Initial Mobility Analysis for ORNL VA-EDH Synthetic Populations

Joseph V. Tuccillo, Angela R. Cunningham, James D. Gaboardi, Whitson Buck

1 Introduction

Travel burdens are a major barrier to healthcare access among US Veteran patient populations, particularly those residing in rural areas (5; 4). Spatial accessibility to points of care for US Veteran populations is commonly assessed in two ways. The first approach uses open data from the US Census to represent collective travel burdens, for example the distance between population-weighted census tract centroids and VHA points of care (7; 8). The second approach uses restricted-access VHA patient data to measure travel costs (e.g., distance, time) for accessing points of care with respect to geolocated patient addresses and real or approximated transportation networks (5; 4). While the advantage of the open data approach lies in its reproducibility, it has notable limitations in its tendency to infer individual travel behavior from aggregate population characteristics, a problem known as ecological fallacy (7). Conversely, while the patient data approach is able to account for individual travel behavior, its ability to account for localized access disparities (e.g., a neighborhood with exceptionally high transportation costs) and patient demographics is limited as protecting individual patient data requires their storage in closed systems with limited capacity for adequately modeling real-world travel patterns or for supplementing patient attributes. Additionally, the patient data approach cannot account for veterans who are not enrolled in the VHA system but who may be eligible for care. These challenges limit the ability to perform "what if" analyses on the effects of place-specific interventions on veteran populations with high access barriers to healthcare.

To address these challenges, we explore the application of realistic synthetic populations to examine travel burdens and spatial accessibility issues among veteran patient populations. Synthetic populations provide a virtual, individually-resolved and cross-sectional representation of the veteran patient population that enables investigation of spatial access to points of care in ways in which aggregate data and patient data do not. First, synthetic populations allow one to directly assess how individuals access points of care, from synthesized residential locations to outpatient facilities on real-world transportation networks (11; 12). Modeling access to points of care at the individual scale addresses the ecological fallacy problem associated with using aggregated census data to represent veteran populations and patterns of movement. Second, synthetic populations provide a means of completely representing an area's veteran population using only publicly available, anonymized census microdata from the American Community Survey (ACS) to ensure the privacy of real-world individuals. Generating synthetic populations from the ACS also expands descriptive characteristics beyond what patient data typically offers to include socio-demographic, economic, housing, and mobility attributes. More detailed profiles of both VHA patient populations and veterans not enrolled in the VA system will provide a comprehensive picture of groups that may

Census Division	VISN Market 2019 Code	VISN Market Description
Pacific	21-c	Central California: Fresno-Merced-Tulare
Mountain	19-c	Central/Southern Front Range and Eastern Colorado
WNC	23-n	North Dakota and Northeast Minnesota
ENC	12-f	Central Illinois
WSC	17-e	Central and South Texas
ESC	09-g	Upper Tennessee Valley and Eastern Kentucky
S Atlantic	08-h	Northern Florida/Big Bend
Mid Atlantic	02-h	New York City, Hudson and Upper Delaware Valleys
New England	01-c	Northern New Hampshire and Vermont

Table 1: Study Areas Crosswalk.

benefit from interventions or outreach.

As an initial exercise for using synthetic populations to measure veteran travel burdens to VA care, we apply Oak Ridge National Laboratory's (ORNL) UrbanPop capability (9) to generate a series of synthetic VHA patient populations for 9 Veterans Integrated Services Networks (VISN) market areas in 9 Census Divisions across the continental United States, which are listed in Table 1. We use UrbanPop to produce synthetic populations for the VISN markets selected for each US Census Division, then assign VA outpatient clinic destinations to synthetic VHA patients based on travel about each VISN market's road network. To demonstrate using the synthetic populations to evaluate healthcare travel burdens, we compare the time-based impedance between simulated home locations and VA outpatient clinics in each VISN market. We then perform validation exercises on the synthetic populations with respect to neighborhood (block group) demographic composition as well as patient mobility, comparing aggregate origin-destination statistics for the synthetic population to outpatient visits available in restricted patient data from the VA's Corporate Data Warehouse (CDW) database.

2 Methods

We generate synthetic populations and simulate patient mobility using Likeness, a Python toolkit that supports UrbanPop's capabilities for population synthesis, transportation network generation, and activity allocation (11; 12). To validate these results, we compare simulated VHA patient mobility to that observed in real-world patient data available from CDW via ORNL's Knowledge Data Infrastructure (KDI) Enclave.

2.1 Population Synthesis

The basis for synthetic populations is the ACS, the United States Census Bureau's primary intercensal dataset which estimates a variety of social, economic, demographic, and other subjects based on an annual 5% sample of the United States population. The ACS provides two components necessary for population synthesis: an anonymized selection of individual survey responses in the Public-Use Microdata Sample (PUMS) and neighborhood profiles in the Summary File (SF). Likeness generates residential synthetic populations from the ACS 5-Year Estimates via its `livelike` utility by estimating occurrence probabilities of households from the PUMS for small census areas (block groups, tracts) (9; 11). This is accomplished by statistically matching PUMS households, observed at the scale of Public-Use Microdata Areas (PUMAs) containing 100,000 or more people, to census block groups (600-3000 people) in a way that preserves aggregate population statistics about those places available from the ACS SF (6). The synthetic population is then generated from a bootstrap sample of PUMS households to small census areas based on residence type (family/nonfamily, group quarters) and household size, which approximately preserves both the total number of people and number of residences in an area. Residential locations for the synthetic population – serving as origin points for travel – are subsequently modeled by conflating synthetic population attributes (dwelling type, household income, number of units in structure) with those of residential buildings in the Federal Emergency Management Agency's (FEMA) public USA Structures dataset (10).

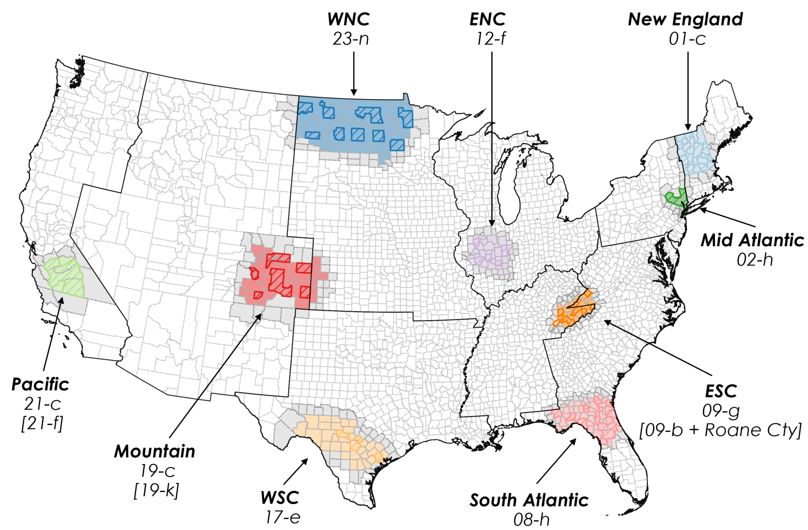

Figure 1: Case study VISN markets for 9 Census Divisions in the continental United States.

We leveraged Likeness residential modeling capabilities to produce synthetic populations for one 2019 VISN market per Census Division in the continental United States (9 in total, see Figure 1). Within the synthetic population models, we controlled for demographic, social, economic, housing, mobility, and worker/student characteristics, as well as veteran status. Expanding upon the approach for US metropolitan areas demonstrated by (12), we generated a synthetic population for each PUMA intersecting the area of interest (AOI) defined by VISN market, then limited the combined population to counties comprising that AOI. This was necessary due to incongruent spatial boundaries between counties and PUMAs: while block groups nest completely within each larger spatial unit, a PUMA is often a collection of block groups from different counties.

We evaluated two AOI definitions in our analysis:

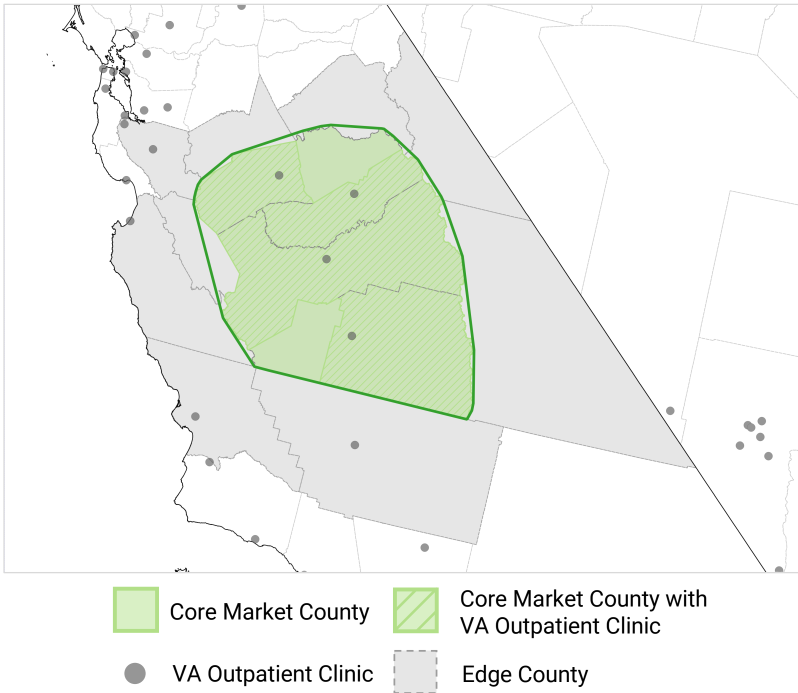

Figure 2: Example of VISN market with edge counties for scenario evaluation.

- **VISN market only**. In this scenario, we assume that VHA patients do not travel beyond the boundaries of the VISN market to seek care. The AOI is limited to counties belonging to the VISN market only.

- **Core VISN market with edge counties**. In this scenario, we assume that VHA patients may seek care at outpatient facilities in portions of VISN markets adjacent to the core market in which they reside. Additionally, we assume that VHA patients in adjacent VISN markets may travel into the core market to seek care. Accounting for both of these factors allows us to assess edge effects in travel to outpatient clinics. These AOIs consist of the core VISN market and edge counties, identified as those intersecting with a convex hull of the core VISN market boundaries (see Figure 2).

Finally, for each AOI, we isolated agents with veteran status and VHA coverage from each VISN market's synthetic population. Because VHA coverage is not available as a modeling constraint for population synthesis, we appended it *post hoc* via record linkage with the PUMS.

2.2 Modeling access to VA points of care

The Likeness `movelike` utility supports generation of real-world transportation networks with OpenStreetMap (OSM) via the `OSMnx` and `pandana` Python packages (1; 3; 11; 12). Using `movelike`, we modeled access to VA outpatient clinics within each VISN market based on time traveled between each synthetic VHA patient residence and clinics on the OSM road network. We compared travel characteristics across VISN markets based on both least-cost paths between residences and clinics and mean travel times from each residence to all clinics.

2.3 Validation

2.3.1 Neighborhood Demographics

Following (9; 12) we validated neighborhood demographics for each AOI based on the percentage of block group synthetic constraints conforming to the 90% Margins of Error (MOEs) on the published constraints from the ACS SF. Because the ACS is a 5% sample of the United States population, every small-area estimate carries a degree of uncertainty (e.g., a block group with 150 people with veteran status $\pm$ 10). Instead of directly minimizing error between a synthetic and published estimate, UrbanPop aims to produce synthetic population estimates that are aligned with the error variances on those estimates (6). Therefore, the 90% MOE fit rate provides a measure of the degree to which the synthetic estimates conform to the bounds of uncertainty on the ACS SF.

2.3.2 VHA Patient Mobility

We validated VHA patient mobility in each AOI by comparing synthetic and observed origin-destination (O-D) matrices representing flows between residential and outpatient facility ZIP codes. O-D flows represent travel to the facility with the least cost (time-based impedance) of access. Cumulative patterns of access to VA clinics among VHA patients in a market form an origin-destination (O-D) matrix between patient and clinic ZIP code locations. We chose ZIP codes as the spatial reference for our initial exercise because they are the finest geographic resolution provided for VA facilities in CDW.

As in (11; 9; 12), we performed this comparison using Canonical Correlation Analysis (CCA), a method that measures the linear association between two multidimensional data matrices (2). A CCA R^2 approaching 1 in our case indicates a high degree of congruence between synthetic and observed patient O-D flows. Any CDW observations of outpatient visits outside the AOI were omitted from the observed O-D matrix (reported in Tables 3, 4, and 5). Additionally, to ensure that the synthetic and observed O-D matrices had the same dimensionality, we padded the synthetic O-D matrix with a row of zeroes to represent no incoming patients for cases where patients were observed in a ZIP code in CDW but not modeled in the synthetic population. We compared the O-D matrices based on relative numbers of patients traveling to each VA patient facility by scaling the destination counts to mean 0 and unit variance (z-score).

3 Results

3.1 Estimated Travel Costs

Table 2 displays the estimated least-cost and mean travel times to outpatient clinics along the road network for both the core market only and core-and-edges AOI definitions. Mean travel times for least-cost routes ranged for core market only AOIs from 9.6 minutes (02-h: Mid-Atlantic) to 30.8 minutes (12-f: East North Central) and for core-and-edges AOIs from 9.1 minutes (02-h: Mid-Atlantic) to 29.7 minutes (23-n: West North Central). Grand mean ("mean of means") travel times are considerably higher than mean least-cost travel times, indicating dispersal of outpatient clinics throughout the VISN markets and associated

	Mean Least-Cost		Grand Mean	
	Core Market	Core Market + Edges	Core Market	Core Market + Edges
Mid-Atlantic (02-h)	9.6	9.1	55.5	59.1
WSC (17-e)	15.7	15.5	43.4	70.9
Mountain (19-c)	16.0	15.6	95.7	119.7
Pacific (21-C)	16.1	17.6	53.0	112.3
ESC (09-g)	18.5	13.3	87.2	109.1
South Atlantic (08-h)	19.5	18.9	101.1	122.1
New England (01-c)	24.7	23.7	110.6	126.5
WNC (23-n)	29.8	29.7	184.1	202.5
ENC (12-f)	30.8	27.3	82.8	113.0

Table 2: Estimated mean least-cost (minimum) and grand mean travel times between synthetic VHA patient residences and VA outpatient clinics on the road network by VISN market.

edge counties. These times range for core-only AOIs from 43.4 minutes (17-e: West South Central) to 184.1 minutes (23-n: West North Central) and for core-and-edges AOIs from 59.1 minutes (02-h: Mid-Atlantic) to 202.5 minutes (23-n: West North Central).

In terms of least-cost paths to outpatient clinics, we find that VISN markets containing large urban centers (02-h: Mid Atlantic, containing portions of the New York City Metropolitan Area; 17-e: West South Central, containing the Houston, Austin, and San Antonio, TX metropolitan areas) offer reduced travel time, whereas predominantly rural markets with smaller urban centers (01-c: New England; 23-n: West North Central; 12-f: East North Central) feature the longest overall travel duration. Travel times for the remainder of markets appear to loosely follow the sizes of their primary cities, with Mountain (12-f, Denver/Colorado Springs, CO) and Pacific (21-c: Fresno/Merced, CA) next in line, followed by markets characterized by smaller to midsized cities (09-g: East South Central, Knoxville, TN; 08-h: South Atlantic, Gainesville/Talahassee, FL). These high and low rankings remain generally consistent with mean travel costs between residences and outpatient facilities. The most pronounced shift between least-cost and mean travel times occurs for the Mountain VISN market (19-c), with a large spatial footprint that encompasses most of Colorado and adjacent states.

The distributions of least-cost path times for all synthetic patients by AOI/VISN market (Figures 3, 4) are all heavily right-skewed, further reflecting differences in access between core urban areas with outpatient clinics and distal peripheral/rural areas. For core market only AOIs, the distribution of times is highly peaked at 30 minutes and below for markets with the shortest travel times (02-h: Mid Atlantic; 17-e: West South Central), but more dispersed and bimodal in the markets with the longest travel times (01-c: New England; 12-f: East North Central; 23-N: West North Central). Least-cost travel time distributions are much the same for the core-and-edges AOIs, with some increased dispersal of longer travel times, particularly in those markets with higher time-based impedance. However, outlying high travel times decrease for core-and-edges AOIs, as evidenced by differences in the upper limits of the x-axes between Figures 3 and 4.

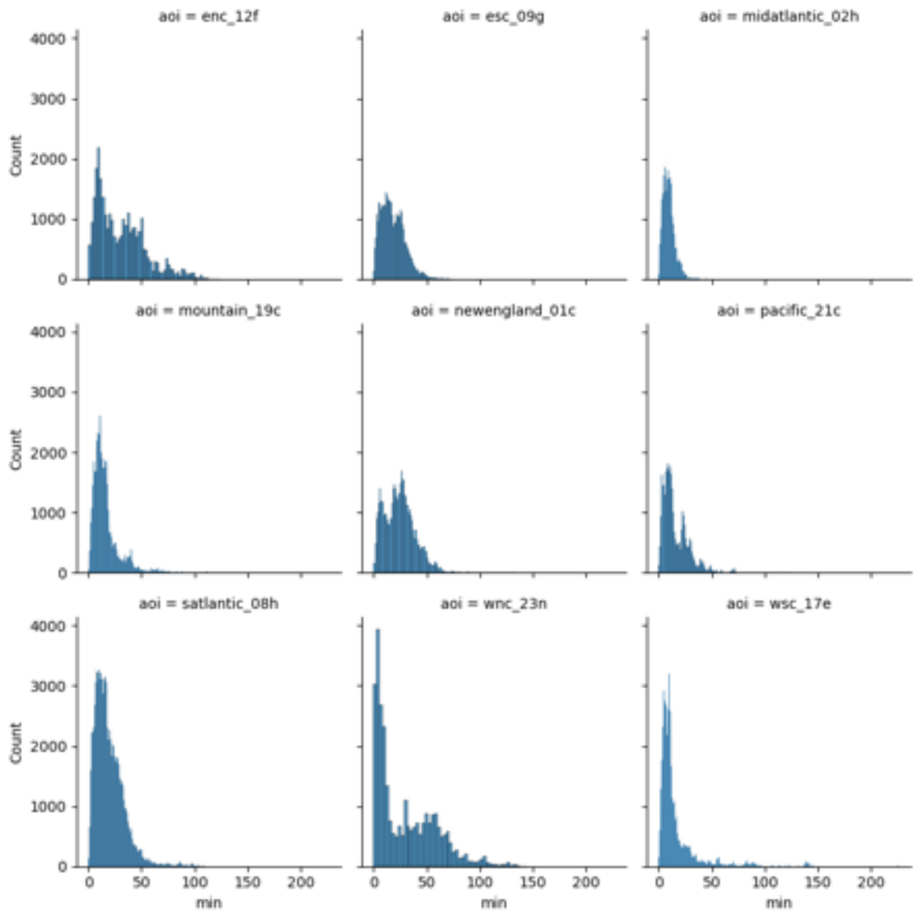

Figure 3: Distribution of least cost travel times in minutes by VISN market, core market only AOIs.

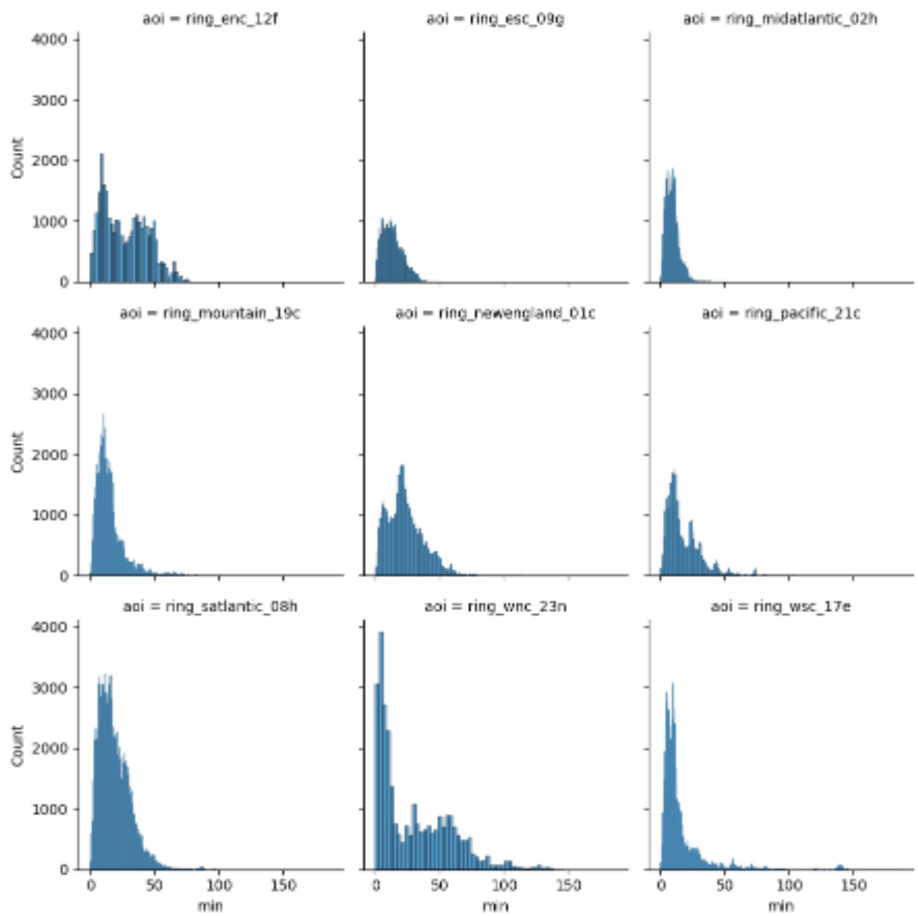

Figure 4: Distribution of least cost travel times in minutes by VISN market, core-and-edges AOIs.

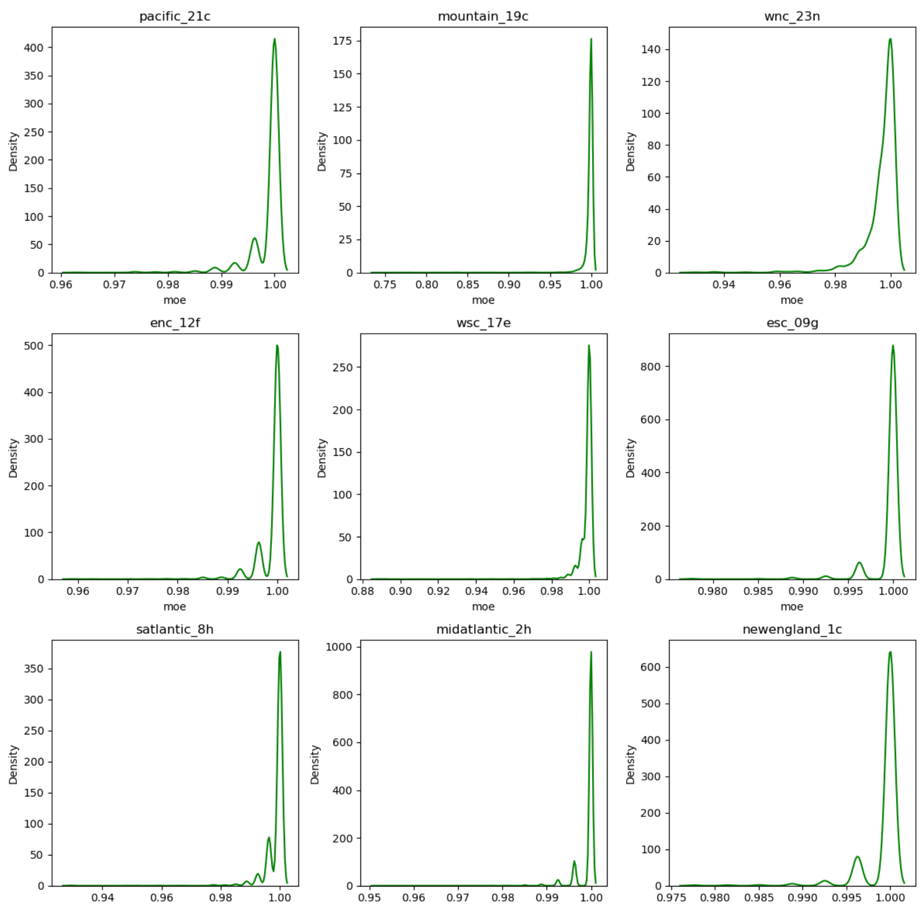

Figure 5: One-dimensional kernel density estimates of block group-level MOE Fit Rates by VISN Market.

Census Division	Origin ZIPs	Destination ZIPs	Observed Trips outside Market %	CCA Adj. R^2	p(Pillai)
Pacific	136	4	8.9	0.878	< 0.0001
Mountain	328	12	5.3	0.758	0
WNC	533	11	11.3	0.856	0
ENC	387	5	18.5	0.869	0
WSC	242	9	5.3	0.648	0
ESC	259	12	7.2	0.768	0
S Atlantic	330	16	10.3	0.745	0
Mid Atlantic	424	18	3.5	0.668	0
New England	496	13	9.9	0.793	0

Table 3: Canonical Correlation Analysis (CCA) results for mobility validation exercise.

3.2 Validation

3.2.1 Neighborhood Composition

In general, the block group-level MOE Fit Rates for the market-only AOIs show between 99%-100% correspondence between the synthetic and reported estimates relative to the MOEs. However, the distributions of MOE Fit Rates (Figure 3.2.1) all feature long left tails, indicating that each VISN Market contains a handful of block groups with populations that conform less neatly to the ACS SF MOEs. We observe MOE Fit Rates below 90% for the Pacific (21-c) and West South Central (17-e) Markets (≈88% each), as well as the Mountain Market (19-c), with a minimum MOE Fit Rate of ≈75%. These cases appear to be consistently tied to block groups where the majority of the population is associated with some large institutional (e.g., prison, nursing home) or military facility (the outlying low value for Mountain 19-c coincides with the US Air Force Academy). This issue is potentially tied to differences in sample size for institutionalized and military populations that make the ACS SF estimates more difficult to fully recreate.

3.2.2 VHA Patient Mobility

For the market-only AOIs, we observed modest to strong association between the synthetic and observed O-D matrices representing ZIP-ZIP flows between residential and VA outpatient facility locations (Table 3). The strongest correspondence between synthetic and observed O-D flows was for the Pacific VISN market (21-c: Fresno-Merced, CA; CCA $R^2 = 0.878$), with similarly high values for the Eastern North Central VISN market (12-f: Central Illinois; CCA $R^2 = 0.869$) and Western North Central VISN market (23-n: North Dakota, Northwest Minnesota; CCA $R^2 = 0.856$). Lower correspondence occurred for the West South Central VISN market (17-e: Central and South Texas; CCA $R^2 = 0.648$) and Mid Atlantic VISN market (02-H: New York City and Hudson/Delaware River Valleys; CCA $R^2 = 0.668$).

For the AOIs defined by a core VISN market and edge counties, we observed a slight shift (mean absolute change in CCA $R^2 = 0.088$) in CCA values, with a tendency toward decreased correspondence between synthetic and observed mobility patterns (Table 5). The largest of these is the Eastern North Central VISN market ($\Delta_{R^2} = -0.247$), with a similarly

Census Division	Origin ZIPs	Destination ZIPs	Observed Trips outside Market %	CCA Adj. R^2	p(Pillai)
Pacific	136	5	8.6	0.805	< 0.0001
Mountain	328	15	4.8	0.761	0
WNC	533	14	10.6	0.690	0
ENC	387	11	15.2	0.621	0
WSC	242	13	4.3	0.561	< 0.0001
ESC	259	12	7.2	0.766	0
S Atlantic	330	26	6.9	0.718	0
Mid Atlantic	424	24	3.5	0.583	0
New England	496	20	9.4	0.692	0

Table 4: Canonical Correlation Analysis (CCA) results for mobility validation exercise, including peripheral counties to the VISN.

Census Division	$\Delta(R^2)$ (vs. Market-Only)
Pacific	-0.073
Mountain	0.004
WNC	-0.165
ENC	-0.247
WSC	-0.087
ESC	-0.002
S Atlantic	-0.028
Mid Atlantic	-0.085
New England	-0.101

Table 5: $\Delta(R^2)$ between core VISN and edge county inclusion for the Canonical Correlation Analysis (CCA).

substantial change for the Western North Central VISN market ($\Delta_{R^2} = -0.165$). The most stable VISN markets were Eastern South Central ($\Delta_{R^2} = -0.002$) and Mountain ($\Delta_{R^2} = 0.004$).

4 Discussion

Together, our demographic and mobility validation exercises demonstrate the effectiveness of using synthetic populations to understand travel burdens among VHA patients. We find that predominantly rural VISN markets with smaller urban core area – often containing outpatient clinics – and larger peripheral areas, tend to have heightened outpatient access barriers. This is likely because patient residential locations are more spatially dispersed rather than concentrated in urban centers. We see this effect across all selected VISN markets, but the degree to which it is expressed appears to depend heavily upon urbanicity. While synthetic neighborhood demographics were more closely aligned with source data (ACS SF) than mobility of synthetic VHA patients (CDW outpatient visits), each captured real-world dynamics with reasonable accuracy.

The general decrease in CCA performance between the market-only and core-and-edges model runs suggests that using VISN markets to bound mobility simulations is more effective than assuming VHA patients seek out-of-market care. That is, despite fewer extreme high travel times observed for the core-and-edges models, it appears less realistic overall to assume that patients would seek care in another VISN market instead of their own.

Differences in mobility validation scores (CCA R^2) among the selected VISN markets reveal some important potential confounding factors for these models that should be addressed in future work. The most notable of these is the travel mode that synthetic patients use to access VA points of care. One of the VISN markets with the weakest relative performance (02-h) contains New York City, for which outpatient mobility is likely influenced more heavily by public transit access than driving. Degree of urbanization may also drive greater differences between synthetic and observed patient mobility. The VISN market with the overall weakest relative performance, 17-e, contains two large metropolitan areas (Houston and San Antonio, TX), and it is possible that factors like traffic congestion affect the ways in patients access points of care in reality.

5 Conclusion

This preliminary analysis applied synthetic populations and simulated mobility about real-world transportation (road) networks to evaluate VHA veteran patient access to outpatient care at VHA facilities. We demonstrated the ability of synthetic populations to provide a complete, close-to-reality approximation of VHA patient populations at large spatial scales (VISN markets) with mobility characteristics that sufficiently recreate those of actual patients while preserving privacy.

In future work, we intend to expand the scope of patient mobility access to encompass multiple treatment types, ranging from those that could be addressed or supplemented by telehealth services (psychotherapy, medication management) to those requiring physical presence (opioid treatment). Toward the former, we will evaluate the potential mitigative effects of internet availability and broadband access on travel time. Toward the latter, we will consider the placement of mobile clinics in different areas of interest, particularly in areas with limited access to brick-and-mortar clinics. To achieve this, we intend to use spatial optimization methods to locate facilities in areas most reachable by patient populations with high access barriers. We will also expand travel modalities to include access barriers relative to walk/bike infrastructure, as well as public transit routes.

References

[1] Geoff Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, environment and urban systems*, 65:126–139, 2017.

[2] Magnus Borga. Canonical correlation: a tutorial. *Online tutorial http://people.imt.liu.se/magnus/cca*, 4(5), 2001.

[3] Fletcher Foti and Paul Waddell. A Generalized Computational Framework for Accessibility: From the Pedestrian to the Metropolitan Scale. In *Transportation Research Board Annual Conference*, pages 1–14, 2012. URL: `https://onlinepubs.trb.org/onlinepubs/conferences/2012/4thITM/Papers-A/0117-000062.pdf`.

[4] Sarah A Friedman, Susan M Frayne, Eric Berg, Alison B Hamilton, Donna L Washington, Fay Saechao, Natalya C Maisel, Julia Y Lin, Katherine J Hoggatt, and Ciaran S Phibbs. Travel time and attrition from vha care among women veterans: how far is too far? *Medical Care*, 53:S15–S22, 2015.

[5] Zachary Hahn, John Hotchkiss, Charles Atwood, Connor Smith, Annette Totten, Eilis Boudreau, Robert Folmer, Priyanka Chilakamarri, Mary Whooley, and Kathleen Sarmiento. Travel burden as a measure of healthcare access and the impact of telehealth within the veterans health administration. *Journal of General Internal Medicine*, 38(Suppl 3):805–813, 2023.

[6] Nicholas N Nagle, Barbara P Buttenfield, Stefan Leyk, and Seth Spielman. Dasymetric modeling and uncertainty. *Annals of the Association of American Geographers*, 104(1):80–95, 2014.

[7] Michael E Ohl, Margaret Carrell, Andrew Thurman, Mark Vander Weg, Teresa Hudson, Michelle Mengeling, and Mary Vaughan-Sarrazin. Availability of healthcare providers for rural veterans eligible for purchased care under the veterans choice act. *BMC Health Services Research*, 18:1–7, 2018.

[8] Eliana Sullivan, Whitney E Zahnd, Jane M Zhu, Erin Kenzie, Mary Patzel, and Melinda Davis. Mapping rural and urban veterans' spatial access to primary care following the mission act. *Journal of General Internal Medicine*, pages 1–7, 2022.

[9] Joseph Tuccillo, Robert Stewart, Amy Rose, Nathan Trombley, Jessica Moehl, Nicholas Nagle, and Budhendra Bhaduri. Urbanpop: A spatial microsimulation framework for exploring demographic influences on human dynamics. *Applied Geography*, 151:102844, 2023. `doi:10.1016/j.apgeog.2022.102844`.

[10] Joseph V. Tuccillo. Downscaling Synthetic Populations to Realistic Residential Locations. In *USRSE23 Conference Proceedings*. Zenodo, 2023. `doi:10.5281/zenodo.10420984`.

[11] Joseph V. Tuccillo and James D. Gaboardi. Likeness: a toolkit for connecting the social fabric of place to human dynamics. In Meghann Agarwal, Chris Calloway, Dillon Niederhut, and David Shupe, editors, *Proceedings of the 21st Python in Science Conference*, pages 125–135, 2022. `doi:10.25080/majora-212e5952-014`.

[12] Joseph V. Tuccillo and James D. Gaboardi. Spatial microsimulation and activity allocation in Python: an update on the Likeness toolkit. In Meghann Agarwal, Chris Calloway, and Dillon Niederhut, editors, *Proceedings of the 22nd Python in Science Conference*, pages 93–100, 2023. `doi:10.25080/gerudo-f2bc6f59-00c`.