A Data-Driven Decision-Making Framework for Spatial Agent-Based Models of Infectious Disease Spread

Emma Von Hoene[1]
Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, USA

Amira Roess
Department of Global and Community Health, George Mason University, Fairfax, VA, USA

Taylor Anderson
Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, USA

Abstract

Agent-based models (ABMs) are powerful tools used for better understanding, predicting, and responding to diseases. ABMs are well-suited to represent human health behaviors, a key driver of disease spread. However, many existing ABMs of infectious respiratory disease spread oversimplify or ignore behavioral aspects due to limited data and the variety of behavioral theories available. Therefore, this study aims to develop and implement a data-driven framework for agent decision-making related to health behaviors in geospatial ABMs of infectious disease spread. The agent decision-making framework uses a logistic regression model expressed in the form of odds ratios to calculate the probability of adopting a behavior. The framework is integrated into a geospatial ABM that simulates the spread of COVID-19 and mask usage among the student population at George Mason University in Fall 2021. The framework leverages odds ratios, which can be derived from surveys or open data, and can be modified to incorporate variables identified by behavioral theories. This advancement will offer the public and decision-makers greater insight into disease transmission, accurate predictions on disease outcomes, and preparation for future infectious disease outbreaks.

2012 ACM Subject Classification Computing methodologies → Modeling methodologies

Keywords and phrases Agent-based model, geographic information science, disease simulation, COVID-19, agent behavior, mask use

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.76

Category Short Paper

Funding National Science Foundation (Award #2030685 and #2109647).

Acknowledgements The survey used in this project was considered exempt by the George Mason University Institutional Review Board (IRB 1684418-3).

1 Introduction

In the twenty-first century, society has faced several infectious disease outbreaks, such as monkeypox, influenza, and the novel COVID-19 [2]. To mitigate these threats, decision-makers rely on epidemiological models for predicting outbreaks and assessing the impact of various interventions [1]. Compartmental models, while computationally efficient, struggle to capture heterogeneous populations, spatial interactions, and individual health behaviors - key drivers of disease trajectories [8]. An alternative approach is agent-based modeling (ABM), which explicitly represents behavior and interactions between individual "agents"

[1] corresponding author

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 76; pp. 76:1–76:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and their environment. ABMs provide the flexibility to assign heterogeneous attributes and decision-making processes to each agent. In addition, spatial data can be used to represent the built environment, facilitating the representation of individual movements and interactions in space [8]. The ability to capture complex behaviors of individuals makes ABM an ideal approach for better understanding the spread of diseases and supporting decision-making.

Health behaviors, actions that affect disease transmission, are often overlooked or simplified in existing ABMs. This is mainly due to lack of data to inform agents' behavioral parameters, competing theories of health behavior to draw from, and institutional challenges limiting interdisciplinary collaboration among modelers and domain experts [5]. For instance, Perez and Dragicevic [14] model the spatial spread of disease without considering behavioral responses like staying home when sick. Other models impose behaviors on a set of agents without considering their individual characteristics, beliefs, or perceptions. For example, Li et al. [10] compare COVID-19 outcomes using scenarios with different percentages of randomly selected agents who are considered vaccinated. More complex representations of health behavior range from the use of social contagion of adopting behaviors along a social network [9], game theory or the rational choice model to inform adoption decisions [17], and fuzzy cognitive maps [12]. However, in general, most models of disease spread either ignore or incorporate health behaviors in an ad hoc manner without leveraging behavioral data or theories. Therefore, a more comprehensive framework that has the potential to leverage both data and theories for simulating health behaviors in ABMs of disease spread is needed.

Realistically incorporating human behaviors into ABMs of infectious disease spread demands a combination of data-driven and theoretical approaches. The objective of this study is to develop and implement a data-driven agent decision framework that improves the representation of health behaviors in geospatial ABMs of infectious disease spread. The framework is intended to leverage behavioral data and theories by implementing a logistic regression model with a geospatial ABM of infectious disease. This is demonstrated by simulating the spread of COVID-19 and mask-usage behavior among the undergraduate student population at George Mason University's (GMU) Fairfax campus in Fall 2021.

2 Methods

2.1 Data

Built Environment Data. The model environment consists of GMU's Fairfax campus, which was created using shapefiles obtained from GMU's open geoportal. Datasets capturing the buildings, parking lots, walkways, and water features were acquired and updated to reflect the built environment on campus in 2021. Within each building are sublocations in which agents interact, for example representing individual classrooms.

Data Informing University Patterns of Life. The agent schedules were generated using GMU's course data. All in-person courses offered during the Fall 2021 semester were collected from PatriotWeb, GMU's openly available database with a schedule of classes. 2,004 courses were collected, along with details on their course number, course section, meeting times, total seats, building and classroom. The data was processed into variables that directed the agents to enroll in courses that corresponded with their undergraduate program, attend class at the scheduled times, and travel to their course's designated building and classroom.

Health Behavior Data. This study utilizes data from a survey conducted in August 2021 that is representative of the United States [16]. The survey gathered information on factors influencing mask-wearing decisions when masks were optional in the U.S. 3,528 respondents

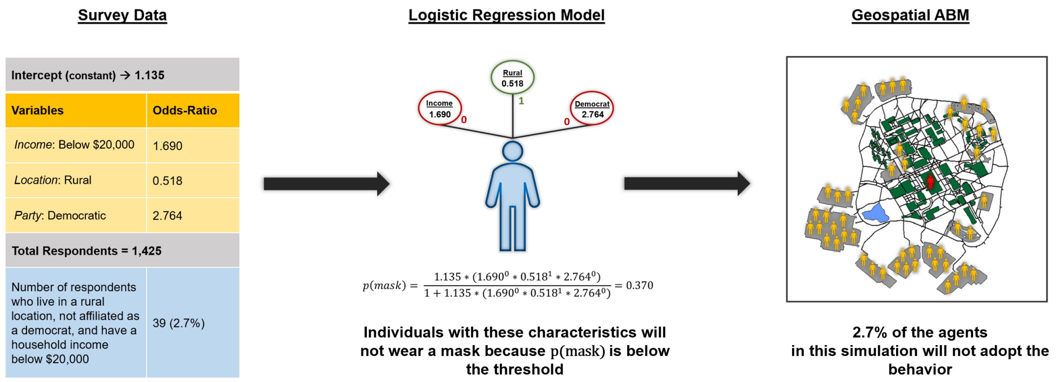

Figure 1 A representation of how the agent decision-making framework works.

participated, providing socio-demographic details such as age, gender, ethnicity, income, political party affiliation, and rural or urban residency. The study exclusively focused on respondents under the age of 40, resulting in a total of 1,425 participants. Respondents were asked how often they decided to wear a mask. Odds ratios were calculated to capture the association between the socio-demographic variables and individuals who would wear a mask 3 or more times a week. The survey results indicated that only income, residence location, and political party affiliation were statistically significant variables for predicting mask use.

Disease Data for Calibration. The model is calibrated using data obtained from GMU's Campus COVID-19 Data Archive. The dashboard includes COVID-19 case, testing, and vaccine data among the Mason community during the Fall 2021 semester, spanning from August 16, 2021 to December 17, 2021.

2.2 Agent Behavioral Framework

The proposed agent decision framework determines how agents in a geospatial ABM make decisions about whether or not to adopt a specific health behavior. A visual representation of how the agent decision-making framework is applied to the geospatial ABM in this study is presented in Figure 1. The framework is adapted from the methods presented by Durham and Casman [4] that uses a standard logistic regression model to calculate the probability that an agent will adopt a behavior based on four antecedents, defined by the Health Belief Model. The logistic regression model is expressed in terms of odd ratios, which are informed by survey data. We modify the methods presented by Durham and Casman [4] to include the variables that we have recognized from our survey data as important predictors of mask usage. These variables are outlined in Section 2.1 and are expressed in Equation 1:

$$p(\text{mask}) = \frac{OR_0 \cdot \prod_i OR_i^{X_i}}{1 + OR_0 \cdot \prod_i OR_i^{X_i}}, \quad i = 1, \dots, 3 \tag{1}$$

The values of $i = 1, \dots, 3$ denotes each of the independent variables that will be used, including income, residence location, and political party. OR_i is the value of the odds ratio. X_i is a binary variable that indicates the state of the independent variable where 1 is true and 0 is false. OR_0 is a constant probability when all X_i variables are low. The probability of behavior, $p(mask)$, is a value from 0 to 1. A threshold is used to determine whether the individual adopts the behavior. When $p(mask) >$ threshold, the individual will uptake the behavior, and if $p(mask) <$ threshold, the individual will not uptake the given behavior.

2.3 Geospatial ABM

The geospatial ABM was developed and integrated with an agent decision framework that can collectively leverage both behavioral theories and data to realistically implement human health behavior during an infectious disease outbreak. The model aims to demonstrate this framework by simulating the spread of COVID-19 and mask-wearing behavior among the undergraduate population at GMU's Fairfax campus in Fall 2021.

Agent Characteristics and Scheduling. The model consists of agents representing undergraduate students who were registered for in-person classes in the Fall 2021 semester and enrolled full-time. University data informed whether the student agent lives on or off campus, as well as the elective and core courses associated with their undergraduate degree program. Based on the joint distributions found in the survey data, agents' demographic profiles are generated to include three binary characteristics: residence location (rural or otherwise), political affiliation (democratic or otherwise), and household income (below $20,000 or otherwise). Given the odds ratios from the survey data and the demographic profile of the agent, the probability of mask-wearing from 0 to 1 for each agent is assigned. Additional agent characteristics that affect the disease transmission include health status, symptomatic status, and quarantine status. The length of the incubation and infectious periods for infectious agents is determined from a normal distribution informed by COVID-19 literature. All agents are assumed vaccinated as it was mandatory to be on campus during the Fall 2021 semester.

The model processes discrete time steps representing one second, captures weekly patterns (Monday-Friday from 7:15am to 11pm), and stops at the end of the semester. Agents follow their class schedule generated at the initialization of the model, beginning each day at their home, represented in the model as either a parking lot or a residential building on campus. They leave for their first class 15 minutes before it starts and travel to each class throughout the day, updating the building and sublocation where they are located. If an off-campus agent has no class following their previous one, they go to a student center or gym until the next class, while on-campus agents have a 50 percent chance of going to a student center/gym or returning to their dorm. After attending all their classes for the day, agents return home.

Disease Transmission. Since COVID-19 was simulated in the Fall 2021 semester, literature was used to define the parameters of the Delta variant. There is a probability that susceptible agents are exposed to the virus if they come in contact with an infected agent, known as the transmission rate, which is calibrated to 0.05 (see Initial Calibration). However, if the agent is wearing a mask, then the transmission rate is reduced by 50% [6], resulting in a transmission rate of 0.025. Additionally, all susceptible agents have a 0.02 off-campus transmission rate, a value obtained from other disease spread models on a university campus [6].

Once an agent becomes exposed, they remain in the exposed stage for an average of 4.41 days, after which they have an 85% chance of becoming symptomatic [18, 13]. Both symptomatic and asymptomatic agents stay in the infectious period for an average of 8 days [7]. However, this infectious period includes a pre-symptomatic stage of two days because individuals can spread the virus 48 hours before symptoms appear [15]. After the pre-symptomatic stage, symptomatic agents begin to quarantine, meaning they do not go to campus until their infectious period has ended, where asymptomatic agents continue to follow their schedules [15]. Individuals have immunity for 90 days, which implies that agents remain in the recovered period for that length of time since they cannot be re-infected [3].

Health Behavior Framework. At the initialization of the model, agents decide to wear a mask for the period of a semester based on their demographic profile. A logistic regression model for each agent determines based on the combination of their characteristics what the probability of mask use is from 0 to 1. Agents will choose to wear a mask if that value is greater than a 0.50 threshold, which may be calibrated in the future. The model currently does not incorporate individual learning, sensing, or prediction. Future work will include agent perception leading to dynamic health behaviors by incorporating different driving variables such as the perceived severity or perceived susceptibility of the disease.

Initialization. The model is implemented with Repast Simphony, a freely available Java-based modeling toolkit, and built upon a Repast Simphony program called RepastCity [11]. The model is currently not accessible online, but it will be made available in the future once the work is completed. Before the model begins running, one agent is selected to be infectious, and 3.5% of agents are set to be recovered. The model is initialized with 5,000 agents, which captures 37.5% of the estimated campus population during the Fall 2021 semester. While the model will be upscaled to include 13,500 agents in future work, for the current study, we limited the simulation to 5,000 agents to test the proposed agent decision-making framework.

Initial Calibration. GMU's Fall 2021 COVID-19 data archive reported 399 symptomatic cases throughout the semester with 100% mask use. Since we simulate roughly 37% of the campus population, we expect around 148 symptomatic cases in our model. We calibrate the model by modifying the transmission rate in the 100% mask-use scenario so that the number of infectious and symptomatic agents throughout the semester corresponds with the data.

3 Results

To address the variation in model results due to randomness within ABM processes, the model was run 50 times for a duration of 105 days to represent a full 15-week semester at GMU. Future work will validate disease outcomes using data that was not used in model calibration prior to running scenarios and generating final results. However, we present initial results for three scenarios here: 1) no mask, 2) 100% mask usage, and 3) mask usage determined by the agent's behavioral framework, where the demographic profiles in the population determine the adoption of masks. The results are presented in Figure 2.

The preliminary results indicate that the number of cumulative cases was on average 253 for no mask-usage, 165 for mask-usage determined by the agent behavioral framework, and 148 for 100% mask-usage. No-mask usage results in a higher peak infection level and greater variation, whereas 100% mask usage leads to less variation across simulation runs and a lower percentage of infections. The findings from the behavioral framework scenario are comparable to those of 100% mask usage as most of the agent population chose to wear masks based on their demographic profile.

4 Discussion and Conclusion

Existing ABMs of infectious respiratory disease spread often overlook or oversimplify the complexities of human health behaviors. To address this gap, this study proposes a novel agent decision making framework with the potential to integrate data-driven and theoretical approaches. Limitations of this work include the use of national survey data to represent a university population. Future work may explore the use of open data that better corresponds

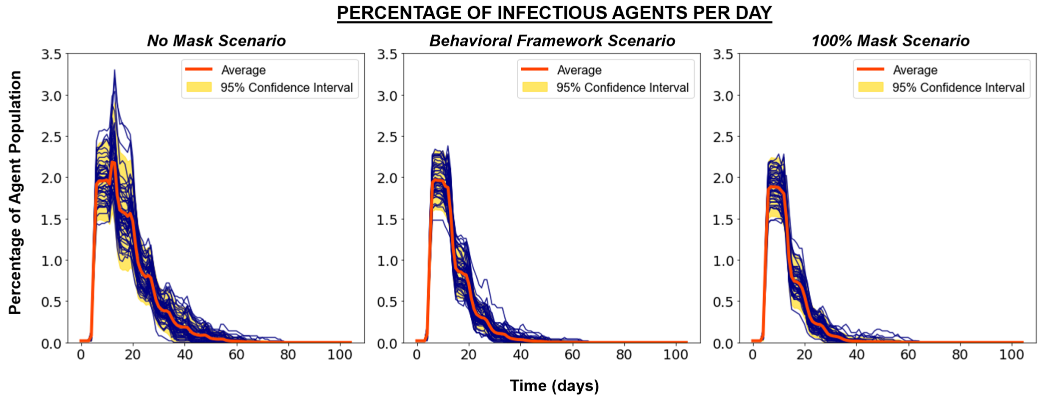

Figure 2 The results for each tested scenario, showing the daily percentage of infectious agents.

with the study area. Although, the agents' health behaviors emerge as a function of each agent's demographic profile, they are unchanging. Future work will explore the effect of agent perceptions on behavior, enabling agents to adapt and respond to new situations. This study aims to advance ABMs of infectious disease spread by improving the representation of how humans respond to diseases, which will ultimately offer the public and decision-makers support with accurate predictions and intervention strategies for future outbreaks.

References

1 L. Berger and et al. Rational policymaking during a pandemic. *PNAS*, 118(4), 2021.

2 D.E. Bloom and D. Cadarette. Infectious disease threats in the twenty-first century: Strengthening the global response. *Frontiers in Immunology*, 10, 2019.

3 CDC. Reinfection: Clinical considerations for care of children and adults with confirmed covid-19, 2020. URL: `https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/clinical-considerations-reinfection.html`.

4 D. Durham and E. Casman. Incorporating individual health-protective decisions into disease transmission models: a mathematical framework. *J R Soc Interface*, 9(68):562–570, 2012.

5 S. Funk, M. Salathé, and V. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface*, 7(50):1247–1256, 2010.

6 P.T. Gressman and J.R. Peck. Simulating covid-19 in a university environment. *Mathematical Biosciences*, 328, 2020.

7 X. He and et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature Medicine*, 26:672–675, 2020.

8 E. Hunter, B. Mac Namee, and J. Kelleher. A taxonomy for agent-based models in human infectious disease epidemiology. *JASS*, 20(3), 2017.

9 D.A. Levy and P.R. Nail. Contagion: a theoretical and empirical review and reconceptualization. *Genetic, social, and general psychology monographs*, 119(12):233–284, 1993.

10 Z. Li, J.L. Swann, and P. Keskinocak. Value of inventory information in allocating a limited supply of influenza vaccine during a pandemic. *PLOS ONE*, 13(10), 2018.

11 N. Malleson. Repastcity, 2012. URL: `https://code.google.com/archive/p/repastcity/`.

12 S. Mei and et al. Individual Decision Making Can Drive Epidemics: A Fuzzy Cognitive Map Study. *IEEE Transactions on Fuzzy Systems*, 22(2):264–273, 2014.

13 T. Ogata and et al. A low proportion of asymptomatic covid-19 patients with the delta variant infection by viral transmission through household contact at the time of confirmation in ibaraki, japan. *ISPRS International Journal of Geo-Information*, 4(3):192–196, 2022.

14 L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *International journal of health geographics*, 8(1):1–17, 2009.

15 M. Reveil and Y.H. Chen. Predicting and preventing covid-19 outbreaks in indoor environments: an agent-based modeling study. *Scientific Reports*, 12, 2022.

16 A. Roess and et al. Predictors of firearm purchasing during the coronavirus pandemic in the united states: a cross-sectional study. *Public health*, 219:159–164, 2023.

17 J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior, 2nd rev.* Princeton university press, 1947.

18 Wu Y. and et al. Incubation period of covid-19 caused by unique sars-cov-2 strains: A systematic review and meta-analysis. *JAMA*, 5(8), 2022.