Agent-Based Modeling of HPV Transmission and Vaccination Impact

Anish Arora*
Georgia Institute of Technology
Atlanta, GA, USA
aarora340@gatech.edu

Hailey Toeppner
Georgia Institute of Technology
Atlanta, GA, USA
htoeppner3@gatech.edu

Scott Watanuki
Georgia Institute of Technology
Atlanta, GA, USA
scottwatanuki@gatech.edu

Abstract

We propose an agent-based model (ABM) of Human Papillomavirus (HPV) transmission tailored to the Atlanta Metropolitan Area to evaluate the impact of vaccination strategies on infection dynamics. Building on the open-source HPVsim framework, our model will simulate individual-level demographics, sexual partnerships, vaccination at routine and catch-up ages, per-partnership transmission, and natural clearance. By integrating local demographic and prevalence data from the U.S. Census and CDC, we will calibrate the model to reflect Atlanta-specific population structure and infection patterns. Through scenario analyses varying vaccine coverage and efficacy, we will estimate reductions in HPV prevalence and incident infections as proximal outcomes and discuss their implications for long-term cervical cancer prevention in the region. Deliverables include a reproducible codebase, scenario results, and a report highlighting policy-relevant insights for public health planning.

Keywords

Epidemiology, Agent-Based Modeling, HPV, Vaccination, STI, Network Models, Public Health

1 Introduction

Motivation. Human papillomavirus (HPV) is one of the most prevalent sexually transmitted infections worldwide and remains a major contributor to cervical and other anogenital cancers. In the United States, widespread vaccination of preadolescents and adolescents has led to a marked reduction in the acquisition of high-risk HPV types (e.g., 16 and 18), which are responsible for the majority of HPV-related cancers. However, the overall population-level impact of vaccination depends on multiple interacting factors, including heterogeneity in sexual behavior networks, vaccine uptake and timing (relative to sexual debut), and the duration and efficacy of vaccine-induced immunity. Agent-based models (ABMs) are well suited for exploring such dynamics because they can explicitly simulate individual-level attributes, partnership formation, and stochastic disease transmission processes within heterogeneous populations. By modeling individuals and their contacts rather than relying solely on aggregated compartmental assumptions, ABMs provide a more realistic representation of how interventions such as vaccination may alter infection trajectories across time and demographic subgroups.

Informal Problem Statement. We aim to study how varying vaccination coverage and efficacy scenarios affect HPV prevalence and incidence over time in a heterogeneous population representative of the Atlanta Metropolitan Area. Specifically, we seek to understand how differences in demographic composition, sexual network structures, and vaccination strategies jointly influence the long-term infection dynamics and herd effects. Our analysis will focus on quantifying both direct protection among vaccinated individuals and indirect protection among unvaccinated individuals through reduced transmission potential.

Research Objectives. The overarching goal of this project is to evaluate the epidemiological outcomes of HPV vaccination programs using an agent-based simulation framework. To achieve this, our milestone objectives are as follows:

(1) **Model Initialization:** Construct a synthetic population reflective of Atlanta's demographic and behavioral structure using publicly available datasets (e.g., ACS, NHANES, and sexual behavior surveys).
(2) **Baseline Simulation:** Implement a preliminary ABM capable of simulating HPV transmission dynamics under baseline vaccination parameters.
(3) **Scenario Analysis:** Define and compare multiple vaccination coverage and efficacy scenarios to assess changes in HPV prevalence and incident infections over time.
(4) **Validation and Sensitivity Exploration:** Perform initial sensitivity analyses to examine the robustness of outcomes to key parameters such as contact rates and vaccine protection duration.

Together, these steps will enable us to evaluate the effects of intervention strategies on HPV transmission and identify the factors most critical to achieving sustained reductions in infection rates within an urban, heterogeneous population.

2 Response to Proposal Feedback

In this section, we summarize the feedback received on our initial project proposal and describe the corresponding changes and improvements we have made for this milestone.

Wider Data Selection: The instructor suggested broadening our data selection beyond a single source. In response, we have expanded our dataset scope to include multiple publicly available epidemiological datasets. We are also preparing to incorporate synthetic data generated using an SEIR-type model to evaluate generalizability.

Related Work Formatting and Discussion: The feedback noted that our selected references, though relevant, were not formatted correctly and should more clearly connect to our project. We have revised the Related Work section to use a consistent citation style and reorganized it to concisely discuss how each cited study (e.g., MDLInfer [6], Rumor Centrality [11]) informs our approach. Each paper is now linked to a specific methodological or conceptual aspect of our work.

*All authors contributed equally to this work.

Clarification of Neural Model Usage: The reviewer recommended describing the role of neural networks in parameter estimation in more detail and considering more interpretable alternatives. We have added a subsection explaining how neural models will be used to learn parameters of the epidemic model (e.g., infection or reporting rates) and how their outputs will be interpreted through visualization and feature importance analysis. We also plan to compare these results against simpler interpretable baselines such as linear regression and MDL-based inference.

Model Establishment and Data Processing Pipeline: It was suggested that model implementation begin earlier and that we clarify how the data will feed into the modeling process. Accordingly, we have started developing the data preprocessing pipeline, including cleaning, normalization, and transformation steps that map directly into the model input structure. The data flow from raw observations to model-ready features is now explicitly described in Section 4.

Division of Work: Refined team contributions section.

3 Related Work and Survey

HPV epidemiology and vaccination. Human papillomavirus (HPV) is the most common sexually transmitted infection, with approximately 80% of sexually active individuals acquiring it during their lifetime. High-risk genotypes such as HPV-16 and HPV-18 are responsible for approximately 70% of cervical cancer cases globally and in the United States [5, 13]. Clinical and population-based studies have characterized HPV acquisition, clearance, and persistence risks by age and type, and vaccine trials have demonstrated high efficacy against incident and persistent infections caused by oncogenic strains [3, 4]. These studies establish the biological foundation for modeling HPV transmission and vaccination impacts.

Modeling approaches. Mathematical models, including both compartmental and individual-based frameworks, have long been used to study HPV dynamics and intervention strategies. Traditional SIS/SIR-type models capture transmission at the population level, but fail to represent heterogeneity in sexual networks and partnership dynamics that drive sexually transmitted infections. Recent research has increasingly leveraged agent-based modeling (ABM) to incorporate this heterogeneity. For example, [8] modeled HPV vaccination uptake among young men who have sex with men (YMSM) in urban networks, demonstrating the importance of behavioral structure and partnership turnover. Similarly, the open-source *HPVsim* platform [12] introduced a flexible ABM for HPV transmission and cervical disease progression across different countries, emphasizing calibration and sexual mixing heterogeneity. While these models provide valuable methodological guidance, none have been tailored to the demographic and behavioral characteristics of the Atlanta Metropolitan Area—an identified gap our work addresses.

Empirical foundations for network structure. To parameterize realistic partnership networks, Anderson et al. [1] used exponential random graph models to characterize HIV transmission potential in San Francisco and Atlanta. Their dynamic network inference methods inform our approach for initializing heterogeneous sexual contact structures within the ABM. However, their study focused exclusively on MSM networks, whereas our model includes both male and female populations to represent broader community-level HPV transmission.

Vaccination scenarios and policy implications. Several modeling studies have evaluated HPV vaccination scale-up and regimen optimization at the population level. Bénard et al. [2] employed an individual-based model (HPV-ADVISE) to examine the effectiveness of single-dose vaccination strategies, showing substantial protection under high coverage. Daniels et al. [7] extended this work by incorporating economic outcomes, revealing favorable cost-effectiveness for one-dose schedules in high-income settings. Complementary studies, such as [10], highlight that consistent, well-targeted vaccination programs can outperform adaptive controls when network effects are properly considered. Together, these results motivate our comparative analysis of vaccination coverage and efficacy scenarios under local heterogeneity.

Summary and gap. Existing literature provides extensive insight into HPV natural history, vaccine efficacy, and large-scale modeling frameworks, yet few studies explicitly integrate regional demographic variation and contact network heterogeneity. Our work contributes a lightweight, transparent ABM calibrated to the Atlanta Metropolitan Area, designed to evaluate vaccination coverage and efficacy scenarios while remaining computationally tractable for policy-relevant sensitivity analysis.

4 Data Collection Process

Overview. Our agent-based model (ABM) requires demographic, behavioral, and epidemiological parameters to simulate HPV transmission and vaccination outcomes in the Atlanta Metropolitan Area. To balance regional specificity with generalizability, we integrate both local and national datasets that together describe population structure, sexual behavior patterns, and vaccination coverage. Table 1 summarizes the primary data sources used in this project.

Demographic data. Baseline demographic characteristics are drawn from the U.S. Census Bureau's *American Community Survey* (ACS, 2022 5-year estimates), which provides age, sex, and race/ethnicity distributions for the Atlanta metropolitan statistical area (MSA). These distributions are used to initialize individual agents and assign them demographic attributes that inform network mixing and vaccination probabilities. To ensure representativeness, we scale the synthetic population to a manageable size (e.g., 50,000 agents) while preserving empirical age and sex proportions.

Behavioral and sexual network data. Sexual behavior parameters—including partnership formation rates, age-assortative mixing, concurrency, and partnership duration—are derived from national surveys such as the *National Health and Nutrition Examination Survey* (NHANES) and the *National Survey of Family Growth* (NSFG). These sources provide stratified estimates by age, gender, and sexual activity level. Because regional-level sexual behavior data for Atlanta are limited, national distributions are used as priors, adjusted with demographic weights to approximate local heterogeneity. We further reference network structure estimates from Anderson et al. [1], who modeled dynamic sexual networks for Atlanta and San Francisco using exponential random graph models.

Vaccination and epidemiological data. Vaccination coverage rates by age and sex are taken from CDC's *National Immunization Survey–Teen* (NIS-Teen, 2023), which reports HPV vaccine uptake

for adolescents aged 13–17 years at both national and state levels. We use Georgia-specific uptake estimates to calibrate baseline vaccination scenarios, and national data to test sensitivity under higher and lower coverage assumptions. Vaccine efficacy parameters are based on results from large clinical trials summarized in [3, 4, 12]. For natural history progression—clearance and persistence probabilities—we draw from meta-analyses incorporated into HPVsim [12].

Data preprocessing and integration. All datasets are processed using `pandas` and `numpy`. Variables are cleaned for missing values, normalized across sources, and aggregated to appropriate time scales. Synthetic individuals are then generated by sampling from these empirical distributions. Each agent receives demographic attributes (age, sex, race), sexual behavior traits, and vaccination status according to conditional probabilities informed by the above data. These attributes collectively initialize the ABM at time $t = 0$.

Future data expansion. To further address feedback encouraging broader data coverage, we plan to incorporate additional datasets from neighboring metro areas and CDC's *Behavioral Risk Factor Surveillance System* (BRFSS) to validate behavioral assumptions. This will help test the robustness of model predictions across varied demographic and behavioral contexts.

Table 1: Primary Data Sources for Model Initialization

Dataset	Purpose
U.S. Census Bureau (2022)	Age, sex, and race/ethnicity distributions for synthetic population generation.
NHANES	National sexual behavior distributions (partner count,concurrency, mixing).
CDC:NIS-Teen (2023)	HPV vaccination coverage by age and sex; Georgia-specific estimates used for baseline.
CDC:BRFSS (2022)	Supplementary behavioral and preventive health indicators (planned integration).
Published literature [1, 4, 12]	Network parameterization, vaccine efficacy, and natural history model calibration.

Data limitations. While our integrated data approach balances regional specificity and national generalizability, several limitations remain. First, local sexual behavior data for the Atlanta Metropolitan Area are sparse, requiring us to rely on national distributions (NHANES, NSFG) adjusted by demographic weights rather than direct regional observations. Second, survey-based estimates may underreport sensitive behaviors, introducing potential bias in partnership rate and concurrency parameters. Finally, vaccination coverage and efficacy are modeled as static probabilities, which may not fully capture temporal or behavioral feedback effects in real populations. Despite these limitations, the combined data sources provide a reasonable foundation for initializing and testing the ABM while maintaining interpret-ability and transparency in model assumptions.

5 Preliminary Findings

We implemented a reproducible baseline ABM to validate the data flow and simulation pipeline. The population was initialized from cleaned demographic distributions (`ga_age_sex_cleaned.csv`), and the model used random-mixing transmission with parameters $\beta = 0.02$, clearance = 0.05, vaccination coverage = 0.7, and efficacy = 0.9.

Over 120 timesteps, infections declined rapidly, with peak prevalence $< 0.1\%$ at initialization and zero at equilibrium, reflecting strong vaccine protection and clearance. These preliminary results confirm correct data integration and stochastic behavior, providing a baseline for future scenario calibration.

6 Mathematical Background

Agent-Based Modeling (ABM) is a bottom-up simulation approach in which the collective dynamics of a system emerge from the interactions of individual agents governed by a set of behavioral rules. Formally, an ABM can be represented as a discrete-time stochastic process:

$$S(t+1) = F(S(t), \Theta, \xi_t), \tag{1}$$

where $S(t)$ denotes the system state at time t, Θ represents a vector of model parameters. The parameters Θ are β (the infection probability), vaccination coverage, vaccination efficacy and recovery probability for infected individuals. ξ_t is a stochastic term capturing random effects or environmental noise. Each agent $i \in \{1, \ldots, N\}$ has its own state vector $x_i(t)$ that evolves according to local rules of the form:

$$x_i(t+1) = f_i(x_i(t), \mathcal{N}_i(t), \Theta, \xi_{i,t}), \tag{2}$$

where $\mathcal{N}_i(t)$ is the neighborhood or interaction set of agent i, and $f_i(\cdot)$ encodes its decision-making or transition dynamics.

Network-level formulation. To bridge population- and individual-level dynamics, the agent-based model adopts a stochastic network-based approach. Each individual i occupies a node in a dynamic sexual contact network $G = (V, E)$, where V is the set of agents and $E(t)$ represents active sexual partnerships at time t. Transmission occurs probabilistically across each edge $(i, j) \in E(t)$ according to:

$$P(\text{transmission}_{i \to j}) = 1 - \exp(-\beta_{ij} \Delta t), \tag{3}$$

where β_{ij} is the per-contact transmission probability (which may vary by agent characteristics and HPV type), and Δt is the simulation time step. As of now the β_{ij} is implemented as a constant and the transmission probability is simply given by a constant β ie $P(\text{transmission}_{i \to j}) = \beta$. This may be changed in the future to incorporate node pair level transmission which may be approximated via a neural network or similar heuristics. Due to the ethical concerns of estimating the node level transmission of an STI we may have to rely more heavily on heuristics due to the black box nature of the neural network. Partnership formation and dissolution are modeled as stochastic processes with empirically derived rates, preserving observed distributions of partner number and duration.

6.1 Neural Network–Based Parameter Calibration

Traditional calibration of agent-based models relies on optimization techniques such as least squares fitting, approximate Bayesian computation, or heuristic search (e.g., genetic algorithms) to minimize the discrepancy between simulated and observed system outputs.

However, these approaches can be computationally expensive and sensitive to stochastic variability.

In this work, parameter calibration is performed using a **neural network surrogate model** trained to approximate the mapping between the ABM parameters and the resulting emergent outcomes. Specifically, let the model output of interest (e.g., aggregated behavior or macro-level statistic) be denoted by:

$$y = G(\Theta), \tag{4}$$

where $G(\cdot)$ represents the simulation process. We train a neural network $\hat{G}_\phi(\Theta)$, parameterized by weights ϕ, such that:

$$\hat{G}_\phi(\Theta) \approx G(\Theta), \tag{5}$$

using a dataset of simulation results generated from sampled parameter configurations. Once trained, the neural network provides a fast differentiable approximation of the ABM, enabling efficient gradient-based optimization of parameters:

$$\Theta^* = \arg\min_{\Theta} \|\hat{G}_\phi(\Theta) - y_{\text{obs}}\|^2, \tag{6}$$

where y_{obs} denotes the empirical data. This approach captures nonlinear dependencies between parameters and emergent dynamics while significantly reducing computational cost during calibration.

Moreover, stochasticity in agent behavior is implicitly learned through the neural network's training process. By incorporating multiple stochastic realizations for each parameter configuration, the surrogate model approximates the expected system response under uncertainty. This hybrid approach thus blends data-driven inference with mechanistic simulation, yielding a robust and adaptive calibration method for complex ABMs.

7 Agent-Based Modeling and Neural Network Algorithms

Our simulation employs an agent-based modeling (ABM) framework to capture the heterogeneous and stochastic nature of HPV transmission within a synthetic population reflective of Georgia. Each agent represents an individual characterized by demographic, behavioral, and epidemiological attributes that evolve over discrete time steps. We also briefly outline the Neural Network algorithm as we are using that to calibrate our parameters for the Agent-Based Modeling simulation.

7.1 Agent Attributes

Each agent a_i in the population $A = \{a_1, a_2, \ldots, a_N\}$ is initialized with: **Demographics:** age, sex, and race/ethnicity, drawn from ACS distributions. **Behavioral traits:** partnership rate, concurrency level, and preferred partner age difference, sampled from NHANES and NSFG data. **Epidemiological state:** one of $\{S, I, R, V\}$ corresponding to susceptible, infected, recovered, or vaccinated

7.2 Model Dynamics

The simulation proceeds in discrete monthly steps ($\Delta t = 1/12$ year), during which the following processes occur:

(1) **Partnership formation:** Each agent may form new partnerships according to an age- and activity-specific probability $p_f(a_i)$, ensuring the network degree distribution matches empirical sexual behavior data. Partnerships are formed through preferential mixing, governed by a mixing matrix $M(a_i, a_j)$ that reflects age and risk assortativity.

(2) **Transmission:** For every active partnership (i, j) where one agent is infected (I) and the other susceptible (S), infection may occur with probability $p_t = 1 - e^{-\beta_{ij}}$, modulated by vaccination status and efficacy:

$$p_t' = p_t(1 - \epsilon_i)(1 - \epsilon_j),$$

where ϵ_i and ϵ_j are the vaccine efficacies for agents i and j respectively.

(3) **Clearance and immunity:** Infected agents recover with probability $\gamma \Delta t$ per time step and transition to the recovered (R) state. Recovered individuals may lose immunity at rate ω and return to the susceptible state.

(4) **Vaccination:** Routine vaccination is applied to agents at target ages (typically 11–12 years) with probability v_{routine}, and catch-up vaccination to eligible adults up to age 26 with probability v_{catchup}. The vaccinated (V) state confers immunity with efficacy ϵ.

(5) **Partnership dissolution:** Existing partnerships end probabilistically at rate μ_d, ensuring dynamic network turnover that reflects observed relationship durations.

7.3 Algorithm Outline

The complete simulation procedure can be summarized as follows:

Algorithm 1 Agent-Based HPV Transmission Simulation

```
Initialize N agents with demographic, behavioral, and vaccination attributes.
Construct initial partnership network G_0(V, E_0).
for t = 1 to T do
    Form new partnerships based on p_f(a_i) and M(a_i, a_j).
    for each active partnership (i, j) ∈ E_t do
        if a_i ∈ I and a_j ∈ S then
            Transmit with probability p_t'; update a_j ← I.
        end if
    end for
    Update infection clearance, immunity loss, and vaccination events.
    Dissolve partnerships with probability μ_d.
    Record infection prevalence, incidence, and vaccination coverage.
end for
```

7.4 Output Metrics

The model records both individual- and population-level outcomes at each time step, including:

- Prevalence and incidence of HPV infection by age and sex group.
- Cumulative infections averted under each vaccination scenario.
- Effective reproduction number (R_t) over time.

Multiple simulation runs are performed to estimate confidence intervals and evaluate sensitivity to key parameters such as vaccine efficacy, coverage, and partnership rate.

This ABM framework enables exploration of how vaccination strategies interact with network heterogeneity and demographic structure to shape HPV transmission trajectories in Georgia.

7.5 Formal Description of the Neural Network Algorithm

A neural network defines a parametric function $f_\phi : \mathbb{R}^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$ that maps an input vector $\mathbf{x} \in \mathbb{R}^{d_{\text{in}}}$ to an output $\hat{\mathbf{y}} \in \mathbb{R}^{d_{\text{out}}}$ through a sequence of L layers. Each layer $l \in \{1, \dots, L\}$ performs an affine transformation followed by a nonlinear activation:

$$\mathbf{h}^{(l)} = \sigma^{(l)}\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right), \tag{7}$$

where $\mathbf{W}^{(l)} \in \mathbb{R}^{n_l \times n_{l-1}}$ and $\mathbf{b}^{(l)} \in \mathbb{R}^{n_l}$ denote the weight matrix and bias vector of layer l, and $\sigma^{(l)}(\cdot)$ is a nonlinear activation function (e.g., ReLU, sigmoid, or tanh). The input to the network is $\mathbf{h}^{(0)} = \mathbf{x}$, and the final output is $\hat{\mathbf{y}} = \mathbf{h}^{(L)}$.

The objective of training is to learn the optimal parameters

$$\phi = \{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}_{l=1}^{L}$$

that minimize a loss function $L(\hat{\mathbf{y}}, \mathbf{y})$, where $\mathbf{y}$ denotes the ground-truth target. This is formulated as:

$$\phi^* = \arg\min_{\phi} \frac{1}{N} \sum_{i=1}^{N} L\big(f_\phi(\mathbf{x}_i), \mathbf{y}_i\big), \tag{8}$$

where $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ are training samples.

To solve this optimization problem, gradient-based methods such as Stochastic Gradient Descent (SGD) or Adam are employed. At each iteration t, the parameters are updated according to:

$$\phi_{t+1} = \phi_t - \eta_t \, \nabla_{\phi_t} L\big(f_{\phi_t}(\mathbf{x}), \mathbf{y}\big), \tag{9}$$

where η_t is the learning rate and $\nabla_{\phi_t} L$ is the gradient of the loss with respect to the parameters, computed efficiently using the *back-propagation* algorithm.

Algorithm 2 Neural Network Training Procedure

Initialize network parameters $\phi = \{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}_{l=1}^{L}$ randomly.
for each epoch $e = 1, \dots, E$ **do**
  for each mini-batch $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{B}$ **do**
    Compute predictions: $\hat{\mathbf{y}}_i = f_\phi(\mathbf{x}_i)$
    Evaluate loss: $\mathcal{L} = \frac{1}{B} \sum_{i=1}^{B} L(\hat{\mathbf{y}}_i, \mathbf{y}_i)$
    Compute gradients: $\nabla_\phi \mathcal{L}$ via backpropagation
    Update parameters: $\phi \leftarrow \phi - \eta \, \nabla_\phi \mathcal{L}$
  end for
end for
Return: Trained parameters ϕ^*

Algorithmic Summary. This procedure enables the neural network to iteratively learn a mapping that minimizes the prediction error on the training data. Once trained, the model generalizes to unseen inputs, providing a nonlinear approximation of the underlying functional relationship between input and output variables.

8 Challenges

One challenge we encountered was working with and building upon the open source HPV Sim model. While it works very well with the pre-loaded data parameters it is very difficult to enter in new parameters to our specific population of the Atlanta Metro Area. However, it will still be a guide in implementing a state of the art model to monitor and better understand cervical cancer, HPV and the HPV vaccine's impacts in Atlanta, GA.

Additionally, some of the data sources for Atlanta specifically are more difficult to find since it is generally less reported. We were able to find enough information within the CDC and Census data to work with our model and make sure our agent based model will be well calibrated. As discussed before, for some parameters like vaccination rates, we will use national data in place of data specific to Georgia.

9 Next Steps

Our team will farther work on the next steps of developing a fully functional model. Specifically we will focus on incorporating age and demographic data from our sources into a sexual dynamic network. This will be then used to predict cervical cancer rates - which we will then validate based on the NIH Cancer institute data [9] for cancer rates in the Atlanta Metropolitan area. We will ensure parameters are correct by using references cited and through real world data to gauge our model.

Finally we will use the neural network as discussed in section 7. Additionally, we will compare these findings to a simpler - more interpretable baseline (such as a linear regression.)

10 Team Contributions

This project is a collaborative effort between all three team members. Detailed individual contributions are listed below.

Scott Watanuki: Led the implementation of the initial ABM prototype (including HPV simulation and vaccination scenarios). Designed the scenario evaluation framework, generated the preliminary prevalence curve, and organized the shared GitHub repository.

Hailey Toeppner: Collected and reviewed relevant epidemiological and demographic datasets (ACS, NHANES, NIS-Teen), verified data suitability, conducted exploratory data inspection/visualization for HPV coverage, compiled/formatted the final milestone report.

Anish Arora: Surveyed and summarized key HPV literature, drafted the Related Work section, and created the mathematical formulation and parameter specification for the proposed model; assisted with technical proofreading.

All team members met regularly to discuss modeling design decisions, interpret feedback from the project proposal, and coordinate responsibilities to ensure timely progress toward the final deliverable.

Code and Data Availability

All scripts and model code for data preprocessing, simulation, and analysis are available at: https://github.com/scottwatanuki/hpv_agent_based_modeling.

References

[1] E. J. Anderson, K. M. Weiss, M. M. Morris, T. H. Sanchez, P. Prasad, and S. M. Jenness. 2021. HIV and Sexually Transmitted Infection Epidemic Potential of Networks of Men Who Have Sex With Men in Two Cities. *Epidemiology* 32, 5 (Sept. 2021), 681–689. doi:10.1097/EDE.0000000000001390

[2] É. Bénard et al. 2023. Potential population-level effectiveness of one-dose HPV vaccination. *The Lancet Global Health* (2023).

[3] E. A. Burger, J. J. Kim, S. Sy, and P. E. Castle. 2017. Age of Acquiring Causal Human Papillomavirus (HPV) Infections: Leveraging Simulation Models to Explore the Natural History of HPV-induced Cervical Cancer. *Clinical Infectious Diseases* 65, 6 (Sept. 2017), 893–899. doi:10.1093/cid/cix475

[4] N. G. Campos, E. A. Burger, S. Sy, M. Sharma, M. Schiffman, A. C. Rodriguez, A. Hildesheim, R. Herrero, and J. J. Kim. 2014. An Updated Natural History Model of Cervical Cancer: Derivation of Model Parameters. *American Journal of Epidemiology* 180, 5 (Sept. 2014), 545–555. doi:10.1093/aje/kwu159

[5] Gary M. Clifford, John S. Smith, Martyn Plummer, Nubia Muñoz, and Silvia Franceschi. 2003. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *British Journal of Cancer* 88, 1 (2003), 63–73. doi:10.1038/sj.bjc.6600738

[6] Jiaming Cui, Bijaya Adhikari, Arash Haddadan, A S M Ahsan-Ul Haque, Jilles Vreeken, Anil Vullikanti, and B. Aditya Prakash. 2025. Accurately Estimating Unreported Infections using Information Theory. arXiv:2502.00039 [cs.SI] https://arxiv.org/abs/2502.00039

[7] V. Daniels et al. 2022. Modeling the health and economic implications of adopting a one-dose HPV vaccination regimen in a high-income country setting. *Vaccine* (2022).

[8] N. D. Goldstein et al. 2019. Modeling HPV vaccination scale-up among urban young men who have sex with men (YMSM). *Vaccine* (2019).

[9] National Cancer Institute and Centers for Disease Control and Prevention. 2025. *Interactive Maps.* https://www.statecancerprofiles.cancer.gov/map/map.withimage.php?13&hsa&001&057&00&2&01&0&1&5&0#results

[10] F. Saldaña, J. A. Camacho-Gutiérrez, I. Barradas, and A. Korobeinikov. 2020. When optimal is not the best: cost-effectiveness analysis for HPV epidemic models. *Mathematical Biosciences* (2020).

[11] Devavrat Shah and Tauhid Zaman. 2010. Detecting Sources of Computer Viruses in Networks: Theory and Experiment. *ACM SIGMETRICS Performance Evaluation Review* 38, 1 (2010), 203–214. doi:10.1145/1811039.1811063

[12] R. M. Stuart, J. A. Cohen, C. C. Kerr, P. Mathur, National Disease Modelling Consortium of India, R. G. Abeysuriya, M. Zimmermann, D. W. Rao, M. C. Boudreau, S. Lee, L. Yang, and D. J. Klein. 2024. HPVsim: An agent-based model of HPV transmission and cervical disease. *PLOS Computational Biology* 20, 7 (July 2024), e1012181. doi:10.1371/journal.pcbi.1012181

[13] {Jan M.M.} Walboomers, {Marcel V.} Jacobs, {M. Michele} Manos, {F. Xavier} Bosch, {J. Alain} Kummer, {Keerti V.} Shah, {Peter J.F.} Snijders, Julian Peto, {Chris J.L.M.} Meijer, and Nubia Muñoz. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology* 189, 1 (1999), 12–19. doi:10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F