Advancing Synthetic Population Generation with Health Characteristics for Agent-Based Models and Small-Area Estimation

Emma Von Hoene, George Mason University, Department of Geography and Geoinformation Science
Dr. Amira Roess, George Mason University, Department of Global and Community Health
Dr. Hamdi Kavak, George Mason University, Department of Computational and Data Sciences
Dr. Taylor Anderson, George Mason University, Department of Geography and Geoinformation Science

College of Science
DEPARTMENT OF GEOGRAPHY AND GEOINFORMATION SCIENCE
George Mason University

John Hopkins University – Invited Presentation for Dr. Gardner's Lab
February 3rd, 2026

MORE ABOUT ME!



- **Education** at George Mason University:
  - Current PhD Candidate in Earth Systems and Geoinformation Science
  - Master of Science in Geospatial Intelligence and Geoinformatics
  - Bachelor of Science in Geography

- **Research** focus:
  - Development of state-of-the-art methodologies
  - Spatial machine learning, data analytics and modeling and simulation
  - Currently working with applications in public health

- Active **internship** with US Department of Defense



Satellite Imagery of George Mason University.

PRESENTATION OUTLINE

1. **Overview** of recent and ongoing work (Brief roadmap of publications and current projects)
2. **Background and motivation** for generating synthetic populations with health characteristics
3. **Paper 1:** A novel application of synthetic population generation with public health characteristics for spatial agent-based models
4. **Paper 2:** Evaluating the transferability of the synthetic population generation approach across public health applications
5. **Paper 3:** Development and operationalization of the synthetic population generation approach into the Spatial Health and Population Estimator (SHAPE) framework
6. **Paper 4:** SHAPE outputs — 2023 and 2024 national synthetic populations and small-area estimates for the United States

1. NOVEL APPLICATIONS OF SYNTHETIC POPULATION GENERATION AND SAE FOR HEALTH

How can we generate synthetic populations with health characteristics (attitudes/perceptions, behaviors, outcomes)?



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Synthetic population generation with public health characteristics for spatial agent-based models

Emma Von Hoene[1]*, Amira Roess[2], Hamdi Kavak[3], Taylor Anderson[1]

1 Department of Geography and Geoinformation Science, George Mason University, Fairfax, Virginia, United States of America, 2 Department of Global and Community Health, George Mason University, Fairfax, Virginia, United States of America, 3 Department of Computational and Data Sciences, George Mason University, Fairfax, Virginia, United States of America

* evonhoen@gmu.edu

Check for updates

OPEN ACCESS

Citation: Von Hoene E, Roess A, Kavak H, Anderson T (2025) Synthetic population generation with public health characteristics for spatial agent-based models. PLoS Comput Biol 21(3): e1012439. https://doi.org/10.1371/journal.pcbi.1012439

Editor: Nicholas Geard, University of Melbourne, AUSTRALIA

Received: August 26, 2024

Accepted: February 10, 2025

Published: March 17, 2025

Copyright: © 2025 Von Hoene et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The individual-level nationally representative survey collected by researchers, the validation dataset, and the

Abstract

Agent-based models (ABMs) simulate the behaviors, interactions, and disease transmission between individual "agents" within their environment, enabling the investigation of the underlying processes driving disease dynamics and how these processes may be influenced by policy interventions. Despite the critical role that characteristics such as health attitudes and vaccination status play in disease outcomes, the initialization of agent populations with these variables is often oversimplified, overlooking statistical relationships between attitudes and other characteristics or lacking spatial heterogeneity. Leveraging population synthesis methods to create populations with realistic health attitudes and protective behaviors for spatial ABMs has yet to be fully explored. Therefore, this study introduces a novel application for generating synthetic populations with protective behaviors and associated attitudes using public health surveys instead of traditional individual-level survey datasets from the census. We test our approach using two different public health surveys to create two synthetic populations representing individuals aged 18 and over in Virginia, U.S., and their COVID-19 vaccine attitudes and uptake as of December 2021. Results show that integrating public health surveys into synthetic population generation processes preserves the statistical relationships between vaccine uptake and attitudes in different demographic groups while capturing spatial heterogeneity at fine scales. This approach can support disease simulations that aim to explore how real populations might respond to interventions and how these responses may lead to demographic or geographic health disparities. Our study also demonstrates the potential for initializing agents with variables relevant to public health domains that extend beyond infectious diseases, ultimately advancing data-driven ABMs for geographically targeted decision-making.

Von Hoene et al. (2025), PLOS Computational Biology: *https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012439*

1. NOVEL APPLICATIONS OF SYNTHETIC POPULATION GENERATION AND SAE FOR HEALTH

Can our synthetic population generation approach be used for other public health applications?

Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

EVALUATING THE TRANSFERABILITY OF A SYNTHETIC POPULATION GENERATION APPROACH FOR PUBLIC HEALTH APPLICATIONS

Emma Von Hoene[1], Aanya Gupta[2], Hamdi Kavak[3], Amira Roess[4], and Taylor Anderson[1]

[1]Dept. of Geography and Geoinformation Science, George Mason University, Fairfax, USA
[2]Thomas Jefferson High School for Science and Technology, Alexandria, USA
[3]Dept. of Computational and Data Sciences, George Mason University, Fairfax, USA
[4]Dept. of Global and Community Health, George Mason University, Fairfax, USA

ABSTRACT

Simulations are valuable in public health research, with synthetic populations enabling realistic policy analysis. However, methods for generating synthetic populations with domain-specific characteristics remain underexplored. To address this, we previously introduced a population synthesis approach that directly integrates health surveys. This study evaluates its transferability across health outcomes, locations, and timeframes through three case studies. The first generates a Virginia population (2021) with COVID-19 vaccine intention, comparing results to probabilistic and regression-based approaches. The second synthesizes populations with depression (2021) for Virginia, Tennessee, and New Jersey. The third constructs Virginia populations with smoking behaviors for 2021 and 2022. Results demonstrate the method's transferability for various health applications, with validation confirming its ability to capture accuracy, statistical relationships, and spatial heterogeneity. These findings enhance population synthesis for public health simulations and offer new datasets with small-area estimates for health outcomes, ultimately supporting public health decision-making.

DOI: 10.1109/WSC68292.2025.11338931

Von Hoene et al. (2025), Winter Simulation Conference:
https://ieeexplore.ieee.org/abstract/document/11338931?casa_token=ylobMLkG90QAAAAA:FqFI9Njan20fkrHIlNisX62tGQZ8fAitsIM-hk71Wgi3pCwOsPXADlNjFu6npxMvzte8Uqv-3Q

1. NOVEL APPLICATIONS OF SYNTHETIC POPULATION GENERATION AND SAE FOR HEALTH

Can we expand our synthetic population generation approach into a fully developed and evaluated framework for synthesizing health data for different outcomes, locations, time periods and spatial scales?

Evaluation of A Spatial Microsimulation Framework for Small-Area Estimation of Population Health Outcomes Using the Behavioral Risk Factor Surveillance System

Emma Von Hoene[1]*, Aanya Gupta[2], Hamdi Kavak[3], Amira Roess[4], Taylor Anderson[1]

Abstract

Public health decision-making, including identifying health inequalities and planning targeted interventions, relies on population health data across multiple geographic scales. However, such ground truth data is often unavailable. To address this, researchers use small area estimation (SAE) techniques to model health outcomes at county and census tract levels. While multilevel regression modeling is commonly applied, spatial microsimulation has emerged as a state-of-the-art capable of generating estimates at the individual level that can be aggregated to the area level. However, applications of microsimulation in health-related SAE remain narrow and its broader potential for producing reliable, multi-outcome health estimates remains underexplored. To address this, we systematically applied and evaluated spatial microsimulation-based SAE across multiple health risk behaviors and health outcomes, settings, and scales. This study introduces the **Spatial Health and Population Estimator (SHAPE)**, a spatial microsimulation framework that applies hierarchical iterative proportional fitting (IPF) to estimate two health risk behaviors and eleven health outcomes across multiple spatial scales. SHAPE was evaluated using county-level direct estimates from the Behavioral Risk Factor Surveillance System (BRFSS) and both county- and census tract–level data from CDC PLACES for New York (2021) and Florida (2019). Results show that SHAPE's SAEs are moderately consistent with BRFSS ($\bar{r} \approx 0.5$), similar to CDC PLACES ($\bar{r} \approx 0.6$), and strongly aligned with CDC PLACES model-based estimates at both county ($\bar{r} \approx 0.8$) and census tract ($\bar{r} \approx 0.7$) levels. SHAPE is an open, reproducible, and transparent framework programmed in R that meets a need for accessible SAE methods in public health. By constructing synthetic populations that preserve joint distributions among individual characteristics, SHAPE enables detailed analysis of health disparities and supports downstream modeling such as microsimulations and agent-based simulations.

Keywords: Small Area Estimation; Spatial Microsimulation; Population Health; Health Disparities; Behavioral Risk Factor Surveillance System (BRFSS); CDC PLACES

Von Hoene et al. (Submitted) to International Journal of Health Geographics: *https://arxiv.org/pdf/2510.22080*

1. NOVEL APPLICATIONS OF SYNTHETIC POPULATION GENERATION AND SAE FOR HEALTH

Evaluation of A Spatial Microsimulation Framework for Small-Area Estimation of Population Health Outcomes Using the Behavioral Risk Factor Surveillance System

Emma Von Hoene[1]*, Aanya Gupta[2], Hamdi Kavak[3], Amira Roess[4], Taylor Anderson[1]

Abstract

Public health decision-making, including identifying health inequalities and planning targeted interventions, relies on population health data across multiple geographic scales. However, such ground truth data is often unavailable. To address this, researchers use small area estimation (SAE) techniques to model health outcomes at county and census tract levels. While multilevel regression modeling is commonly applied, spatial microsimulation has emerged as a state-of-the-art capable of generating estimates at the individual level that can be aggregated to the area level. However, applications of microsimulation in health-related SAE remain narrow and its broader potential for producing reliable, multi-outcome health estimates remains underexplored. To address this, we systematically applied and evaluated spatial microsimulation-based SAE across multiple health risk behaviors and health outcomes, settings, and scales. This study introduces the **Spatial Health and Population Estimator (SHAPE)**, a spatial microsimulation framework that applies hierarchical iterative proportional fitting (IPF) to estimate two health risk behaviors and eleven health outcomes across multiple spatial scales. SHAPE was evaluated using county-level direct estimates from the Behavioral Risk Factor Surveillance System (BRFSS) and both county- and census tract–level data from CDC PLACES for New York (2021) and Florida (2019). Results show that SHAPE's SAEs are moderately consistent with BRFSS ($\bar{r} \approx 0.5$), similar to CDC PLACES ($\bar{r} \approx 0.6$), and strongly aligned with CDC PLACES model-based estimates at both county ($\bar{r} \approx 0.8$) and census tract ($\bar{r} \approx 0.7$) levels. SHAPE is an open, reproducible, and transparent framework programmed in R that meets a need for accessible SAE methods in public health. By constructing synthetic populations that preserve joint distributions among individual characteristics, SHAPE enables detailed analysis of health disparities and supports downstream modeling such as microsimulations and agent-based simulations.

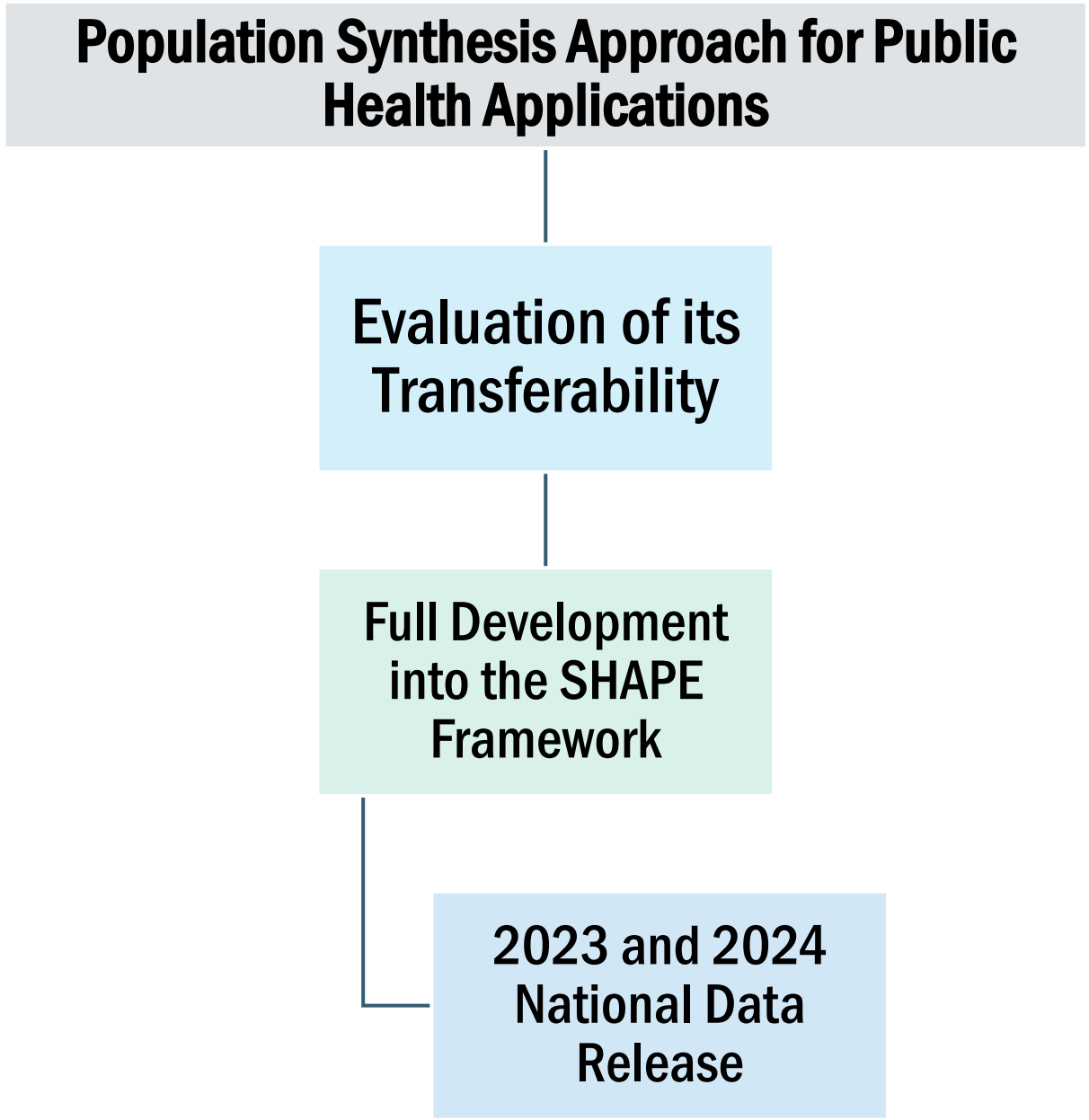
Keywords: Small Area Estimation; Spatial Microsimulation; Population Health; Health Disparities; Behavioral Risk Factor Surveillance System (BRFSS); CDC PLACES

Nature Scientific Data – *Manuscript Being Prepared*

Can we apply our SHAPE framework to generate recent health-relevant synthetic populations and SAEs for the entire U.S.?

Von Hoene et al. (Submitted) to International Journal of Health Geographics: *https://arxiv.org/pdf/2510.22080*

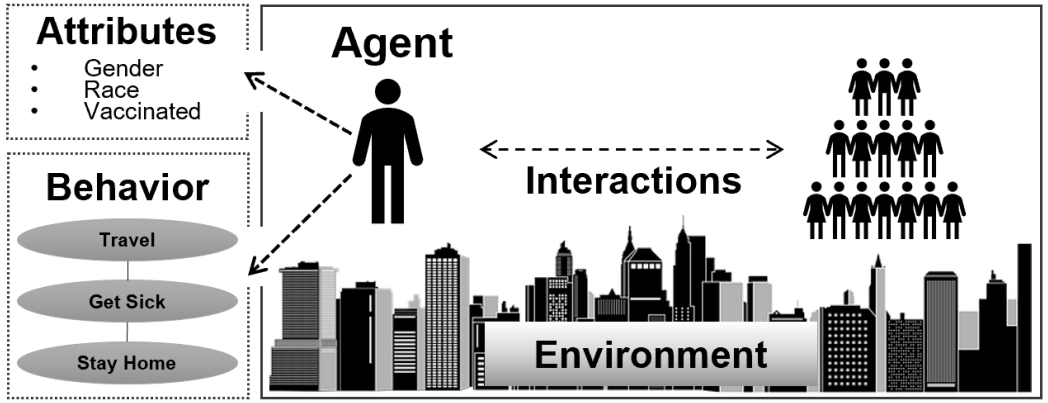
1. NOVEL APPLICATIONS OF SYNTHETIC POPULATION GENERATION AND SAE FOR HEALTH



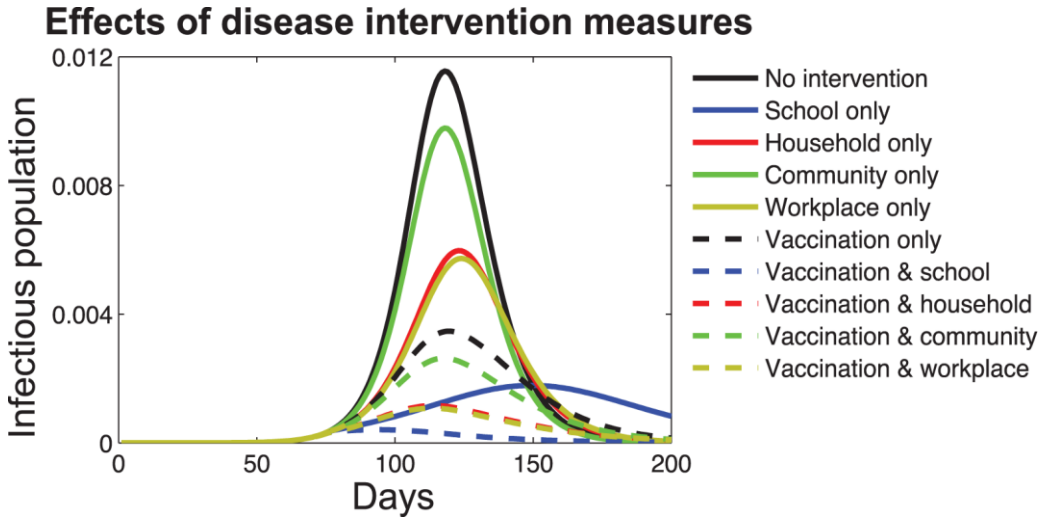
2. BACKGROUND – AGENT-BASED MODELING AND DISEASE SPREAD

- Simulations, such as **agent-based models (ABMs)**, are essential for studying infectious disease spread

- **Health behaviors** are key drivers of disease trajectories



Conceptual representation of an ABM.

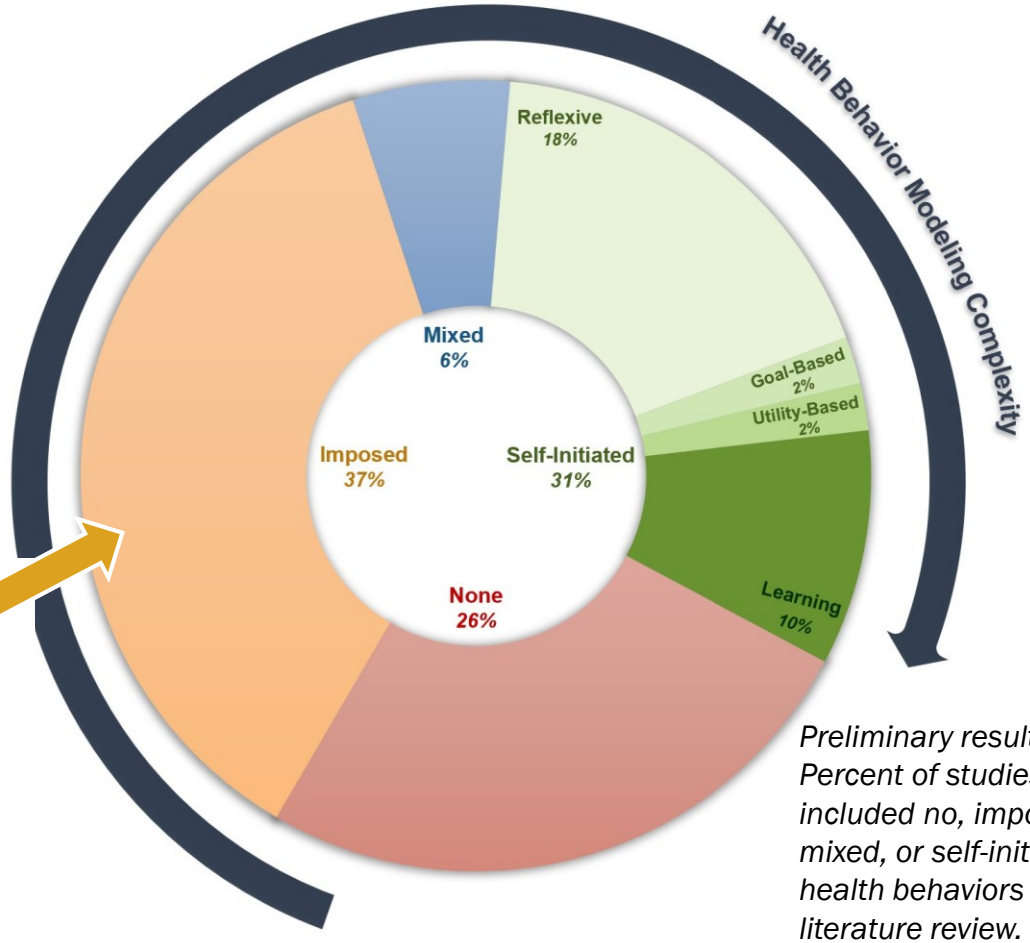


Infectious disease dynamics with vaccination and varying contact reduction measures (Xia et al., 2013).

Xia, S., Liu, J., & Cheung, W. (2013). Identifying the relative priorities of subpopulations for containing infectious disease spread. PloS one, 8(6), e65271.

2. BACKGROUND – BEHAVIORAL LIMITATIONS IN ABM OF DISEASE SPREAD

- ABMs across a range of public health applications traditionally ***ignore*** or and continue to ***oversimplify*** the human behavior component





Preliminary results: Percent of studies that included no, imposed, mixed, or self-initiated health behaviors in our literature review.



JASSS

THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

All Models Are Wrong, But Can They Be Useful? Lessons from COVID-19 Agent-Based Models: A Systematic Review

Emma Von Hoene [1], Sara Von Hoene [1], Szandra Péter [1], Ethan Hopson [1], Emily Csizmadia [1], Faith Fenyk [1], Kai Barner [1], Timothy Leslie [1], Hamdi Kavak [3], Andreas Züfle [4], Amira Roess [2], Taylor Anderson [1]

[1] *Department of Geography and Geoinformation Science, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[2] *Department of Global and Community Health, College of Human and Health Services, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[3] *Department of Computational and Data Sciences, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[4] *Department of Computer Science, Emory University, 201 Dowman Drive, Atlanta, Georgia, 30322*
Correspondence should be addressed to *tander6@gmu.edu*

Journal of Artificial Societies and Social Simulation 29(1) 7, 2026
Doi: 10.18564/jasss.5937 Url: http://jasss.soc.surrey.ac.uk/29/1/7.html

Received: 12-09-2025 Accepted: 16-01-2026 Published: 31-01-2026

Abstract: The COVID-19 pandemic prompted a surge in computational models to simulate disease dynamics and guide interventions. Agent-based models (ABMs) are well-suited to capture population and environmental heterogeneity, but their rapid deployment raised questions about utility for health policy. We systematically reviewed 536 COVID-19 ABM studies published from January 2020 to December 2023, retrieved from Web of Science, PubMed, and Wiley on January 30, 2024. Studies were included if they used ABMs to simulate COVID-19, where reviews were excluded. Studies were assessed against nine criteria of model usefulness, including transparency and re-use, interdisciplinary collaboration and stakeholder engagement, and evaluation practices. Publications peaked in late 2021 and concentrated in a few countries. Most models explored behavioral or policy interventions ($n = 294$, 54.85%) rather than real-time forecasting ($n = 9$, 1.68%). While most described model assumptions ($n = 491$, 91.60%), fewer disclosed limitations ($n = 349$, 65.11%), shared code ($n = 219$, 40.86%), or built on existing models ($n = 195$, 36.38%). Standardized reporting protocols ($n = 36$, 6.72%) and stakeholder engagement were less commonly reported (13.62%, $n = 73$). Only 2.24% ($n = 12$) described a comprehensive validation framework, though uncertainty was often quantified ($n = 407$, 75.93%). Over time, reporting of stakeholder engagement and evaluation increased. Studies that claimed policy relevance ($n = 354$, 66.05%) more often included some evaluation ($n = 283$, 79.94% vs. $n = 125$, 68.68%) and stakeholder engagement ($n = 61$, 17.23% vs. $n = 12$, 6.59%), though they were less likely to re-use models or share code. Limitations of this review include under-representation of non-English studies, subjective data extraction, variability in study quality, and limited generalizability. Overall, COVID-19 ABMs advanced quickly, but lacked transparency, accessibility, and participatory engagement. Stronger standards are needed for ABMs to serve as reliable decision-support tools in future public health crises.

Keywords: COVID-19, Agent-Based Models, Infectious Disease Modeling, Systematic Literature Review

Von Hoene et al. (2026), JASSS: https://www.jasss.org/29/1/7.html.

2. MOTIVATION – IMPROVING THE REPRESENTATION OF HEALTH BEHAVIOR IN ABM

A Framework for Simulating Emergent Health Behaviors in Spatial Agent-Based Models of Disease Spread

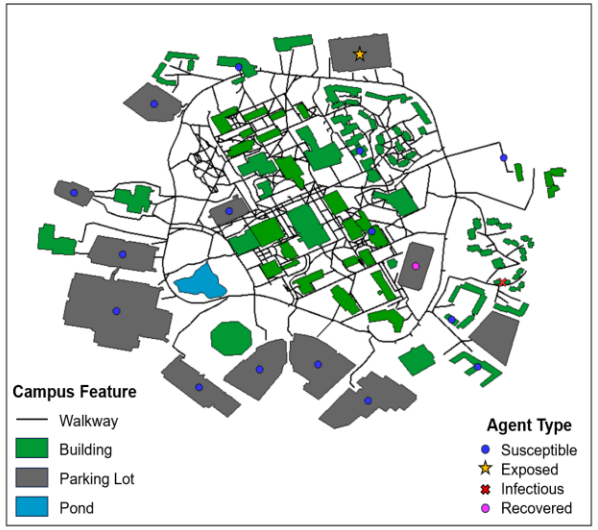
Emma Von Hoene
evonhoen@gmu.edu
Department of Geography and Geoinformation Science
George Mason University
Fairfax, Virginia, USA

Amira Roess
aroess@gmu.edu
Department of Global and Community Health
George Mason University
Fairfax, Virginia, USA

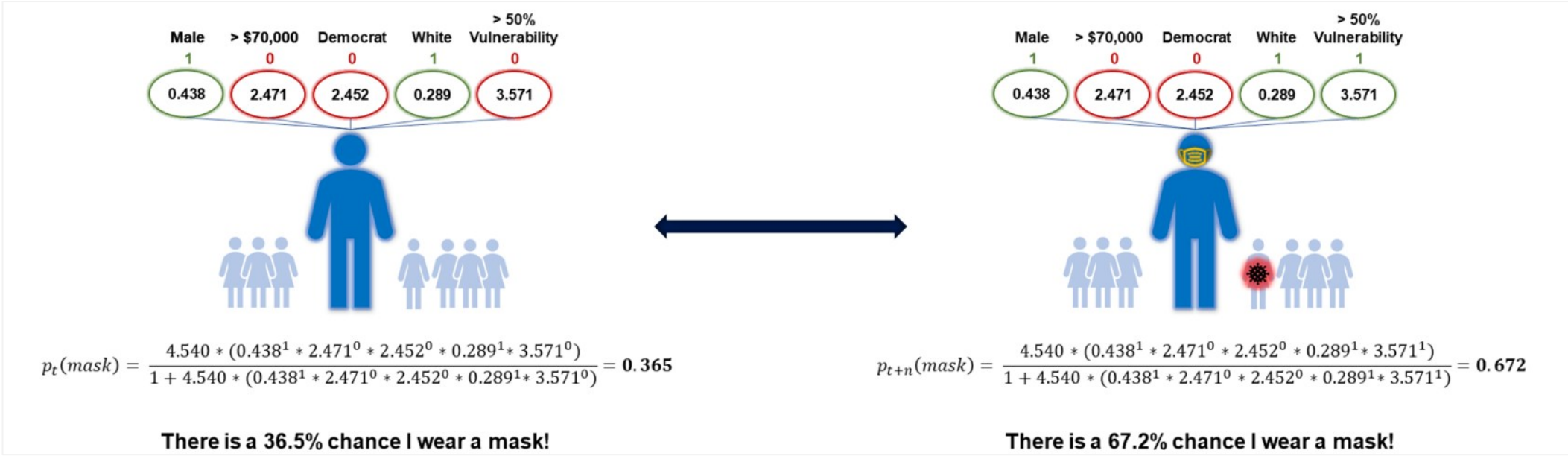
Shivani Achuthan
James Madison High School
Vienna, Virginia, USA
shivani@achuthans.com

Taylor Anderson
tander6@gmu.edu
Department of Geography and Geoinformation Science
George Mason University
Fairfax, Virginia, USA

Von Hoene et al. (2024), SIGSPATIAL International Workshop on GeoSim: *https://dl.acm.org/doi/abs/10.1145/3615891.3628010*



Environment of geospatial ABM simulating COVID-19 spread and mask behavior.

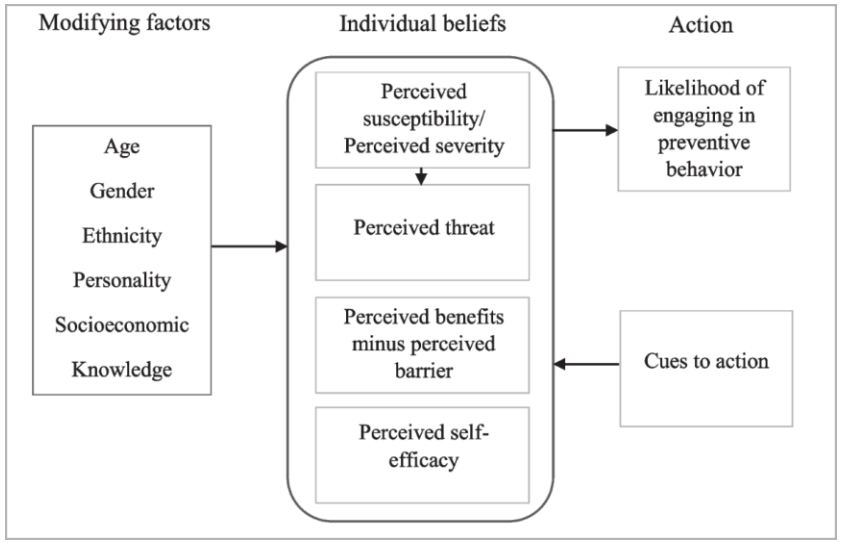


Representation of the decision-making framework using odds ratios corresponding to agent attributes in a logistic regression model to predict dynamic mask-use.

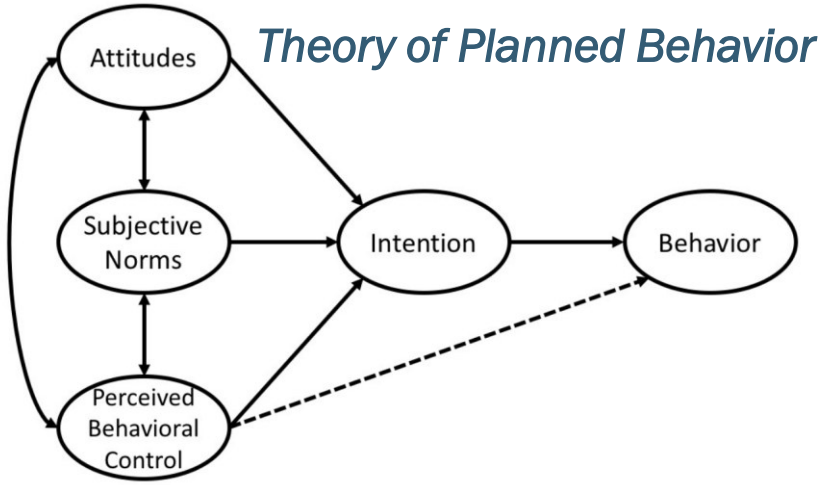
2. MOTIVATION – INITIALIZING POPULATIONS WITH DOMAIN-SPECIFIC ATTRIBUTES

Domain Specific Frameworks

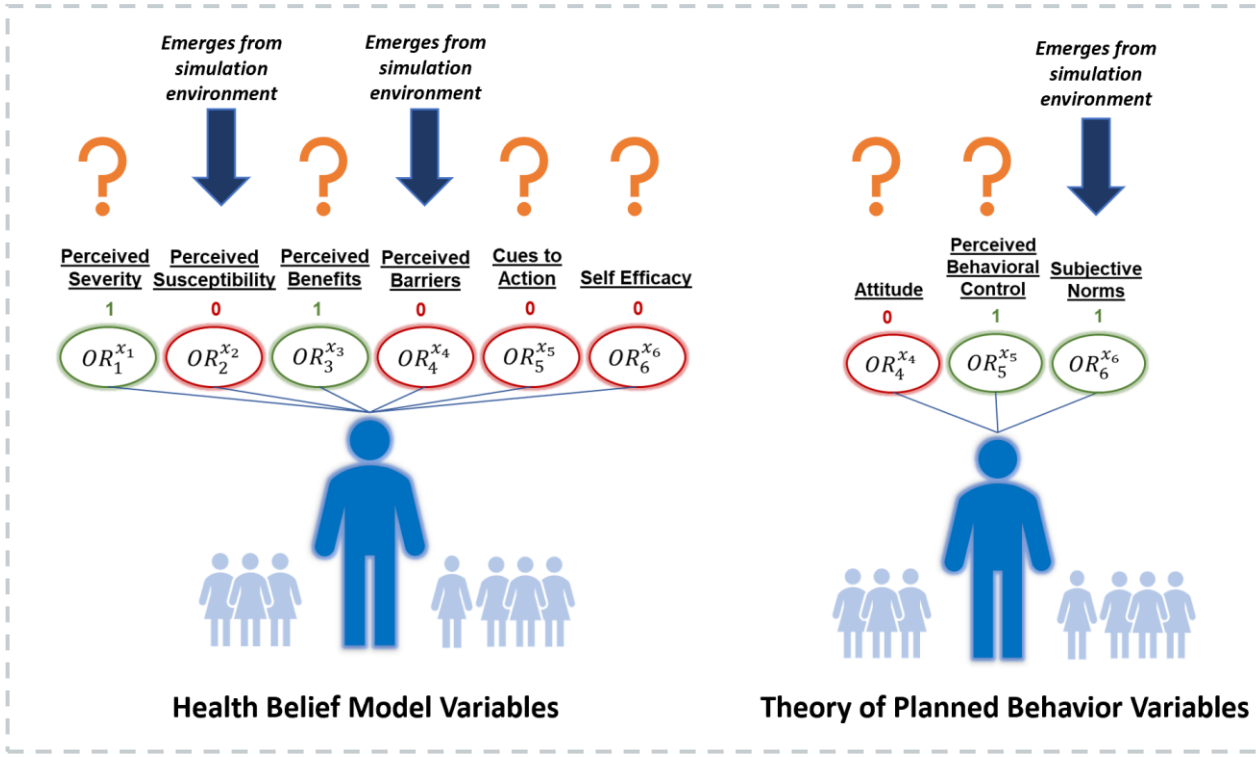
Health Belief Model



Theory of Planned Behavior



How can we realistically initialize agents with characteristics (other than demographics) that strongly influence their decision to adopt a health behavior (e.g., vaccines)?



2. EXISTING RESEARCH – SYNTHESIZING POPULATIONS WITH HEALTH CHARACTERISTICS

- Behaviors, attitudes, or perceptions are often assigned probabilistically using hypothetical scenarios or real-world aggregate data.
  - Example: *Rafferty et al. (2020)* simulate chickenpox with vaccination attitudes from aggregate data — ignores statistical relationships with individual characteristics.
  - Example: *Pandey et al. (2023)* model COVID-19 booster uptake using age-specific rates from flu data— assumes spatially uniform uptake.

Synthetic population methods to initialize agents with behavior, attitudes, or perceptions in infectious disease models remain underexplored.

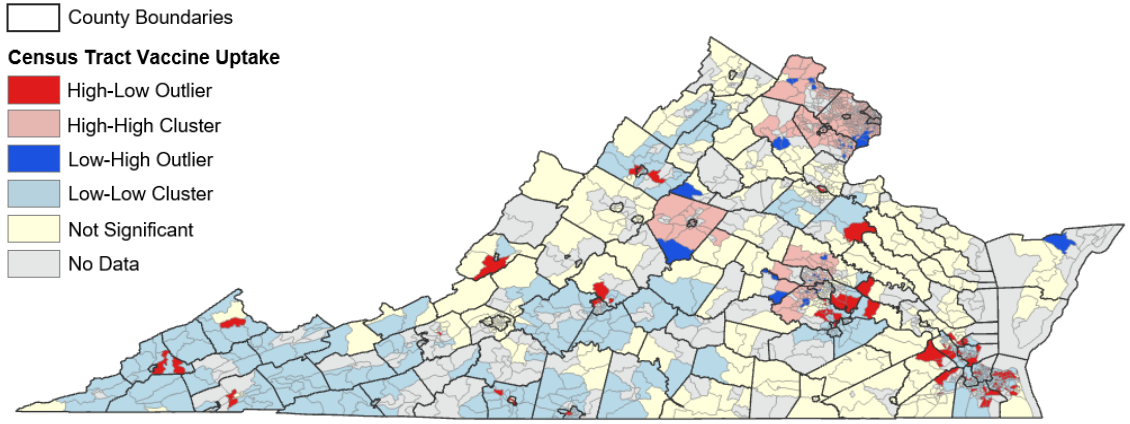
Rafferty E, McDonald W, Osgood N, Qian W, Doroshenko A. Seeking the optimal schedule for chickenpox vaccination in Canada: Using an agent-based model to explore the impact of dose timing, coverage and waning of immunity on disease outcomes. Vaccine. 2020;38(3):521–9.

Pandey A, Fitzpatrick MC, Moghadas SM, Vilches TN, Ko C, Vasan A, et al. Modelling the impact of a high-uptake bivalent booster scenario on the COVID-19 burden and healthcare costs in New York City. Lancet Reg Health Am. 2023;24:100555. pmid:37554154

3. OUR INITIAL RESEARCH OBJECTIVES

- **Research Question:** How can the **direct integration of survey data** into existing synthetic population generation approaches realistically reflect domain-specific characteristics, such as vaccine status and associated attitudes, and their spatial patterns in the agent population?

- **Research Goals:** To develop and validate a population synthesis approach for initializing agents with behavioral and related attitude or perceptions characteristics, specifically aiming to:
  - Capture the true **spatial heterogeneity** of such attributes
  - Preserve the **statistical relationship** of such attributes across demographic groups



Clusters and outliers of observed vaccine uptake in December 2021.

Relationships between demographic groups and attitudes towards COVID-19 vaccine intention [UAS 282, February – March 2021].

Independent Variable: Descriptor	Dependent Variable	Odds Ratio
Education: Bachelor's Degree	Positive Attitude	3.214
Gender: Male	Positive Attitude	1.855
Age: 65+	Positive Attitude	1.583
Income: Less than $25,000	Positive Attitude	0.443
Race: Black	Positive Attitude	0.410

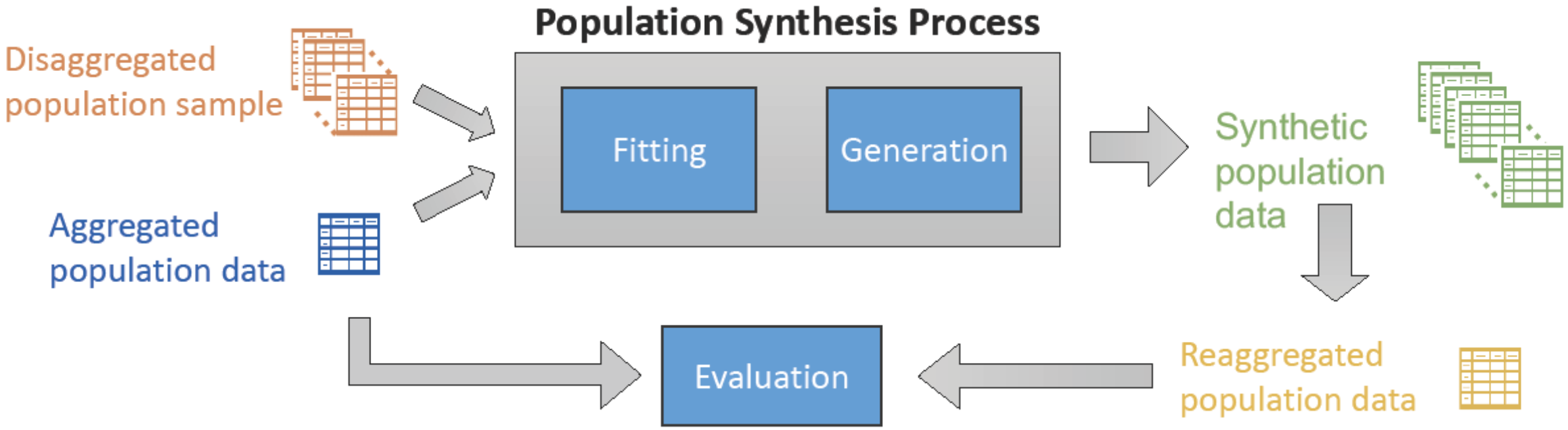
3. METHODOLOGY SUMMARY

- **To achieve our research objectives:**

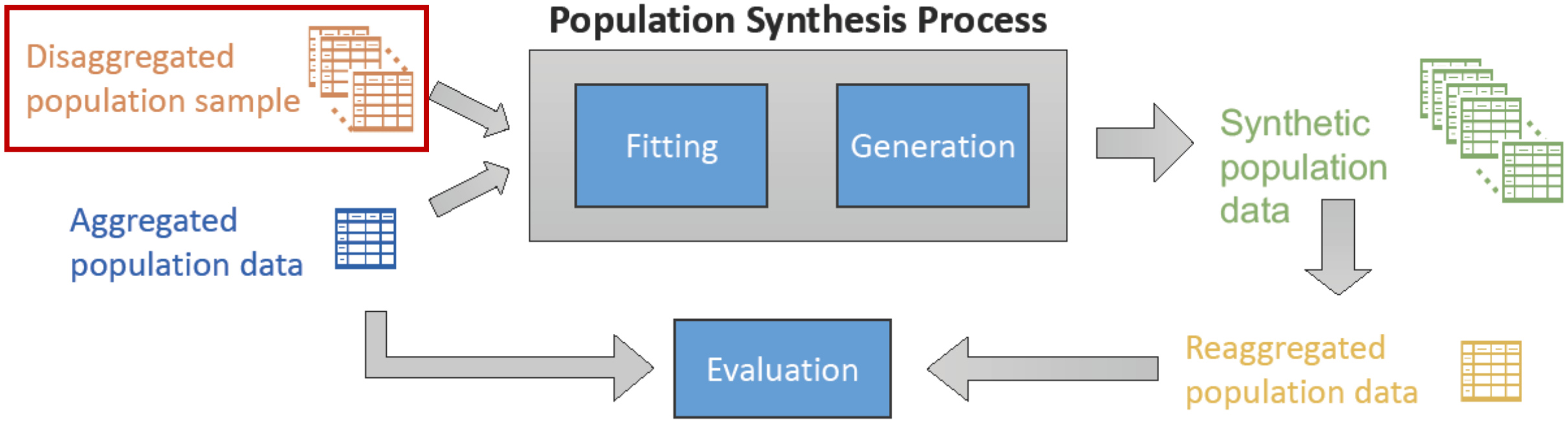
  - Extended an iterative proportional fitting (IPF) approach by replacing traditionally used sources (e.g., PUMS) with domain-specific surveys.

  - Used COVID-19 vaccination as a case study
    - Tested using two surveys to generate synthetic populations (age 18+) in Virginia, U.S.
    - Captures COVID-19 vaccine attitudes and uptake as of December 2021.
    - Validated using census tract-level vaccine uptake data for the same period.

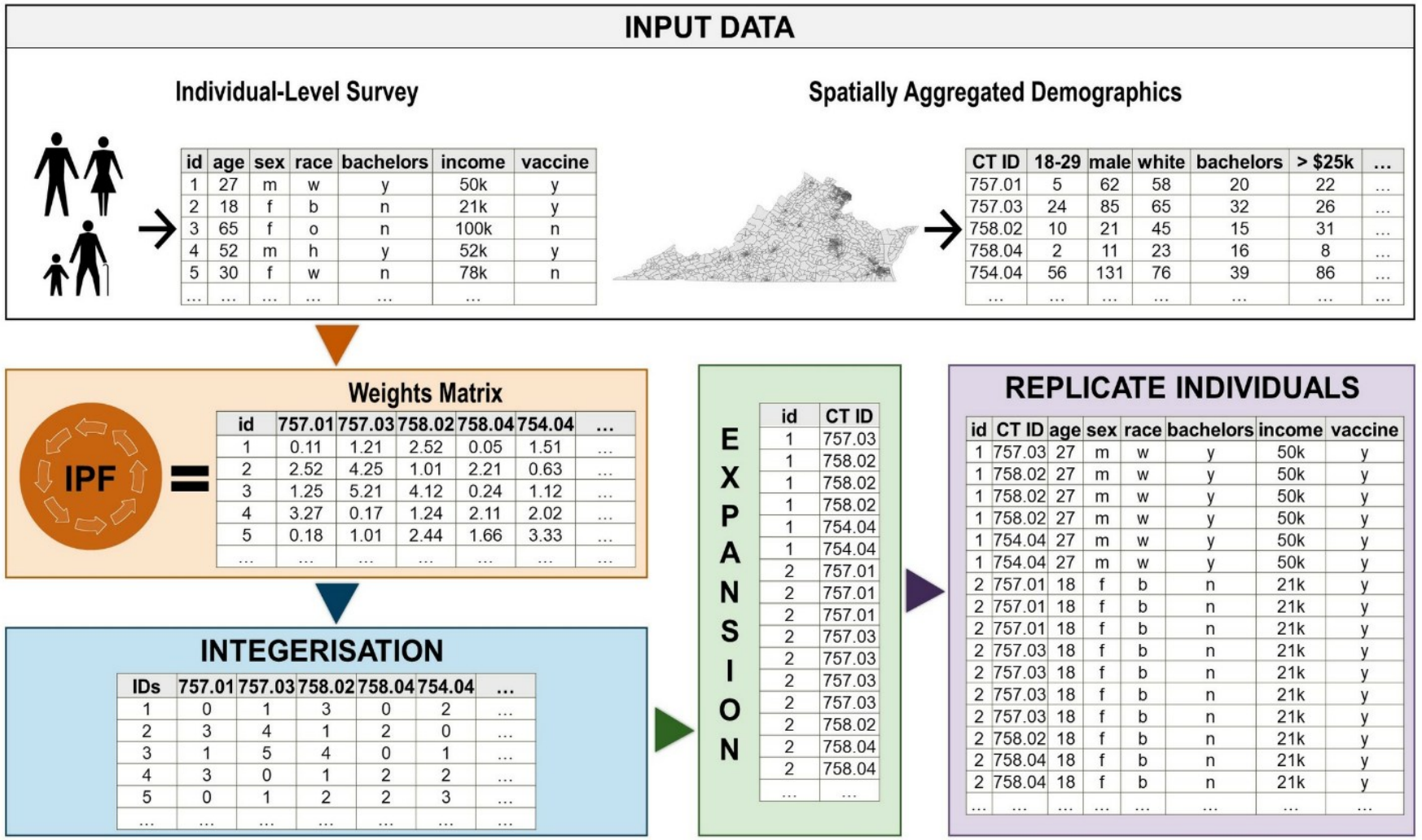
3. POPULATION SYNTHESIS PROCESS OVERVIEW



3. POPULATION SYNTHESIS PROCESS OVERVIEW



3. PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS



id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
...	...	...	...	...	...	

CT ID	18-29	male	white	bachelors	> $25k	...
757.01	5	62	58	20	22	...
757.03	24	85	65	32	26	...
758.02	10	21	45	15	31	...
758.04	2	11	23	16	8	...
754.04	56	131	76	39	86	...
...	...	...	...	...	...	...

id	757.01	757.03	758.02	758.04	754.04	...
1	0.11	1.21	2.52	0.05	1.51	...
2	2.52	4.25	1.01	2.21	0.63	...
3	1.25	5.21	4.12	0.24	1.12	...
4	3.27	0.17	1.24	2.11	2.02	...
5	0.18	1.01	2.44	1.66	3.33	...
...	...	...	...	...	...	...

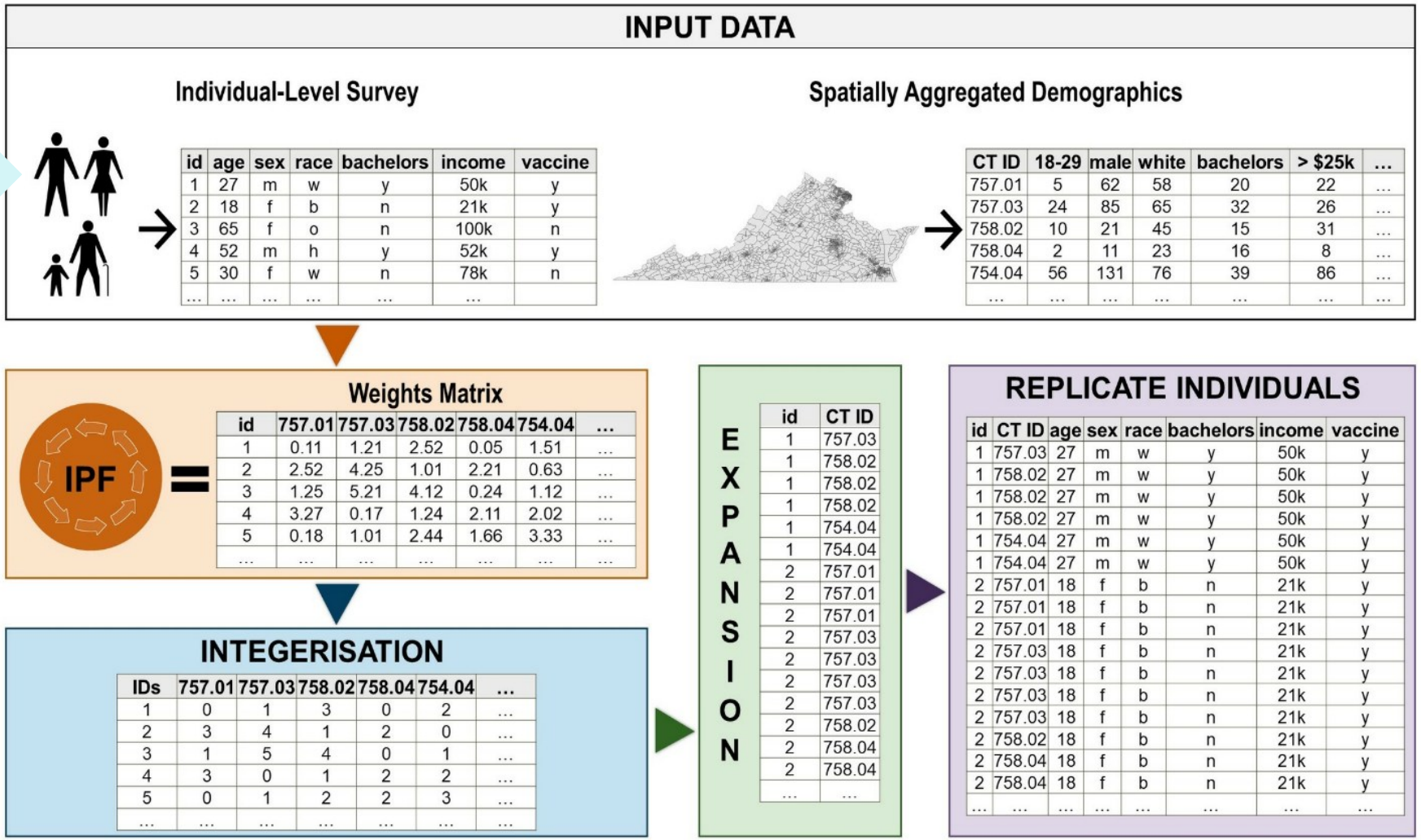
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

id	CT ID
1	757.03
1	758.02
1	758.02
1	758.02
1	754.04
1	754.04
2	757.01
2	757.01
2	757.01
2	757.03
2	757.03
2	757.03
2	757.03
2	758.02
2	758.04
2	758.04
...	...

id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	758.02	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
...	...	...	...	...	...	...	...

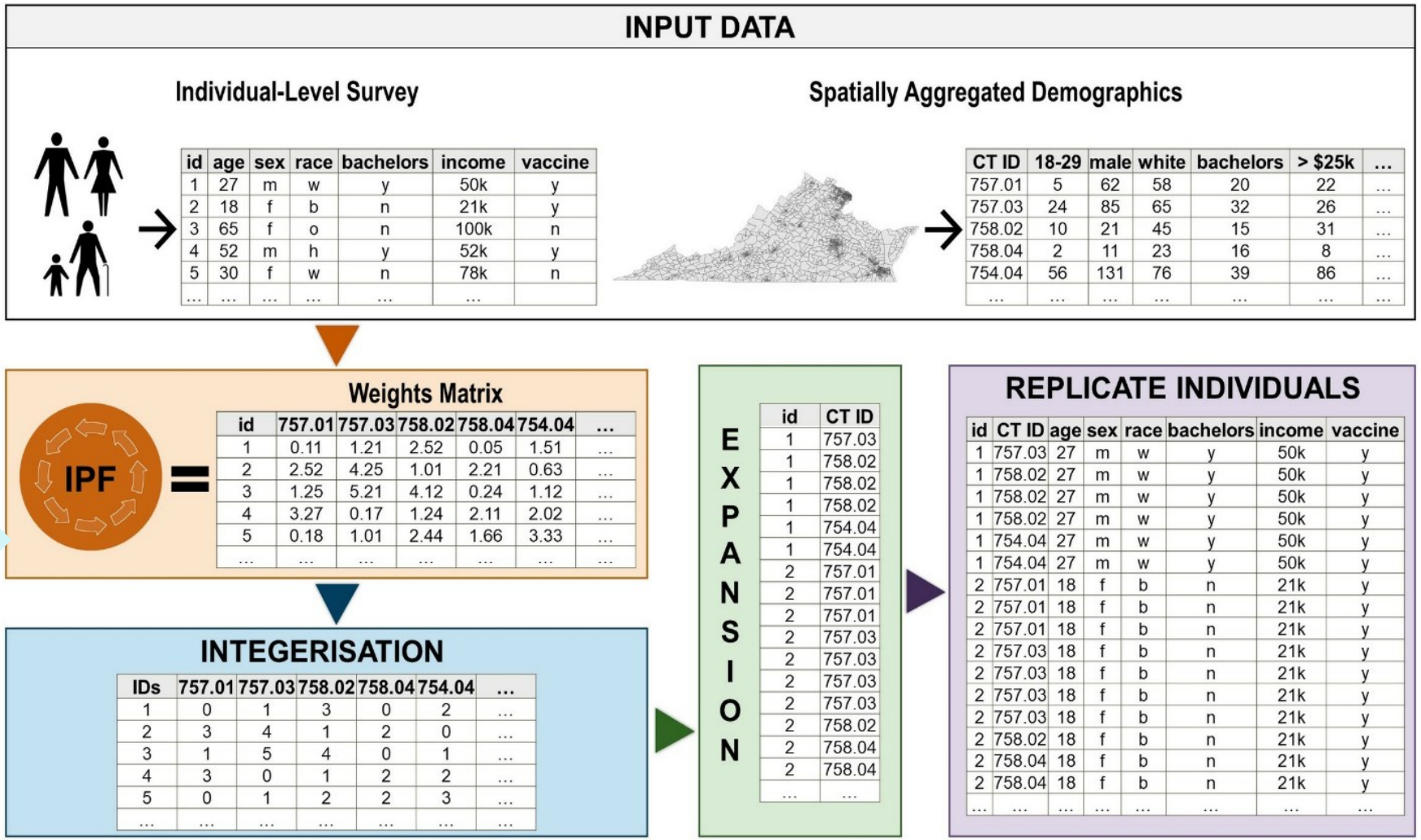
3. PROPOSED METHOD: INPUT DATA

- Focused on gender, race, age, education and income for individuals aged 18+
- **Spatially aggregate demographic data**
  - US Census Bureau, census tract level
- **Individual-level survey data**
  - HPS→ Household pulse survey (national, but not nationally representative)
  - NRS→ Nationally representative survey, collected by Dr. Amira Roess



3. PROPOSED METHOD: ITERATIVE PROPORTIONAL FITTING (IPF)

- IPF assigns a weight to each individual from the survey in each geographic zone (in this case, census tracts) to match the overall population distribution.
- These weights represent how much each individual contributes to a census tract based on their demographic characteristics.
- The goal is to adjust the weights so that the final distribution of individuals across census tracts aligns with the observed or expected geographic data.



id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
...	...	...	...	...	...	

CT ID	18-29	male	white	bachelors	> $25k	...
757.01	5	62	58	20	22	...
757.03	24	85	65	32	26	...
758.02	10	21	45	15	31	...
758.04	2	11	23	16	8	...
754.04	56	131	76	39	86	...
...	...	...	...	...	...	...

id	757.01	757.03	758.02	758.04	754.04	...
1	0.11	1.21	2.52	0.05	1.51	...
2	2.52	4.25	1.01	2.21	0.63	...
3	1.25	5.21	4.12	0.24	1.12	...
4	3.27	0.17	1.24	2.11	2.02	...
5	0.18	1.01	2.44	1.66	3.33	...
...	...	...	...	...	...	...

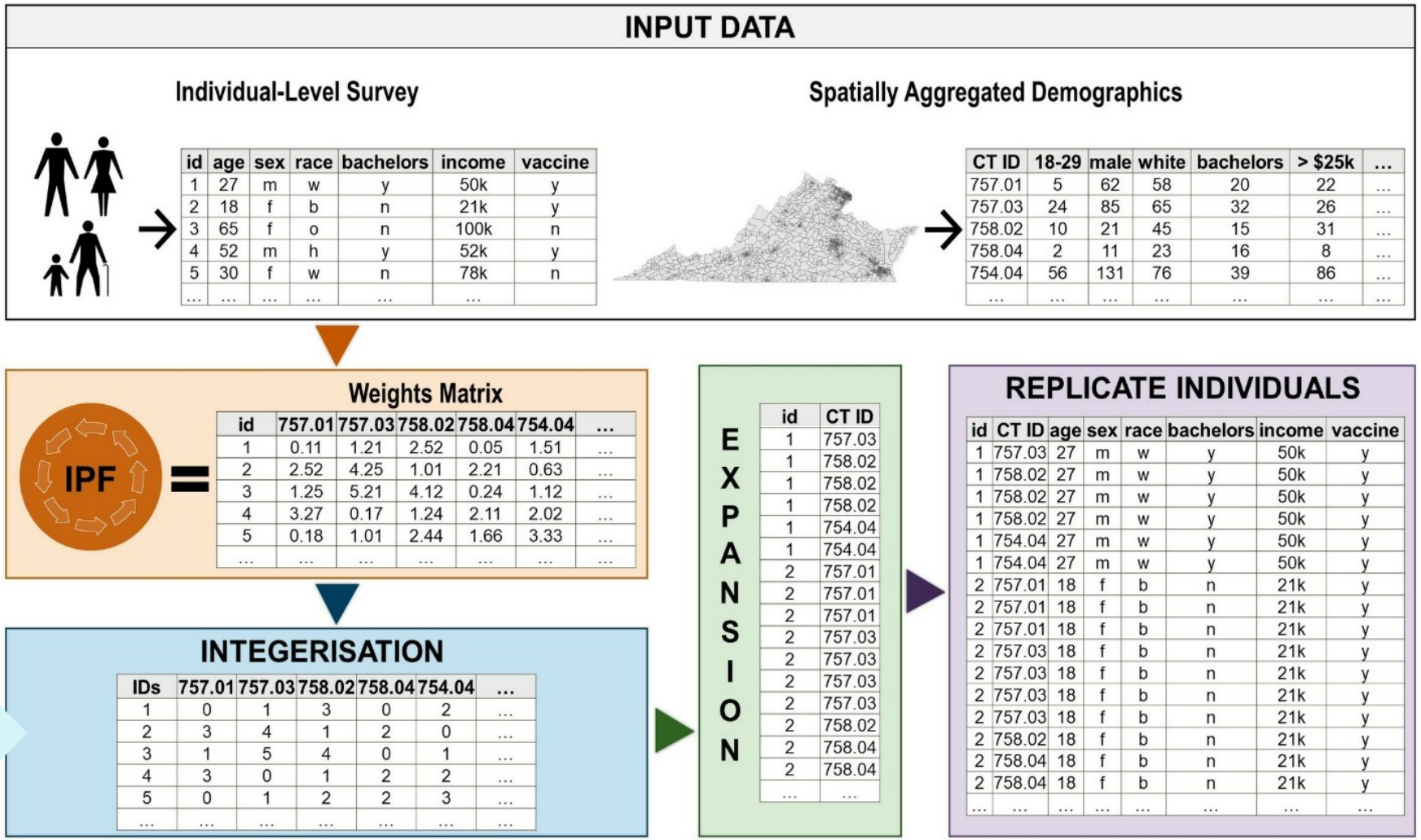
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

id	CT ID
1	757.03
1	758.02
1	758.02
1	758.02
1	754.04
1	754.04
2	757.01
2	757.01
2	757.01
2	757.03
2	757.03
2	757.03
2	757.03
2	758.02
2	758.04
2	758.04
...	...

id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	758.02	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
...	...	...	...	...	...	...	...

3. PROPOSED METHOD: INTEGERISATION

- After IPF, a weight matrix is generated where individuals from the survey are represented as rows and geographic zones (i.e., census tracts) are columns
- Each cell in the matrix contains the weight assigned to the corresponding individual for the respective zone, which requires a *conversion to counts*



id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
...	...	...	...	...	...	

CT ID	18-29	male	white	bachelors	> $25k	...
757.01	5	62	58	20	22	...
757.03	24	85	65	32	26	...
758.02	10	21	45	15	31	...
758.04	2	11	23	16	8	...
754.04	56	131	76	39	86	...
...	...	...	...	...	...	...

id	757.01	757.03	758.02	758.04	754.04	...
1	0.11	1.21	2.52	0.05	1.51	...
2	2.52	4.25	1.01	2.21	0.63	...
3	1.25	5.21	4.12	0.24	1.12	...
4	3.27	0.17	1.24	2.11	2.02	...
5	0.18	1.01	2.44	1.66	3.33	...
...	...	...	...	...	...	...

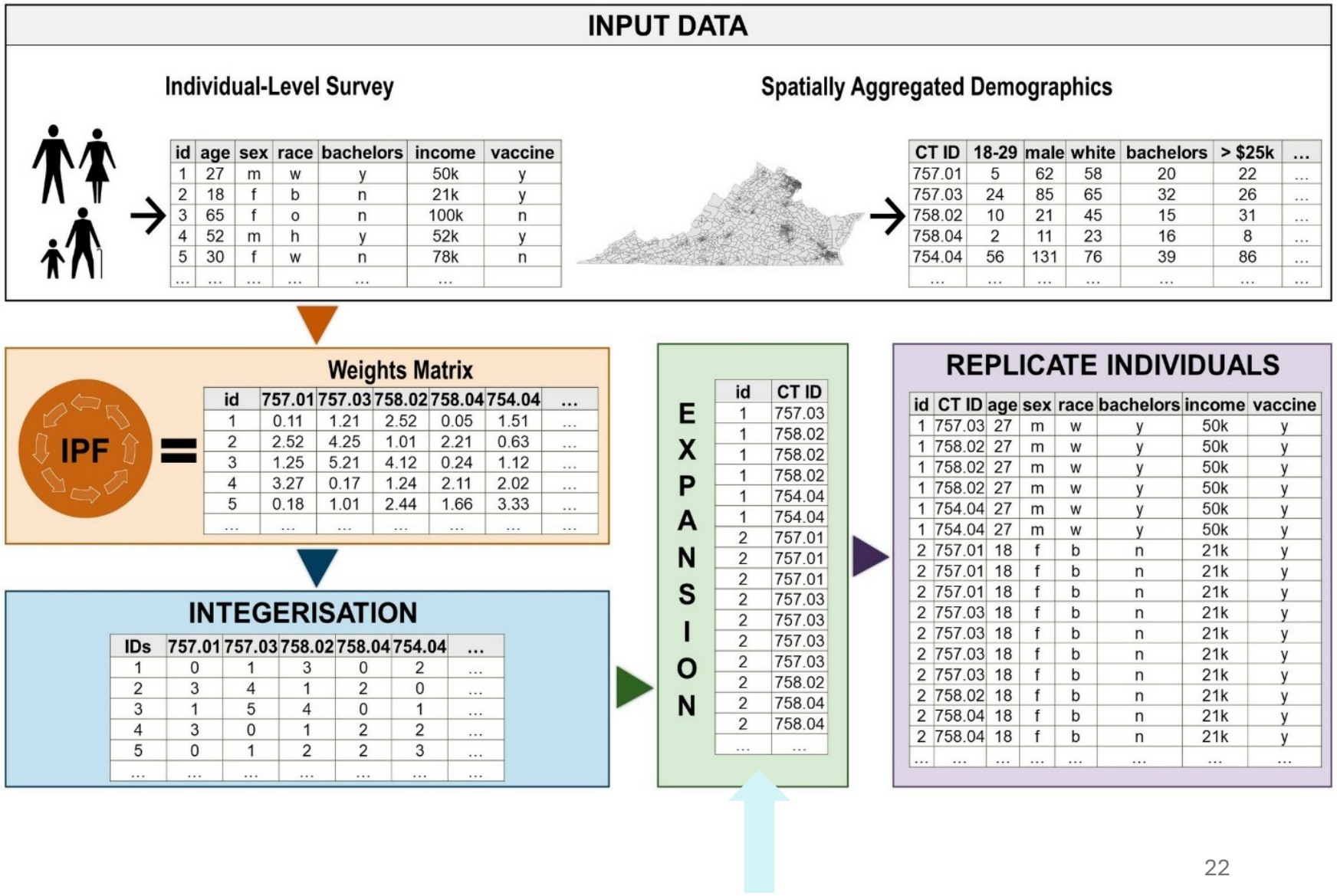
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

id	CT ID
1	757.03
1	758.02
1	758.02
1	758.02
1	754.04
1	754.04
2	757.01
2	757.01
2	757.01
2	757.03
2	757.03
2	757.03
2	757.03
2	758.02
2	758.04
2	758.04
...	...

id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	758.02	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
...	...	...	...	...	...	...	...

3. PROPOSED METHOD: EXPANSION

- Integerisation produces a matrix representing the number of times each individual needs to be replicated in a given geographic zone (i.e., census tract)
- Expansion is required to 'expand' the data based on the integer counts
- Transformation into a synthetic population output, which contains a row per individual within the whole population and a geographic zone for each person



id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
...	...	...	...	...	...	

CT ID	18-29	male	white	bachelors	> $25k	...
757.01	5	62	58	20	22	...
757.03	24	85	65	32	26	...
758.02	10	21	45	15	31	...
758.04	2	11	23	16	8	...
754.04	56	131	76	39	86	...
...	...	...	...	...	...	...

id	757.01	757.03	758.02	758.04	754.04	...
1	0.11	1.21	2.52	0.05	1.51	...
2	2.52	4.25	1.01	2.21	0.63	...
3	1.25	5.21	4.12	0.24	1.12	...
4	3.27	0.17	1.24	2.11	2.02	...
5	0.18	1.01	2.44	1.66	3.33	...
...	...	...	...	...	...	...

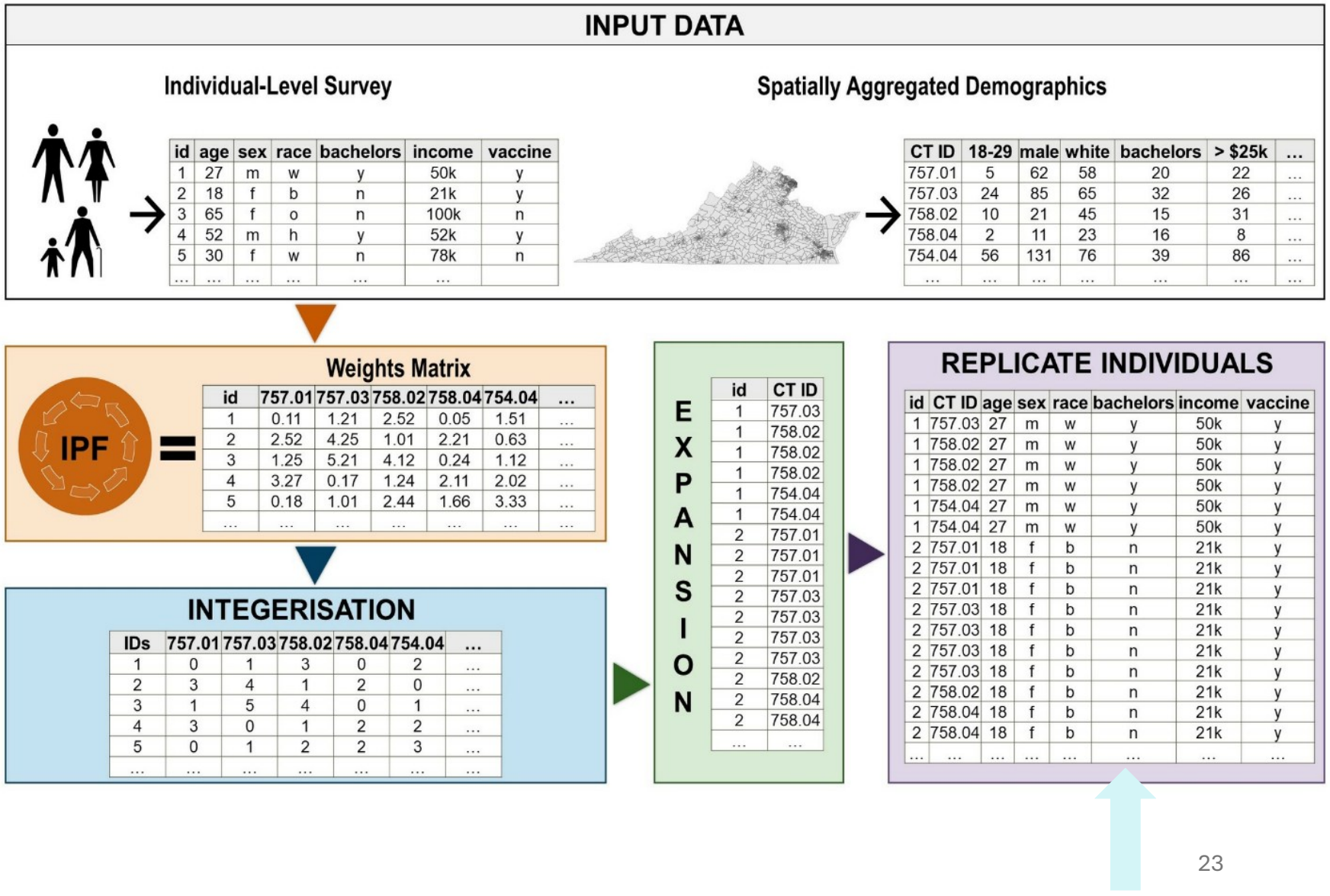
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

id	CT ID
1	757.03
1	758.02
1	758.02
1	758.02
1	754.04
1	754.04
2	757.01
2	757.01
2	757.01
2	757.03
2	757.03
2	757.03
2	757.03
2	758.02
2	758.04
2	758.04
...	...

id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	758.02	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
...	...	...	...	...	...	...	...

3. PROPOSED METHOD: REPLICATION OF INDIVIDUALS

- Any selected attributes from the survey data can be joined using the survey respondents' ID to the expanded dataset



id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
...	...	...	...	...	...	

CT ID	18-29	male	white	bachelors	> $25k	...
757.01	5	62	58	20	22	...
757.03	24	85	65	32	26	...
758.02	10	21	45	15	31	...
758.04	2	11	23	16	8	...
754.04	56	131	76	39	86	...
...	...	...	...	...	...	...

id	757.01	757.03	758.02	758.04	754.04	...
1	0.11	1.21	2.52	0.05	1.51	...
2	2.52	4.25	1.01	2.21	0.63	...
3	1.25	5.21	4.12	0.24	1.12	...
4	3.27	0.17	1.24	2.11	2.02	...
5	0.18	1.01	2.44	1.66	3.33	...
...	...	...	...	...	...	...

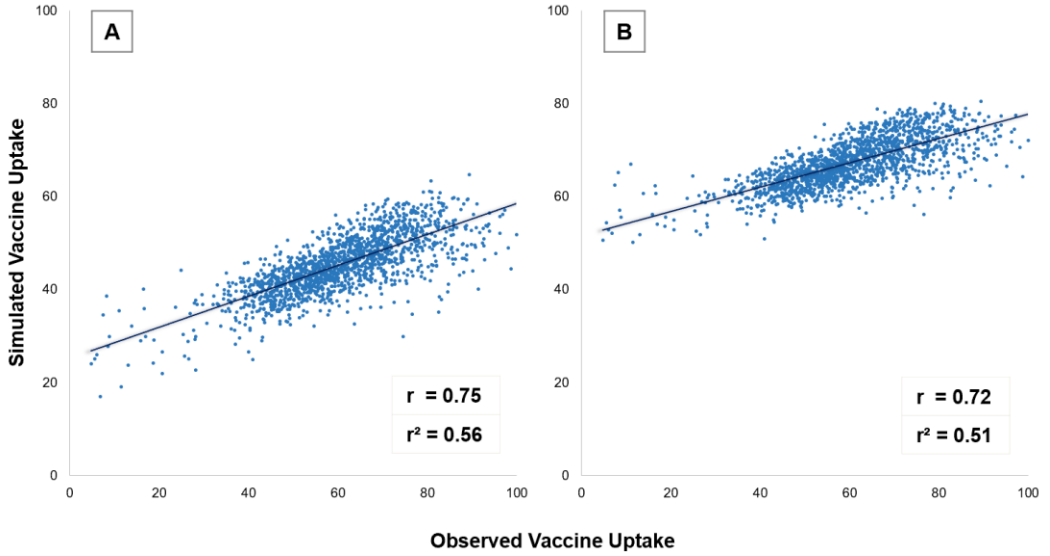
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

id	CT ID
1	757.03
1	758.02
1	758.02
1	758.02
1	754.04
1	754.04
2	757.01
2	757.01
2	757.01
2	757.03
2	757.03
2	757.03
2	757.03
2	758.02
2	758.04
2	758.04
...	...

id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	758.02	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
2	758.04	18	f	b	n	21k	y
...	...	...	...	...	...	...	...

3. RESULTS: OVERALL ACCURACY

- **Validation**
- Synthetic populations aggregated to census tract level
  - Internally against input US Census data
  - External data → Census tract vaccine uptake for Virginia (Virginia Department of Health – by request)



Scatterplots comparing observed and simulated vaccine uptake within Virginia census tracts: A) percentage from the HPS, and B) percentage from the NRS.

Variable: Descriptor	HPS Dataset				NRS Dataset			
	Pearson's r	*r^2*	*RMSE*	*MAE*	*Pearson's r*	*r^2*	*RMSE*	*MAE*
Gender: Male	0.9899	0.9799	0.7784	0.5867	0.9855	0.9712	0.9450	0.7307
Gender: Female	0.9899	0.9799	0.7784	0.5867	0.9855	0.9712	0.9450	0.7307
Race/Ethnicity: White	0.9991	0.9981	1.3229	1.0064	0.9985	0.9969	1.5779	1.2491
Race/Ethnicity: Black	0.9993	0.9987	1.0111	0.6690	0.9990	0.9980	1.1379	0.8145
Race/Ethnicity: Hispanic	0.999	0.9980	0.4696	0.3461	0.9989	0.9978	0.4965	0.3769
Race/Ethnicity: Other	0.9982	0.9965	0.9257	0.5872	0.9985	0.9969	0.5381	0.3612
Age: 18-29	0.9982	0.9963	0.758	0.5076	0.9830	0.9663	2.9901	2.3066
Age: 30-49	0.9948	0.9896	0.9428	0.7392	0.9625	0.9264	3.4143	2.5824
Age: 50-64	0.9945	0.9891	0.7982	0.6127	0.9927	0.9854	0.8424	0.6546
Age: 65 and over	0.9928	0.9857	1.2256	0.9309	0.9963	0.9926	0.7648	0.5728
Education: Bachelor's degree or higher	0.9969	0.9938	1.7922	1.5619	0.9971	0.9941	2.1185	1.8389
Education: No Bachelor's degree	0.9969	0.9938	1.7922	1.5619	0.9971	0.9941	2.1185	1.8389
Income: Less than $25,000	0.9992	0.9983	0.5555	0.4177	0.9984	0.9967	1.3171	1.0332
Income: $25,000 - $49,999	0.9984	0.9967	0.6097	0.4674				
Income: $50,000 - $99,999	0.9974	0.9948	0.7428	0.5484	0.9966	0.9933	0.8222	0.6086
Income: Greater than $100,000	0.9994	0.9988	1.0354	0.7246	0.9993	0.9985	1.1931	0.8400
Vaccine Uptake: Received	0.7489	0.5608	18.2780	15.6970	0.7151	0.5113	13.2810	10.6457

Evaluation metrics comparing observed and simulated percentages of demographic and vaccine status variables across Virginia census tracts.

3. RESULTS: STATISTICAL RELATIONSHIPS ACROSS DEMOGRAPHIC GROUPS

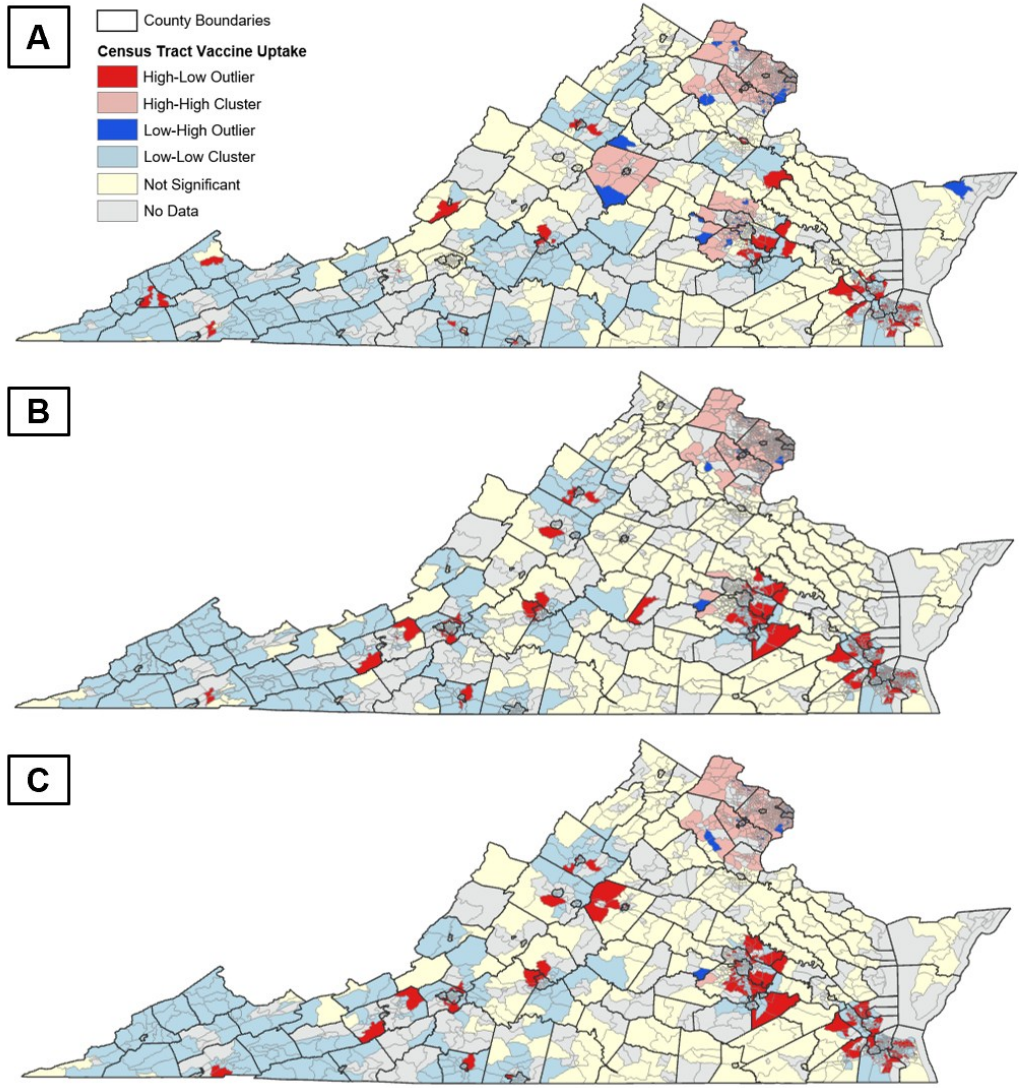
Coefficients from logistic regression models describing the relationship between demographic variables and vaccine uptake from the HPS, along with the synthetic population generated from the HPS. Significant coefficients are indicated with an asterisk (*) at the 90% confidence level (p-value < 0.10).

Variable: Descriptor	HPS Data: $R^2 = 0.08$ Coefficient β	Synthetic Population: $R^2 = 0.11$ Coefficient β
Intercept	1.8615*	- 0.3245*
Gender: Male	- 0.0595*	- 0.1872*
Race/Ethnicity: White	- 0.2277*	- 0.2758*
Age: 65 and over	1.1670*	1.4072*
Education: Bachelors or Higher	1.2429*	1.3435*
Income: $25,000 or less	- 0.5776*	- 0.8446*

Coefficients from logistic regression models describing the relationship between demographic variables and vaccine uptake from the NRS, the synthetic population generated from the NRS. Significant coefficients are indicated with an asterisk (*) at the 90% confidence level (p-value < 0.10); Gender was not included due to its insignificance (p-value greater than 0.10) in the NRS dataset.

Variable: Descriptor	NRS Data: $R^2 = 0.05$ Coefficient β	Synthetic Population: $R^2 = 0.05$ Coefficient β
Intercept	0.2341*	0.4020*
Race/Ethnicity: White	- 0.1511*	- 0.3036*
Age: 65 and over	1.0524*	1.1722*
Education: Bachelors or Higher	0.6248*	0.3839*
Income: Greater than $100,000	0.4058*	0.6669*

3. RESULTS: SPATIAL HETEROGENIETY IN BEHAVIOR

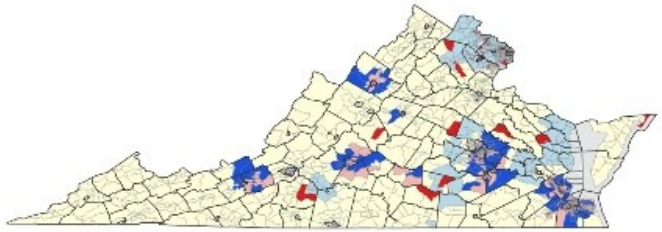


Clusters and outliers of vaccination uptake in Virginia Census Tracts: A) observed, B) synthesized population with HPS data, and C) synthesized population with nationally representative survey collected by researchers.

3. RESULTS: SPATIAL PATTERNS OF INDIVIDUAL ATTITUDES, BELIEFS AND PERCEPTIONS

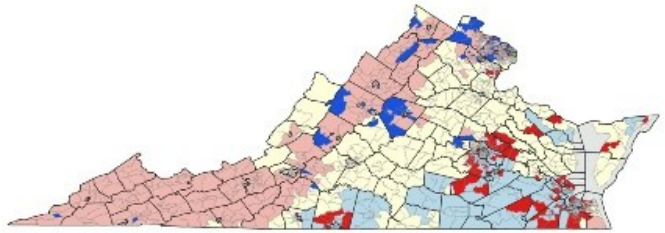
Perceived Behavioral Control/Barrier

A. I plan to wait and see if the COVID-19 vaccine is safe and may get it later



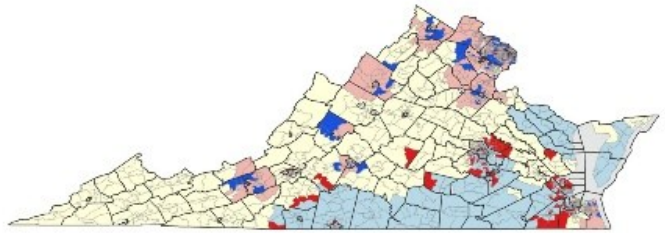
Social Norms

B. My doctor has not recommended the COVID-19 vaccine



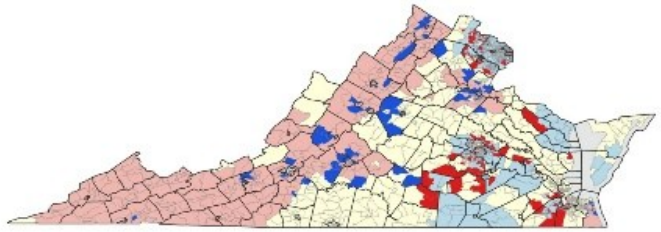
Attitude

C. I don't think COVID-19 is that big of a threat



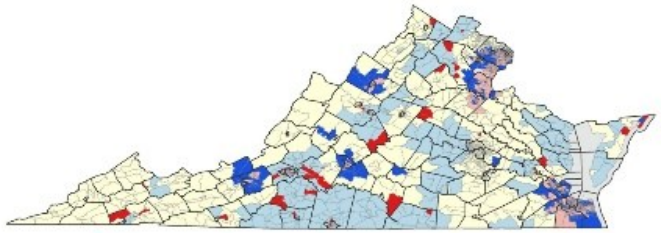
Attitude

D. I don't trust COVID-19 vaccines



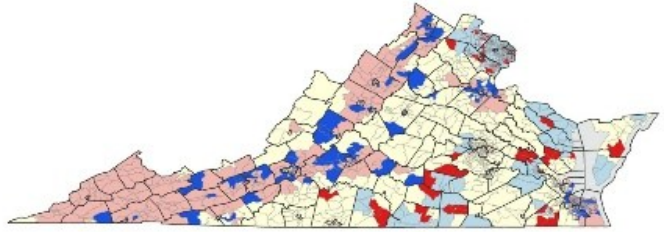
Perceived Behavioral Control/Barrier

E. I am concerned about the cost of a COVID-19 vaccine



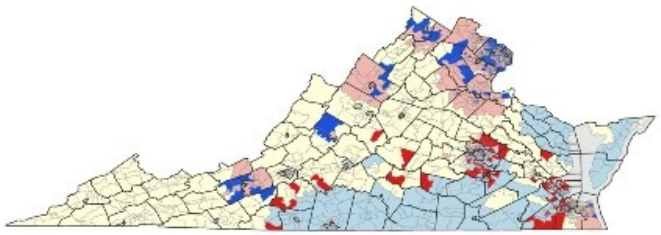
Perceived Behavioral Control/Barrier

F. I am concerned about possible side effects of the COVID-19 vaccine



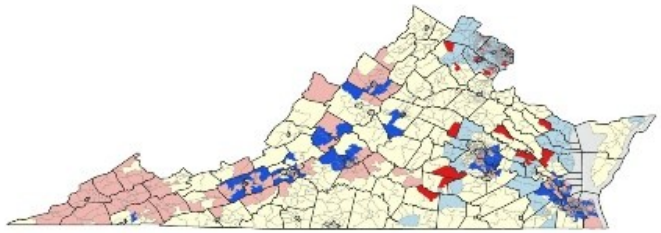
Attitude

G. I don't believe I need a COVID-19 vaccine



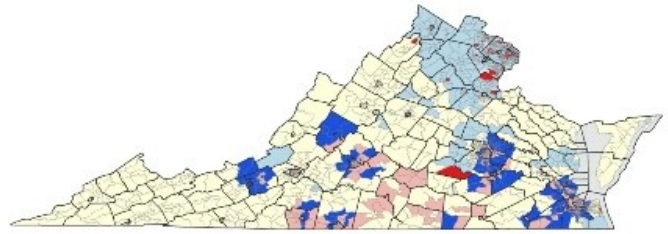
Perceived Behavioral Control/Barrier

H. I don't know if a COVID-19 vaccine will protect me



Perceived Behavioral Control/Barrier

I. It's hard for me to get a COVID-19 vaccine

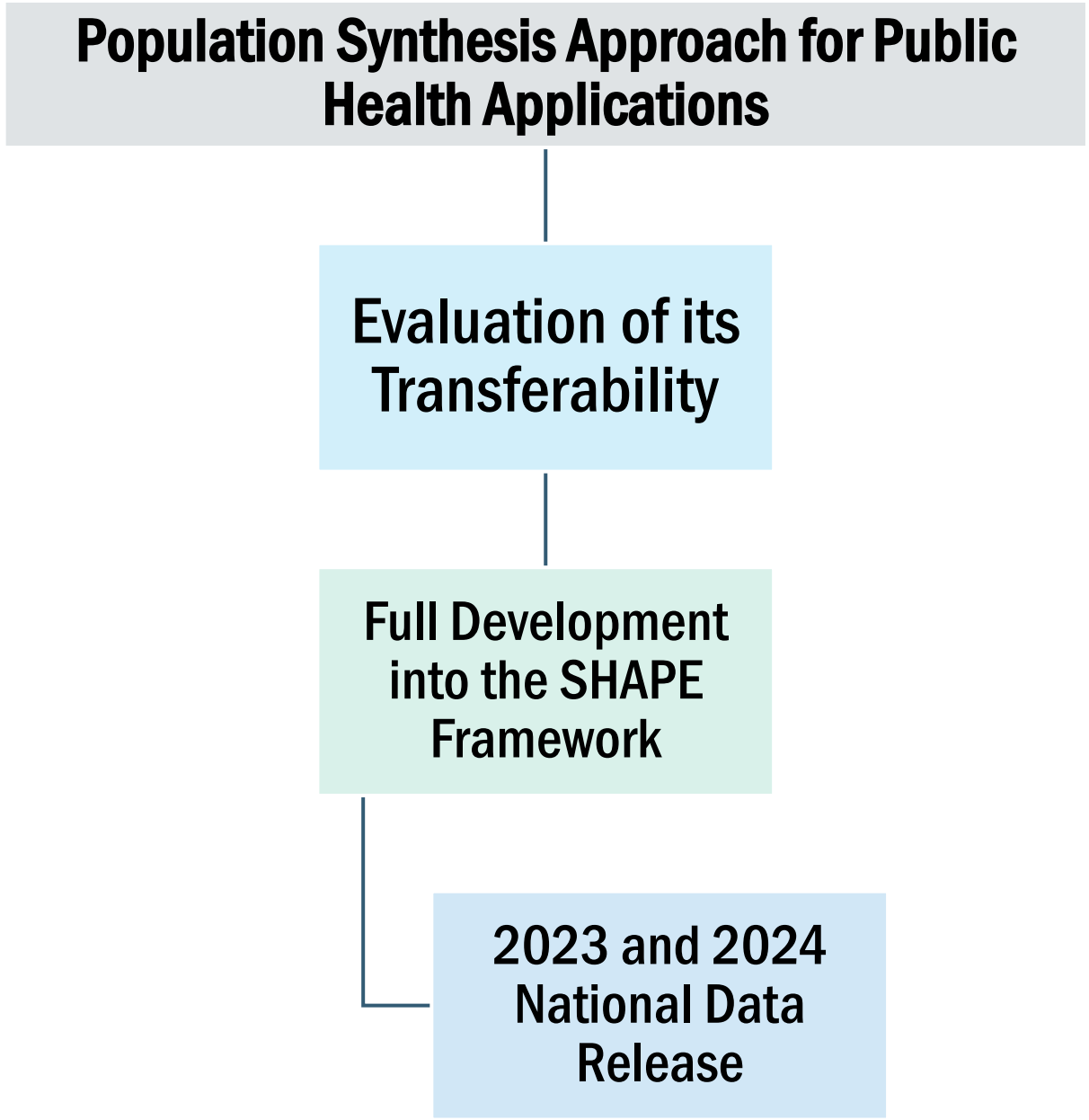


County Boundaries | Low-Low Cluster | Low-High Outlier | High-High Cluster | High-Low Outlier | Not Significant | No Data

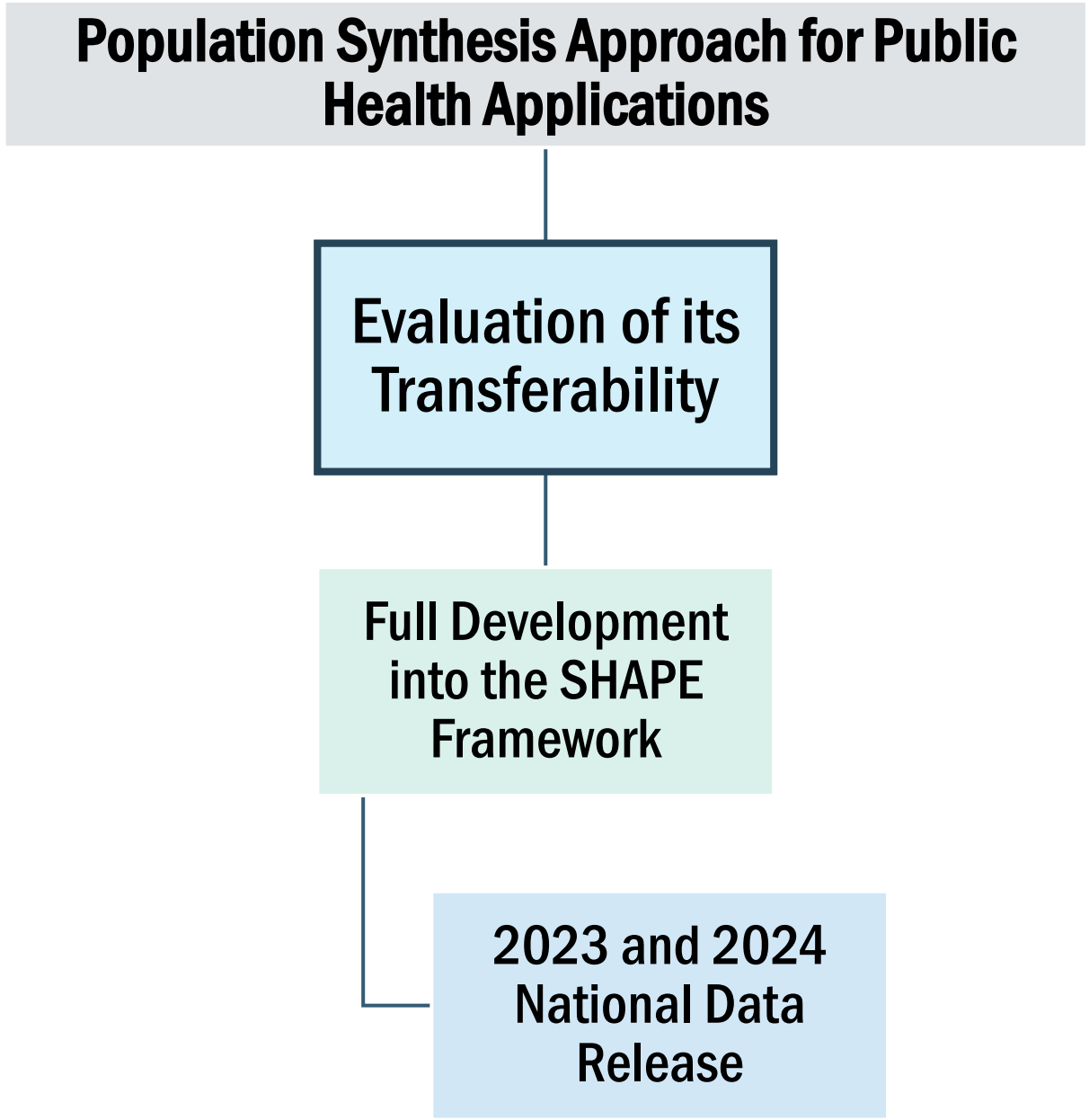
3. SUMMARY OUR NOVEL APPLICATION OF POPULATION SYNTHESIS FOR PUBLIC HEALTH

- We adapt traditional population synthesis approaches, particularly IPF, by introducing a novel application within the context of public health.
- To our knowledge, this work marks one of the first attempts to extend synthetic population generation approaches using public health surveys to initialize agents with protective behaviors and attitudes
- This approach reasonably reflects real-world populations, captures spatial heterogeneity and preserves statistical relationships across demographic groups
- Allows modelers already using synthetic population approaches to easily and directly incorporate domain-specific attributes in simulations, offering computational efficiency
- The approach is flexible, with the potential to generate populations with characteristics from any survey, for any application (***but… how well does the approach perform for other public health applications?***)

3. RELATED AND EXTENDED WORK



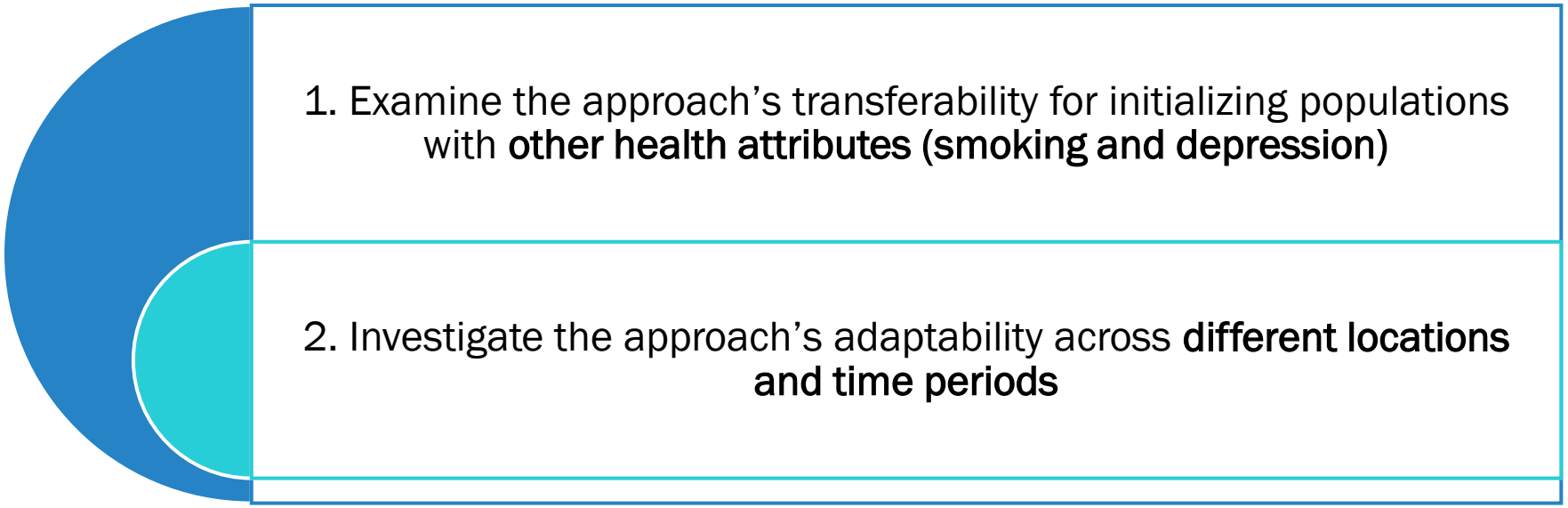
3. RELATED AND EXTENDED WORK



4. MOTIVATION: EVALUATING ADAPTABILITY OF OUR PROPOSED APPROACH

- The transferability of our previously developed population synthesis approach for initializing individual with health characteristics remained unexplored.

- We addressed this by developing a series of synthetic populations that:



4. DATA AND EXPERIMENTS FOR EVALUATING TRANSFERABILITY

Outline of each case study, its purpose, and study settings, along with corresponding input and validation datasets used.

Case Study/ Synthesized Outcome	Experiment Purpose	Time Period (s) and Study Area (s)	Disaggregated Individual-Level Survey Data	Spatially Aggregate Data	Validation Data
Depression	Demonstrating Spatial Adaptability	*2021:* Virginia, Tennessee and New Jersey	*March 2021*: Understanding America Survey – Understanding Coronavirus in America (**UAS-UCA**)	*2017-2021 5YR* **Census Tract** Estimates: US Census American Community Survey for age, race, gender, income, and education	*2024 Release*: CDC PLACES Depression ***Model-Based*** Estimates
Smoking	Modeling Temporal Change	*2021 and 2022:* Virginia	**UAS-UCA** + *March 2022*: Understanding America Survey – Drug Use Supplement (UAS-DUS)		*2024 Release*: CDC PLACES Smoking ***Model-Based*** Estimates

4. EVALUATING TRANSFERABILITY: DEPRESSION ACROSS LOCATIONS

Overall Accuracy:

Metrics comparing synthesized **depression (2021)** rates for different states against CDC PLACES estimates.

Location	Pearson r	R^2	RMSE	MAE
Virginia	0.759	0.576	4.533	4.093
New Jersey	0.531	0.282	3.349	2.947
Tennessee	0.744	0.553	8.775	8.458

Preservation of Statistical Relationships:

Logistic regression coefficients measuring the association between demographics and **depression (2021)** from the original UAS-UCA survey and synthesized populations for different states. Education is not included due to insignificance.
All variables are significant at 99% confidence.

Variable: Descriptor	UAS-UCA	Virginia	New Jersey	Tennessee
Age: 65+	-0.322	-0.295	-0.240	-0.415
Gender: Male	-0.837	-0.771	-0.833	-0.866
Income: $100,000+	-0.624	-0.672	-0.589	-0.808
Race: Black	-0.564	-0.646	-0.567	-0.839
Intercept	-0.722	-0.795	-0.860	-0.599

4. EVALUATING TRANSFERABILITY: DEPRESSION ACROSS LOCATIONS

Spatial Heterogeneity: Clusters and outliers of **depression** across census tracts for:

A) Virginia CDC PLACES

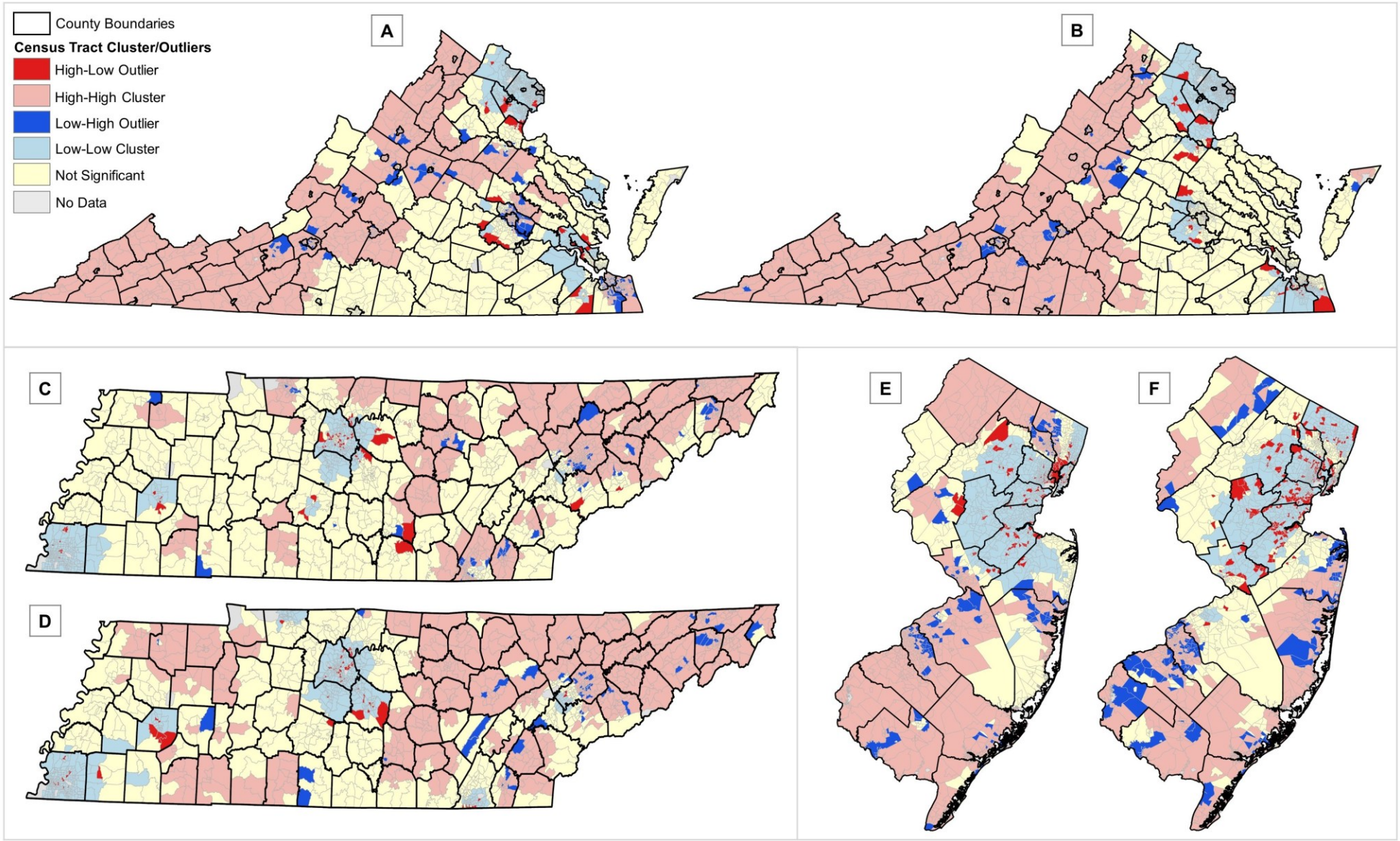
B) Virginia synthesized population

C) Tennessee CDC PLACES

D) Tennessee synthesized population

E) New Jersey CDC PLACES

F) New Jersey synthesized population



4. EVALUATING TRANSFERABILITY: SMOKING AND TEMPORAL CHANGE

Overall Accuracy:

Metrics comparing synthesized **Virginia** smoking rates from different time periods against CDC PLACES estimates.

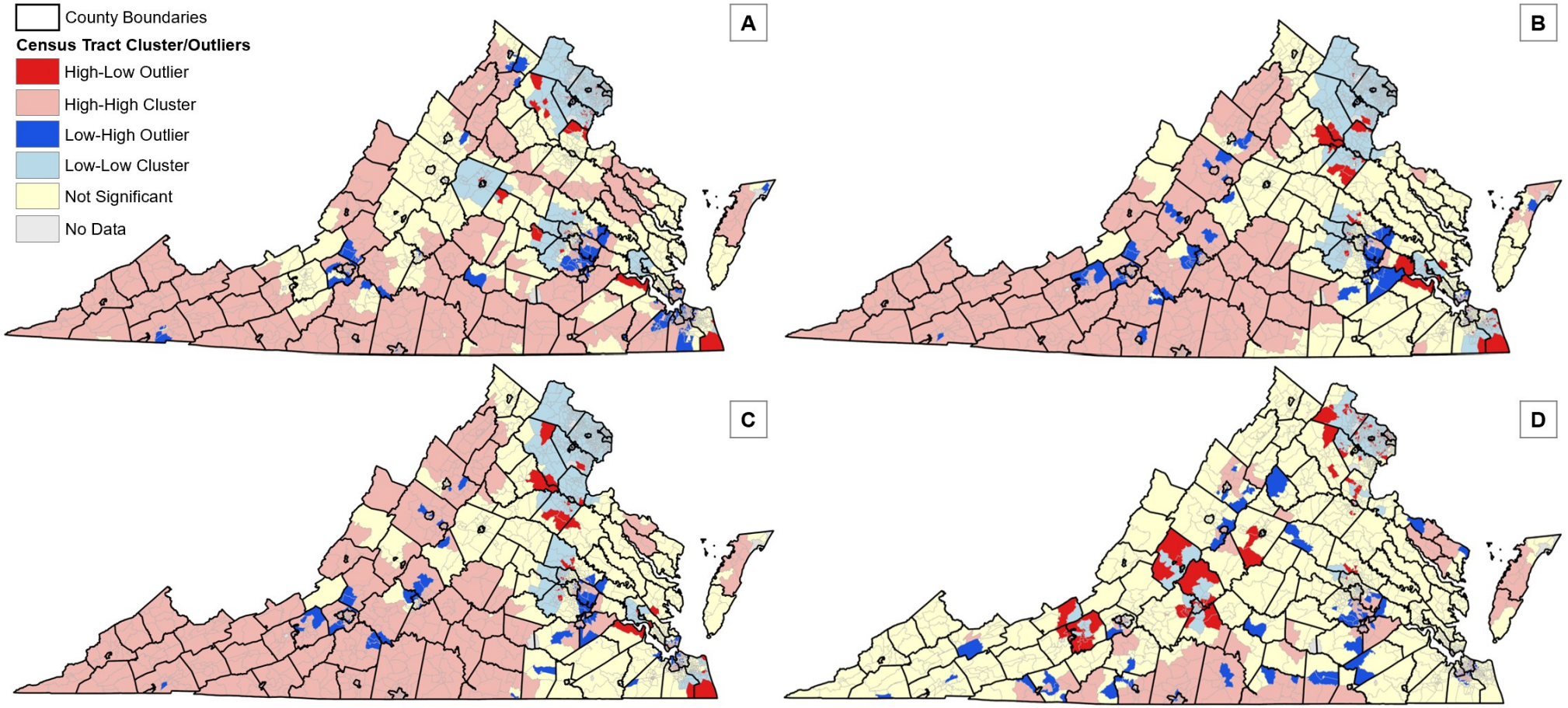
Year	Pearson r	R^2	RMSE	MAE
2021	0.762	0.580	3.763	2.890
2022	0.816	0.665	4.103	3.206

Preservation of Statistical Relationships:

Logistic regression coefficients measuring the association between demographics and **smoking** from the originally used UAS-UCA and UAS-UAS-DUS surveys and the synthesized Virginia populations for different time periods. Gender is not included due to insignificance.
All variables are significant at 99% confidence.

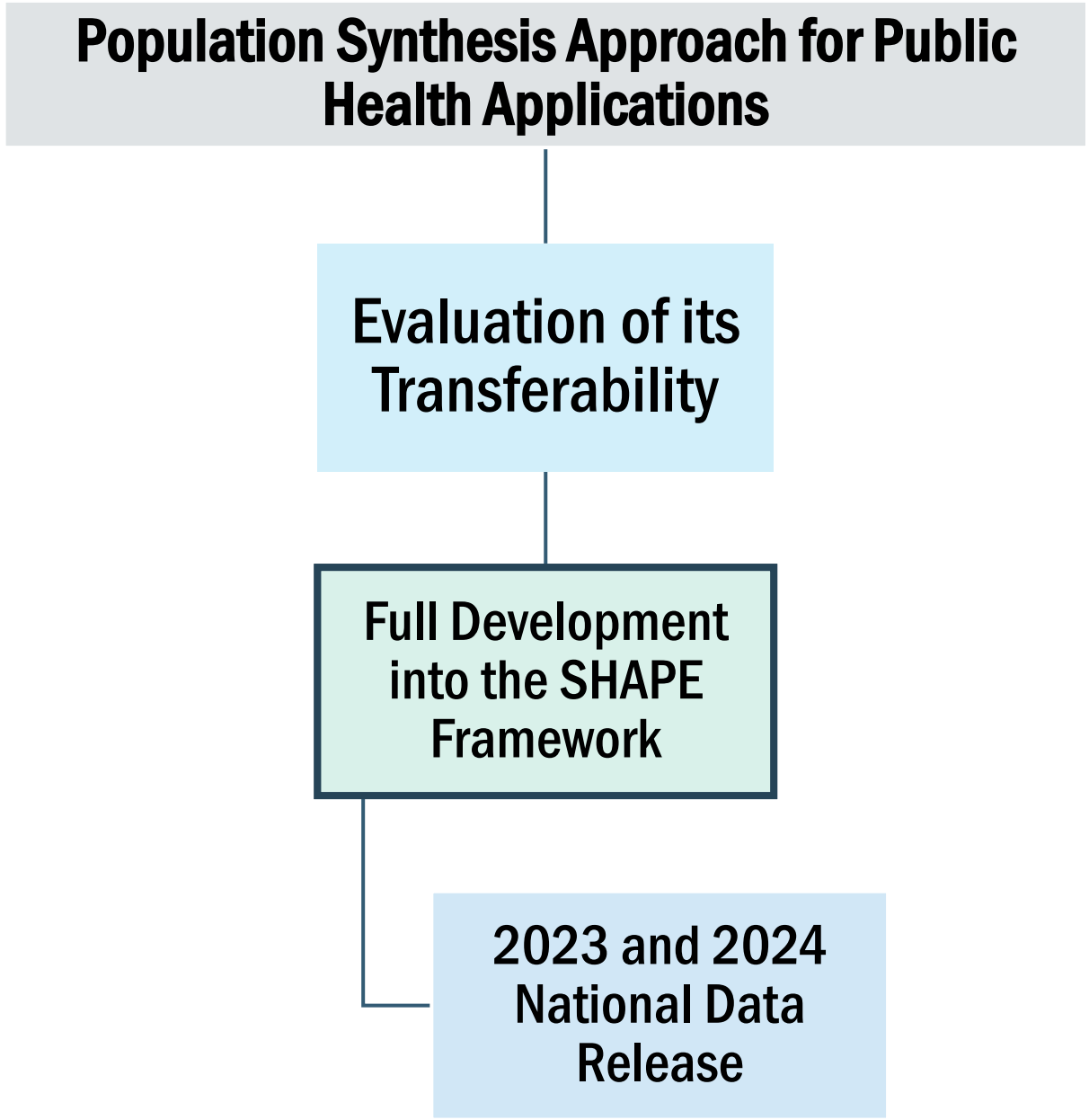
Variable: Descriptor	UAS-UCA	2021	UAS-DUS	2022
Age: 65+	-1.048	-0.760	-1.017	-0.788
Education: Bachelor's or higher	-1.028	-1.030	-1.138	-1.264
Income: $100,000+	-1.020	-0.772	-1.316	-0.987
Race: White	0.576	0.361	0.575	0.397
Intercept	-1.710	-1.723	-1.680	-1.775

4. EVALUATING TRANSFERABILITY: SMOKING AND TEMPORAL CHANGE



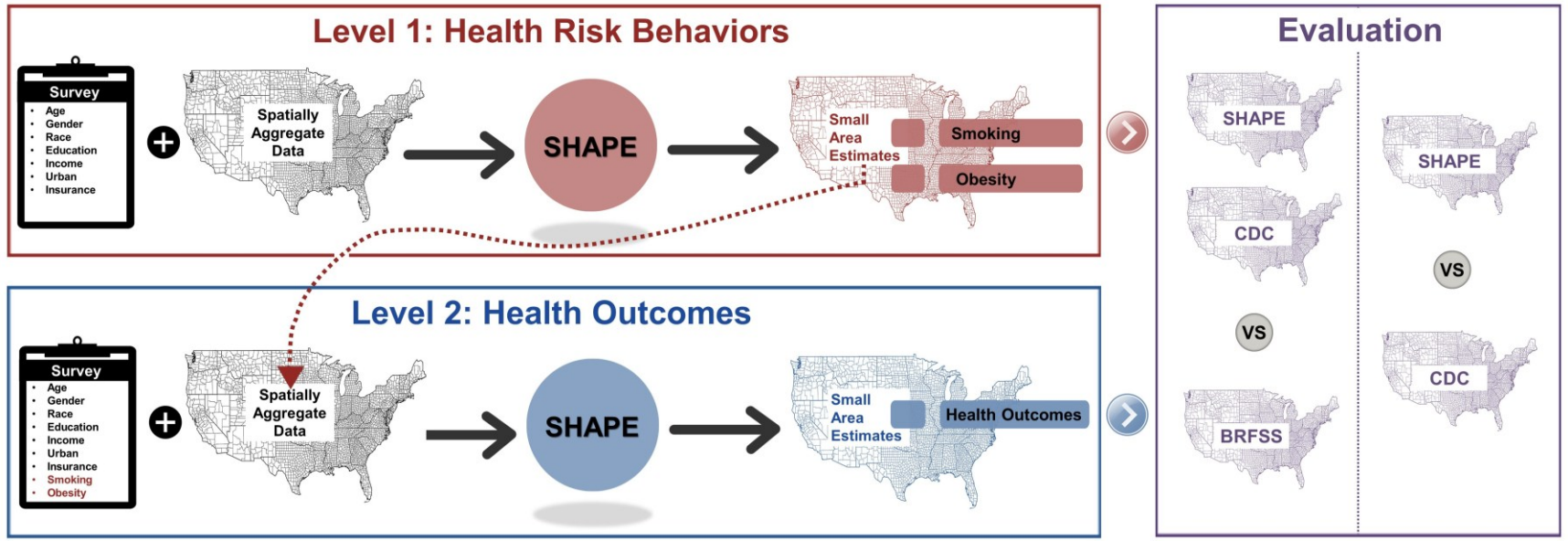
Clusters and outliers of the following: A) CDC PLACES estimates of current cigarette smokers, B) 2021 synthesized population with smoking behavior C) 2022 synthesized population with smoking behavior D) Change in smoking from 2021 to 2022

4. RELATED AND EXTENDED WORK



5. DEVELOPMENT INTO THE SPATIAL HEALTH AND POPULATION ESTIMATOR (SHAPE)

- Our proposed population synthesis approach for initializing individuals with health characteristics was fully **operationalized into the Spatial Health and Population Estimator (SHAPE)**
  - An open, reproducible and flexible spatial microsimulation framework, using publicly available data
    - BRFSS health surveys and sociodemographic spatially aggregated census data
  - Leverages hierarchical IPF to generate health-relevant synthetic populations and small area estimates (SAEs)

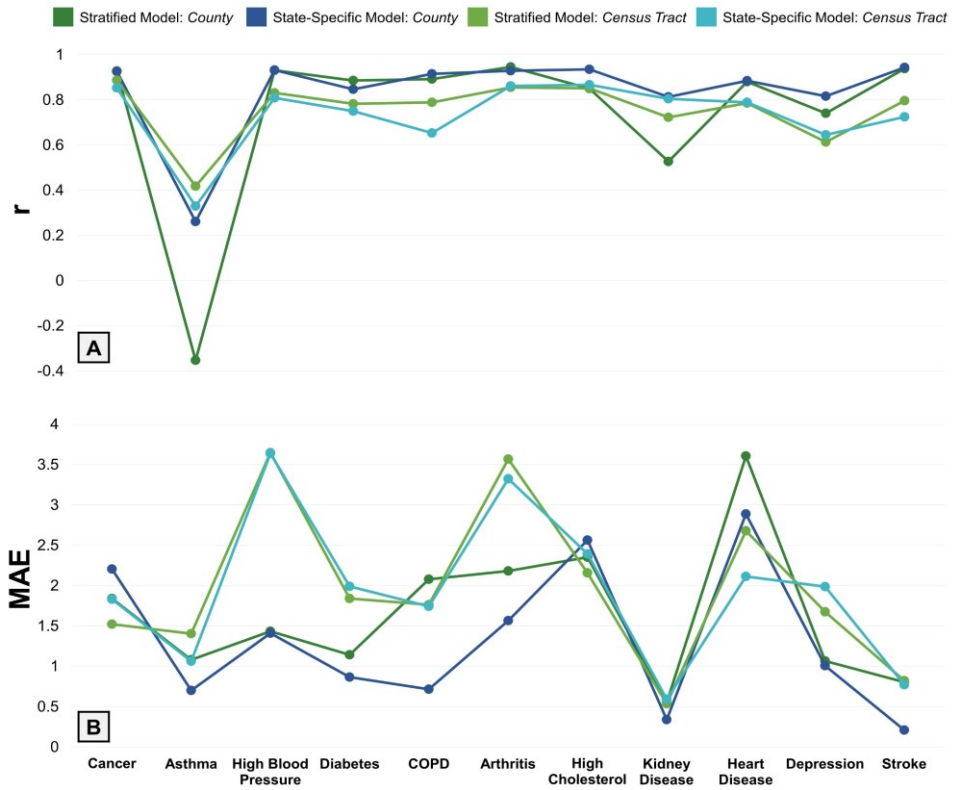


Overview of the two-leveled hierarchical IPF-based process used within the SHAPE framework.

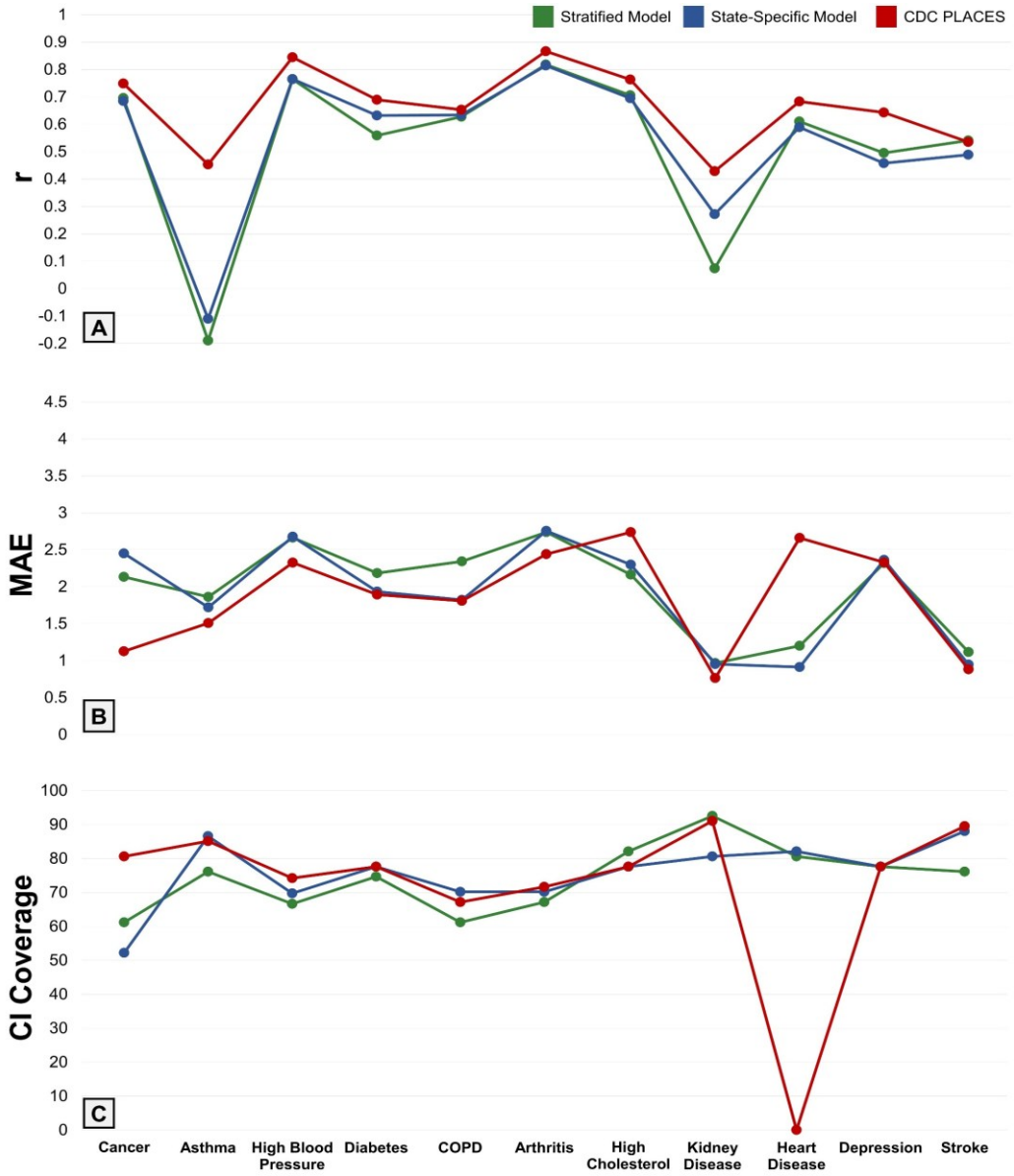
tinyurl.com/fy8rk6u6

5. SHAPE EVALUATION

- **Applied SHAPE to generate SAEs for:**
  - 2 health risk behaviors, 11 health outcomes
  - Florida (2019) and New York (2021)
  - Counties and census tracts
- **Evaluated** against BRFSS direct estimates & CDC PLACES
- *SHAPE estimates were moderately consistent with BRFSS ($\bar{r} \approx 0.54$), and strongly aligned with CDC PLACES – county ($\bar{r} \approx 0.81$) and census tracts ($\bar{r} \approx 0.74$).*



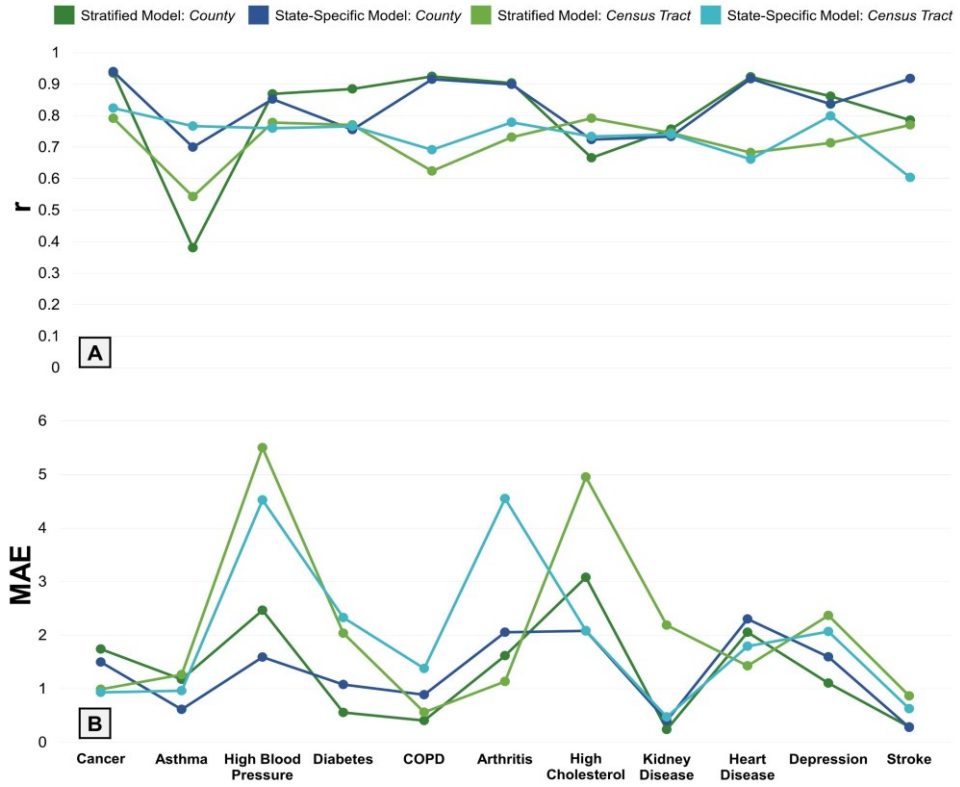
*Metrics evaluating SHAPE and CDC PLACES SAEs against direct estimates across health outcomes in **Florida**, 2019.*



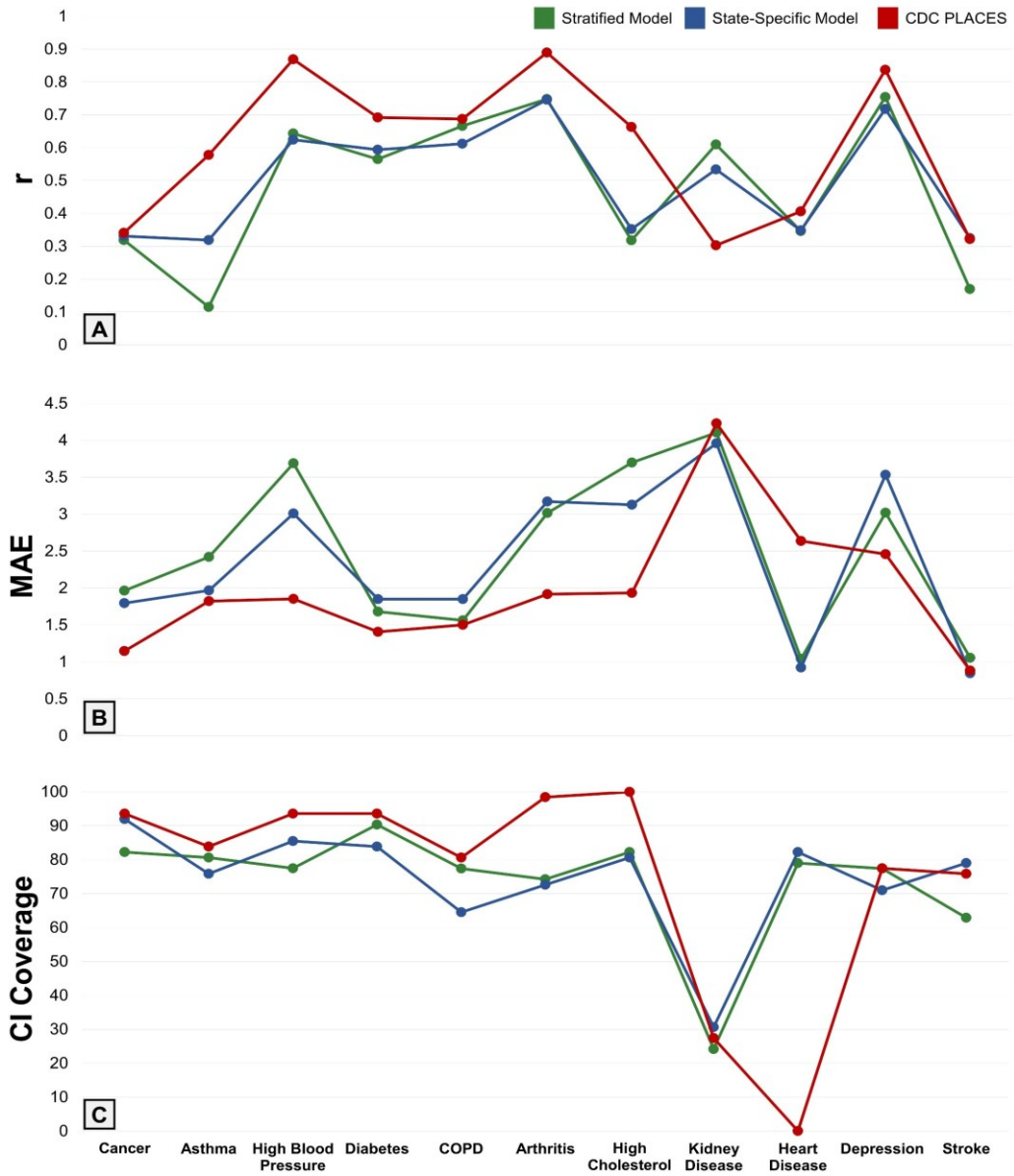
*Metrics evaluating SHAPE SAEs against CDC PLACES across health outcomes in **Florida**, 2019.*

5. SHAPE EVALUATION

- **Applied SHAPE to generate SAEs for:**
  - 2 health risk behaviors, 11 health outcomes
  - Florida (2019) and New York (2021)
  - Counties and census tracts
- **Evaluated** against BRFSS direct estimates & CDC PLACES
- *SHAPE estimates were moderately consistent with BRFSS ($\bar{r} \approx 0.54$), and strongly aligned with CDC PLACES – county ($\bar{r} \approx 0.81$) and census tracts ($\bar{r} \approx 0.74$).*



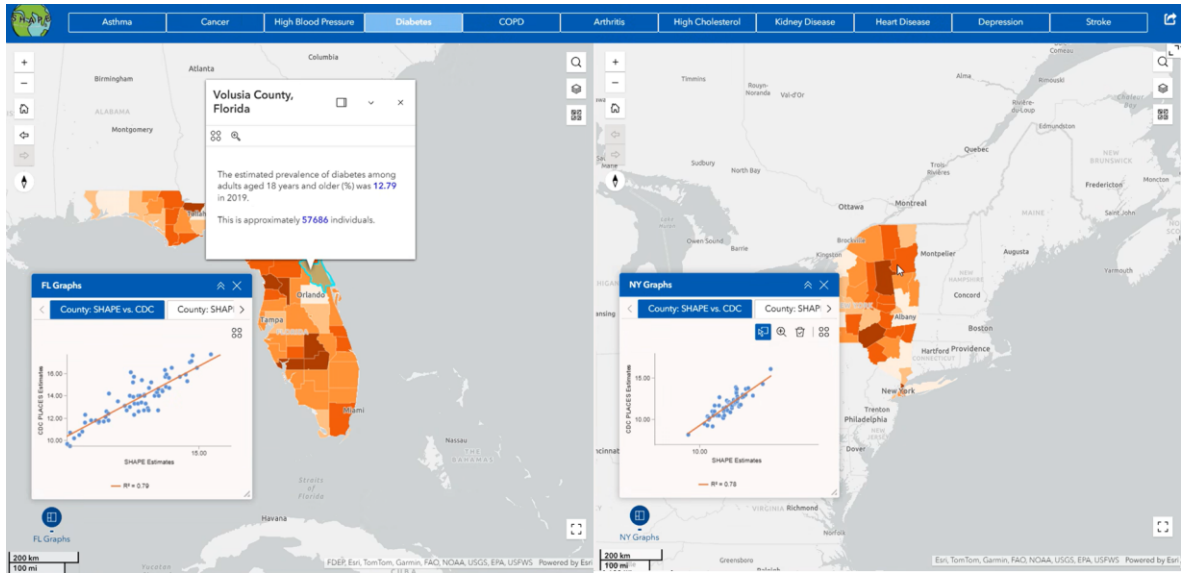
*Metrics evaluating SHAPE and CDC PLACES SAEs against direct estimates across health outcomes in **New York**, 2021.*



*Metrics evaluating SHAPE SAEs against CDC PLACES across health outcomes in **New York**, 2021.*

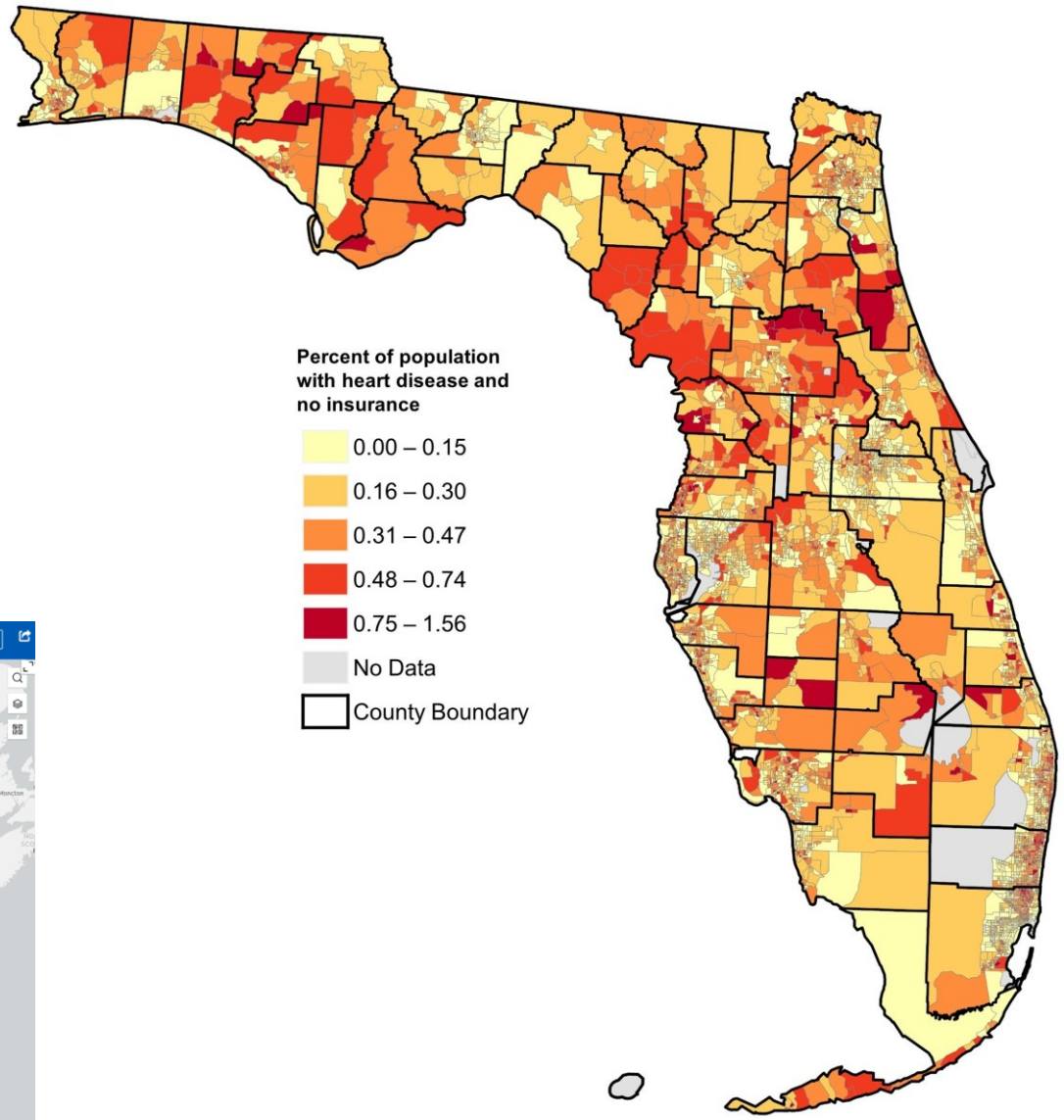
5. SHAPE IMPLICATIONS

- SHAPE generates individual-level synthetic populations that preserves joint socio-demographics and health characteristics
  - Enables **detailed health policy analyses** for identifying disparities by race, income, and location, etc.
  - Supports **downstream applications**, such as ABMs, for exploring disease dynamics, or policy scenarios
- Offers a **timely and sustainable tool** and datasets for public health researchers and stakeholders



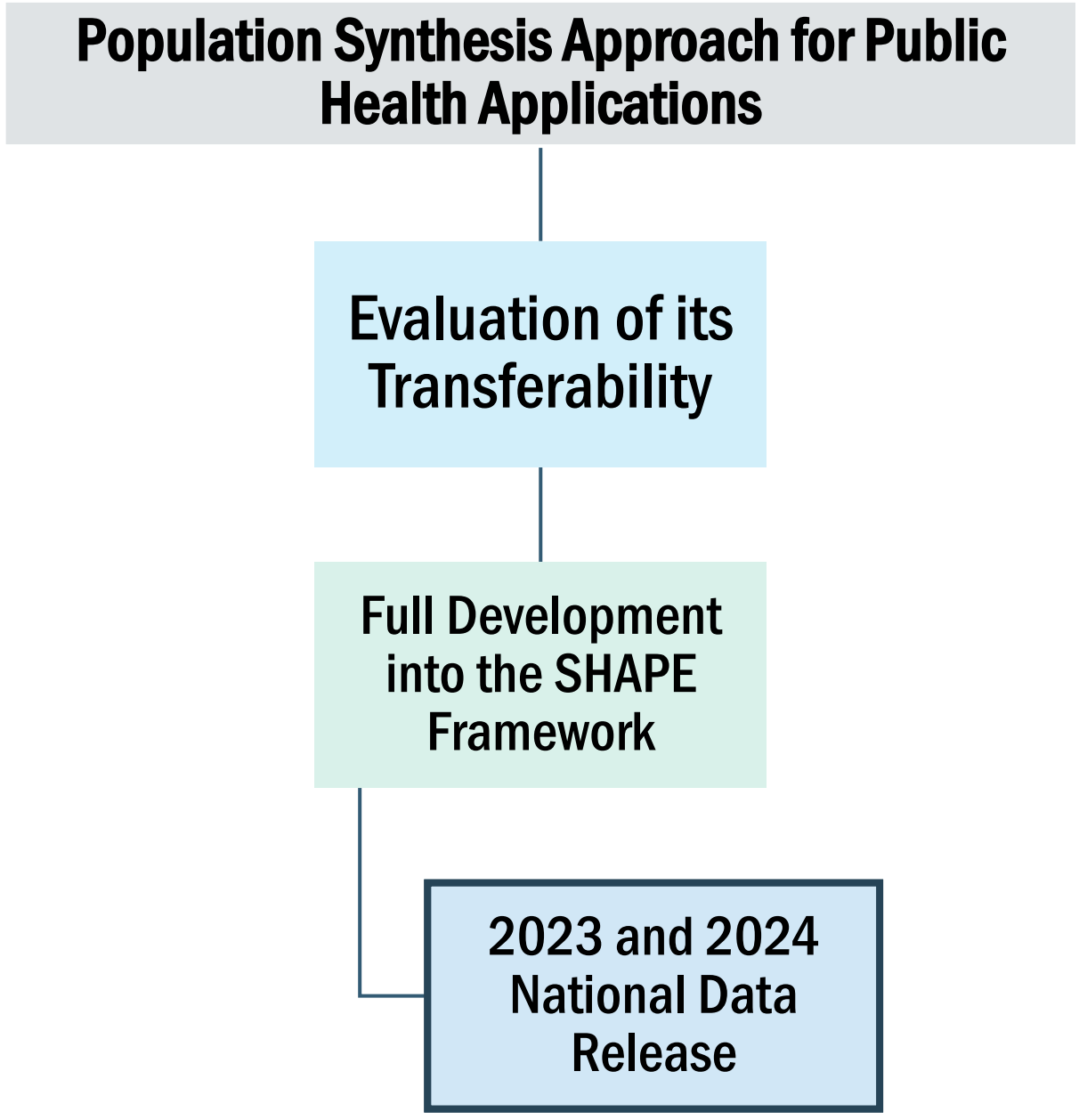
tinyurl.com/SHAPE-GIS

Interactive ArcGIS Online application for viewing SHAPE estimates.



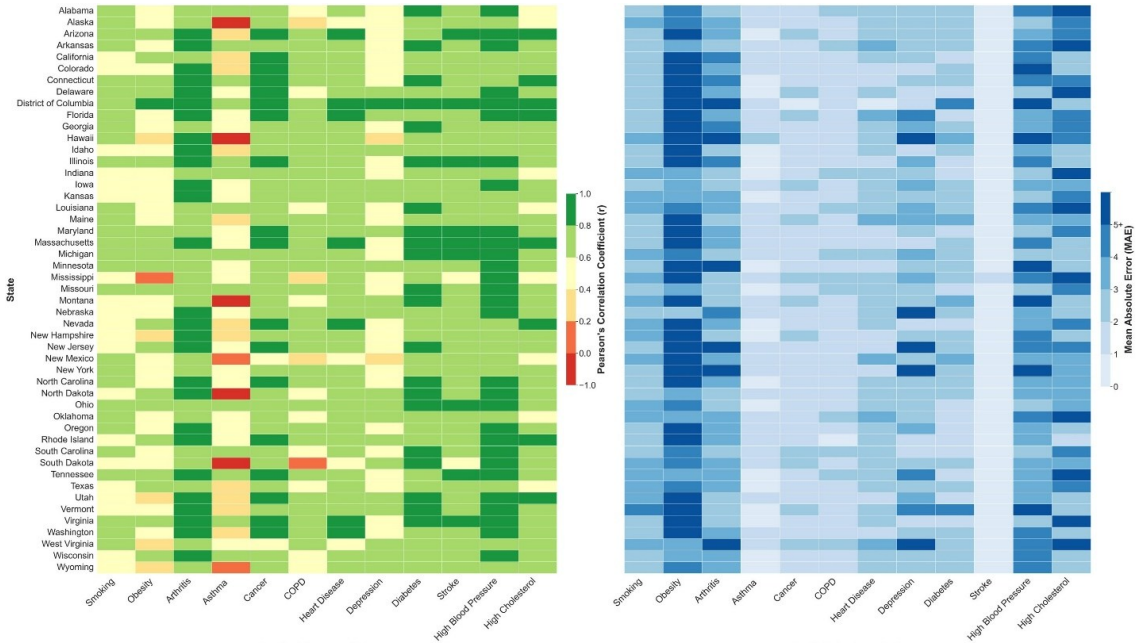
SHAPE estimates of percentage of population with heart disease and no health insurance across Florida census tracts (2019).

5. RELATED AND EXTENDED WORK



6. LATEST SHAPE OUTPUTS: A NATIONAL DATASET FOR THE U.S.

- *To our knowledge, no open-source, national individual-level synthetic population and SAE dataset exists for the U.S. that integrates sociodemographic attributes with multiple health behaviors and outcomes using transparent, reproducible methods and publicly available data sources*
  - **We offer the latest SHAPE outputs: a 2023 and 2024 spatially-explicit synthetic population (aged 18+) and corresponding SAEs at county and census tract scales for the entire U.S. (≈258 million agents)**
    - Seven sociodemographic attributes, two health behaviors, and eleven health outcomes
  - Evaluation of SHAPE's 2023 estimates with the latest CDC PLACES estimates (2023) demonstrates strong agreement, with variation by outcome and location

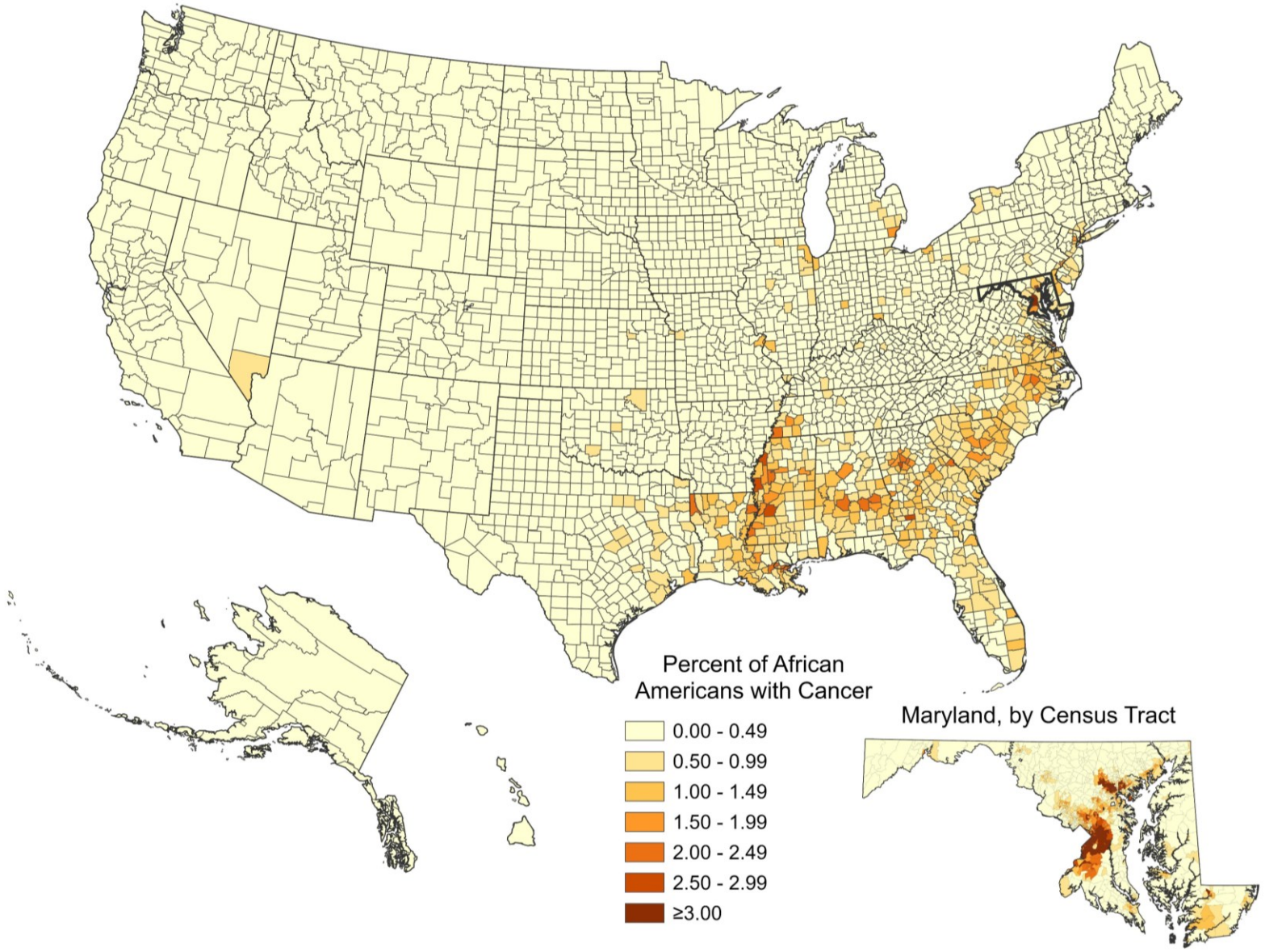


Heat map with the agreement between 2023 SHAPE and CDC PLACES estimates by state.

Health Behavior/Outcome	Pearson's r	MAE
Arthritis	0.862	1.732
Heart Disease	0.838	3.159
High Blood Pressure	0.819	2.527
Diabetes	0.815	1.885
Cancer	0.790	2.359
Smoking	0.771	1.703
COPD	0.635	1.781
Stroke	0.603	0.680
Depression	0.510	2.263
High Cholesterol	0.492	4.943
Obesity	0.420	3.503
Asthma	0.335	0.752

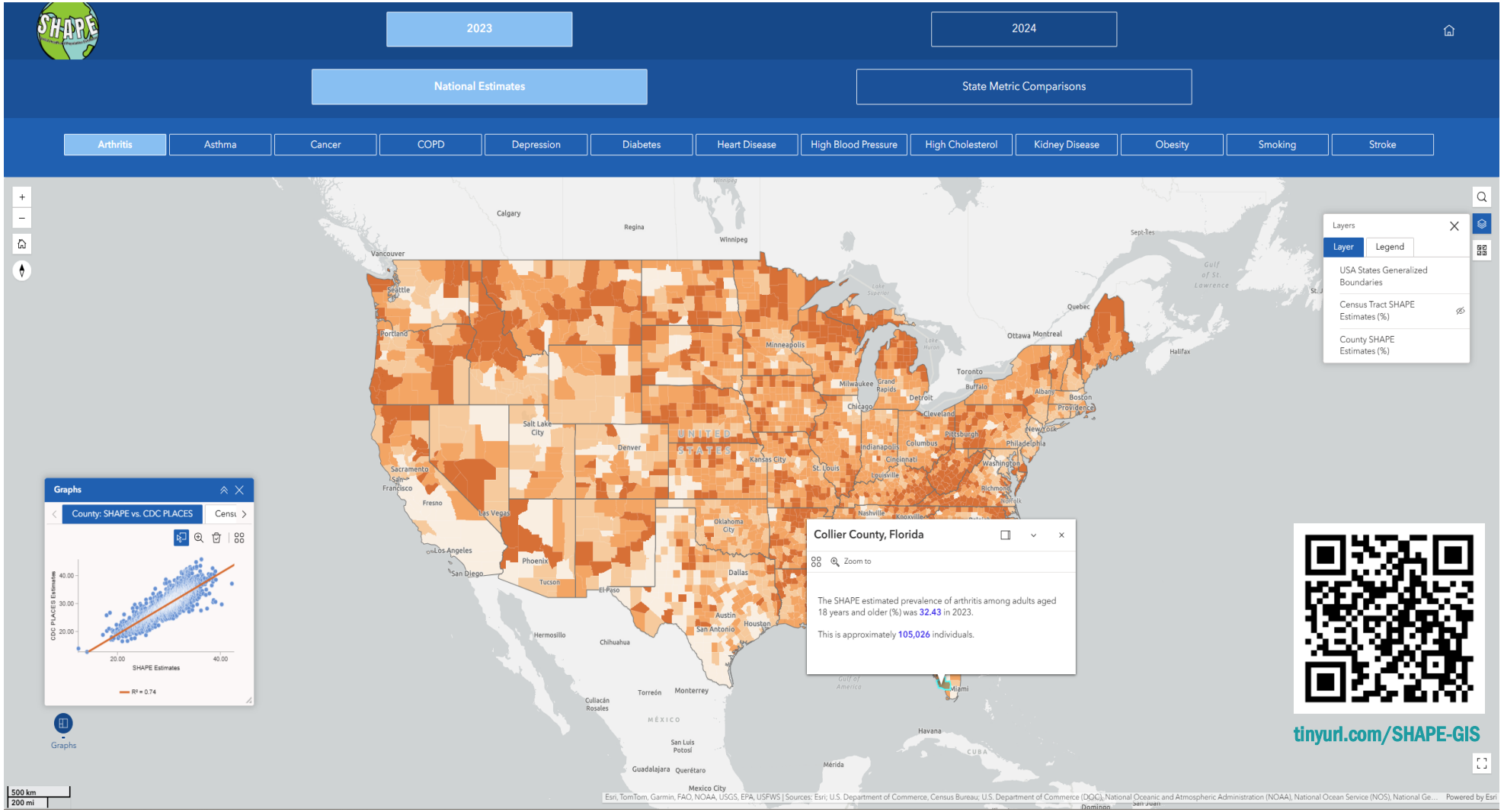
Metrics comparing the 2023 national SHAPE SAEs at the county level for each health behavior and outcome with the corresponding CDC PLACES estimates.

6. LATEST SHAPE OUTPUTS: A NATIONAL DATASET FOR THE U.S.



County-level map of the estimated percentage of African American individuals diagnosed with cancer, with an inset at the census-tract level for Maryland. These estimates were calculated by aggregating joint sociodemographic and health attributes from the 2023 national synthetic population.

6. LATEST SHAPE OUTPUTS: A NATIONAL DATASET FOR THE U.S.



SHAPE's GitHub:

https://github.com/evonhoene/SHAPE

- Input Data
- Validation Data
- R Scripts
- National SAEs

Synthetic Populations on OSF
→ *Coming soon*

Screenshot of our GIS dashboard that contains interactive maps of SHAPE's SAEs for 2023 and 2024. The example shown illustrates the spatial distribution of 2023 arthritis estimates, along with dynamic scatterplots, interactive pop-ups, and layer controls that allow users to toggle between county- and census tract–level estimates.

Thank you!

Questions?

Contact Us:

Emma Von Hoene (evonhoen@gmu.edu)

Amira Roess (aroess@gmu.edu)

Hamdi Kavak (hkavak@gmu.edu)

Taylor Anderson (tander6@gmu.edu)

Acknowledgements:

This research is funded by the National Science Foundation (Award #2030685 and #2109647).

College of Science
DEPARTMENT OF GEOGRAPHY AND GEOINFORMATION SCIENCE
George Mason University