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 have developed 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 sexual network framework by Walker et Al. [20], our model simulates 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 have calibrated the model to reflect Atlanta-specific population structure and infection patterns. Through scenario analyses varying vaccine coverage and efficacy, we have estimated 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 Motivation and Problem Statement

Motivation. HPV is a highly prevalent sexually transmitted infection. Vaccination of preadolescents and adolescents substantially reduces acquisition of high-risk types in Atlanta (e.g., 16/18), which in turn lowers downstream risks of precancerous lesions and cervical cancer. Population-level impact depends on heterogeneity in sexual networks, coverage, timing (pre-exposure), and vaccine efficacy. ABMs naturally capture these features by simulating individuals and partnerships.

Informal problem statement. How do vaccination coverage and efficacy scenarios change HPV prevalence and incident infections in the heterogeneous population over time in the Atlanta Metropolitan Area?

2 Related Work and Survey

HPV natural history and vaccination. Prior studies describe HPV acquisition, spontaneous clearance, and persistence risks by age and type; vaccine trials quantify efficacy against incident and persistent infection for high-risk types (16/18). HPV is a commonly spread STD infecting about 80 % of the population. It is important to note that while harmless in most cases, HPV (16/18) can lead to cervical cancer. In fact, more than 90 % [6] of cervical cancer is caused by HPV. Cervical cancer is one of the leading causes of cancer deaths among women in the United States making it a prevalent matter and an adequate research topic.

Modeling approaches. Compartmental models (SIS/SIR variants for STIs) and microsimulation/ABMs have been used to evaluate HPV vaccination and screening policies, with heterogeneous mixing (degree, age-assortativity) and partnership turnover shown to influence impact estimates. Recent work highlights the role of pre-exposure vaccination timing, single-dose effectiveness, and catch-up strategies.

Gap. Many classroom-scale projects fit ODE models to respiratory pathogens or perform forecasting; few tackle STIs with explicit individual-level contact heterogeneity. We contribute a lightweight ABM isolating proximal infection outcomes under vaccination, with sensitivity and transparent assumptions—tractable for a semester while remaining epistemologically relevant. Additionally, the Atlanta Metropolitan Area does not have specific studies integrating how HPV affects women and men in the area.

2.1 Survey of Key Literature

2.1.1 Paper 1: HIV and Sexually Transmitted Infection Epidemic Potential of Networks of Men Who Have Sex with Men in Two Cities. In this paper [1] Anderson et Al. use empirical data from San Francisco and Atlanta to estimate the generative parameters of the forward reachable path and compare results to to the HIV epidemics in each city. The methods exist of using exponential random graph models to estimate each city's dynamic sexual network from survey data. After simulating the stochastic dynamic networks from the fitted models the results were stratified by partnership type an demographics. The paper does a nice job of explain the forward-reachable path (shown below in Figure 1), and why it is important to consider to quantify the upper threshold for epidemic potential.

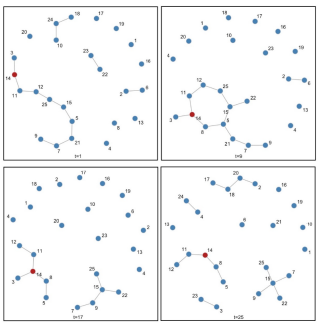

Figure 1: Forward Reachable Path Example

*All authors contributed equally to this work.

Figure 1 specifically demonstrates the The Underlying Network that Led to the FRP for Node 14 at Time Steps 1, 9, 17, and 25. In addition to FRP explanation, their model also links survey data to racial and age relations - crucial for designing the agent based model. A shortfall of the paper is that there is no in depth explanation of modeling sexual network structure in sexually transmitted disease prevention. Additionally, the paper largely focuses on male to male sexual interaction - where as dealing with HPV, it is important to build a model which focused on the entire population which is discussed in the following publication.

2.1.2 Paper 2: HPVsim: An agent-based model of HPV transmission and cervical disease. Stuart et al. [18], recently published a paper discussing an open source model which is used specifically to model HPV in any given country. This paper does a great job of evaluating agent based simulations and applying them specifically to how HPV spreads. It analyzes different sexual relationships in different countries and discusses how that was important to creating HPVSim. The paper also does a great job of explaining calibration and how the authors designed their model. However an obvious shortfall is the model has not been calibrated and fit specifically to Georgia, therefore our research project will fill in these gaps.

2.1.3 Paper 3: Age of Acquiring Causal Human Papillomavirus (HPV) Infections: Leveraging Simulation Models to Explore the Natural History of HPV-induced Cervical Cancer. A publication by Burger et Al. [4] discusses an empirically calibrates model simulating cervical cancer. The paper does not go in depth in the model specifics but more goes into a deeper analysis of the implications of when a female gets the vaccine. A lack of description of the model and model parameters was a major shortcoming of this article however it links a more in depth explanation by similar authors discussed below.

2.1.4 Paper 4: An Updated Natural History Model of Cervical Cancer: Derivation of Model Parameters. Campos et Al. discuss their HPV model parameters and design in this article [5]. This paper discusses how they implemented a Monte Carlo simulation to track disease progression from HPV infection to cervical cancer diagnoses. A major shortcoming of this approach is the lack of a dynamic model in a sexually communicable disease. Their parameters will be potentially helpful as we build and calibrate our model, however we hope to expand upon their approach by including an agent based model.

2.1.5 Paper 5: Potential Population-Level Effectiveness of One-Dose HPV Vaccination. Bénard et al. [2] use an individual-based transmission model (HPV-ADVISE) to explore the population-level effectiveness of a single-dose HPV vaccine. The paper demonstrates that one-dose regimens could achieve substantial protection in high-coverage settings. This work informs our sensitivity analysis on vaccine efficacy and dosing schedules.

2.1.6 Paper 6: Modeling HPV Vaccination Scale-Up Among Urban Young Men Who Have Sex with Men (YMSM). Goldstein et al. [9] construct an agent-based model focused on HPV spread in urban MSM populations, simulating vaccination uptake and behavioral heterogeneity. While specific to a subgroup, the model's design for dynamic partnerships and vaccine rollout provides valuable insight into modeling sexual network structure and intervention timing.

2.1.7 Paper 7: Modeling the Health and Economic Implications of Adopting a One-Dose HPV Vaccination Regimen. Daniels et al. [7] evaluate health and economic outcomes of shifting from two-dose to one-dose HPV vaccination strategies using a dynamic transmission model calibrated to high-income settings. The methodology offers guidance for comparing alternative vaccination scenarios in our study, particularly when examining long-term policy trade-offs.

2.1.8 Paper 8: When Optimal is Not the Best—Cost Effectiveness Analysis for HPV Epidemic Models. Saldaña et al. [14] explore cost-effectiveness in HPV control strategies using mathematical optimization. The study shows that constant, well-targeted vaccination programs can outperform complex adaptive controls. Their findings provide an analytical framework for interpreting our simulation outcomes and potential public health recommendations.

2.1.9 Paper 9: A revision of sexual mixing matrices in models of sexually transmitted infection. The paper by Walker et al. [20] presents a revised mathematical model for sexual mixing, which is a critical component of dynamic STI models, and can be applied to HPV. This model explicitly calculates the probability of sexual interaction between individuals based on two non-independent criteria: age group and sexual activity class

3 Data and Parameters

Four key focuses for our model to match current publications and findings in the Atlanta Metropolitan Area are listed as: 1. Current Infection (Transmission): Age-specific HPV Prevalence in women [11]. 2. Screening Rates (Precursor): How screening rates vary by demographic make up[12]. 3. Final Outcome (Cancer): Cervical Cancer Rates in Metro Atlanta counties [16]. 4. Demographic data pulled from the US Census data for the Atlanta Metropolitan area.

4 Proposed Method

The project utilizes an Agent-Based Model (ABM) framework, which is designed to simulate the complex, microscopic interactions that drive disease spread over time. This methodology allows us to create a high-fidelity projection of the long-term effects of Human Papillomavirus (HPV) within the Atlanta Metropolitan region by focusing on individual-level variation in behavior and risk.

For the model initialization and customization, we have framed the model to Atlanta- Specific Data. Our framework is based upon the sexual network methodology by Walker et Al. [20] leveraging their structure for sexual networks - specifically with HPV. We will manually load in population and demographic data from US Census data, specifically molding our model to the Atlanta Metropolitan Area.

After fitting and building our initial model we will then work on calibrating and the model, to ensure the parameters used are sufficient to represent HPV 16/18 cases in the populations, and cancer rates in the population.

Finally we will display experiments taking a deeper dive into how the HPV vaccine and screening strategies can be used within

the Atlanta community to potentially save lives and contribute to health equity.

4.1 Agent Structure

The initial population of agents in the model are based specifically to fit the population structure of the Atlanta Metropolitan Area (which involves Fulton and surrounding counties.)

Demographics: Pulling data from the census data [19] we were able to stratify the agents by age, gender, and race.

Intervention and Risk Assignment: The probability of being vaccinated is assigned using our vaccination mapping function, which contains age-, sex-, and race-specific coverage rates sourced from CDC and Ricon et Al. 2024 [13]. This allows the model to capture the reality of differential coverage rates (e.g., higher rates for young females versus lower, race-specific rates for older cohorts). The agent's monthly probability of undergoing screening is assigned only to females 25 and older. The annual screening uptake is pulled from the screening map (e.g., White: 63.5%, Black: 53.2% etc), reflecting known health disparities in preventative care access, the data was found by a study done by Spencer et Al. [17].

The agent is assigned one of four activity classes (Low, Mod, Mod-High, High) based on the sex activity array. This framework is based on the work by Walker et Al. and was supported on the study by Santelli et Al.[15] which supplies sexual activity data on the US population.

4.2 Sexual Network Structure: The Walker Mixing Matrix

To accurately reflect the partially non-random nature of STI transmission, our group implemented and used a simplified methodology developed by Walker et Al. We stratified the population into 20 distinct subgroups. The two groups we used were $n_A = 5$, separating the population by age and $n_S = 4$ which divided the population by sexual behavioral groups. The probability $p_{i,h\to j,m}$ represents the probability of partner from target age i, behavior group h selecting a partner from age j and behavior group m. We chose this framework since it was relatively simple to implement, yet still added a level of complexity improving accurate sexual behaviors of adults in the Atlanta Metropolitan Area. Associativity parameters ϵ_A and ϵ_S represents the probability of a partner 1 determining partner 2 based on either age or sexual activity. The equation implemented is shown below:

$$\begin{aligned}\rho_{i,h\to j,m} =& \epsilon_A\epsilon_S\delta_{ij}\delta_{hm} \\ &+ \epsilon_A(1-\epsilon_S)\delta_{ij}\frac{P_{j,m}}{\sum_\beta P_{j,\beta}} \\ &+ (1-\epsilon_A)\epsilon_S\delta_{hm}\frac{P_{j,m}}{\sum_\alpha P_{\alpha,m}} \\ &+ (1-\epsilon_A)(1-\epsilon_S)\frac{P_{j,m}}{\sum_\alpha\sum_\beta P_{\alpha,\beta}}\end{aligned}$$

Where $\epsilon_A = 0.7$ (age associativity), $\epsilon_S = 0.3$ (activity associativity), and δ is the Kronecker delta. Each row is normalized to sum to 1. Partner selection applies this mixing probability with an opposite-sex constraint, sampling partners without replacement. Activity classes determine the number of partners per timestep via multipliers scaled to a baseline rate, reflecting heterogeneity in sexual behavior observed in NSFG data.

The heterosexual mixing structure is based on age and activity preference. Race and other factors were not included. In order to calibrate we ran a simulation calculating the correct parameters to give partners per year to be similar to a study done by Santelli et Al. [15]. Ensuring that the sexual contact matrix was representative of real data.

4.3 Disease Progression States

The three core dynamic states are Susceptible (S), Infected (I), and Cancer (C). Cleared agents (if the HPV infection clears) return immediately to the Susceptible state, reflecting the lack of lasting type-specific immunity following natural clearance. There is also one absorbing state of cancer that occurs with small probability of the HPV16/18 cases.

(1) **Transmission (S → I):** Susceptible agents acquire HPV through sexual contact with infected partners. The per-contact transmission probability β_i for agent i is:

$$\beta_i = C \cdot r_a \cdot (1 - \epsilon \cdot v_i)$$

where C is the calibrated baseline infectivity, r_a is the age-specific relative contact rate, ϵ is vaccine efficacy (0.9), and v_i is the agent's vaccination status. Age-specific contact rates to reflect peak risk in young adults. This is supported by Haderxhanaj et Al.'s findings [10].

(2) **Screening and Treatment (I → S):** Women aged 25+ undergo cervical cancer screening with race-specific monthly probabilities. Detected infections are cleared with **90**% effectiveness, representing successful treatment. These rates are from a study by Spencer et Al. [17]

(3) **Viral Clearance (I → S):** Infected agents clear HPV naturally each month with age-dependent probability γ_a (e.g., 0.008/mo for 15-24 year-olds). This captures faster clearance in younger individuals. This data was found and is supported by a study by Bulkmans et Al. [3].

(4) **Cancer Progression (I → C):** Persistent infections progress to cervical cancer with duration-dependent probability p_{cancer}:

$$p_{\text{cancer}} = p \cdot \ln(t+1)$$

The logarithmic scaling captures elevated risk from long-term persistent infections, with t being the infection duration in months, necessary to model the multi-decade latency period. Where p is the probability of progressing to cancer.

5 Evaluation Methodology

We evaluated the model's accuracy, robustness, and policy relevance along three dimensions:

5.1 Calibration:

Compare model-predicted prevalence and incidence to empirical Atlanta data. The model was calibrated in many different settings. The cancer progression rate was calibrated in order to match similar frequencies seen across published studies mentioned earlier in the paper. Cancer outcomes are not be modeled directly, but

we estimated potential long-term implications using infection-to-cancer transition probabilities from literature. The cancer outcomes still have potential to interpret and understand for long and term interventions. The number of sexual partners was calibrated to be similar to a published study by Santelli et Al. [15] in order to ensure the model is an accurate reflection of sexual partners that could happen in the Atlanta Metropolitan Area. Viral clearance and transmission rates were calibrated in order to ensure the final population infected was similar to rates found in a study by McQuillan et Al.[11].

5.2 Scenario Analysis:

We simulated multiple vaccination coverage and efficacy levels to evaluate reductions in prevalence and infections averted. This is implemented and farther discussed in the next section.

5.3 Sensitivity:

Test robustness to key parameters (transmission rate, clearance probability, partnership duration). The sensitivity analysis for partnership data showed our model performed well against published findings by Haderxhanaj et Al. [10]. The model consistently had similar partnership patterns to those found in the literature.

For beta we used a scaling factor to make sure our results were similar to findings about HPV in the population. By analyzing our scaling factor with many iterations we were able to ensure that HPV rates in the population were aligned with literature by McQuillan et Al. [11]

6 Experiments and Results

6.1 Questions and Setup

In our simulation setup we chose to have two main parameter types: parameters that had to be calibrated with real world data and parameters that we could experiment with. The parameters that we chose to calibrate were age distribution data, sexual activity data, vaccination uptakes based on race, and cancer screening by race and age. The parameters that we decided to experiment with were the rate of transmission β, the efficacy of the vaccine and the vaccination coverage percentage. For each parameter we changed we measured the HPV prevalence and the cancer incidence rate over time. The data that are calibrated are found in the data folder in the github repository. We sought to answer three questions that are presented below:

(1) How do changes in transmission rate β affect the prevalence of HPV and cervical cancer rates among females in Georgia?
(2) How do changes in the HPV vaccine coverage rate affect the prevalence of the disease and the rates of cervical cancer among females in Georgia?
(3) How do changes in the efficacy of the HPV vaccine affect the prevalence of the disease and the rates of cervical cancer among females in Georgia?

While experimenting with the other questions the parameter β is calibrated to 0.3, the vaccine coverage rate is set to 0.6 and the efficacy of the vaccine to 0.9. The simulation is otherwise ran with the calibrated population distribution, 1000 individual agents designed to mimic the actual population distribution of Georgia, and time steps of 300 months. The β was varied from 0.1 to 0.9 in steps of 0.1. The vaccine coverage rate was varied from 0.1 to 1.0 in steps of 0.1. The efficacy rate was varied from 0.5 to 0.95. This was chosen since the vaccine would not be approved if it was below 50 percent effective. Additionally, there were three models that we tested on for each parameter we varied. The first model was the baseline model which included basic vaccine coverage, vaccine efficacy and age and sex distribution data. The second model was the race stratified vaccination rate model. This included additional data over how the vaccination rate was varied based on each race. Finally the third model included both the race stratified vaccination coverage as well as the screening data. The screening data helped to clear those women who were screened and infected from the infection earlier than the normal progression of the disease.

6.2 HPV Prevalence Results

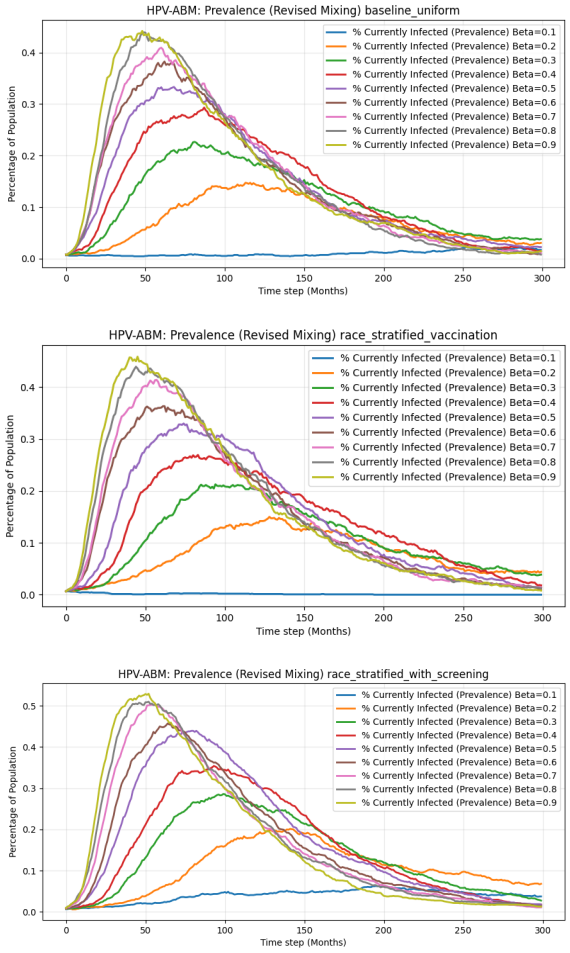

Figure 2: Varying β in the baseline model, race stratified vaccination model, and race stratified with screening model

6.2.1 Varying β. The experiment in figure 2 demonstrated that HPV prevalence is highly sensitive to the transmission probability

(β), confirming an epidemic threshold between $\beta = 0.1$ (where the infection quickly dies out) and $\beta = 0.2$ (where the infection sustains itself). For higher values of β, the peak prevalence increased rapidly (reaching over 40% for $\beta = 0.9$) and occurred much earlier in the simulation (around months 40-50). The prevalence drops sharply after the peak in all scenarios with the exception of the race stratified with screening scenario which has a much flatter curve, consistent with the epidemic burning through the susceptible population until it stabilizes at a low endemic level around the 300-month mark.

Comparing the three intervention scenarios, the models with Race Stratified Vaccination exhibited dynamics very similar to the Baseline Model, suggesting the high transmission rate may overwhelm the implemented vaccination coverage in the short term. Intriguingly, the Race Stratified with Screening model showed the highest peak prevalence (over 50%). This suggests that while screening is intended to treat and reduce prevalence, the model may be capturing increased case ascertainment (better detection of existing cases), making the measured prevalence appear higher than in the non-screening scenarios.

6.2.2 Varying Vaccine Coverage. The experiment in figure 3 investigated the impact of varying vaccine coverage rates on HPV prevalence across three modeling scenarios. In all models, increasing the vaccine coverage rate consistently reduced the peak prevalence and delayed the epidemic's peak time. For the highest coverage rates (0.9 and 1.0, represented by the purple and light blue lines, respectively), the infection curves were significantly blunted, with the peak prevalence for 100% coverage hovering around 5% in the Race Stratified Vaccination model, and around 10% in the Screening and Baseline model. This confirms the critical role of high vaccination rates in controlling and ultimately preventing large-scale epidemics.

Comparing the three scenarios, the Race Stratified Vaccination model behaved similarly to the Baseline Uniform model, achieving substantial control at high coverage rates. However, the Race Stratified with Screening model consistently exhibited the highest prevalence across all non-zero coverage rates. Even with 100% coverage, this model's peak prevalence was double that of the other two models (approximately 10% versus 5%). This reinforces the finding from the previous analysis that, in this specific model structure, the introduction of screening protocols leads to a higher measured prevalence, likely due to increased case ascertainment (detecting more existing, perhaps asymptomatic, infections) rather than an uncontrolled epidemic, although the control offered by vaccination is clearly visible in the steep reduction of the curve compared to low coverage rates.

Further we see results where the Race Stratified versions have higher endemic rates than the Baseline Uniform models at the end. This is due to the varied vaccination rates of different populations. Notably, there are much lower rates of coverage in the Black and Hispanic communities. This can lead to higher endemic rates of prevalence due to the lower coverage rates in these communities.

6.2.3 Varying Vaccine Efficacy. This experiment in figure 4 varied vaccine efficacy from 50% to 95% across baseline, race-stratified vaccination, and screening scenarios, demonstrating a clear inverse relationship: higher efficacy significantly reduced peak prevalence

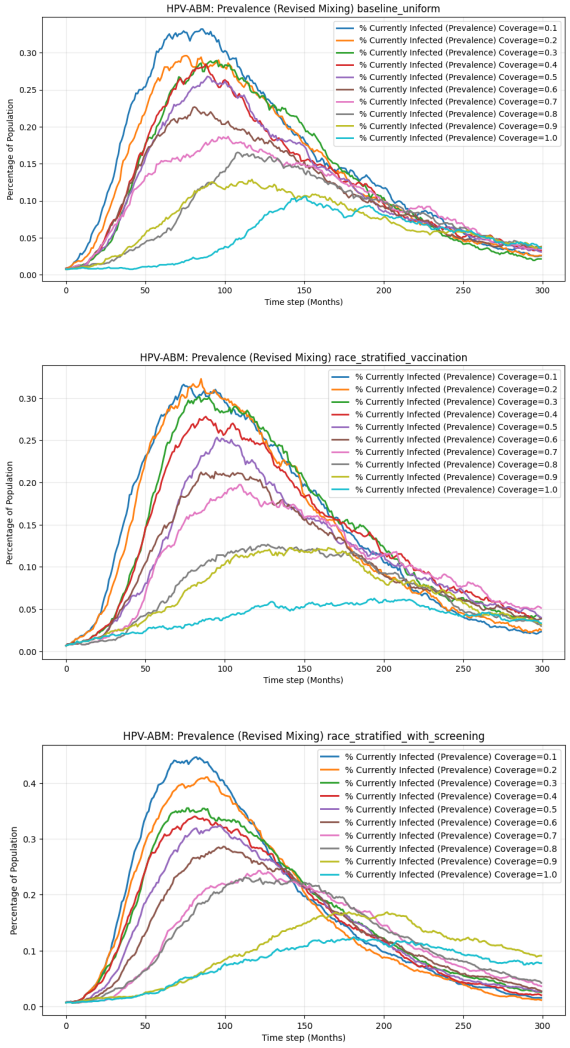

Figure 3: Varying vaccine coverage in the baseline model, race stratified vaccination model, and race stratified with screening model

and delayed the epidemic's peak time. Specifically, increasing efficacy from 50% to 95% led to a notable reduction in peak infection rates (from $\sim$ 30% down to $\sim$ 18% in the baseline model). The dynamics of the Baseline and Vaccination models remained very similar, suggesting race stratification alone did not alter the overall outcome when efficacy was the only varied parameter. Crucially, the Screening model consistently reported higher prevalence across all efficacy levels, peaking at $\sim$ 40% for 50% efficacy and $\sim$ 28% for 95% efficacy, a finding interpreted as evidence of increased case ascertainment (better detection of existing cases) due to the screening protocol.

6.2.4 Comparing β, Coverage, and Efficacy. The three sets of experiments systematically analyzed how the transmission probability

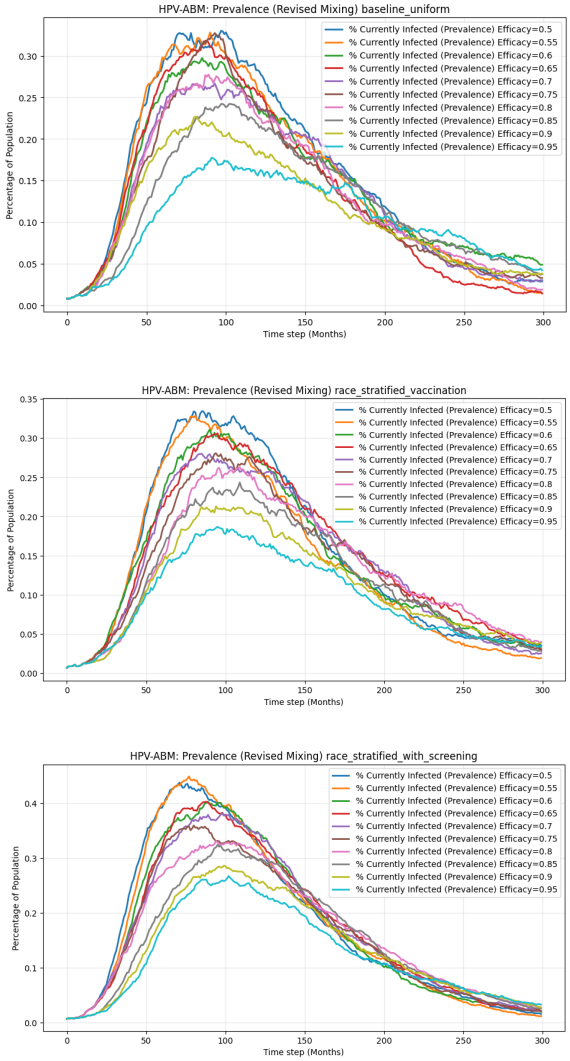

Figure 4: Varying vaccine efficacy in the baseline model, race stratified vaccination model, and race stratified with screening model

(β), vaccine coverage, and vaccine efficacy influence Human Papillomavirus (HPV) prevalence dynamics across three model structures: Baseline Uniform, Race Stratified Vaccination, and Race Stratified with Screening.

Varying β demonstrated its role as the primary driver of epidemic magnitude and speed, where high β resulted in the highest and quickest peak prevalence (over 40%) before the infection subsided due to population immunity. Conversely, increasing both coverage and efficacy successfully dampened the epidemic curve. Achieving maximum vaccine coverage (100%) proved to be the most effective intervention, reducing peak prevalence to its lowest overall point (around 5% in the baseline model). High efficacy (90%) was also crucial, but its impact (peak prevalence around 20%) was secondary to high coverage, highlighting that implementation (getting the vaccine to the population) is a more critical control lever than vaccine formulation alone.

A consistent and notable finding across all three experiments was the behavior of the Race Stratified with Screening model. In every comparison, this model reported a higher peak prevalence than the Baseline and Vaccination models, even when controlling for high coverage or high efficacy. This strongly suggests that the inclusion of screening in this specific ABM increases case ascertainment (detection and measurement) of the currently infected population. All models eventually converged to a low endemic state by the 300-month mark, indicating long-term population control regardless of the initial surge.

6.3 Cancer Incidence Results

The cancer incidence graphs are all very similar across the three different models due to the very low incidence rate of cancer (13,000 cases in the US every year). They all show similar patterns to the HPV Prevalence rate graphs but the differences are difficult to confidently differentiate due to the very low incidence rates and the fact that we only have 1000 agents in this model. Perhaps, we could get more detailed results and nuanced results from running the simulation a 100 times but the code is not parallelized for this. This is shown in the graphs in figure 5 which feature improbable patterns. In the first graph in figure 5 the cancer incidence is highest for $\beta = 0.5$. This is quite improbable in a real world scenario. Additionally, the lowest $\beta = 0.1$ is at the second lowest cancer incidence rate. In the second graph, we see similarly improbable patterns with the middle of the pack coverage leading in cancer incidence and a coverage of 0.3 in second. Finally, in the third image we similarly have the middle efficacy of 0.7 with the highest cancer incidence and also have an efficacy of 0.9 in second. This highlights our limited compute budget in what we are able to accomplish especially when the cancer incident rate is so low. Having a significantly higher trials and a more parallel process (easier GPU access) would allow more discernible patterns to be discovered but both require much more compute.

7 Conclusion and Discussion

7.1 Risks, Ethics, and Feasibility

HPV is difficult to model because it is the most prevalent STI in the world, is frequently asymptomatic and is difficult to test for especially in men[8]. Additionally, it is possible for a person to be infected with multiple of the nearly 200 strains of HPV. Further complicating the picture is the heterogeneous population of the Atlanta Metropolitan Area, and the presence of multiple HPV vaccines. Despite all of these challenges we believe that an agent based model is feasible. This is due to its innate ability to handle extremely complicated and heterogeneous systems. Additionally, our projects feasibility is complemented by the fact that our scope is only the Atlanta Metro and not a wider area. This is because agent based modeling gets significantly more computationally complicated the larger the population that the model attempts to model.

However, there are some ethical risks in agent based modeling and the data that we use. Namely, the fact that we are using an agent based model limits our scope to a smaller area like the Atlanta Metro area. This smaller area could have small census tracts with small

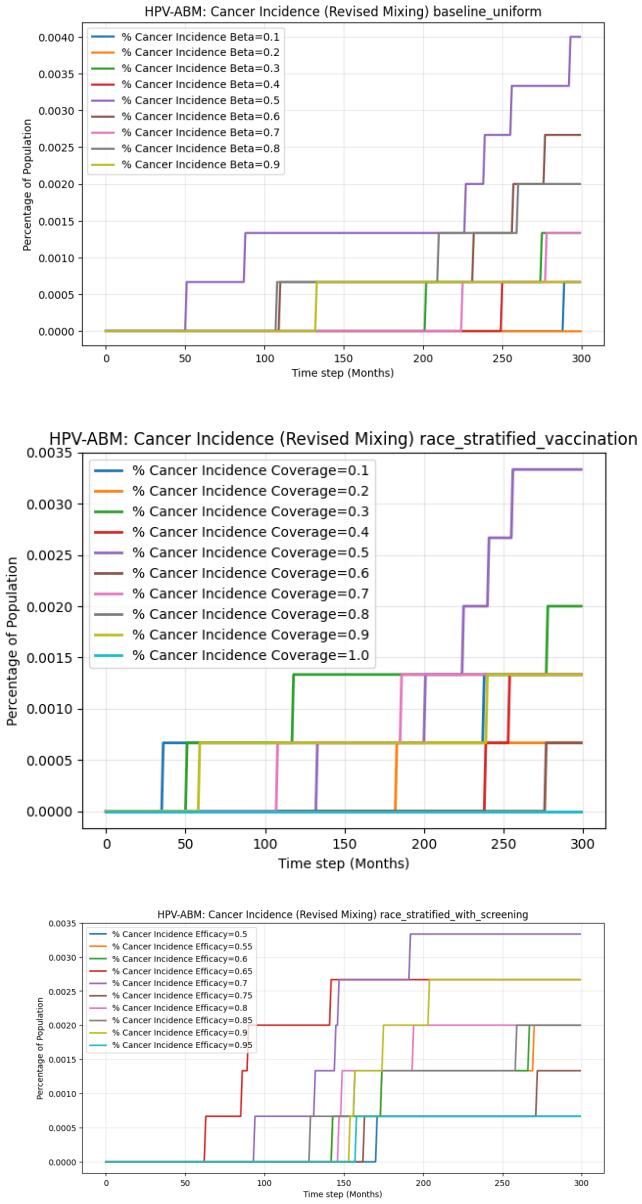

Figure 5: Lack of Patterns in Cancer incidence rate

amounts of data which could allow identification of individuals. To avoid this risk we focus on the most populated areas of the Atlanta Metro with an ample number of infections and population.

Some other risks include the fact that there is no HPV test for men and we only have data for women. This lack of data could lead to detrimental results for the model. Additionally, modeling sexual behaviors is usually difficult because of poor survey responses. Finally, a large amount of data that we collect is only on the occurrences of cancer which happen in rare cases. If the rate of HPV infections that cause cancer is misunderstood this can lead to a poor model.

7.2 Future Work

This study could be improved if there were ways to better link HPV infections to all its possible associated cancer types. While there is little data on current HPV types screening by age and gender for the Atlanta Metropolitan Area we had to rely on sparse data or national data. This would improve the model, in addition to more definitive numbers of what percentage of HPV types turn into cancer and what types of cancer. Although this is hard to measure and difficult to study it would have led to a much more straight-froward model for Atlanta specifically.

Another improvement on our current model would be to add parallelization to the model so that the agent based model can be utilized more efficiently. Agent-based modeling is very compute intensive since we have to iterate over all agents at every step. By increasing the parallelization of our code we can get more simulation data that will result in more interpretable and confident outcomes. This will help especially with very rare and improbable events like modeling the incidence of cervical cancer in a smaller population.

7.3 Public Health Implications

Our simulation results can inform real-world public health policy and vaccination strategies. By quantifying how changes in vaccine coverage, timing, and efficacy affect HPV prevalence, this model highlights the thresholds required for herd immunity in a local population such as Atlanta. These insights can guide public health agencies and school-based vaccination programs on where to focus outreach and funding to maximize impact.

Moreover, the model framework can be reused by other municipalities or researchers to evaluate HPV or other STI vaccination programs under different demographic and behavioral assumptions. Even though we do not explicitly model cervical cancer progression, the strong empirical link between persistent HPV infection and cancer incidence allows our infection-based metrics to serve as early indicators of long-term prevention success. Ultimately, this work demonstrates how computational modeling can translate epidemiological data into actionable strategies for improving community health outcomes.

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] N. W. Bulkmans, J. Berkhof, S. Bulk, M. C. Bleeker, F. J. van Kemenade, L. Rozendaal, P. J. Snijders, C. J. Meijer, and POBASCAM Study Group. 2007. High-risk HPV type-specific clearance rates in cervical screening. *British Journal of Cancer* 96, 9 (2007), 1419–1424. doi:10.1038/sj.bjc.6603653

[4] 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

[5] 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

[6] CDC. 2025. Cancers Caused by HPV. *CDC* (2025).

[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] Julia Gargano, Elissa Meites, Meg Watson, Elizabeth Unger, and Lauri Markowitz. 2022. Chapter 5: Human papillomavirus. *Centers for Disease Control and Prevention* (2022). https://www.cdc.gov/surv-manual/php/table-of-contents/chapter-5-human-papillomavirus.html

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

[10] L. T. Haderxhanaj, J. S. Leichliter, S. O. Aral, and H. W. Chesson. 2014. Sex in a lifetime: Sexual behaviors in the United States by lifetime number of sex partners, 2006–2010. *Sexually Transmitted Diseases* 41, 6 (2014), 345–352. doi:10.1097/OLQ.0000000000000132

[11] G. McQuillan, D. Kruszon-Moran, L. E. Markowitz, E. R. Unger, and R. Paulose-Ram. 2017. Prevalence of HPV in Adults Aged 18–69: United States, 2011–2014. https://www.cdc.gov/nchs/data/databriefs/db280.pdf NCHS Data Brief No. 280.

[12] National Cancer Institute (NCI) and Centers for Disease Control and Prevention (CDC). 2023. Cervical Cancer (Invasive) Incidence Rates for Georgia. State Cancer Profiles interactive data map. Data collected through SEER and NPCR.

[13] Natalie L Rincon, Kelsey Rae McDowell, Darien Weatherspoon, Tiarney D Ritchwood, Daniel J Rocke, Eric Adjei Boakye, and Nosayaba Osazuwa-Peters. 2024. Racial and ethnic disparities in human papillomavirus (HPV) vaccine uptake among United States adults, aged 27–45 years. *Human Vaccines & Immunotherapeutics* 20 (2024), 2313249. Issue 1. doi:10.1080/21645515.2024.2313249

[14] 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).

[15] John S. Santelli, Nancy D. Brener, Richard Lowry, Eve Smith, and Laurie S. Zabin. 2017. Sexual Behaviors in the United States by Lifetime Number of Sex Partners, 2006–2010. *Public Health Reports* 132, 1_suppl (2017), 83S–94S. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5795598/

[16] SEER. 2025. Rates of Cancer in GA from NCI. *GOV* (2025).

[17] Jennifer C Spencer, Jane J Kim, Jasmin A Tiro, Sarah J Feldman, Sarah C Kobrin, Celette Sugg Skinner, Lei Wang, Anne Marie McCarthy, Steve J Atlas, Sandi L Pruitt, Michelle I Silver, and Jennifer S Haas. 2023. Racial and Ethnic Disparities in Cervical Cancer Screening From Three U.S. Healthcare Settings. *American Journal of Preventive Medicine* 65, 4 (2023), 667–677. doi:10.1016/j.amepre.2023.04.016

[18] 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

[19] U.S. Census Bureau. 2025-12-01. Atlanta city, Georgia. *US Census* (2025-12-01). https://data.census.gov/all?q=Atlanta+city,+Georgia

[20] Robert Walker, Carolyn Nickson, Jie-Bin Lew, Megan Smith, and Karen Canfell. 2012. A revision of sexual mixing matrices in models of sexually transmitted infection. *Statistics in Medicine* 31, 30 (2012), 3419–3432. doi:10.1002/sim.5545