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 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 % [**?**] 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.

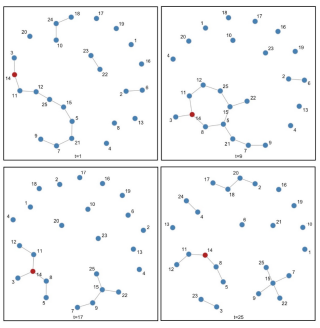

Figure 1: Forward Reachable Path Example

*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. [12], 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. [3] 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 [4]. 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. [8] 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. [5] 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. [11] 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.

3 Proposed Approach

The proposed project will incorporate the innovative hybrid approach of an agent based model and a neural network to predict and better understand the long term effects of HPV in a specific high priority area - the Atlanta Metropolitan region.

For the model initialization and customization, we will need to frame the model so that it is changed to Atlanta- Specific Data. Our framework will consist of using the HPVsim Python package [12] leveraging its structure for sexual networks - specifically with HPV. We will manually enter in population and demographic data from US Census data.

After fitting and building our initial model we will then work on calibrating the model by using a Neural Network, as discussed briefly in class, to learn the relationship between parameters, discussed in the following section.

4 Data and Parameters

Three key targets 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 [9] 2. Intermediate Outcome (Precursor): CIN2+/CIN3+ Incidence rates [7] 3.Final Outcome (Cancer): Cervical Cancer Rates in Metro Atlanta counties [10]

By evaluating our model with relevant parameters from current and relevant publications we will be able to evaluate how the model is performing. Additionally, another important data set to consider for the agent based model are the initial demographics for the Atlanta Metropolitan Area which will be retrieved through US Census Data.

5 Evaluation Methodology

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

- Calibration: Compare model-predicted prevalence and incidence to empirical Atlanta data.

- Scenario Analysis: Simulate multiple vaccination coverage and efficacy levels to evaluate reductions in prevalence and infections averted.
- Sensitivity: Test robustness to key parameters (transmission rate, clearance probability, partnership duration).

Cancer outcomes will not be modeled directly, but we will estimate potential long-term implications using infection-to-cancer transition probabilities from literature. We will interpret reductions in persistent infections as proxies for reduced cervical cancer risk.

6 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[6]. 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 are like the Atlanta Metro area. This smaller area could have small census tracts with small 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 Team Plan and Timeline

All three members will work equally on the project proposal, milestone report, final report and final presentation. Similarly the data processing and the implementation of the agent based model and its corresponding neural networks will be done equally by all three team members.

In total we have 8 weeks to work on the project. The first step is to process and clean all of the data that we have. Due to the varied sources of data that we have this will take a longer than expected time and will probably take 3 weeks. There are three sources of data of the infections. This will be equally divided with Anish handling the Transmission data, Scott the precursor data and Hailey the cancer data. Finally, we will all handle the more well understood Census data together.

After data processing we will spend the next 4 weeks on building the model. The modeling step will be worked on jointly and equally by all three members. The first model that we will build is using the HPVSim model as a baseline with which we can compare our new model with. This will take approximately half of a week to do. Additionally, at this time the milestone report is due which will also take half a week to work on. The validation framework will also be developed at this time so that we can make ample comparisons for calibration or sensitivity. For the next 3 weeks we will work on our innovative approach of using a neural network to better estimate the relationship between parameters for the agent based model. This will involve a large amount of experimentation and dealing with unknown unknowns so it deserves the large amount of time allotted to it.

Finally, the last week will be spent on summarizing the results of our new model in the form of the final report and the final presentation. This time will include rehearsing for the final presentation and will be done jointly and equally. We will seek assistance from the TAs or the professor as needed.

8 Expected Outcomes and Further Modifications

(1) A concise open-source ABM codebase with documented parameters; (2) quantitative scenario analysis of vaccination effects on HPV prevalence and incidence with uncertainty; (3) a clear discussion linking infection reductions to expected long-term cancer benefits (qualitatively); and (4) an ACM-formatted report and slides.

9 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] 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] 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).

[6] 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

[7] J. W. Gargano, R. Stefanos, R. M. Dahl, J. L. Castilho, E. A. Bostick, L. M. Niccolai, I. U. Park, S. Blankenship, M. M. Brackney, K. Chan, E. L. Delikat, S. Ehlers, K. G. Barrera, R. Kurtz, J. I. Meek, E. Whitney, M. Vigar, E. R. Unger, L. E. Markowitz, and HPV-IMPACT Working Group. 2025. Trends in Cervical Precancers Identified Through Population-Based Surveillance – Human Papillomavirus Vaccine Impact Monitoring Project, Five Sites, United States, 2008–2022. *MMWR Morbidity and Mortality Weekly Report* 74, 6 (Feb. 2025), 96–101. https://www.cdc.gov/mmwr/volumes/74/wr/pdfs/mm7406a4-H.pdf

[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] 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.

[10] 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.

[11] 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).

[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