Synthetic Population Generation with Public Health Characteristics for Spatial Agent-Based Models

Modeling and Analysis Innovation Center (MAIC) Brown Bag Invited Presentation

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



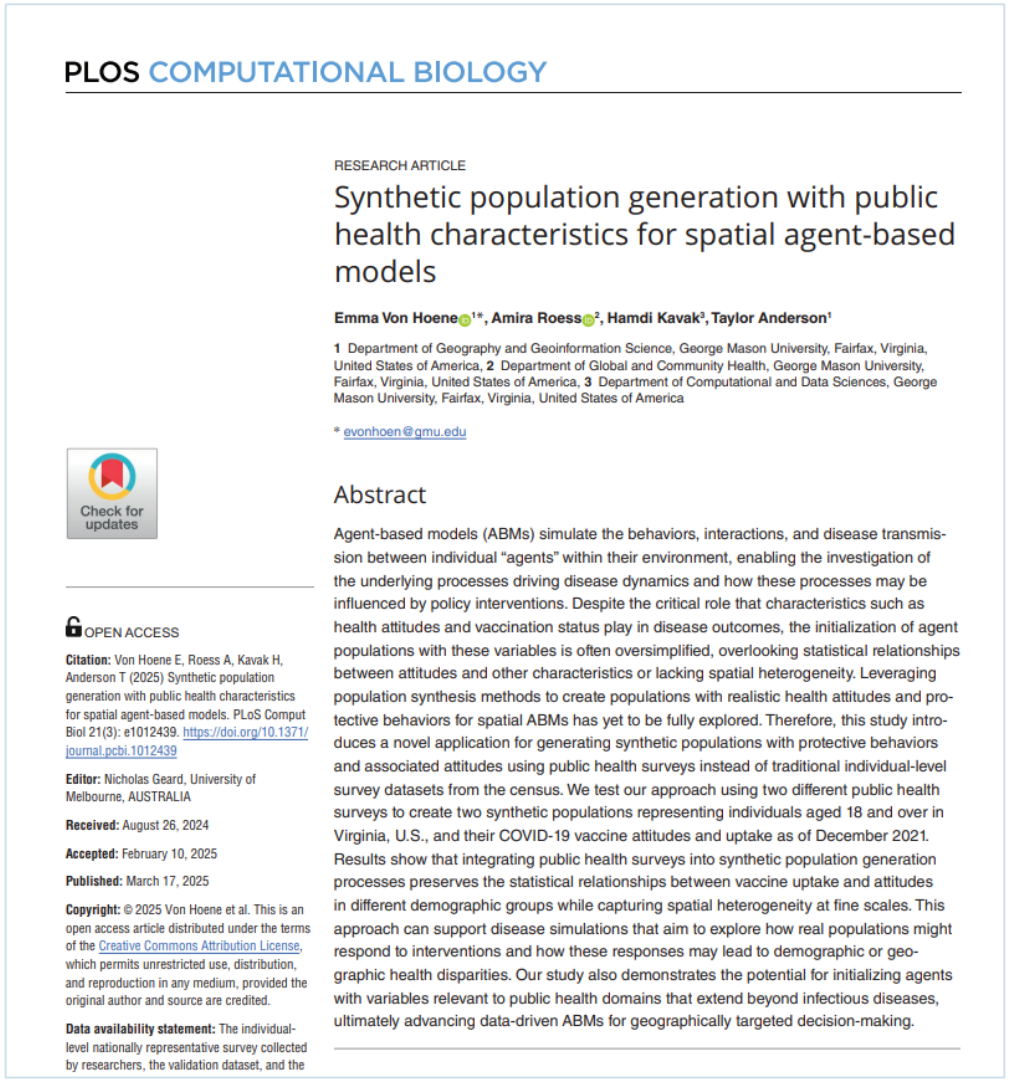
MORE ABOUT ME

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

- **Research** focus:
  - Spatial data science, data-driven modeling and geospatial analysis
  - Currently working with applications in public health

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

NOVEL APPLICATION FOR SYNTHETIC POPULATION GENERATION



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

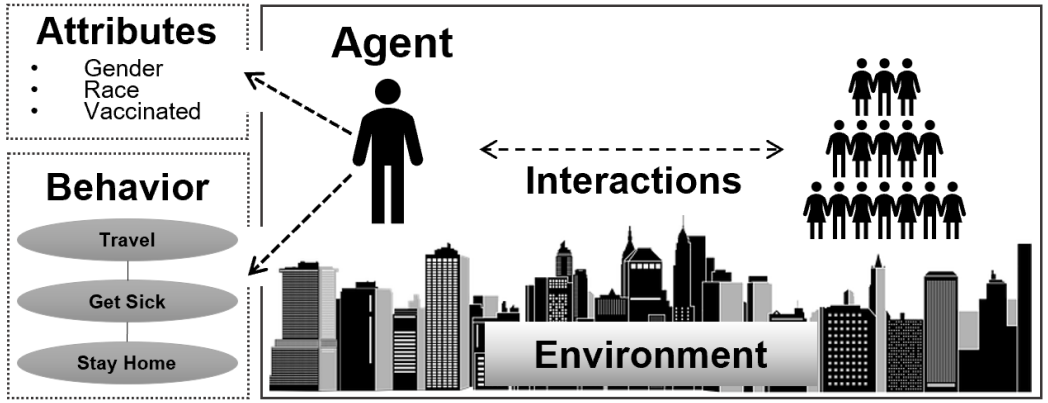
Abstract

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.

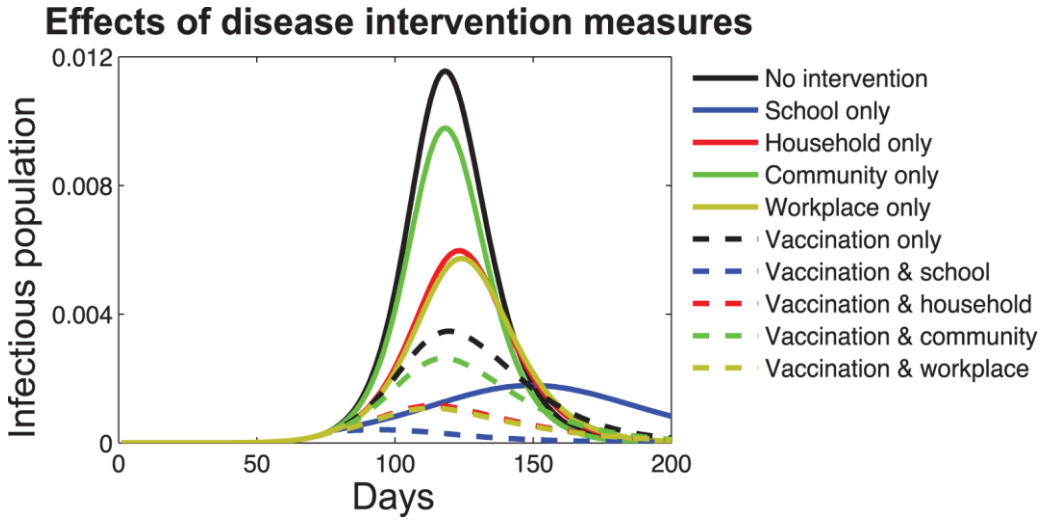
Von Hoene et al. (2025), PLOS Computational Biology: *https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012439*

BACKGROUND

- Simulations, such as **agent-based models (ABMs)**, are essential for studying infectious disease spread
- **Health behaviors** are key drivers of disease trajectories
- ABMs across a range of public health applications traditionally ***ignore*** or and continue to ***oversimplify*** the human behavior component

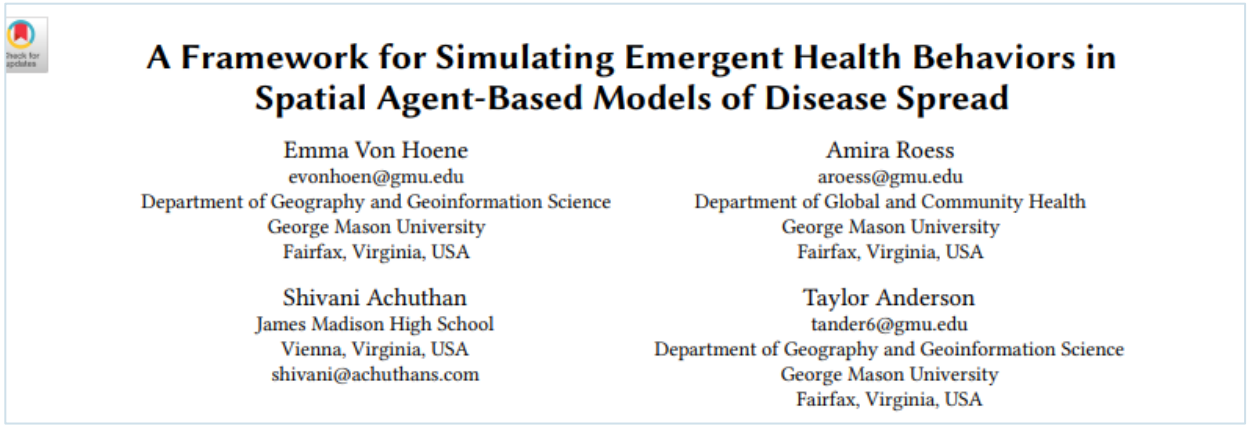


Conceptual representation of an ABM.



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

MOTIVATION



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

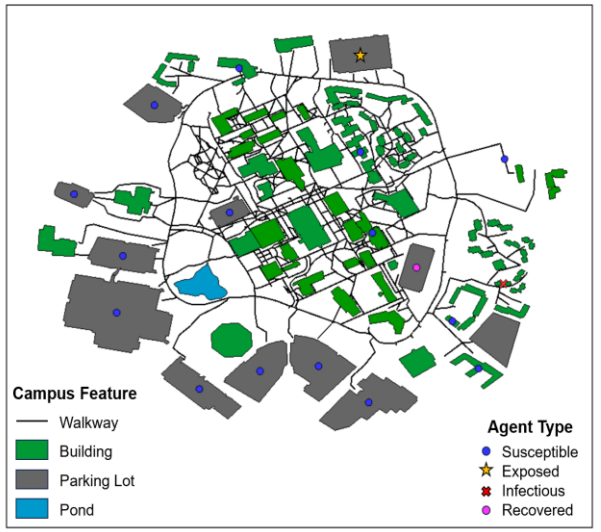
Emma Von Hoene
evonhoen@gmu.edu
Department of Geography and Geoinformation Science
George Mason University
Fairfax, Virginia, USA

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

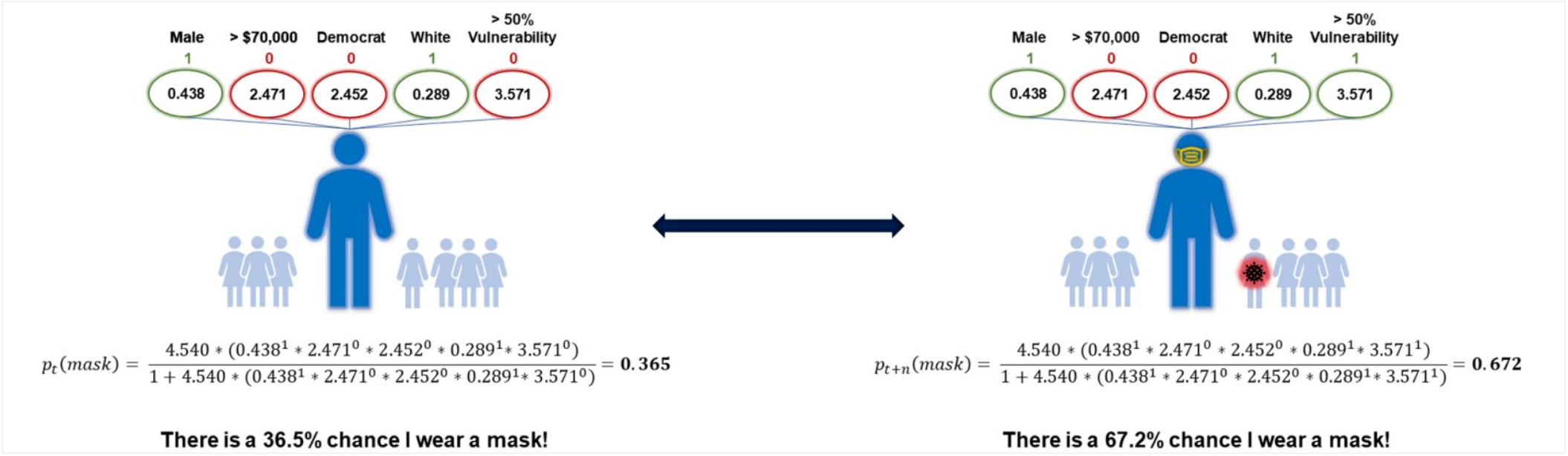
Shivani Achuthan
James Madison High School
Vienna, Virginia, USA
shivani@achuthans.com

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

Von Hoene et al. (2024), SIGSPATIAL Conference:
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012439



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

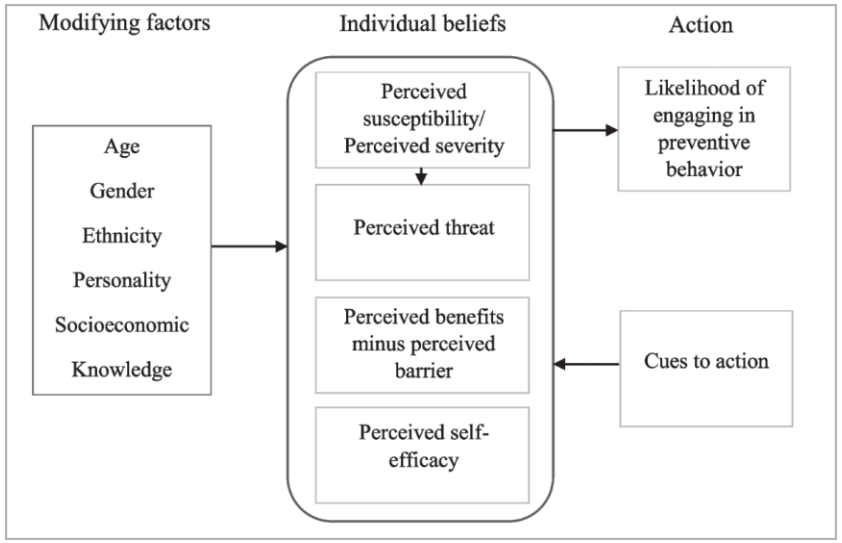


Representation of the decision-making framework using odds ratios corresponding to agent attributes in a logistic regression model to predict dynamic mask-use.

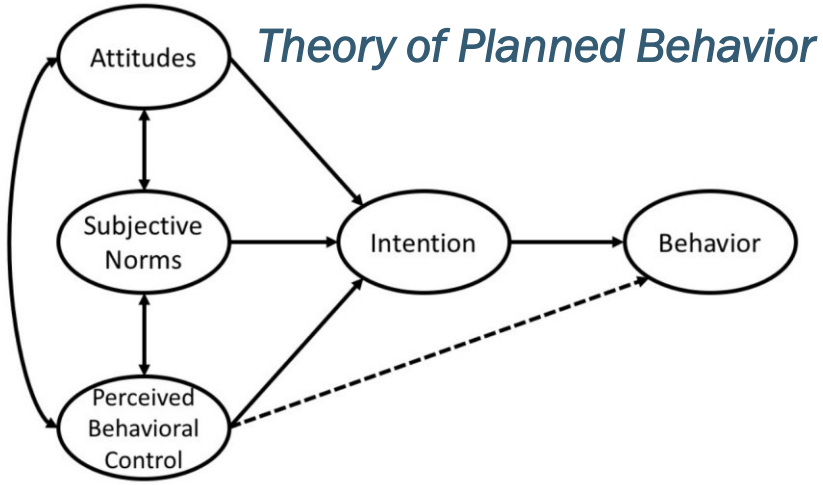
MOTIVATION

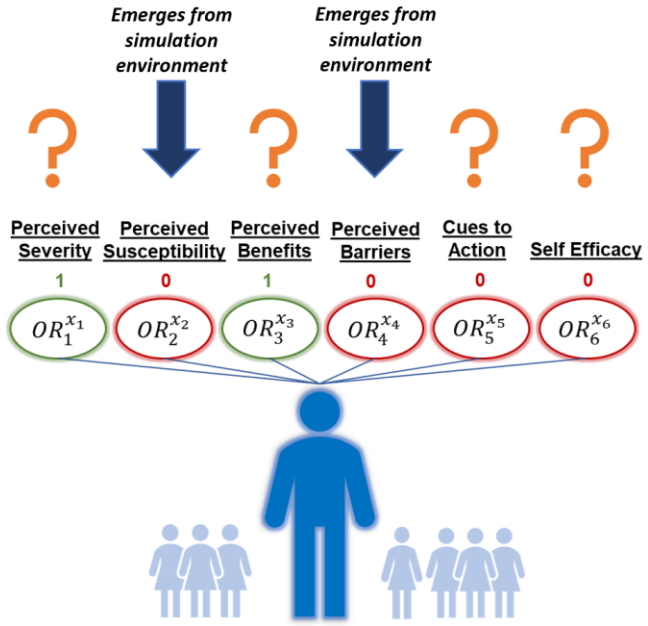
Domain Specific Frameworks

Health Belief Model

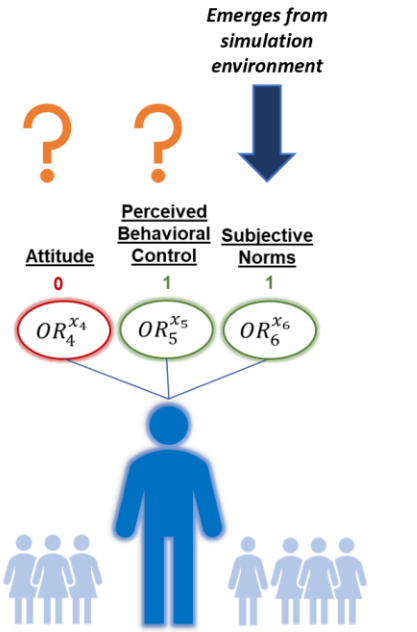


Theory of Planned Behavior





Health Belief Model Variables



Theory of Planned Behavior Variables

EXISTING RESEARCH

- 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.

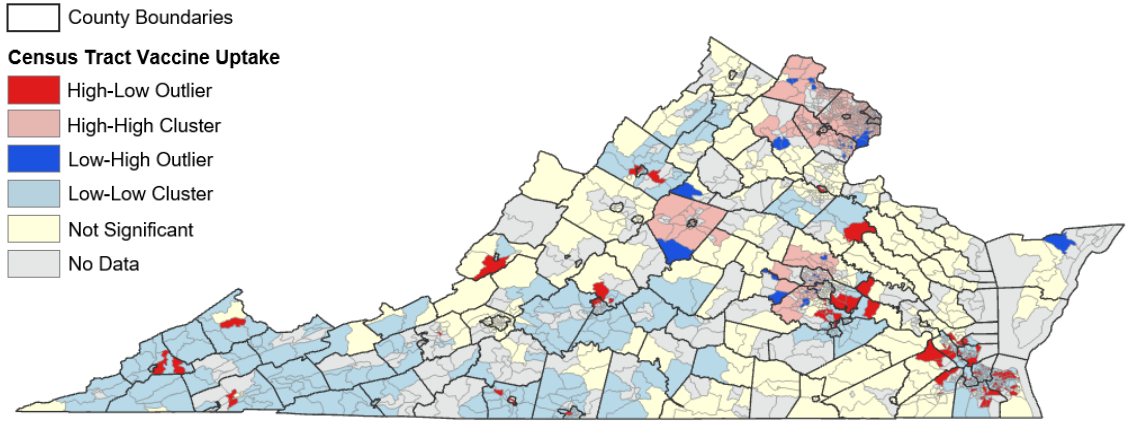
Rafferty E, McDonald W, Osgood N, Qian W, Doroshenko A. Seeking the optimal schedule for chickenpox vaccination in Canada: Using an agent-based model to explore the impact of dose timing, coverage and waning of immunity on disease outcomes. Vaccine. 2020;38(3):521–9.

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

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.

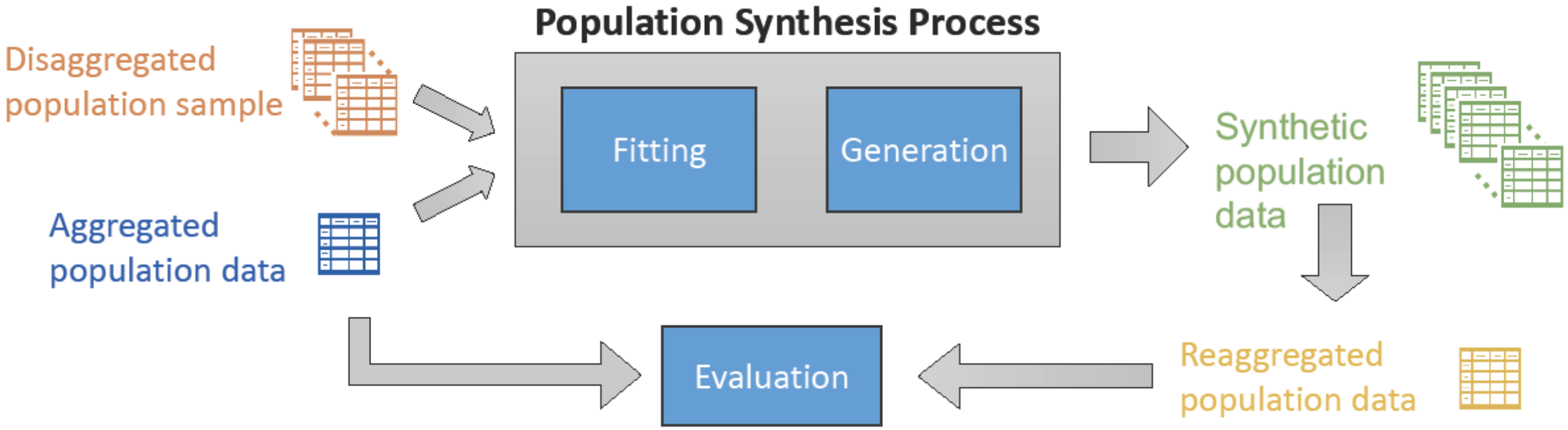
Relationships between demographic groups and attitudes towards COVID-19 vaccine intention [UAS 282, February – March 2021].

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

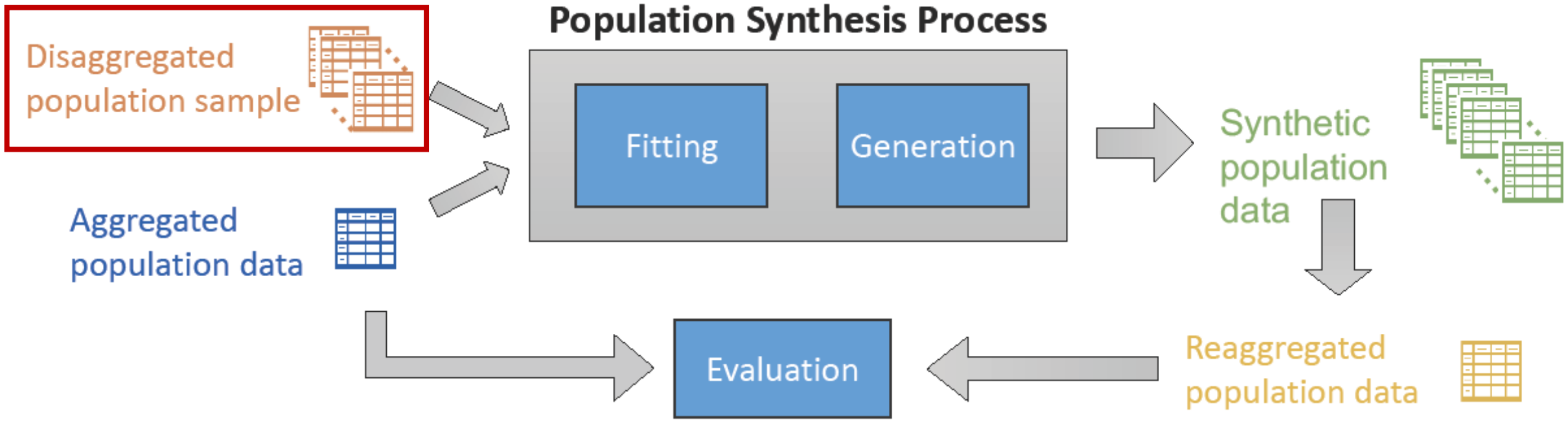
METHODOLOGY SUMMARY

- **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.

POPULATION SYNTHESIS PROCESS OVERVIEW

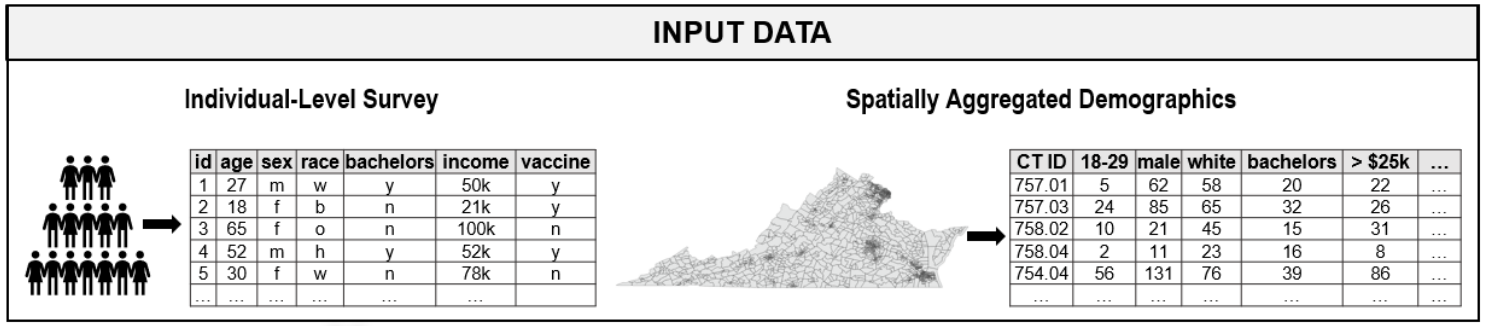


POPULATION SYNTHESIS PROCESS OVERVIEW



DATASETS

- **Input data**
  - **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

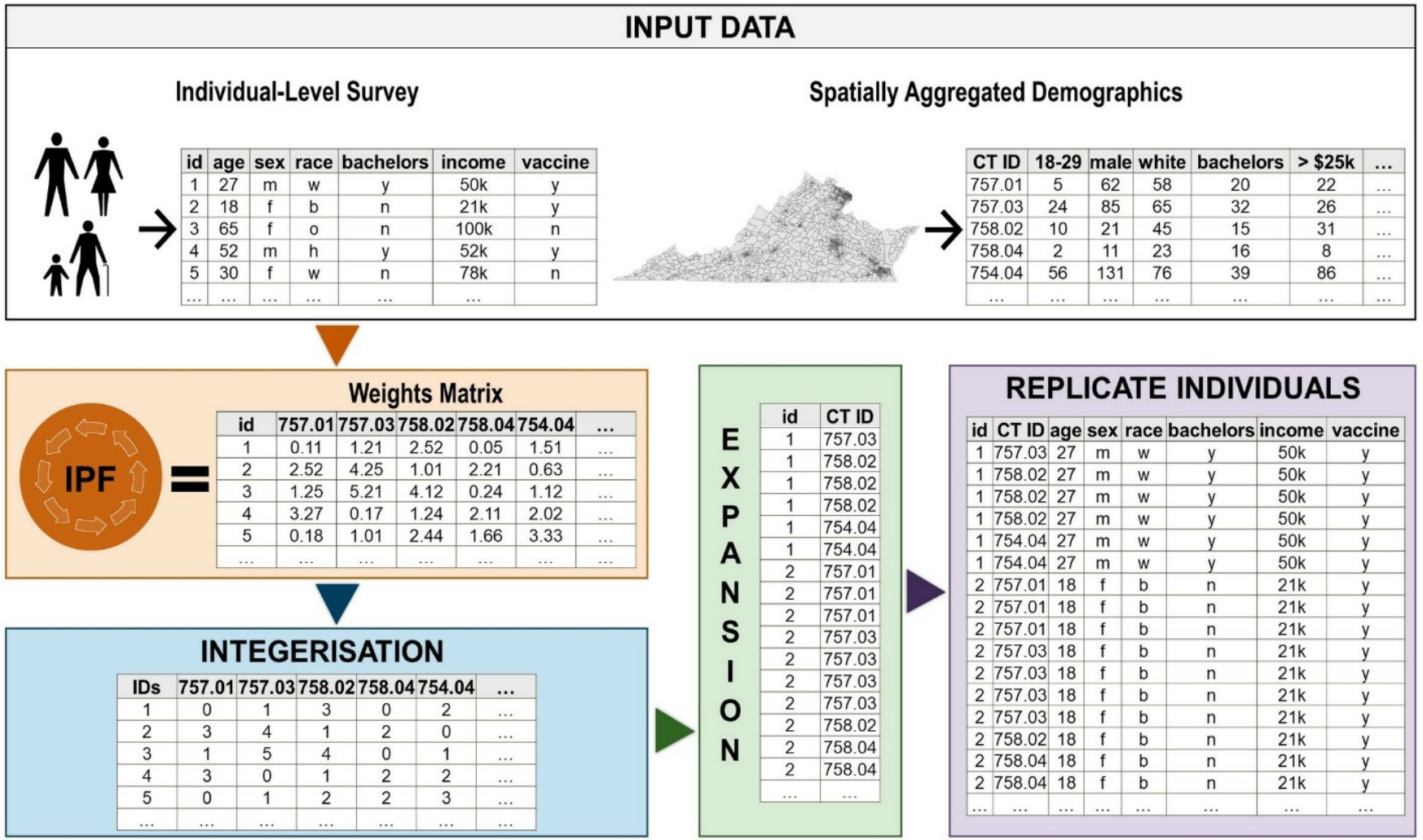


id	age	sex	race	bachelors	income	vaccine
1	27	m	w	y	50k	y
2	18	f	b	n	21k	y
3	65	f	o	n	100k	n
4	52	m	h	y	52k	y
5	30	f	w	n	78k	n
…	…	…	…	…	…	

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	…
…	…	…	…	…	…	…

- **Validation data**
  - Census tract vaccine uptake for Virginia (Virginia Department of Health – by request)

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	…
…	…	…	…	…	…	…

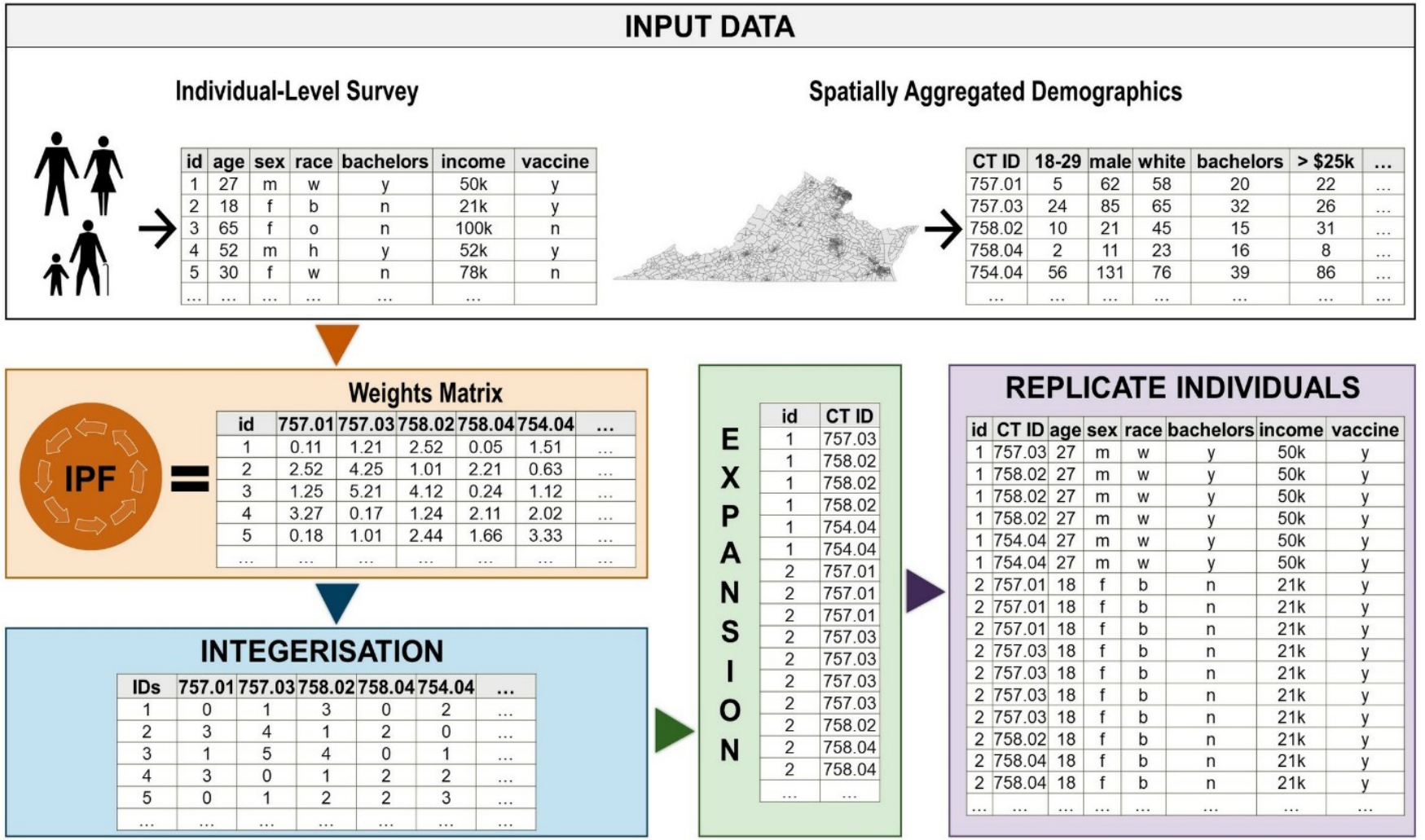
IDs	757.01	757.03	758.02	758.04	754.04	…
1	0	1	3	0	2	…
2	3	4	1	2	0	…
3	1	5	4	0	1	…
4	3	0	1	2	2	…
5	0	1	2	2	3	…
…	…	…	…	…	…	…

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
…	…	…	…	…	…	…	…

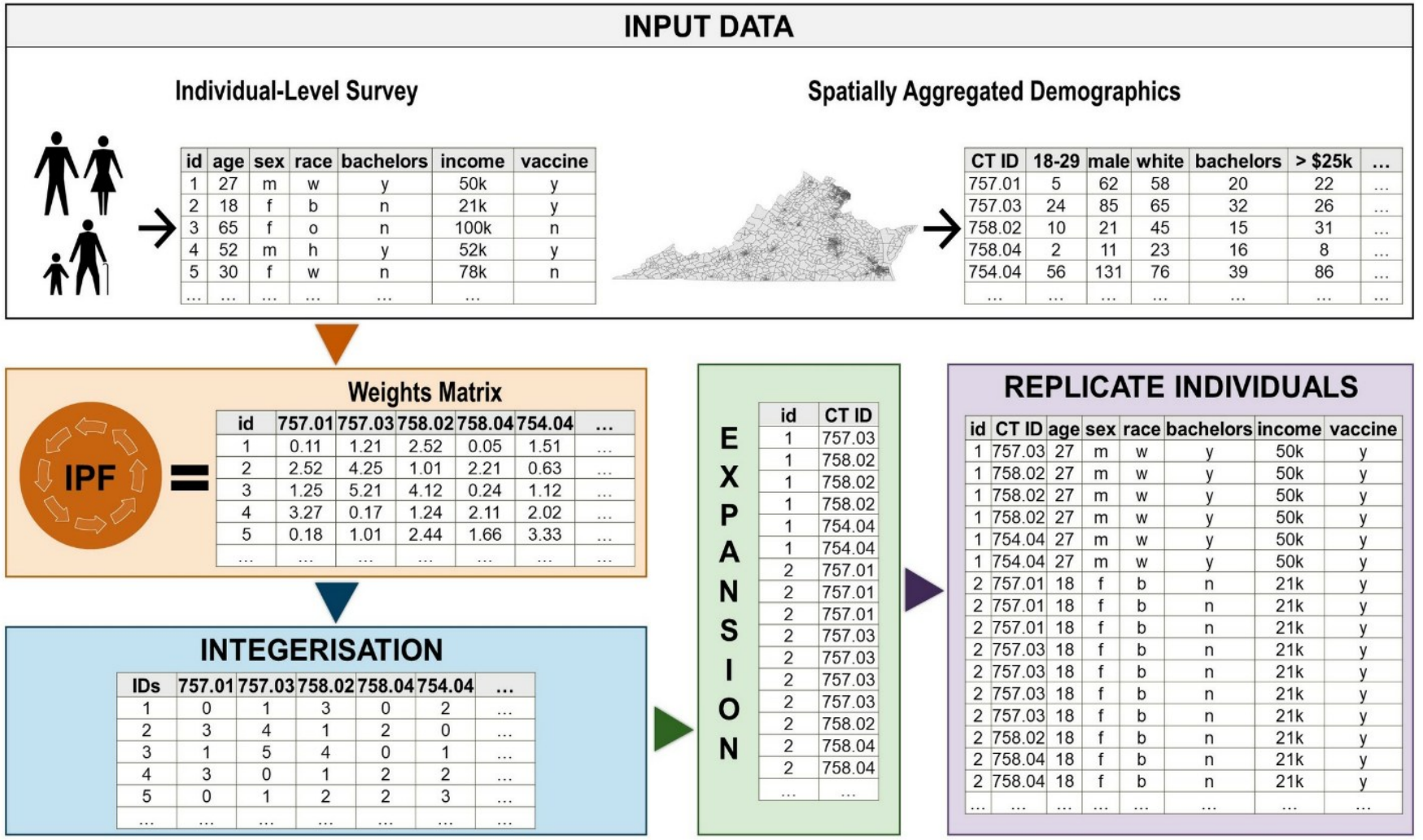
PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS

- **Input Data**
- Input and process survey and spatially aggregated data
- Focus on gender, race, age, education, and income variables for individuals aged 18 and over, since these demographics influence COVID vaccine uptake and beliefs



PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS

- **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	...
...	...	...	...	...	...	...

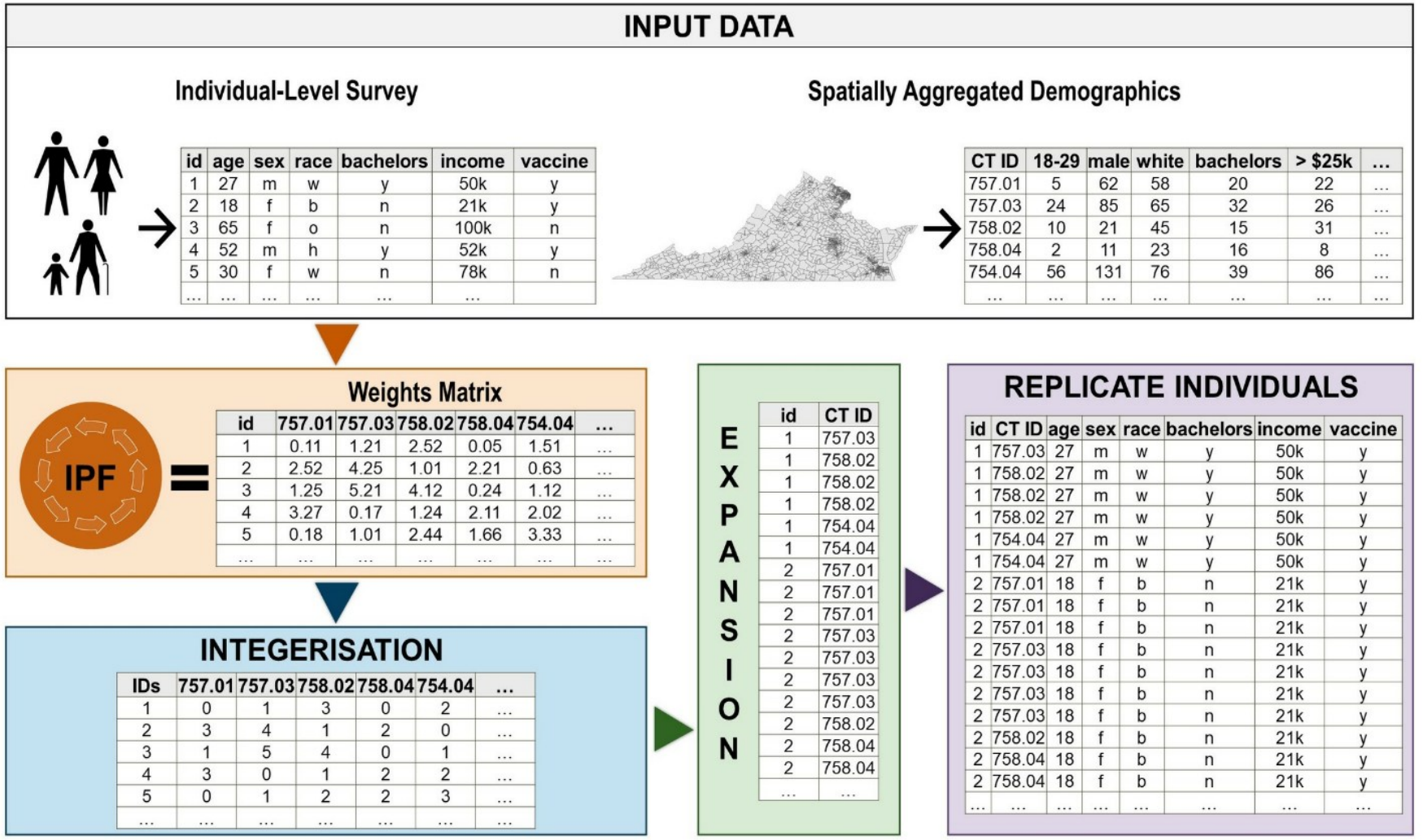
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

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
...	...	...	...	...	...	...	...

PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS

- **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	...
...	...	...	...	...	...	...

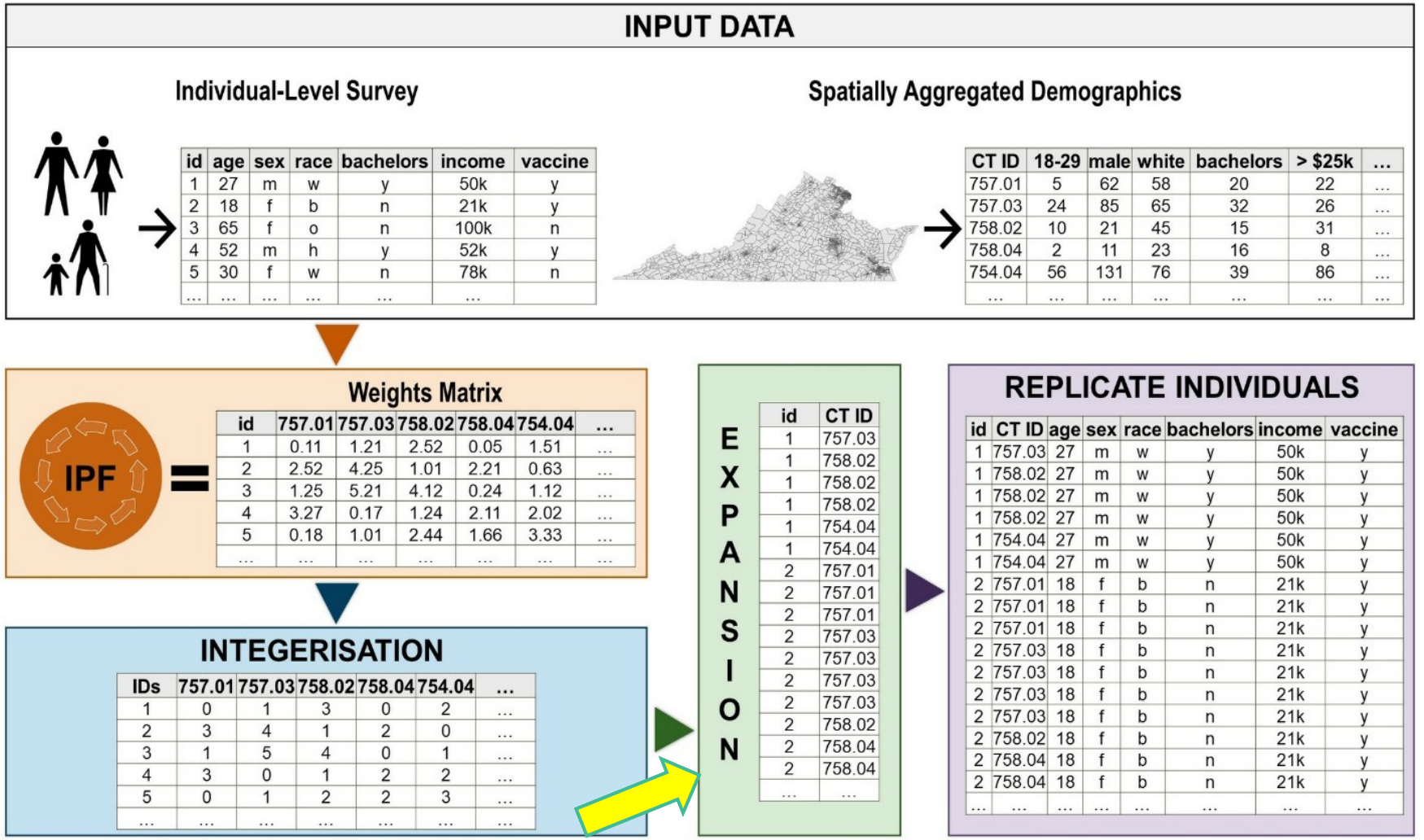
IDs	757.01	757.03	758.02	758.04	754.04	...
1	0	1	3	0	2	...
2	3	4	1	2	0	...
3	1	5	4	0	1	...
4	3	0	1	2	2	...
5	0	1	2	2	3	...
...	...	...	...	...	...	...

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
...	...	...	...	...	...	...	...

PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS

- **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	…
…	…	…	…	…	…	…

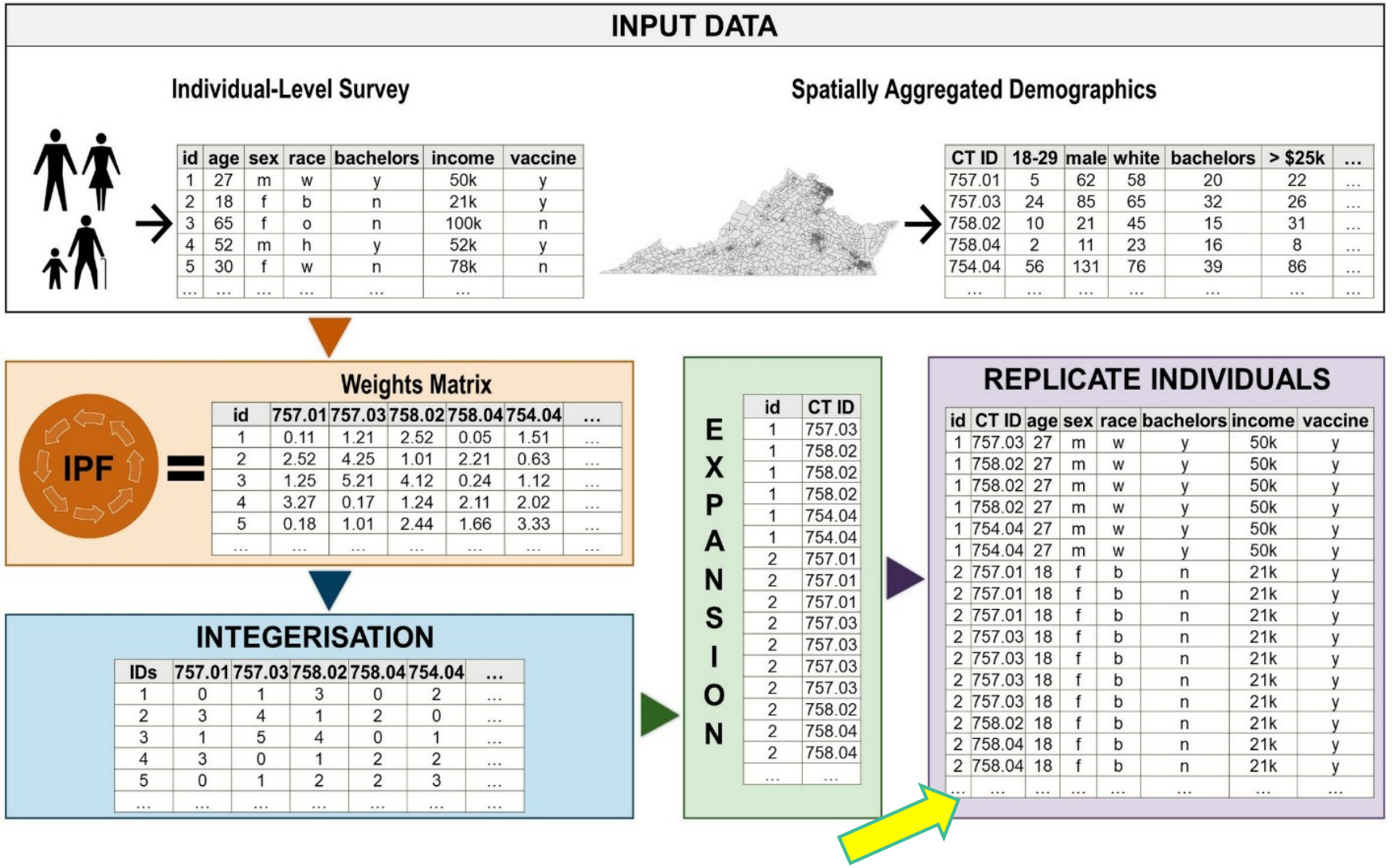
IDs	757.01	757.03	758.02	758.04	754.04	…
1	0	1	3	0	2	…
2	3	4	1	2	0	…
3	1	5	4	0	1	…
4	3	0	1	2	2	…
5	0	1	2	2	3	…
…	…	…	…	…	…	…

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
…	…

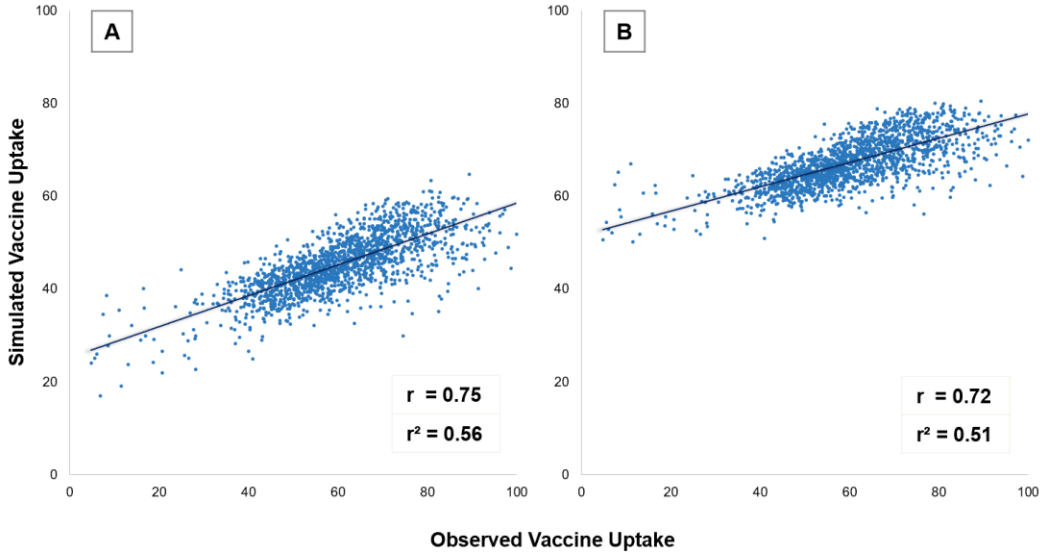
id	CT ID	age	sex	race	bachelors	income	vaccine
1	757.03	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	758.02	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
1	754.04	27	m	w	y	50k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.01	18	f	b	n	21k	y
2	757.03	18	f	b	n	21k	y
2	757.03	18	f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

PROPOSED METHOD: NOVEL APPLICATION OF POPULATION SYNTHESIS

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



RESULTS: OVERALL ACCURACY



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

	HPS Dataset				NRS Dataset			
Variable: Descriptor	***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.

RESULTS: STATISTICAL RELATIONSHIPS ACROSS DEMOGRAPHIC GROUPS

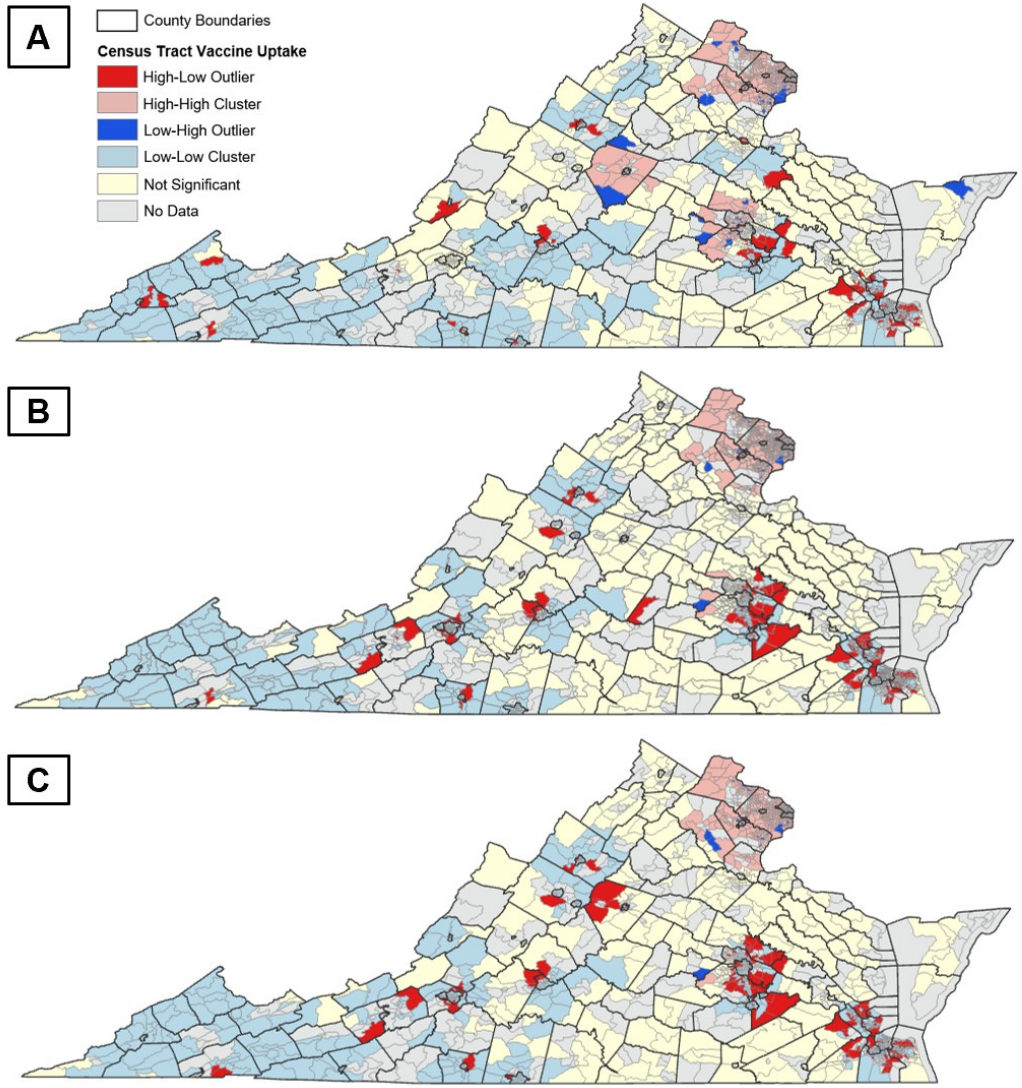
Coefficients from logistic regression models describing the relationship between demographic variables and vaccine uptake from the HPS, along with the synthetic population generated from the HPS. Significant coefficients are indicated with an asterisk (*) at the 90% confidence level (p-value < 0.10).

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*

RESULTS: SPATIAL HETEROGENIETY IN BEHAVIOR

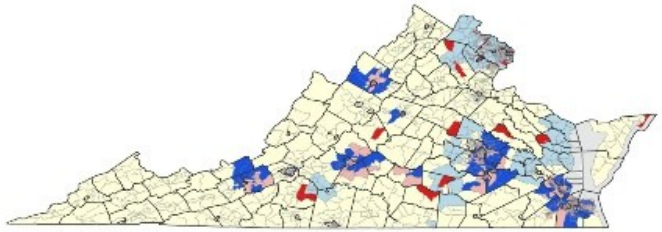


Clusters and outliers of vaccination uptake in Virginia Census Tracts: A) observed, B) synthesized population with HPS data, and C) synthesized population with nationally representative survey collected by researchers.

RESULTS: SPATIAL PATTERNS OF INDIVIDUAL ATTITUDES, BELIEFS AND PERCEPTIONS

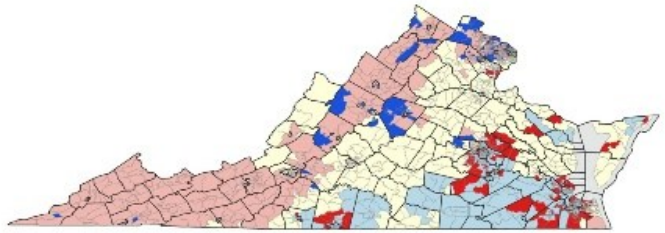
Perceived Behavioral Control/Barrier

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



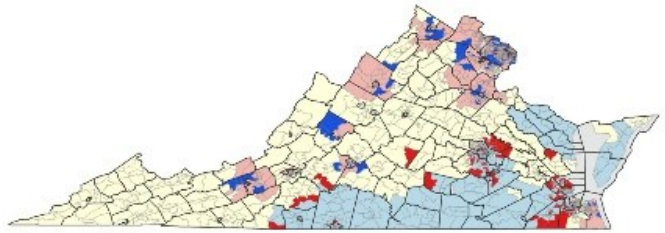
Social Norms

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



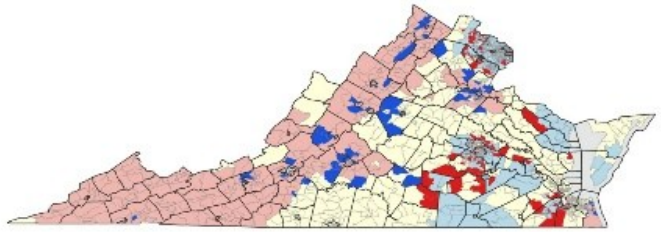
Attitude

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



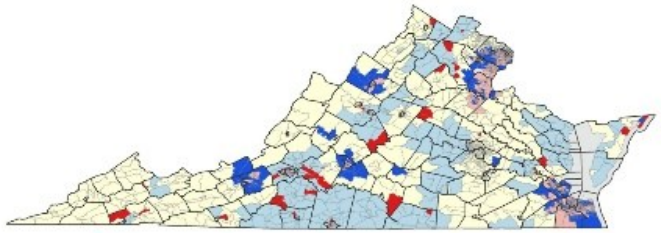
Attitude

D. I don't trust COVID-19 vaccines



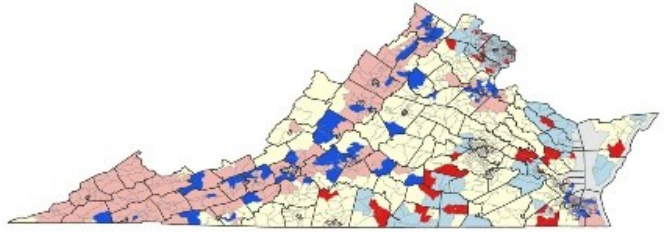
Perceived Behavioral Control/Barrier

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



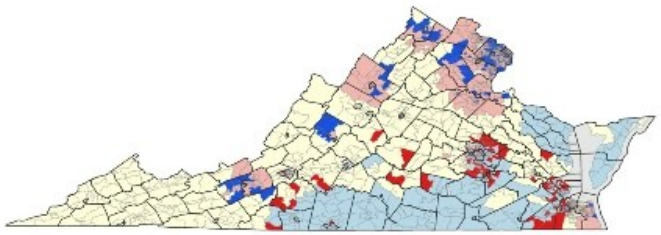
Perceived Behavioral Control/Barrier

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



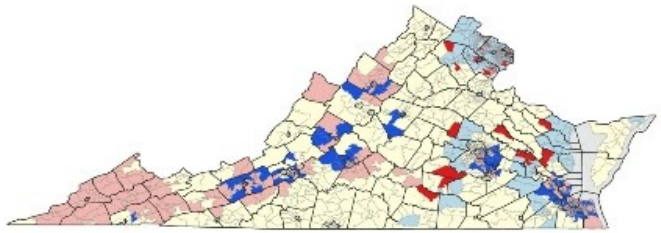
Attitude

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



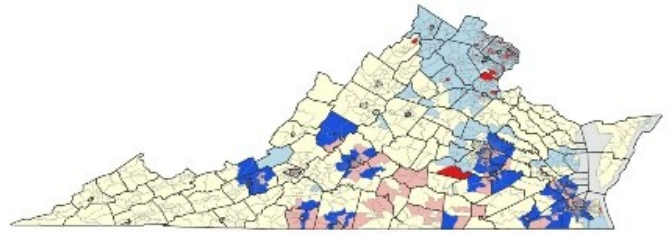
Perceived Behavioral Control/Barrier

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

DISCUSSION

SUMMARY

- 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
- The approach is flexible, with the potential to generate populations with characteristics from any survey, for any application

BROADER IMPACTS

- Generates synthetic spatial data at finer scales (e.g., census tracts) that are not available
- Offers policymakers new data for better informing geographically targeted interventions and focusing on vulnerable populations
- Allows modelers already using synthetic population approaches to easily incorporate domain-specific attributes in simulations

FUTURE RESEARCH BEYOND OUR PUBLICATION…

Exploring the adaptability of our proposed population synthesis approach:

- Connecting initialization of population to domain-specific frameworks that explain human behavior
- Generating and validating populations with other behavioral characteristics, and across different locations and time periods
- Comparing the method to other commonly used approaches used for generating synthetic populations with health-related attributes
- Generating a synthetic population and small area estimates with vaccine-related attributes for entire US

GROUNDING POPULATION SYNTHESIS IN THEORY AND DATA

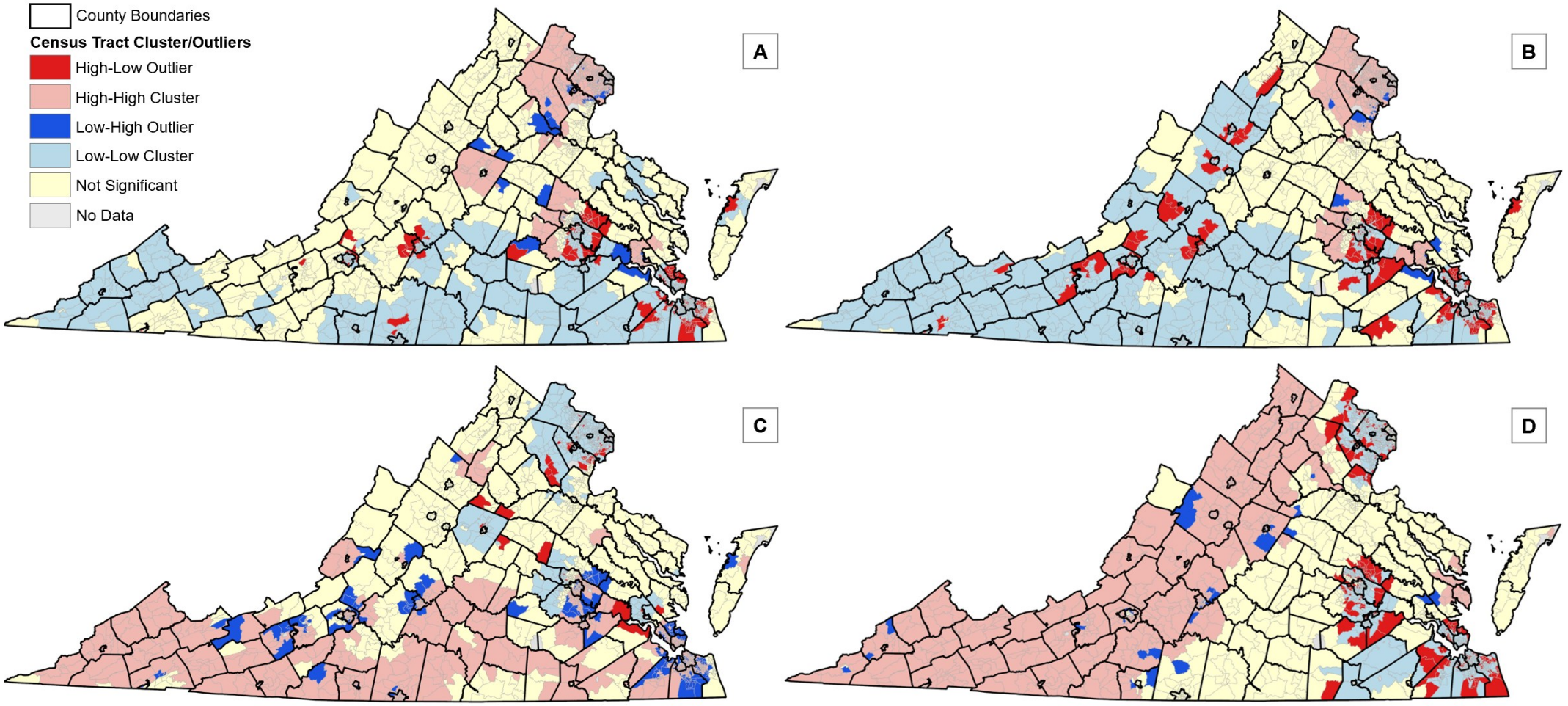
Connecting initialization of population to domain-specific frameworks that explain human behavior

- Using March 2021 nationally representative survey data from the Understanding America Survey (UAS), we mapped survey questions to theoretical constructs from the TPB and HBM
- This showcases how the method can be tailored to generate domain-specific populations that are grounded in relevant theory and data

Survey questions from UAS 282 (March 2021) connected to public health theories

Theory	Theoretical Construct	Code	Survey Question
TPB	*Attitude*	*ucv005*	How much do you trust the process in general (not just for COVID-19) to develop safe vaccines for the public?
	Attitude	*ucv006*	How much do you trust the governmental approval process to ensure the COVID-19 vaccine is safe for the public?
	Attitude	*cr073*	On a scale of 0 to 100, what is the percent chance that someone who is vaccinated against the coronavirus could still catch it? If you're not sure, please give your best guess.
	Attitude	*cr074*	On a scale of 0 to 100, what is the percent chance that a coronavirus vaccine will cause serious side effects or long-term health problems for someone who has been vaccinated? If you are unsure, please give your best guess.
	Social Norms	*vd003*	What percentage of your friends and family members do you think will get the COVID-19 vaccine when it becomes available to them? Include individuals who are already vaccinated.
	Perceived behavioral control	*cr013*	Do you currently have health insurance?
HBM	*Perceived Benefit*	*cr072b*	Do you agree or disagree with the following statements? Vaccines for the coronavirus (COVID-19): 2) Provide important benefits to society
	Perceived Susceptibility	*cr023*	On a scale of 0 to 100 percent, what is the chance that you will get the coronavirus in the next three months? If you are not sure, please give your best guess.
	Perceived Severity	*cr023b*	If you do get the coronavirus, what is the percent chance you will be hospitalized (spend at least one night in the hospital) from it? If you are not sure, please give your best guess.
	Perceived Barrier	*cr072a*	Do you agree or disagree with the following statements? Vaccines for the coronavirus (COVID-19): 1) Have many known harmful side effects

GROUNDING POPULATION SYNTHESIS IN THEORY AND DATA



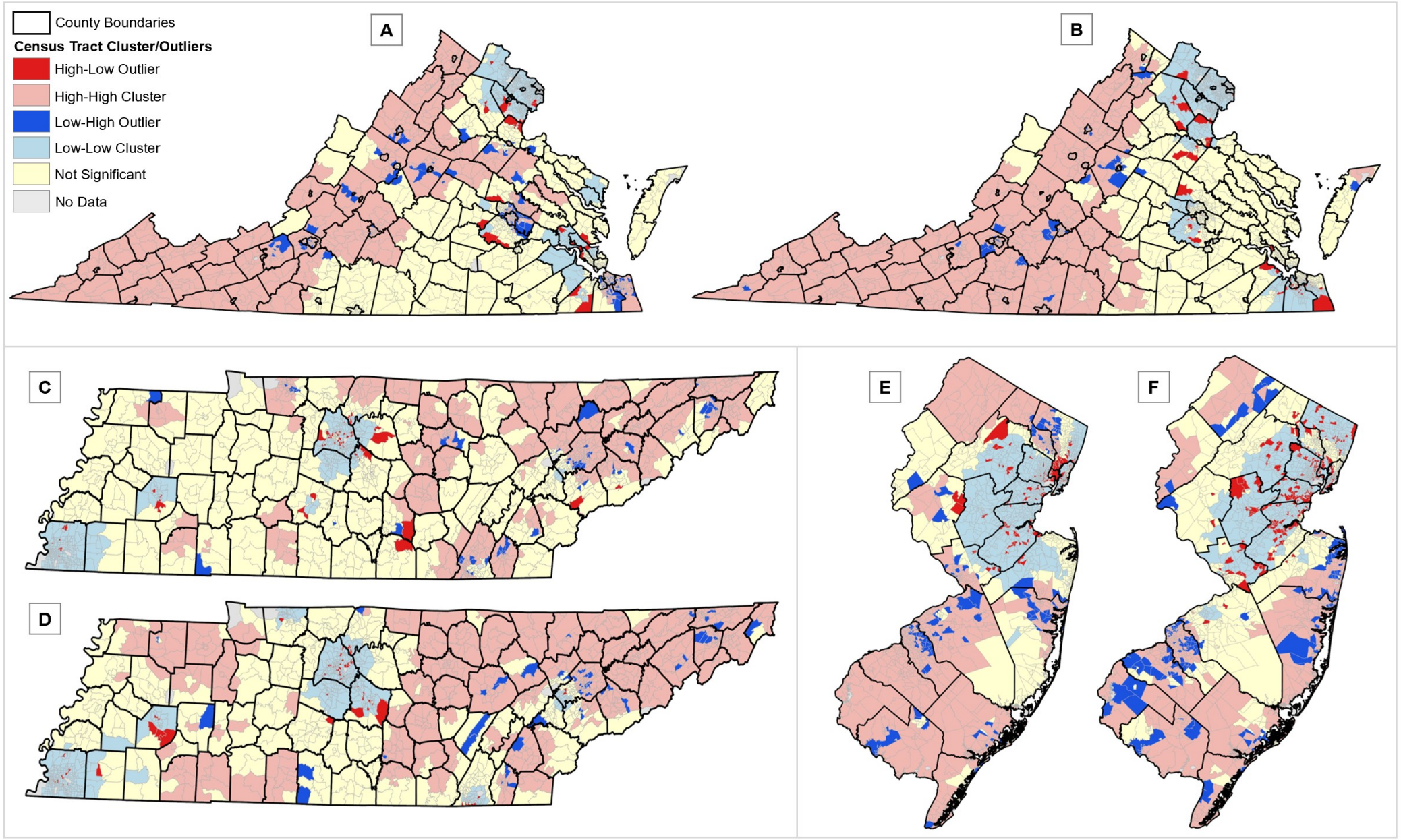
Clusters and outliers of the following attitudes and perceptions within the synthesized population from the UAS across Virginia census tracts: A) attitude, B) perceived benefits, C) perceived barriers, D) perceived susceptibility.

EVALUATING ADAPTABILITY

Investigating the transferability of our proposed approach in generating populations with other health outcomes, and across different locations and timeframes

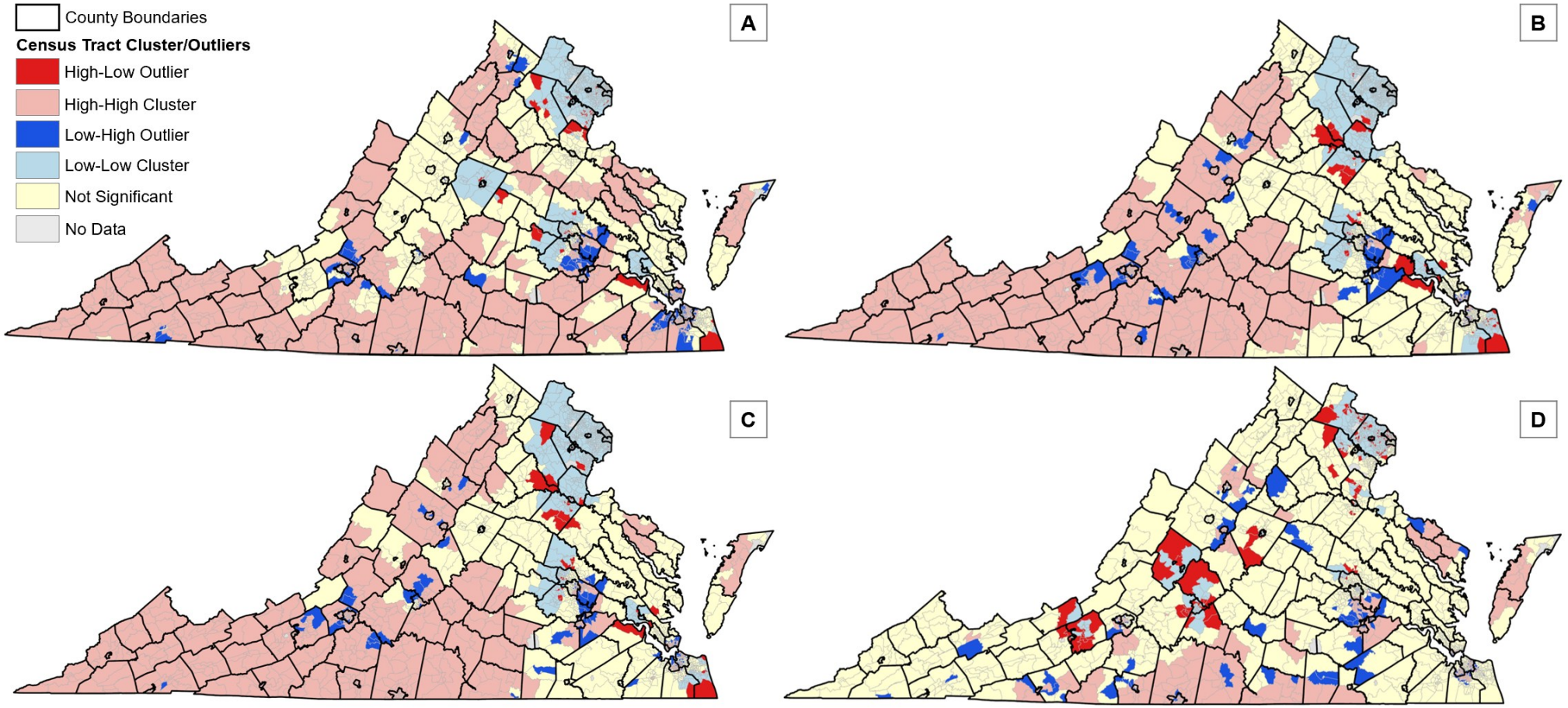
- Using nationally representative survey data from the Understanding America Survey (UAS), we generated populations with smoking behavior and diagnosed depression
- These populations were generated for various states, and for different points in time

EVALUATING ADAPTABILITY: DEPRESSION ACROSS LOCATIONS



Clusters and outliers of the following: A) VA CDC places depression estimates, B) VA synthesized population with depression, C) TN CDC places depression estimates, D) TN synthesized population with depression, E) NJ CDC places depression estimates, F) NJ synthesized population with depression

EVALUATING ADAPTABILITY: 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

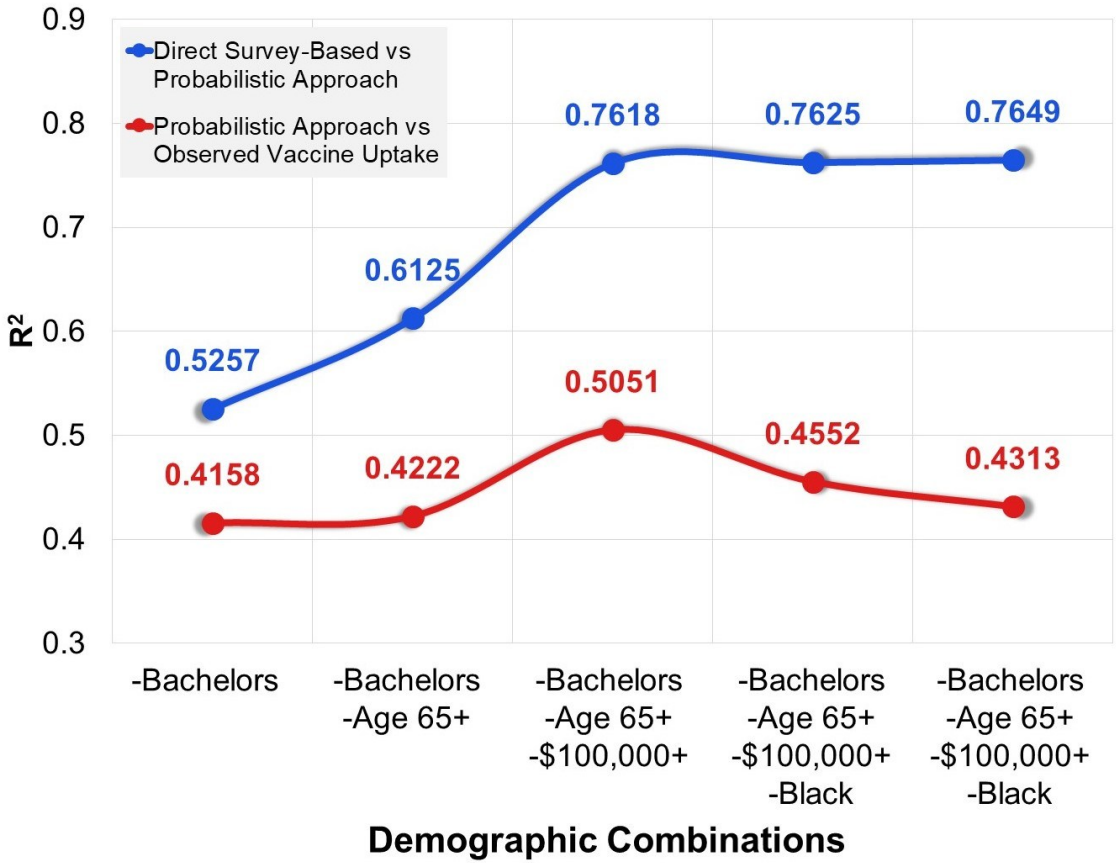
COMPARISION TO OTHER POPULATION SYNTHESIS APPROACHES

Comparing our proposed method to other commonly used approaches used for generating synthetic populations with health-related attributes

- Using nationally representative survey data from the Understanding America Survey (UAS), we use IPF to generate base populations with only demographics
- COVID-19 vaccine intention is then assigned to agents with two different approaches:
  - *Regression-based*
  - *Probabilistic-based*

Evaluation metrics comparing COVID-19 vaccine intention in each of the synthesized populations, using three different approaches, to validation data.

Method	Pearson's r	R^2	RMSE	MAE
COMPARING SIMULATED AND REAL COVID-19 VACCINE INTENTION				
Our Approach	*0.737*	*0.543*	*11.790*	*8.991*
Regression-Based	0.690	0.476	13.829	10.672
Probabilistic-Based	0.657	0.431	12.997	9.966



Computed R2 values for each iteration of the probabilistic approach, comparing (1) simulated COVID-19 vaccine intent from our direct survey-based and probabilistic approach (blue) with (2) COVID-19 vaccine intent from the probabilistic approach with actual vaccine uptake in VA (Dec. 2021, red).

CONCLUSIONS AND FUTURE DIRECTIONS

- Demonstrated survey-based population synthesis for public health outcomes across time and space.
- First approach to directly integrate domain-specific surveys into synthetic population generation.
- Future work should focus on handling bias introduced in the population synthesis process
- We also encourage other scientists to apply this method beyond public health, such as in marketing, disaster management, urban/regional science, or national security domains, to improve behavior modeling in large-scale simulations.

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