha\gamma)\) individuals.

### 4.2.1. Temporal model description - first age stratum

The \emph{susceptibles} are described by the difference equations for the first \( n \) activity groups in Eq. 14 and remaining activity groups in Eq. 19.
(a) First age stratum (the first $n$ subgroups): The population enters this portion of the model in either the susceptible or vaccinated compartment and a proportion($v$) of each compartment will transfer to the next contiguous age stratum.

(b) The remaining age strata (remaining $l - n$ subgroups): The population enters the model from the previous age-stratum in to their corresponding compartment and exits to the next contiguous age stratum.
\begin{equation}
\Delta S_{kl} = .5\mu (1 - \phi_{kl}) \frac{\omega_i}{\sum_{i=1}^n \omega_i} \eta - (\lambda_{kl} + \nu) S_{kl} + \sigma V_{kl}
\end{equation}

The change in \textit{susceptible} individuals (\(\Delta S_{kl}\)) of gender \(k\) and activity group \(l\) for the first age group’s \(n\) activity classes in Eq. 14 contains three main components. Unvaccinated \((1-\phi)\) individuals in the first activity classes \((\omega_i/ \sum_n \omega_n)\) of the sexually active population become susceptible. Susceptibles also age-in to the next contiguous age-group \((\nu)\) and become infectious with infectivity \(\lambda\). Finally, vaccinated individuals in which the vaccine ceases to be effective become susceptible \((\sigma)\).

\begin{equation}
\Delta I_{kl} = \lambda_{kl} S_{kl} - (\gamma + \nu) I_{kl}
\end{equation}

The change in \textit{infectious} individuals (\(\Delta I_{kl}\)) of gender \(k\) and activity group \(l\) for the first age group’s \(n\) activity classes (Eq. 15) is derived from two components. The susceptible individuals who become infectious at a force of infection \(\lambda_{kl}\) enter this compartment. The individuals who leave the compartment, age-in to the next contiguous age group at rate \(\nu\) or they recover from the disease at rate \(\gamma\).

\begin{equation}
\Delta R_{kl} = \gamma I_{kl} + \alpha \gamma V_{kl} - \nu R_{kl}
\end{equation}

The change in \textit{recovered/immune} individuals (\(\Delta R_{kl}\)) of gender \(k\) and activity group \(l\) for the first age group’s \(n\) activity classes (Eq. 16) is determined by three components. Infectious individuals enter this compartment at the rate of recovery from the disease \(\gamma\); similarly, vaccinated and infectious individuals enter this compartment at the recovery rate \(\gamma\) modified by the difference in recovery time from natural versus breakthrough infection \((\alpha)\). Finally, recovered individuals leave this age stratum by moving into the next contiguous age group rate \(\nu\).
\[
\Delta V_{kl} = 0.5\mu \phi_{kl} \frac{\omega_i}{\sum_{i=1}^{n} \omega_i} \eta - (v + \sigma + \psi \lambda_{kl})V_{kl}
\]

The change in vaccinated individuals (\(\Delta V_{kl}\)) of gender \(k\) and activity group \(l\) for the first age group’s \(n\) activity classes (Eq. 17) is determined by three components. Individuals in the first activity classes (\(\omega_i/ \sum_n \omega_i\)) of the sexually active population enter the vaccinated compartment at rate \(\phi\). Vaccinated individuals leave this compartment via aging-in to the next contiguous age-group (\(\nu\)), and become infectious with infectivity \(\lambda\) modified by the effectiveness of the vaccine or vaccine degree \(\psi\). Finally, vaccinated individuals in which the vaccine ceases to be effective become susceptible (\(\sigma\)).

\[
\Delta V_{I_{kl}} = \psi \lambda_{kl} V_{kl} - (\nu + \alpha \gamma) V_{I_{kl}}
\]

The change in vaccinated and infectious individuals (\(\Delta V_{I_{kl}}\)) of gender \(k\) and activity group \(l\) for the first age group’s \(n\) activity classes (Eq. 18) is determined by the following components. Individuals become infectious after vaccination at a rate \(\psi \lambda_{kl}\) where \(\psi\) is the vaccine degree and \(\lambda\) is the force of infection. Individuals leave this compartment by aging in-to the next contiguous age group at rate \(\nu\) or recover from the disease at rate \(\alpha \gamma\) where \(\gamma\) is the recovery rate and \(\alpha\) is the effect of recovering from breakthrough infection as compared to recovery from natural infection.

4.2.2. Temporal model description - remaining age strata

The susceptible, infectious, removed, vaccinated and vaccinated-infectious individuals in the second set of equations that describe the remaining age-strata in the temporal model are quantified similarly. The infection dynamics remain the same; however, individuals in the \(l-n\) groups flow into the next contiguous age group and same activity class. The second set of equations in eh temporal model is completely described below.
\[\Delta S_{kl} = (1 - \phi_{kl})vS_{k(l-n)} - (\lambda_{kl} + v)S_{kl} + \sigma V_{kl}\]

The change in susceptible individuals \((\Delta S_{kl})\) for the remaining \(l - n\) activity groups are described in Eq. 19. A proportion of susceptibles who are unvaccinated \((1 - \phi)\) age-in \((v)\) from the preceding age group \((S_{k(l-n)})\). Similarly, susceptibles age-in to the next contiguous age-group \((v)\) and become infectious with infectivity \(\lambda\). Finally, vaccinated individuals in which the vaccine ceases to be effective become susceptible \((\sigma)\).

\[\Delta I_{kl} = vI_{k(l-n)} + \lambda_{kl}S_{kl} - (\gamma + v)I_{kl}\]

The change in infectious individuals \((\Delta I_{kl})\) for the remaining \(l - n\) activity groups are described in Eq. 20. Infectious individuals enter this compartment from the preceding infectious individual age-group \((v)\) or susceptible individuals in the current age group become infectious at a force of infection \(\lambda_{kl}\). Infectious individuals exit this compartment by recovering from the disease at rate \(\gamma\) and individuals exit by moving in to the next contiguous age-group or exiting the modeled population at rate \(v\).

\[\Delta R_{kl} = vR_{k(l-n)} + \gamma I_{kl} + \alpha \gamma VI_{kl} - vR_{kl}\]

The change in recovered/immune individuals \((\Delta R_{kl})\) for the remaining \(l - n\) activity groups are described in Eq. 21. Individuals enter this compartment by aging-in from the preceding recovered individuals age group at rate \(v\). Individuals also enter this compartment from infectious and vaccinated infectious individuals in the current age stratum at rates \(\gamma\) and
\( \alpha \gamma \) respectively. Ultimately, individuals exit this compartment and move in to the next age-stratum or exit the population at rate \( v \).

\begin{equation}
\Delta V_{kl} = vV_{k(l-n)} + \phi_{kl}vS_{k(l-n)} - (\nu + \sigma + \psi \lambda_{kl})V_{kl}
\end{equation}

The change in vaccinated individuals (\( \Delta V_{kl} \)) for the remaining \( l - n \) activity groups are described in Eq. 22. Individuals enter this compartment from the preceding age-stratum vaccinated compartment at rate \( v \). Individuals also enter this compartment through vaccination of susceptible individuals in the preceding age-stratum at rate \( v\phi_{kl} \). Individuals exit this compartment via vaccine duration (\( \sigma \)), becoming infectious (\( \psi \lambda \)) or move in to the next age-stratum or exit the population at rate \( v \).

\begin{equation}
\Delta VI_{kl} = vVI_{k(l-n)} + \psi \lambda_{kl}V_{kl} - (\nu + \alpha \gamma)VI_{kl}
\end{equation}

The change in vaccinated and infectious individuals (\( \Delta VI_{kl} \)) for the remaining \( l - n \) activity groups are described in Eq. 23. Individuals enter this compartment from the preceding age-stratum and enter via infection at rate \( \psi v \). Vaccinated infectious individuals exit this compartment by recovery from disease (\( \alpha \gamma \)) or via moving in to the next age-stratum or exit the population at rate \( v \).

4.3. Non-Temporal Model

The non-temporal model allows for population stratification by demographic traits that are time-independent (e.g., race, ethnicity or income). The non-temporal model is a special case of the temporal model, with the strata age-range equal to the age range of the population modeled; hence, \( v = 1 \). This model is described by a discretized form of the model developed by Hughes et al. with slight modification to the population compartments; specifically, each demographic stratum is divided into sexual activity classes (\( \omega \)) defined by contact rates (\( c \)).
\[ \Delta S_{kl} = .5\mu(1 - \phi_{kl})\eta - (\lambda_{kl} + \mu)S_{kl} + \sigma V_{kl} \]

The change in susceptible individuals (\(\Delta S_{kl}\), Eq. 24) of gender \(k\) and activity group \(l\) is the new susceptibles who age-in, the individuals who become infectious or age-out, and those in which the vaccine has ceased to be effective. The new susceptibles are one half (approximate proportion of each gender) the product of individuals that age-in, at rate \(\mu\), are not vaccinated (at rate \(1 - \phi_{kl}\)) and belong to activity class \(l\) in the total sexually active population (\(\omega/\eta\)). The negative change in this compartment occurs when susceptibles become infectious, with force of infection \(\lambda_{kl}\), and when individuals age-out of the sexually active population, at rate \(\mu\). The final change that occurs is due to the influx of individuals in which the vaccine has ceased to be effective, at a rate \(\sigma\).

\[ \Delta I_{kl} = \lambda_{kl}S_{kl} - (\gamma + \mu)I_{kl} \]

The change in infectious individuals (\(\Delta I_{kl}\), Eq. 25) of gender \(k\) and activity group \(l\) is the new individuals who have become infectious and those who are no longer infectious or age-out of the sexually active population. The new infectious are those susceptibles who acquire infection and become infectious at rate (force) \(\lambda_{kl}\). The negative change occurs when infectious individuals clear the virus (at rate \(\gamma\)) and the population ages-out of this compartment, at rate \(\mu\).

\[ \Delta R_{kl} = \gamma I_{kl} + \alpha\gamma V_{kl} - \mu R_{kl} \]

The change in recovered/immune individuals (\(\Delta R_{kl}\), Eq. 26) of gender \(k\) and activity group \(l\) is the individuals who clear the infection, and those who age-out of the sexually active population. The new recovered individuals are those infectious individuals who clear
the infection (at rate $\gamma$) and those vaccinated and infectious individuals who clear the infection at rate $\alpha \gamma$ where $\alpha$ is the relative recovery rate from breakthrough infection compared to natural infection.

\begin{equation}
\Delta V_{kl} = 0.5 \mu \phi_{ki} \eta - (\mu + \sigma + \psi \lambda_{ki}) V_{kl}
\end{equation}

The change in vaccinated individuals ($\Delta V_{kl}$, Eq. 27) of gender $k$ and activity group $l$ is the new vaccinated who age-in, the individuals who age-out, those in who the vaccine ceases to be effective and the individuals who become infectious while vaccinated. The new vaccinated are one half (approximate proportion of each gender) the product of individuals that age-in, at rate $\mu$, are vaccinated (at rate $\phi_{ki}$) and belong to activity class $l$ in the total sexually active population ($\omega_i \eta$). The negative population flow occurs when vaccinated individuals age-out (at rate $\mu$) and when the efficacious vaccine period expires (at rate $\sigma$); also, vaccinated individuals become infectious at a rate $\psi \lambda$, where $\psi$ is the relative susceptibility of breakthrough infection compared to natural infection or degree and $\lambda$ is the rate of infection.

\begin{equation}
\Delta V_{kl} = \psi \lambda_{ki} V_{kl} - (\mu + \alpha \gamma) V_{kl}
\end{equation}

The change in vaccinated and infectious individuals ($\Delta V_{kl}$, Eq. 28) of gender $k$ and activity group $l$ is the new vaccinated and infectious individuals who become infectious while vaccinated and those who exit the sexually active population or clear the infection. Vaccinated individuals become infectious at a rate $\psi \lambda$, where $\psi$ is the relative susceptibility of breakthrough infectious compared to natural infection. A proportion of the population exits this group by aging-out at a rate ($\mu$) and by clearing the infection at rate $\alpha \gamma$, where $\gamma$ is the recovery (clearance) rate and $\alpha$ is the relative recovery rate of breakthrough infection compared with natural infection.
4.4. Sensitivity Analysis

Simple disease transmission models can be solved analytically; however, understanding the behavior of more interesting and complex models require the use of techniques in numerical analysis. Two types of analysis performed on complex epidemic models include uncertainty and sensitivity analysis. The structural complexity, coupled with a high degree of uncertainty associated with choosing appropriate input parameters necessitate the use of model sensitivity and uncertainty analysis [49]. Uncertainty analysis are useful to determine the prediction imprecision in the outcome variable due to estimating the input parameters. Sensitivity analysis will assign a numerical value relating to the impact an input parameter has on an output variable. A sensitivity analysis can identify which input parameters are important in contributing to the prediction imprecision of the outcome variable; hence, a sensitivity analysis evaluates the effect changes in the value of input parameters have on the outcome variable[50]. Due to the limited availability of clinical trial data, this thesis will only perform sensitivity analysis on temporal model and leave the uncertainty analysis for future work.

The temporal model can be thought of as a vector function:

\[
(29) \quad y = y(x)
\]

where

\[
(30) \quad x = [\epsilon, \phi, \sigma, \psi, u]
\]

and

\[
(31) \quad y = [\text{Infectious, Vaccinated\\& Infectious, Total\text{Infectious}]
\]

35
with sensitivity analysis, we will answer the question "How important are the individual elements of \( \mathbf{x} \) with respect to the uncertainty of \( \mathbf{y}(\mathbf{x}) \)" [51]. Due to the computational efficiency of the model, a nearly full factorial sampling of the input parameter space was chosen for our sensitivity analysis. Sampling-based methods for sensitivity analysis are based on a sample

\[
\mathbf{x}_k = [\epsilon_k, \phi_k, \sigma_k, \psi_k, u_k], \; k = 1, 2, \ldots, nS,
\]

of size \( nS \) from the possible values for \( \mathbf{x} \) characterized by a uniform distribution in this analysis and on the corresponding evaluations

\[
\mathbf{y}(\mathbf{x}_k) = [y_1(\mathbf{x}_k), y_2(\mathbf{x}_k), \ldots, y_{nY}(\mathbf{x}_k)], \; k = 1, 2, \ldots, nS,
\]

of \( \mathbf{y} \). The pairs

\[
[\mathbf{x}_k, \mathbf{y}(\mathbf{x}_k)], \; k = 1, 2, \ldots, nS,
\]

define a map containing the vector of uncertain inputs (\( \mathbf{x}_k \)) to the corresponding uncertain analysis results (\( \mathbf{y}_k \)).

Correlation coefficients provide a formal and basic description of the relationship between input parameters and model outcomes. The sample or Pearson correlation \( r_{x_i,y} \) between \( x_i \) and \( y \) for the sequence of outcomes in eqn. 34 is defined by

\[
r_{x_i,y} = \frac{\sum_{k=1}^{nS}(x_{ki} - \bar{x}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{nS}(x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{nS}(y_k - \bar{y})^2}}
\]

where
\begin{equation}
\mathcal{V} = \frac{\sum_{k=1}^{nS} y_k}{nS} \quad \mathcal{X} = \frac{\sum_{k=1}^{nS} x_{ki}}{nS}
\end{equation}

The correlation coefficient evaluates to a value between -1.0 and 1.0 and provides a measure of the linear relationship between \( x_i \) and \( y \) with 0 meaning no correlation. The correlation coefficient measures the effect of one variable at a time on \( y \) and does not consider for the possible effects of on the outcome variables of other uncertain variables [51]. However, we need to evaluate the effect of a single parameter regardless of the synergistic or antagonistic effects of other inputs; hence, we use the partial correlation coefficient (PCC). The partial correlation coefficient provides a measure of the linear relationship between two variables after a correction has been made to remove the linear effects of all other variables in the analysis. The partial correlation coefficient between an input parameter \( x_i \) and outcome variable \( y \) is obtained from the use of a sequence of regression models. First, the following two regression models are constructed:

\begin{equation}
\bar{x}_i = c_0 + \sum_{p=1}^{nX} c_p x_p \quad \bar{y} = b_0 + \sum_{p=1}^{nX} b_p x_p
\end{equation}

Now, the partial correlation coefficient \( p_{x_i y} \) between \( x_i \) and \( y \) is the Pearson Correlation (eqn. 35) between \( x_i - \bar{x}_i \) and \( y - \bar{y} \). Hence, the partial correlation coefficient provides an evaluation metric for the strength of linear relationships between \( x_i \) and \( y \) with all other variables removed [51].

The sensitivity analysis in the thesis is implemented with partial correlation coefficients. A prerequisite for calculating partial correlation coefficients is the model input parameters have a linear relationship to one another and they be monotonic. The temporal model is monotonic and has linear relationships among the input parameters, this is validated by the lack of periodicity in the model. Scatter plots can also be used to detect anomalies by plotting input parameters against outcome variables, and can be shown by a scatter plot of
the evaluations [50, 49]. The input parameter sampling was chosen as follows: 20 values uniformly distributed from values in a reasonable range of five parameters ($\epsilon, \phi, \sigma, \psi, \nu$) were sampled to compose $\mathbf{x}$, the model was then executed 3,200,000 ($5^{20}$) times, evaluating a set of outcome variables ($\mathbf{y}(\mathbf{x})$) each iteration. The partial correlation coefficient is then computed from the vector spaces of $\mathbf{x}$ and $\mathbf{y}(\mathbf{x})$.

The partial correlation coefficient describes the effect of one of the five input parameters on the prevalence of infectious individuals, vaccinated and infectious individuals and the total HPV prevalence in the population. Table 4.1 displays the resulting PCCs; the results demonstrate the model operates within acceptable limits. The infectious PCC of $\phi$ evaluates to -0.776, meaning with increased vaccination the prevalence of infectious individuals decreases. Interestingly, the vaccinated infectious PCC of $\phi$ is 0.389, meaning with increased vaccination coverage there is also an increase in the number of vaccinated and infectious individuals; this increase is due to a greater vaccinated population pool hence, more people will have breakthrough infection. Overall, when the level of vaccination is increased, the total infectious population prevalence decreases, demonstrated by a PCC of -0.601. Analysis of $\nu$ demonstrates an artifact of this model, that with an increased granularity of population strata, the evaluated endemic prevalence decreases, due to this artifact combining temporally and non-variably dependent demographics would infeasible.
Table 4.1. Temporal model sensitivity analysis.

<table>
<thead>
<tr>
<th></th>
<th>Infectious</th>
<th>Vaccinated</th>
<th>Total Infectious</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>0.350</td>
<td>0.110</td>
<td>0.321</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-0.776</td>
<td>0.389</td>
<td>-0.601</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.448</td>
<td>-0.282</td>
<td>0.241</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.162</td>
<td>0.597</td>
<td>0.427</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-0.771</td>
<td>0.134</td>
<td>-0.656</td>
</tr>
</tbody>
</table>
CHAPTER 5

POPULATION-LEVEL IMPACT : MODELING PREVALENCE THROUGH DEMOGRAPHICS

5.1. Integrating Demographics

Through demographic population stratification and analysis of risk behavior surveys, effective vaccination strategies can be defined for a given geographic area. Demographic analysis is critical for appropriate and effective policies in regions with different demographics, including high-risk communities such as youth or minority groups. The model developed by Hughes et al stratifies the population into sexual activity classes; however, it is lacking a demographic component which our model(s) add. We incorporate different demographic parameters that are critical to determine the most effective vaccination strategy for a population. Population demographics from 6 U.S. counties selected for their demographic variance, obtained from the U.S. census in 2000 are used in our analysis. County abbreviations are found in Table 5.1 and Table 5.2 describes population demographics by age and race.

Table 5.1. U.S. county abbreviations

<table>
<thead>
<tr>
<th>County</th>
<th>Abbrev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orange County, CA</td>
<td>OCA</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>WDC</td>
</tr>
<tr>
<td>Miami-Dade County, FL</td>
<td>MFL</td>
</tr>
<tr>
<td>Fulton County, GA</td>
<td>FGA</td>
</tr>
<tr>
<td>Polk County, IA</td>
<td>PIA</td>
</tr>
<tr>
<td>Denton County, TX</td>
<td>DTX</td>
</tr>
</tbody>
</table>
### Table 5.2. U.S. census data for select U.S. counties

<table>
<thead>
<tr>
<th></th>
<th>OCA</th>
<th>WDC</th>
<th>MFL</th>
<th>FGA</th>
<th>PIA</th>
<th>DTC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population by age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-19</td>
<td>Males</td>
<td>0.157</td>
<td>0.124</td>
<td>0.159</td>
<td>0.139</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.158</td>
<td>0.140</td>
<td>0.171</td>
<td>0.136</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.314</td>
<td>0.264</td>
<td>0.329</td>
<td>0.276</td>
<td>0.314</td>
</tr>
<tr>
<td>20-24</td>
<td>Males</td>
<td>0.166</td>
<td>0.167</td>
<td>0.161</td>
<td>0.167</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.142</td>
<td>0.196</td>
<td>0.143</td>
<td>0.158</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.308</td>
<td>0.363</td>
<td>0.304</td>
<td>0.325</td>
<td>0.319</td>
</tr>
<tr>
<td>25-29</td>
<td>Males</td>
<td>0.194</td>
<td>0.178</td>
<td>0.183</td>
<td>0.205</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.184</td>
<td>0.195</td>
<td>0.184</td>
<td>0.194</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.378</td>
<td>0.373</td>
<td>0.367</td>
<td>0.399</td>
<td>0.367</td>
</tr>
<tr>
<td><strong>Population by race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Males</td>
<td>0.197</td>
<td>0.162</td>
<td>0.087</td>
<td>0.201</td>
<td>0.402</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.192</td>
<td>0.187</td>
<td>0.082</td>
<td>0.187</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.388</td>
<td>0.349</td>
<td>0.170</td>
<td>0.388</td>
<td>0.829</td>
</tr>
<tr>
<td>Hispanic</td>
<td>Males</td>
<td>0.219</td>
<td>0.057</td>
<td>0.289</td>
<td>0.066</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.192</td>
<td>0.045</td>
<td>0.279</td>
<td>0.036</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.411</td>
<td>0.102</td>
<td>0.568</td>
<td>0.101</td>
<td>0.065</td>
</tr>
<tr>
<td>African-American</td>
<td>Males</td>
<td>0.009</td>
<td>0.220</td>
<td>0.108</td>
<td>0.215</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.007</td>
<td>0.260</td>
<td>0.118</td>
<td>0.239</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.016</td>
<td>0.480</td>
<td>0.226</td>
<td>0.454</td>
<td>0.050</td>
</tr>
<tr>
<td>Other</td>
<td>Males</td>
<td>0.092</td>
<td>0.030</td>
<td>0.018</td>
<td>0.030</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0.093</td>
<td>0.039</td>
<td>0.018</td>
<td>0.027</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.185</td>
<td>0.069</td>
<td>0.036</td>
<td>0.057</td>
<td>0.055</td>
</tr>
</tbody>
</table>

### 5.1.1. Sexual Mixing Patterns

Fusing multiple-demographic subgroups with sexual activity classes is critical for determining effective vaccination strategies. Two surveys conducted by the Centers for Disease
Control and Prevention: Youth Risk Behavior Surveillance and College Risk Behavior Surveillance [6, 7] are incorporated in our demographically stratified model. These surveys state that reported sexual activity varies by age and demographics. We consider sexual mixing characteristics in varied demographic groups, in addition to the sexual activity variance across the entire population provided by the Hughes et al. model and the contact rates are based upon contact tracing from a particular disease outbreak [42]. Additionally, we claim that sexual activity rates vary by age and by race [52, 6]. Our model(s) derive contact rates from data obtained from the Youth Risk Behavior Surveillance: National College Health Risk Behavior Survey [6] indicate students aged greater than or equal to 25 years (97.8%) were more likely to report sexual intercourse in their lifetime than students aged 18–24 years (79.9%). African-American students (92.8%) were more likely than White (86.7%) and Hispanic (85.2%) students to report sexual activity. Sexual activity classes for our models were created artificially to reflect this data. The sexual activity class proportions in the demographic groups are 3%, 15% and 82% correspond to high, moderate and low partner change rates. The contact rates used in our analysis are listed in Table 3 by age and race. It should be noted that the contact rates for each demographic stratum are created artificially; however, the variance in contact rates is in line with published reports [53, 6].

5.2. Tool Interface

We have created an application interface to the HGAK and temporal models. Our application interface is a tool to facilitate quantification of the effectiveness of different vaccination policies. This tool, illustrated in Figure 5.1 and Figure 5.2, accepts the necessary vaccine, disease and demographic parameters of HPV to predict the endemic prevalence in a demographically stratified population. This tool is designed for epidemiologists to determine the effect of a vaccination policy, given a particular demographic makeup. The parameter input is intuitive and resulting model output simplifies the tool operation and increases its accessibility. The application interface is publicly available on the Internet, providing a central application
(a) Web-based interface to models. The user can manually input parameters or use the configuration wizard.

(b) Configuration wizard, the user can input disease, vaccine and population parameters
Figure 5.2. HPV tool interface

(a) Configuration wizard, the user can enter demographic strata proportions, contact rates and arbitrary vaccine coverage

(b) Results from the web-based interface include level of endemic prevalence and plots over the duration of the evaluation.
Table 5.3. Sexual activity group contact rates per demographic stratum

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion in each demographic stratum</td>
<td>0.03</td>
<td>0.15</td>
<td>0.82</td>
</tr>
<tr>
<td>Contact Rates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ages 15-19</td>
<td>8</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>ages 20-24</td>
<td>9</td>
<td>3</td>
<td>1.25</td>
</tr>
<tr>
<td>ages 25-29</td>
<td>9.5</td>
<td>3.5</td>
<td>1.5</td>
</tr>
<tr>
<td>White</td>
<td>9</td>
<td>3.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.25</td>
<td>3</td>
<td>1.25</td>
</tr>
<tr>
<td>African-Am</td>
<td>10</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>3</td>
<td>1.4</td>
</tr>
</tbody>
</table>

location and availability to users in disparate locations. The front-end to the application consists of an internet portal to the models coded in Perl/CGI and the back-end is composed of the computational application is written in C++. A visualization of the model output can be found in Figure 5.3, this figure illustrates the model output in Denton county, TX. The x-axis is the HPV prevalence in Denton county, TX of the susceptible, infectious, vaccinated, vaccinated infectious, and recovered population over time (y-axis) in the temporal model, with no vaccine coverage (Fig. 5.3(a),5.3(b)) and 90% vaccine coverage (Fig. 5.3(c),5.3(d)).

Our tool allows the user to choose to stratify the population (Figure 5.1(a)) by age or other demographic and to define the number of demographic groups and sexual activity subgroups. Sexual activity subgroup proportions can be predefined (0.03, 0.15 and 0.82) for each group or the user can define the proportions. The user can choose to input model parameters manually with a configuration file or use the configuration wizard we provide. The configuration wizard (Figures 5.1 and 5.2a) displays the model variables the user can change,
such as the population count, transmission risk, duration of inclusion in the sexually active population, vaccine efficacy, and et cetera. Further, it allows the user to input information about contact/partner change rates and proportions of each demographic and sexual activity subgroup (Figure 5.2(a)). The key component of the configuration wizard is the ability for the user to specify the vaccine coverage for each demographic and sexual activity subgroup which allows for evaluation of a demographically tailored vaccination strategy. Our tool produces (Figure 5.2(b)) the endemic prevalence and temporal flow of the population in each state.
for the chosen demographic stratification and allows quantification of the effectiveness of
different vaccination strategies.

5.3. Results

To evaluate the population-level impact of a specific vaccination policy we measure the
relative reduction in endemic prevalence. Endemic prevalence with no vaccination coverage
is the baseline to measure the impact of vaccination and we define the relative reduction in
prevalence ($\pi_p$) as change in prevalence due to a vaccine policy ($p$) compared with the baseline
prevalence ($\theta_0$) (Equation 38). Evaluations are performed in the temporal and non-temporal
demographic settings, both experimental settings utilize the same disease and vaccine param-
eters and a demographic 1 composition that corresponds to a particular county in the US. In
each model evaluation, the predicted endemic prevalence is measured and compared and the
potential impact on endemic prevalence by targeting vaccine coverage in certain sub-groups
is analyzed.

\[
\pi_p = \frac{\theta_0 - \theta_p}{\theta_0}
\]

(38)

We chose to demonstrate the reduction in prevalence in six U.S. counties listed in Table 5.1
and their corresponding demographic composition, obtained from the 2000 U.S. Census, is
displayed in Table 5.2. Disparate vaccination practices are implemented for the two models.
The temporal model analysis considers vaccination of specific age-strata and gender; whereas,
the non-temporal model implements a policy of vaccinating the stratum with the highest
average contact rate, in these evaluations it was African-Americans; however, this group
could be set as any arbitrary demographic stratum. The endemic prevalence with no vaccine

---

1Population demographics are obtained from the U.S. Census data in 2000 and the disease and vaccine
parameters are the initial values in the work by Hughes et al. [5]
and the relative reduction in prevalence of various arbitrary vaccination strategies are shown in Tables 5.4 and 5.5.

5.4. Discussion

The non-temporal model (Table 4) evaluates a baseline endemic prevalence\(^2\) of close to 0.038 across the select counties; however, Washington D.C. and Fulton county, GA have a slightly greater endemic prevalence due to their larger composition of high-risk groups, approximately 0.48 and 0.454 respectively. When vaccinating 90\% of the entire population, a difference in the relative reduction in prevalence is present in the counties presented; such as the 6\% difference between Orange County California and Washington D.C. Table 4 also demonstrates the impact of targeting vaccination with 90\% coverage of both males and females in high-risk groups and spread targeting (90\% coverage in high-risk stratum and 10\% in all remaining strata) males and females. Additionally, there is a clear difference in impact from a vaccine targeted at high-risk groups in Fulton County, GA compared to Orange County, CA with the reduction in Fulton County 22 times greater than in Orange County.

The temporal model (Table 5) evaluates an endemic prevalence of 0.046 and 0.049 for males and females, respectively across all the counties. The low variance in endemic prevalence for the evaluations is due to the homogeneous age-strata compositions in the selected counties and the low variance in contact rates per age stratum. The combined effects of these two factors limit the endemic prevalence variability in the temporal model. Measuring the effectiveness of vaccine coverage on a specific age group illustrates a beneficial age to begin vaccination. Table 5 also demonstrates vaccinating females, ages 15-19 (25.5\%) has lower impact than targeting all females in the population (37\%); however, this is still a significant reduction in prevalence and starting vaccination at an earlier age will have the greatest impact with a prevalence reduction compared to vaccinating at a later age. Both

\(^2\)The baseline endemic prevalence for the two models vary due to evaluations with different demographic parameters and contact rates.
experimental settings demonstrated that female vaccination nearly doubled the reduction in prevalence compared to men. Although full vaccine coverage in the population would be desirable, costs associated with the vaccine makes it infeasible [3]. Table 5 demonstrates that vaccinating only females reduces the prevalence by 47% which is competitive to the impact on prevalence by vaccinating males and females ages 15-19. Clearly, these two policies will have the same impact in the population; however, vaccinating a particular age stratum is more cost-effective than blanketing an entire gender, this is due to the difference in demographic population levels.

Also, the baseline endemic prevalence varies in each experimental setting due to the disparate contact rates and demographic proportions. Incorporating a temporal flow with age-stratification clearly improves the reduction in prevalence for all policies evaluated compared to the non-temporal evaluations. Tables 4 and 5 demonstrate a disparity among reductions across settings; specifically, the temporal model evaluated much higher reductions in prevalence due to the greater demographic proportion in its target vaccination stratum.

5.4.1. Analysis of Denton County, TX

The following describes the effects of several intervention strategies in Denton County, Texas. In our county, high-risk groups account for 7.0% of the population, when demographically stratified by race. The effect on prevalence of targeting vaccination of both males and females in high-risk groups (π = 3.26%) is significantly lower than vaccination strategies targeting the total population (π = 42.41%) and vaccine coverage targeting females in high-risk groups (π = 1.72%) is significantly less effective than targeting all females (π = 21.01%).

The temporal model evaluates an endemic prevalence of 4.6% and 4.9% for males and females, respectively. Vaccine coverage targeting both males and females, ages 15-19 (π = 46.28%) is less effective than targeting the total population (π = 68.7%). Vaccinating females, ages 15-19 (π = 25.25%) has lower impact than targeting all females in the population (π = 37.31%).
Table 5.4. Temporal model: endemic prevalence and relative reduction in prevalence per vaccine policy

<table>
<thead>
<tr>
<th>Vaccine Period</th>
<th>Males (Prevalence)</th>
<th>Females (Prevalence)</th>
<th>Total (Prevalence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Vaccine*</td>
<td>0.0463</td>
<td>0.0487</td>
<td>0.0475</td>
</tr>
<tr>
<td>Vaccinate M&amp;F</td>
<td>0.6885</td>
<td>0.6850</td>
<td>0.6869</td>
</tr>
<tr>
<td>Vaccinate F</td>
<td>0.2767</td>
<td>0.4650</td>
<td>0.3731</td>
</tr>
<tr>
<td>Vaccinate 15-19 M&amp;F</td>
<td>0.4644</td>
<td>0.4606</td>
<td>0.4625</td>
</tr>
<tr>
<td>Vaccinate 15-19 F</td>
<td>0.1808</td>
<td>0.3215</td>
<td>0.2527</td>
</tr>
<tr>
<td>Vaccinate 20-24 M&amp;F</td>
<td>0.3666</td>
<td>0.3627</td>
<td>0.3646</td>
</tr>
<tr>
<td>Vaccinate 20-24 F</td>
<td>0.1381</td>
<td>0.2497</td>
<td>0.1953</td>
</tr>
<tr>
<td>Vaccinate 25-29 M&amp;F</td>
<td>0.1744</td>
<td>0.1717</td>
<td>0.1730</td>
</tr>
<tr>
<td>Vaccinate 25-29 F</td>
<td>0.0629</td>
<td>0.1225</td>
<td>0.0934</td>
</tr>
</tbody>
</table>

* The low variance in prevalence is due to the composite age uniformity in select counties and the low variance in contact rate among the age strata.
Table 5.5. Non-temporal model: endemic prevalence and relative reduction in prevalence per vaccine policy

<table>
<thead>
<tr>
<th></th>
<th>OCA</th>
<th>WDC</th>
<th>MFL</th>
<th>FGA</th>
<th>PIA</th>
<th>DTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence – No Vaccine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.0369</td>
<td>0.0394</td>
<td>0.0377</td>
<td>0.0390</td>
<td>0.0369</td>
<td>0.0367</td>
</tr>
<tr>
<td>Females</td>
<td>0.0389</td>
<td>0.0411</td>
<td>0.0396</td>
<td>0.0410</td>
<td>0.0386</td>
<td>0.0385</td>
</tr>
<tr>
<td>Total</td>
<td>0.0379</td>
<td>0.0402</td>
<td>0.0387</td>
<td>0.0400</td>
<td>0.0377</td>
<td>0.0376</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Reduction in Prevalence per Vaccine Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccinate M&amp;F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.4486</td>
<td>0.3917</td>
<td>0.4271</td>
<td>0.3921</td>
<td>0.4278</td>
<td>0.4262</td>
</tr>
<tr>
<td>Females</td>
<td>0.4457</td>
<td>0.3855</td>
<td>0.4233</td>
<td>0.3887</td>
<td>0.4233</td>
<td>0.4221</td>
</tr>
<tr>
<td>Total</td>
<td>0.4471</td>
<td>0.3885</td>
<td>0.4252</td>
<td>0.3904</td>
<td>0.4255</td>
<td>0.4241</td>
</tr>
<tr>
<td>Vaccinate F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.1463</td>
<td>0.1337</td>
<td>0.1408</td>
<td>0.1330</td>
<td>0.1424</td>
<td>0.1399</td>
</tr>
<tr>
<td>Females</td>
<td>0.2835</td>
<td>0.2715</td>
<td>0.2788</td>
<td>0.2725</td>
<td>0.2794</td>
<td>0.2768</td>
</tr>
<tr>
<td>Total</td>
<td>0.2165</td>
<td>0.2043</td>
<td>0.2114</td>
<td>0.2043</td>
<td>0.2126</td>
<td>0.2101</td>
</tr>
<tr>
<td>High Risk* M&amp;F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.0091</td>
<td>0.2008</td>
<td>0.1018</td>
<td>0.1895</td>
<td>0.0239</td>
<td>0.0327</td>
</tr>
<tr>
<td>Females</td>
<td>0.0090</td>
<td>0.1970</td>
<td>0.1004</td>
<td>0.1874</td>
<td>0.0235</td>
<td>0.0324</td>
</tr>
<tr>
<td>Total</td>
<td>0.0091</td>
<td>0.1989</td>
<td>0.1011</td>
<td>0.1884</td>
<td>0.0237</td>
<td>0.0326</td>
</tr>
<tr>
<td>High Risk* F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.0031</td>
<td>0.0722</td>
<td>0.0358</td>
<td>0.0679</td>
<td>0.0085</td>
<td>0.0114</td>
</tr>
<tr>
<td>Females</td>
<td>0.0063</td>
<td>0.1430</td>
<td>0.0711</td>
<td>0.1355</td>
<td>0.0164</td>
<td>0.0226</td>
</tr>
<tr>
<td>Total</td>
<td>0.0048</td>
<td>0.1085</td>
<td>0.0538</td>
<td>0.1025</td>
<td>0.0125</td>
<td>0.0172</td>
</tr>
<tr>
<td>Spread Targeting M&amp;F †</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.0548</td>
<td>0.2213</td>
<td>0.1357</td>
<td>0.2112</td>
<td>0.0687</td>
<td>0.0754</td>
</tr>
<tr>
<td>Females</td>
<td>0.0542</td>
<td>0.2171</td>
<td>0.1339</td>
<td>0.2090</td>
<td>0.0677</td>
<td>0.0746</td>
</tr>
<tr>
<td>Total</td>
<td>0.0545</td>
<td>0.2191</td>
<td>0.1347</td>
<td>0.2101</td>
<td>0.0682</td>
<td>0.0751</td>
</tr>
<tr>
<td>Spread Targeting F †</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>0.0178</td>
<td>0.0790</td>
<td>0.0469</td>
<td>0.0752</td>
<td>0.0233</td>
<td>0.0254</td>
</tr>
<tr>
<td>Females</td>
<td>0.0376</td>
<td>0.1578</td>
<td>0.0946</td>
<td>0.1513</td>
<td>0.0471</td>
<td>0.0435</td>
</tr>
<tr>
<td>Total</td>
<td>0.0279</td>
<td>0.1194</td>
<td>0.0713</td>
<td>0.1141</td>
<td>0.0355</td>
<td>0.0392</td>
</tr>
</tbody>
</table>

* High-risk group is defined as the demographic stratum with the highest average contact rate. In these evaluations it was the African-American stratum with vaccine coverage of 90%.

†Spread targeting implements a 90% vaccine coverage with the high-risk stratum and 10% coverage in the remaining strata.
Through analysis of predictive endemic prevalence of HPV types in Denton County, we show that targeting vaccination at age group 15-19 will have a 46.28% decrease in the prevalence compared to targeting high-risk minority groups with a decrease in prevalence of 1.72%. It should be noted that the contact rates for each group are created artificially; however, the variance in contact rates is in line with published reports [53, 6].
CHAPTER 6

CONCLUSION

6.1. Discussion

The discrete-time models and the web-based application focus on predicting HPV transmission in heterogeneous populations and measures the effectiveness of a HPV vaccine. Both models vaccinate target demographic subgroups and simulate assortative population interaction where the non-temporal model incorporates time-independent demographics such as race, ethnicity or income and our temporal model stratifies a population by time-dependent demographics such as age. This thesis proves that applying a uniform vaccination policy across disparate demographic regions would have varying impact on the prevalence of a disease. Tailoring vaccination policy and information dissemination should vary by the demographic composition of a community. We provide this tool for epidemiologists to define solutions with the greatest impact on prevention of HPV in a community.

By modeling multiple local regions incorporating sub-population demographics, the cost-effectiveness of a HPV vaccine can be increased. Simulating these scenarios is critical for determining appropriate and effective policies and what-if analysis. Information available such as risk-behavior surveys for adults and youth and demographic data are processed by our model to prioritize demographic subgroups for vaccination. Targeted vaccination strategies reduce the incidence and prevalence of cervical disease, thereby enhancing the health of the general population.

6.2. Limitations and Future work

Given the broad application of these epidemic-type models, there are several components that can be expanded and improved upon to provide a greater variability in their use.
The current version of the web-interface requires the user manually enter values for levels of vaccination and mandates the user ultimately determine through implementing vaccination levels by a trial-and-error method to classify the “best” vaccination solution. To automate this process, use of evolutionary learning algorithms could be implemented to evaluate a strategy that will obtain a predetermined goal; such as, maximizing reduction in prevalence in the entire population.

A subset of the above improvement could be to not only maximize population-level impact of a vaccine but to also minimize costs associated with the marketing and vaccine dissemination. For example, a state government allocates two million dollars for a vaccination campaign what strategy should be formed for a given region? Future work could include handling situations such as this by assigning a cost/benefit value to each person/vaccine and use the cost/benefit analysis to maximize the population-level impact and minimize costs associated with not only vaccination but marketing. This thesis evaluates impact of an intervention on endemic prevalence; however, to properly forecast the economic impact of a human papilloma virus vaccine, the models need to include the development of carcinogenesis from infection

Upon the near future conclusion of FDA drug trials, a greater amount of clinical data will be available that can be used to decrease uncertainty in parameter estimation. Given a higher precision in parameter estimation, the cost-effectiveness and population benefit of viable vaccine can be improved. Also, the current form of the models only consider a mono-valent intervention; however, the vaccines in development are bi-valent and quad-valent. The models defined in this thesis need to be enhanced to include the impact and intervention of multiple virus strains.

One characteristic of the models presented in this thesis is the assumption of a male:female ratio of 1:1; however, this is not completely accurate and a separate demographic parameter can be introduced to account for the region specific gender ratio.
BIBLIOGRAPHY


