

COVASIM-G: A METHOD FOR PROJECTING COVID-19 HEALTH BURDEN BY DEMOGRAPHIC AND GEOGRAPHIC GROUPS

Alisa Hamilton^{1,2}, Geoffrey Clapp¹, Colton Ragland³, Alexander Tulchinsky⁴, Eili Klein^{4,5},
and Gary Lin¹

¹Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA

²Dept. of Civil and Systems Eng., Johns Hopkins University, Baltimore, MD, USA

³Dept. of Computer Science, University of Texas, Austin, TX, USA

⁴One Health Trust, Washington, DC, USA

⁵Dept. of Emergency Medicine, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Estimating infectious disease burden across social stratifiers with defined spatial scale is critical for designing policy responses to reduce overall burden and health disparities (Buckee, Noor, and Sattenspiel 2021). In this study, we developed a method (Covasim-G) for using the open-source agent-based model, Covasim, with a geographically realistic synthetic population of Maryland. Covasim-G was able to simulate cumulative COVID-19 deaths per 100,000 by age, gender, race, and county similar to observed data. Covasim-G can be used for any US location in the Census, offering a useful tool for scenario modeling.

1 INTRODUCTION

While many mathematical models projected health burden during the COVID-19 pandemic, only a minority stratified populations by geographic sub-unit and demographic groups (Hamilton et al. 2024). In this study, we developed a method (Covasim-G) for using the open-source agent-based model, Covasim (Kerr et al. 2021), with a geographically realistic synthetic population of Maryland, USA.

2 METHODS

Covasim (COVID-19 Agent-based Simulator) is an open-source model developed by the Institute for Disease Modeling in which agents transition through epidemic states (e.g., susceptible, infected, recovered). Covasim includes information on age structure and population size for US states and transmission networks for community, household, school, workplace, and long-term care facilities; however, it does not include any spatial data. GREASYPOP (Geographically Realistic Synthetic Population) is an open-source software developed by One Health Trust that produces synthetic populations with spatial data (Tulchinsky et al. 2024). GREASYPOP generates a population of households using US Census data and assigns individuals to workplaces and schools using data on commute patterns, employment statistics, and school enrollment. GREASYPOP agents also have attributes for race, workplace categories, and income levels.

We processed GREASYPOP files generated for a population of Maryland, District of Columbia, and northern Virginia to be readable by Covasim. We then used Covasim, with minor adjustments to the default parameters (manually adjusted transmission and death rates), to simulate an epidemic between 1 February and 30 June 2020 (the first wave of the pandemic). We recorded the epidemic state of each agent throughout the simulation and compared outcomes by subgroup on 11 May 2020, the date roughly corresponding to the day of peak deaths. Deaths are considered a more reliable metric than cases as many cases are not reported early in an epidemic. We compared simulated deaths per 100,000 Maryland residents with observed data (Council on Open Data 2024), disaggregating by age, gender, race, and county (Table 1).

3 RESULTS

Table 1. Simulated and observed deaths per 100k by demographic group and county on 11 May 2020 in Maryland, USA

Demographic Group	Simulated deaths per 100k	Observed deaths per 100k	County	Simulated deaths per 100k	Observed deaths per 100k	County	Simulated deaths per 100k	Observed deaths per 100k
Age 0-9	0.1	0.0	24001	38.8	17.0	24027	26.3	9.5
Age 10-19	0.3	0.0	24003	21.8	21.4	24029	37.2	66.9
Age 20-29	0.7	1.4	24005	30.1	23.4	24031	27.3	36.9
Age 30-39	1.5	2.5	24009	23.6	11.9	24033	24.3	39.5
Age 40-49	6.3	5.7	24011	28.1	0.0	24035	31.2	17.9
Age 50-59	15.2	13.4	24013	29.1	34.4	24037	15.4	7.9
Age 60-69	38.7	34.4	24015	33.9	12.6	24039	15.6	0.0
Age 70-79	110.6	93.2	24017	23.6	33.1	24041	47.9	2.7
Age 80+	342.4	319.3	24019	32.1	6.3	24043	30.5	6.0
Gender Female	28.6	26.0	24021	24.9	28.5	24045	21.5	15.4
Gender Male	23.7	28.9	24023	24.1	0.0	24047	40.0	7.7
Race Black alone	23.0	36.6	24025	28.8	9.8	24510	22.7	28.8
Race Hispanic	9.1	18.8						
Race White non-Hispanic	31.6	22.8						

4 DISCUSSION

Covasim-G was able to simulate COVID-19 deaths per 100,000 by age, gender, race and county within the same order of magnitude as observed data for Maryland with few exceptions. Parameters were manually adjusted, and transmissions by layer are currently a function of the number of agents and their contacts. Future iterations of this work will include formal calibration and vary the probability of infection per contact by layer. Covasim-G offers a useful tool for retrospective analyses and scenario modeling because it can be used for any US location in the Census, and researchers can calibrate to data and modify the disease model according to their research questions. During the next pandemic, it will be important to leverage existing open-source tools, as we have done in this study, to inform resource and policy decisions that could reduce overall burden as well as disparities between geographic regions and sociodemographic groups.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Health and Human Services Office of the Administration for Strategic Preparedness and Response.

CODE AVAILABILITY

All code is publicly available at <https://github.com/JHUAPL-Matador/Covasim-G>.

REFERENCES

- Council on Open Data. 2024. "Maryland's Open Data Portal." Maryland, USA. <https://opendata.maryland.gov/>.
- Hamilton, Alisa, Fardad Haghpahan, Alexander Tulchinsky, Nodar Kipshidze, Suprena Poleon, Gary Lin, Hongru Du, Lauren Gardner, and Eili Klein. 2024. "Incorporating Endogenous Human Behavior in Models of COVID-19 Transmission: A Systematic Scoping Review." *Dialogues in Health* 4 (June):100179. <https://doi.org/10.1016/j.dialog.2024.100179>.
- Kerr, Cliff C., Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, et al. 2021. "Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions." *PLOS Computational Biology* 17 (7): e1009149. <https://doi.org/10.1371/journal.pcbi.1009149>.
- Tulchinsky, Alexander, Fardad Haghpahan, Alisa Hamilton, Nodar Kipshidze, and Eili Klein. 2024. "Generating Geographically and Economically Realistic Large-Scale Synthetic Contact Networks: A General Method Using Publicly Available Data." *arXiv*, June. <https://doi.org/10.48550/arXiv.2406.14698>.