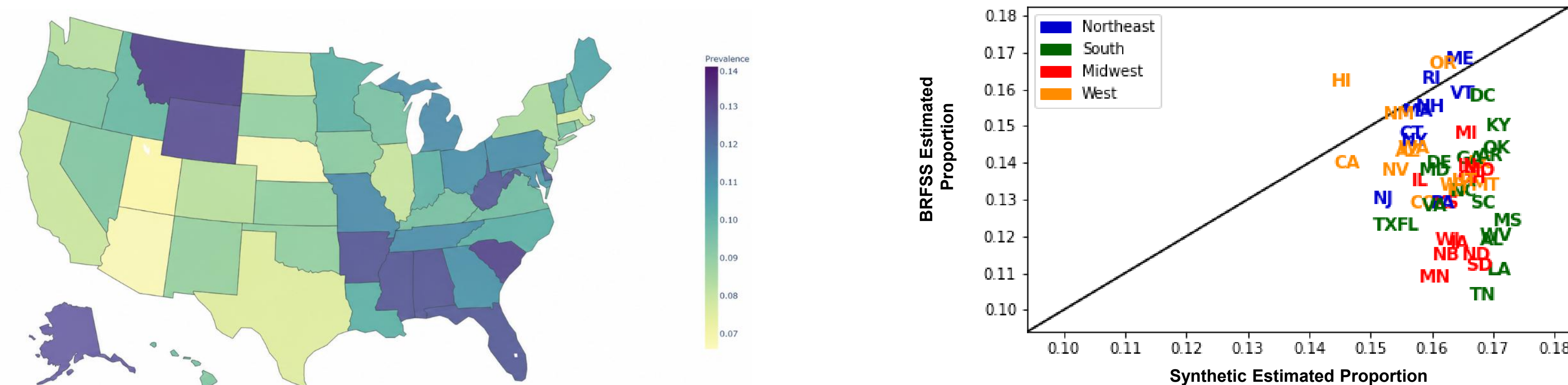
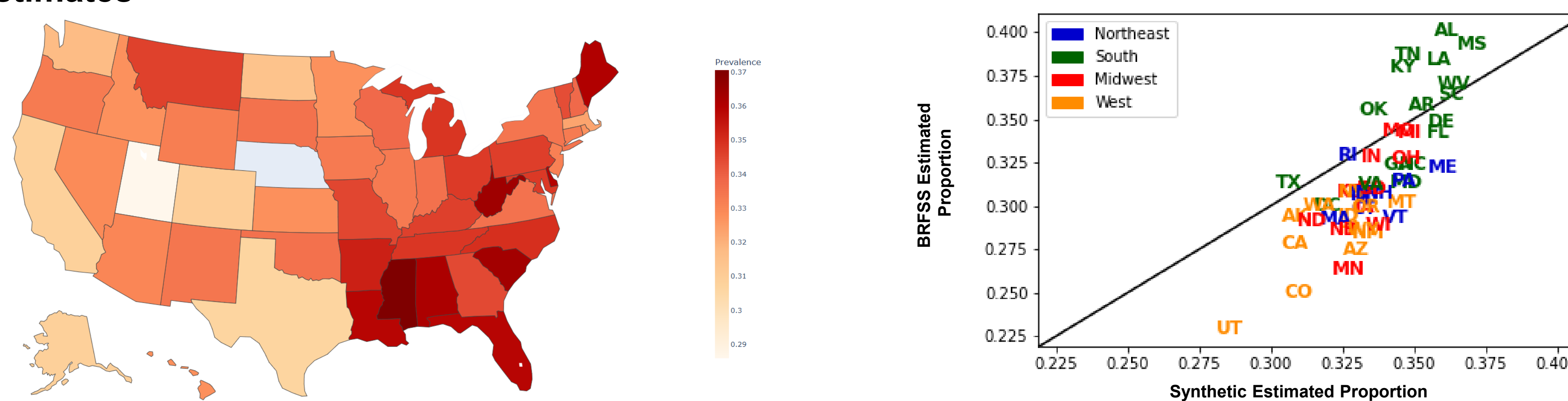
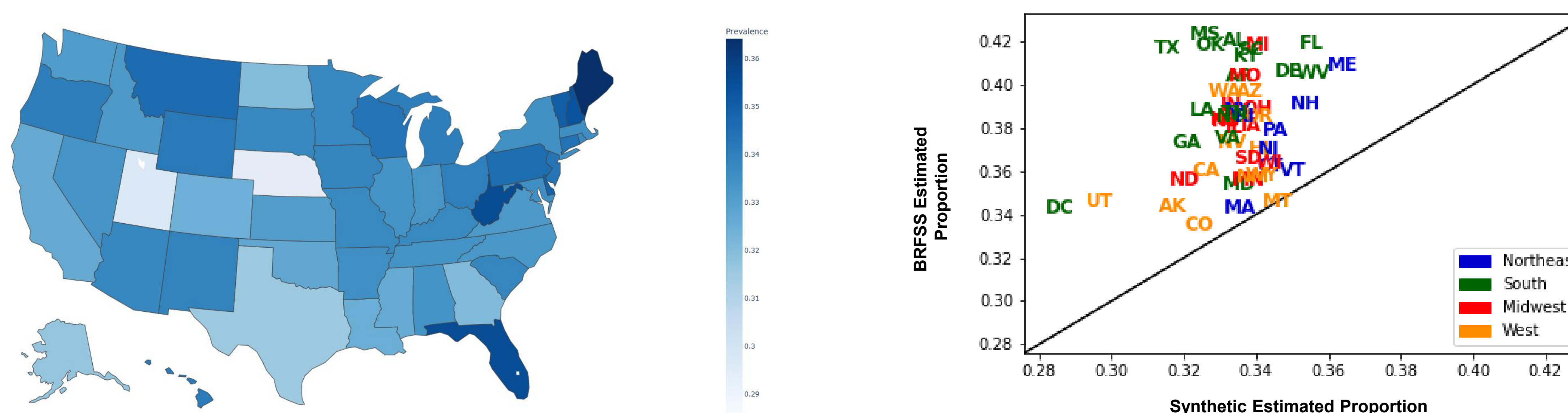


Background

- Understanding the prevalence of health conditions within specific geographies is critical to guiding healthcare decisions
- However, accurately estimating the prevalence at the census-tract or county-level is challenging
- Small area estimation is frequently used but faces challenges related to sparse or nonexistent samples for rural counties, sensitivity to survey bias or nonresponse, and dependence on correctly specifying the base model's covariates and structure
- Prevalence estimates based on claims data is subject to limitations including underrepresentation of specific populations (e.g., uninsured, low-income) and only capturing diagnosed conditions
- Our objective was to generate and validate synthetic individual-level datasets containing biomarker and sociodemographic data for any county or state in the United States using synthetic population reconstruction approach**

Methods

- Using data from the 2022 U.S. Census American Community Survey and three combined waves of the National Health and Nutrition Examination Survey (NHANES, 2013-2018), we used a synthetic reconstruction approach to generate individual-level populations representative of all U.S. counties, states, and Washington D.C.
- Our analysis focused on three common health conditions: asthma, hypertension, and high cholesterol based on survey responses within NHANES (**Table 1**)
- We used state-level prevalence estimates from the Behavioral Risk Factor Surveillance System (BRFSS) to externally validate our synthetic population-based estimates using two one-sided *t*-test (TOST) to assess statistical equivalence using an equivalence margin of ± 0.05

Figure 1. State-level estimates for asthma and comparison between BRFSS and synthetic estimates

Figure 2. State-level estimates for hypertension and comparison between BRFSS and synthetic estimates

Figure 3. State-level estimates for high cholesterol and comparison between BRFSS and synthetic estimates


Results

- Synthetic individual-level datasets featuring NHANES health data (including biomarkers, nutrition and physical activity profiles, and risk factor data) were successfully generated for all U.S. counties and states and representative of 2022 ACS demographics
- Within external validation, statistically equivalent estimates were observed for 40 (78%) of 51 tests for asthma, 39 (76%) for hypertension, and 17 (33%) for high cholesterol
- Absolute mean percentage differences between synthetic and BRFSS state-level estimates were 2.6% for asthma, 2.0% for hypertension, and 4.7% for high cholesterol
- Validation against BRFSS demonstrated strong agreement for asthma and hypertension across most states (**Figures 1 and 2**) but lower equivalence for high cholesterol (**Figure 3**)

Limitations

- Discrepancies observed between BRFSS and synthetic population state-level estimates could reflect differences in survey design, as evidenced by differences between NHANES and BRFSS at the national-level, or inability to incorporate spatially varying factors such as state-level policies on smoking or environmental factors into the synthetic reconstruction process
- The difference between estimates for high cholesterol between the synthetic population and BRFSS was greatest of the three conditions; likely, because the BRFSS estimate was conditional on having received a cholesterol check
- Health conditions that are heavily influenced by spatially varying factors may not be well-suited for this approach
- External validation of county-estimates is needed; however, challenging due to sparse data
- Next steps include validating to county-level estimates, if data are available, and evaluating the ability of this methodology to inform clinical trial recruitment efforts and improve prevalence estimates

Conclusions

- This synthetic population reconstruction approach successfully addresses a critical gap not filled by existing methods or data
- By incorporating biomarker data, sociodemographic characteristics, and underrepresented populations within individual-level datasets for any U.S. state or county, these estimates provide a comprehensive view of disease burden
- Synthetic individual-level datasets can be used directly in spatial temporal analyses and disease models as demonstrated by its current use within SMILES, a chronic disease microsimulation (poster MSR202)
- Policy makers can use these data to better assess prevalence, allocate resources, and design interventions that reach populations typically excluded from claims-based surveillance and to augment small area estimates

Table 1. NHANES and BRFSS Survey Questions and National Prevalence of asthma, hypertension, and high cholesterol

Disease/risk factor	NHANES Survey Question	NHANES Prevalence	BRFSS Survey Question Wording	BRFSS Prevalence	Difference
Asthma	"Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP has} asthma?"	16%	"Have you ever been told by a doctor, nurse, or other health professional that you had asthma?" and "Do you still have asthma?"	15%	~1%
Hypertension	"{Have you/Has SP} ever been told by a doctor or other health professional that {you/s/he} had hypertension, also called high blood pressure?"	34%	"Have you ever been told by a doctor, nurse, or other health professional that you have high blood pressure?"	33%	~1%
High cholesterol	"{Have you/Has SP} ever been told by a doctor or other health professional that {your/his/her} blood cholesterol level was high?"	35%	Adults who have had their cholesterol checked and responded "yes" to the following: "Have you ever been told by a doctor, nurse or other health professional that your blood cholesterol is high?"	38%	~3%

EPH186

Presenting author: Joëlla Adams, MPH, PhD
jadams@rti.org

www.rti.org

RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.