Discovering Hidden Patterns in County-Level Diagnosed Diabetes Prevalence in The United States Using Neural Networks: A Spatio-Temporal Analysis From 2011 to 2020

Hui Xie and Deborah B. Rolka

FCSM 2023 @ Maryland

National Center for Chronic Disease Prevention and Health Promotion
Division of Diabetes Translation, CDC

Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention

Outline

- Introduction
- Spatio-Temporal Analysis: A Machine Learning Approach
- Results
- Discussion

Introduction

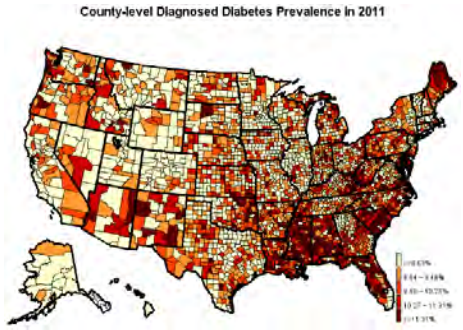
- Despite a slowdown in national diabetes prevalence over the past decade in the United States, trends in the prevalence of diagnosed diabetes are highly variable at lower geographic levels.

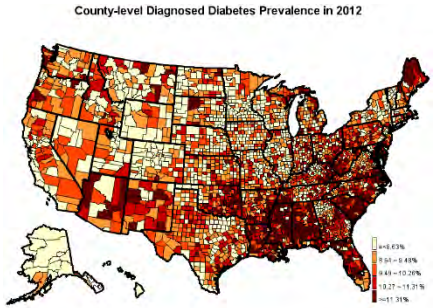
- This study aims to unveil intricate patterns and relationships in diabetes prevalence at the county level.

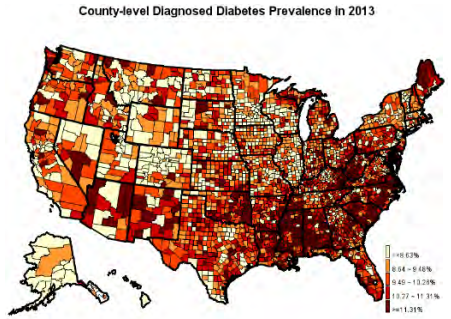
- We applied small area estimation (SAE) techniques to estimate county-level diagnosed diabetes incidence rates, using data sourced from the Behavioral Risk Factor Surveillance System (BRFSS).

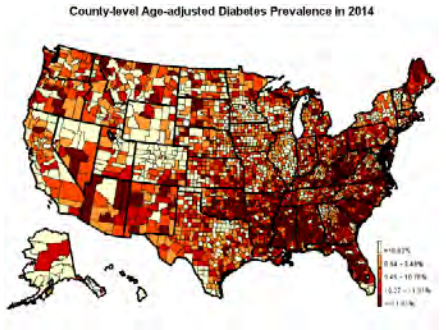
- We thoroughly analyzed ten years of county-level diagnosed diabetes prevalence data spanning from 2011 to 2020 to unravel the hidden patterns. The data is expressed as changes from prevalence in 2011

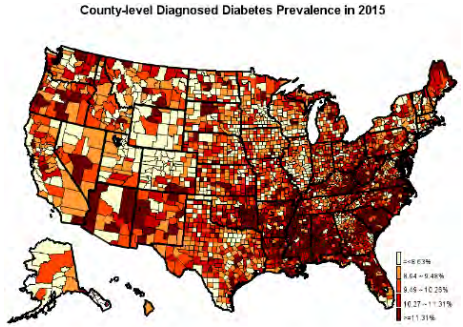
County-level Diagnosed Diabetes Prevalence in the U.S., 2011- 2022

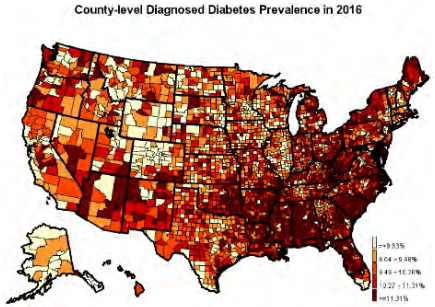


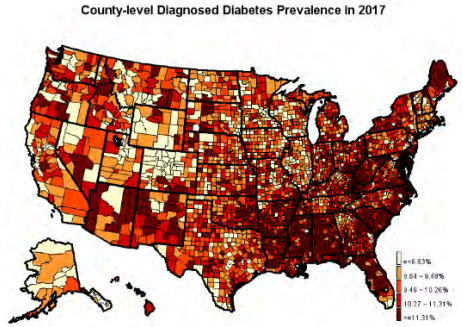


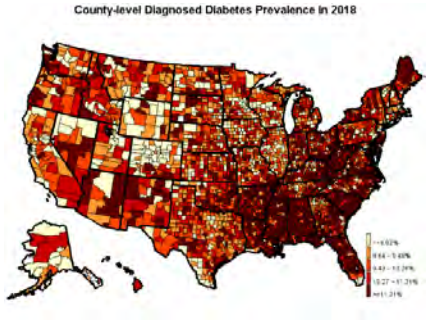


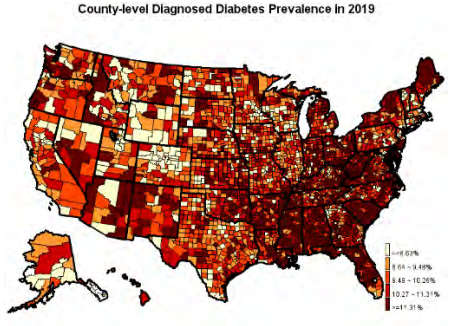


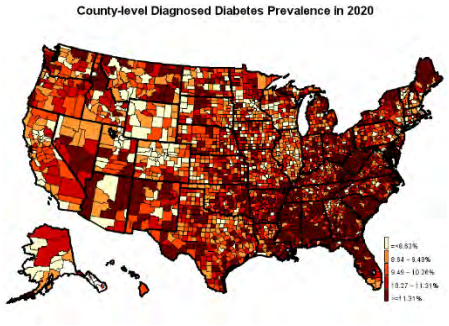


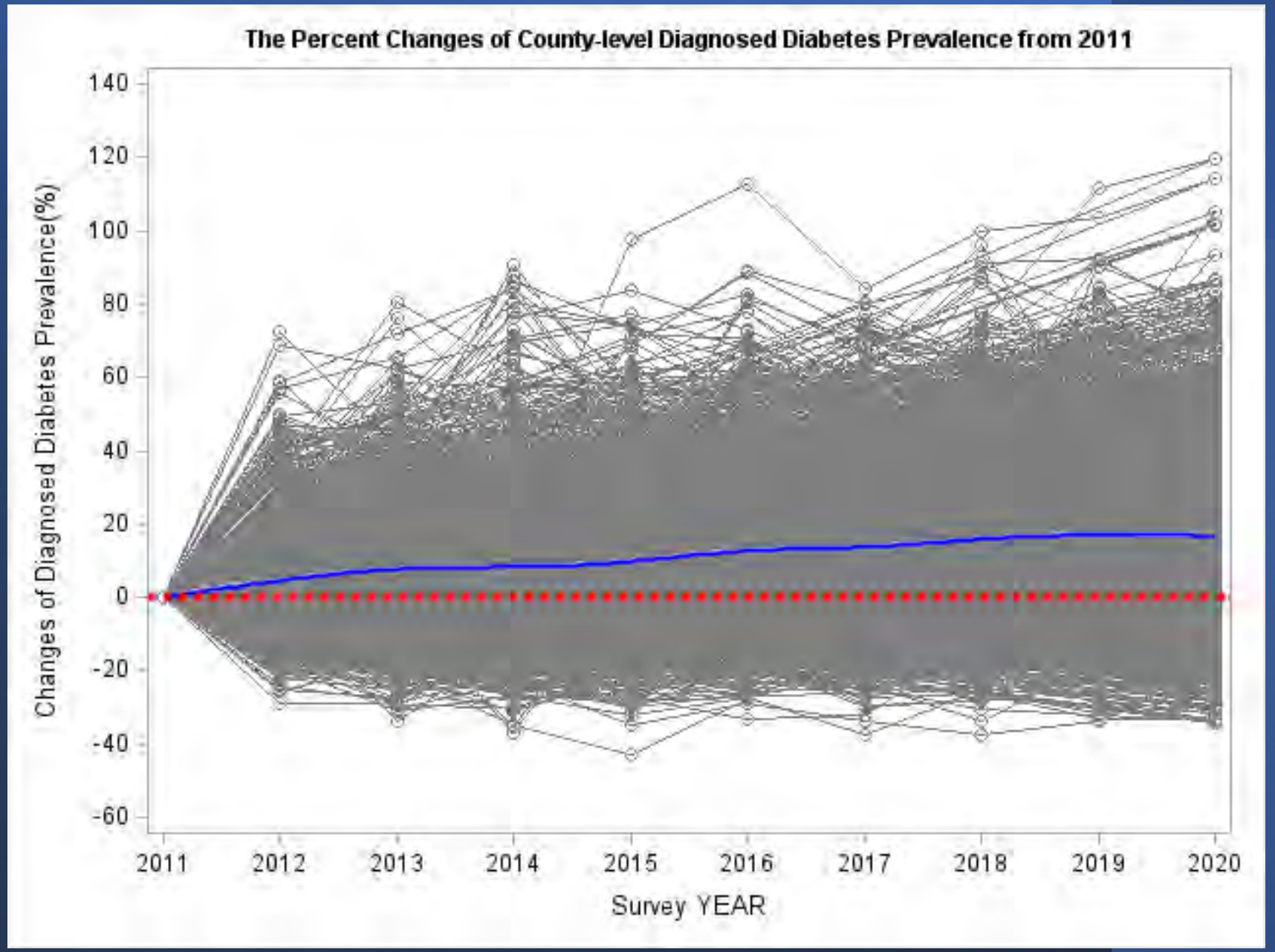











The Percent Changes of County-level Diagnosed Diabetes Prevalence from 2011
Changes of Diagnosed Diabetes Prevalence(%)
140
120
100
80
60
40
20
0
-20
-40
-60
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR

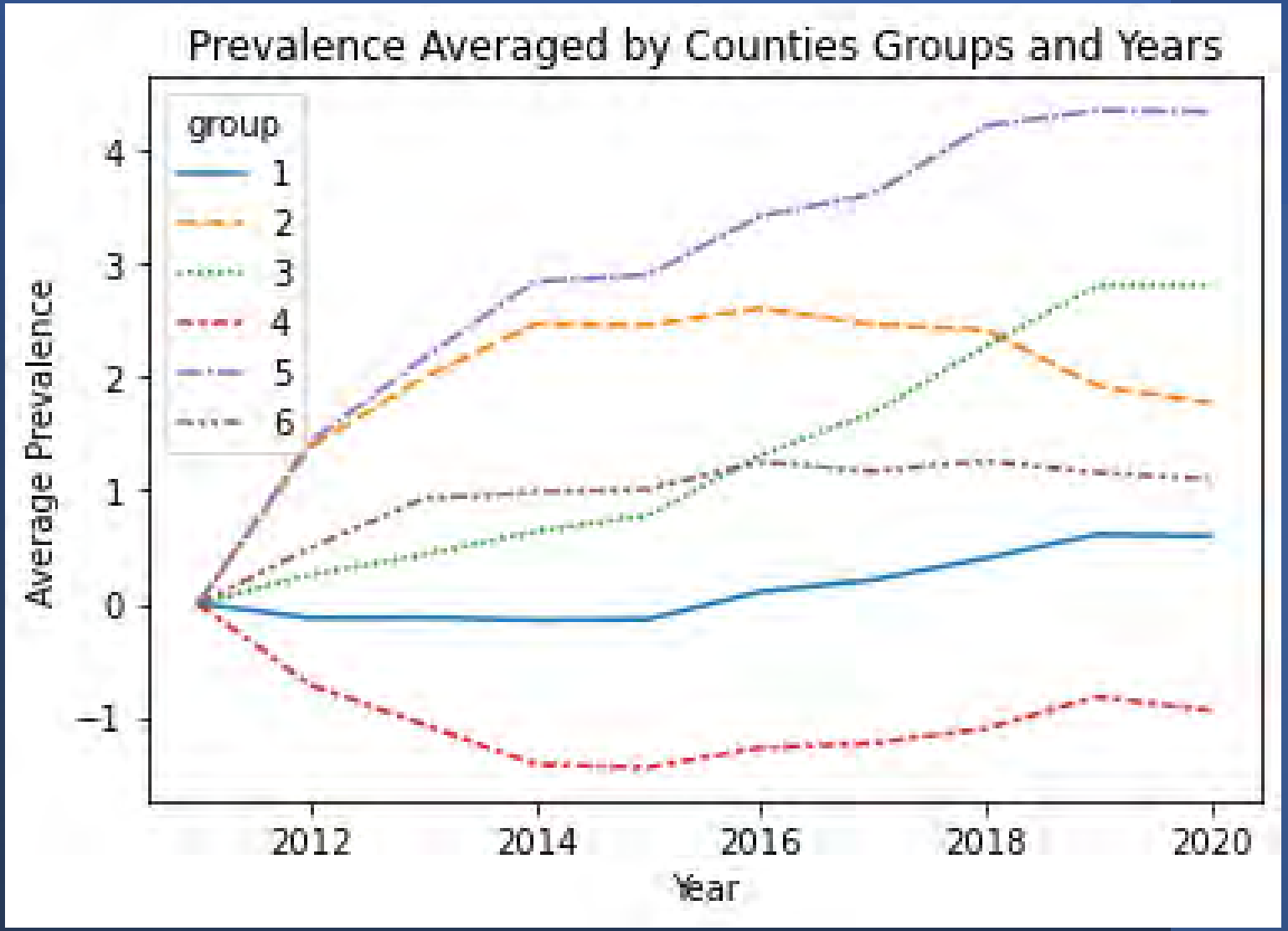
Methods

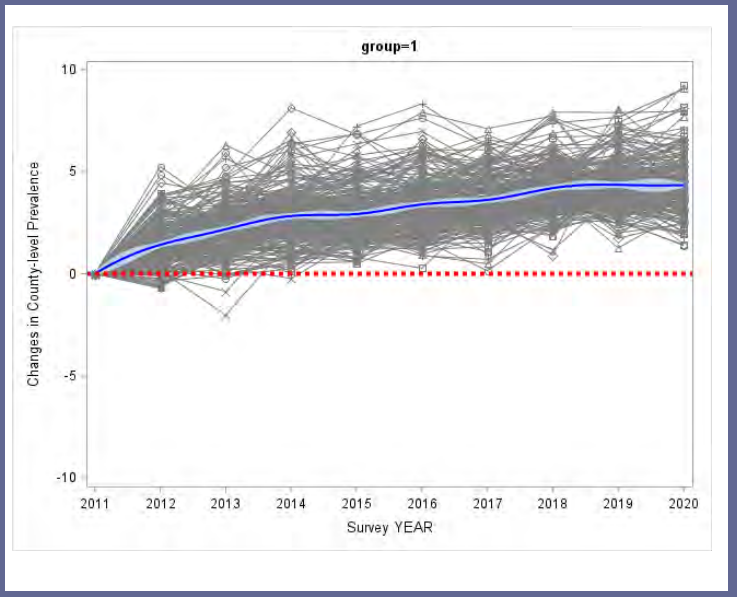
- We applied unsupervised classification, a Neural Network architecture, to identify the patterns within county-level trends across ten survey years.

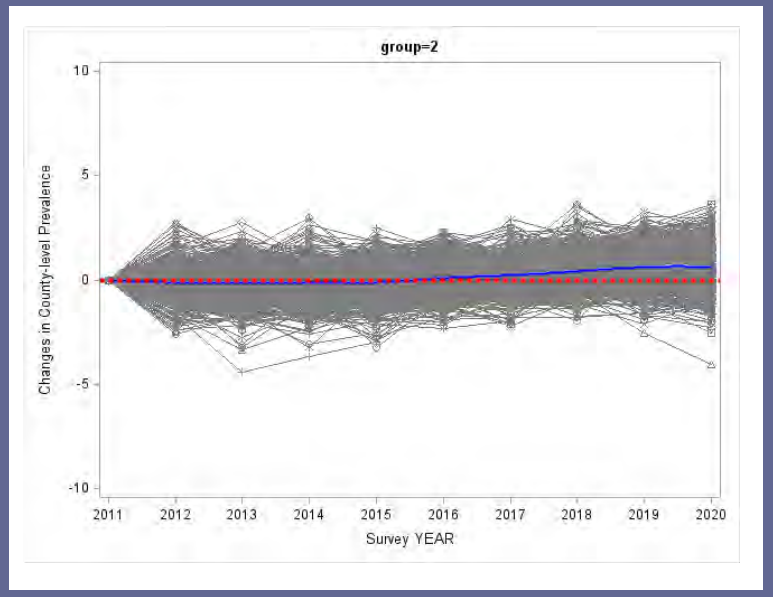
- We then employed a Spatial Autoregressive model to capture the spatial patterns and relationships across counties

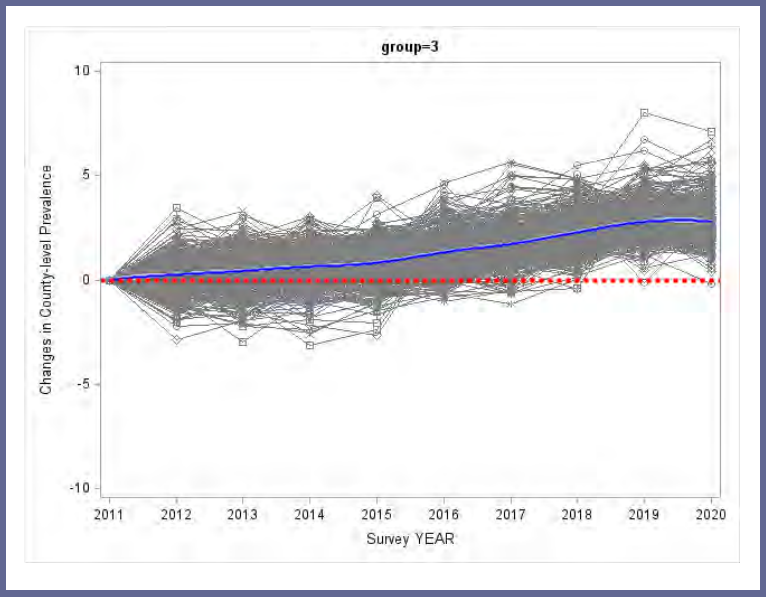
- We assessed the model's performance using appropriate spatial evaluation metrics, including Moran's I, Local Indicators of Spatial Associations (LISA), and spatial cross-validation metrics

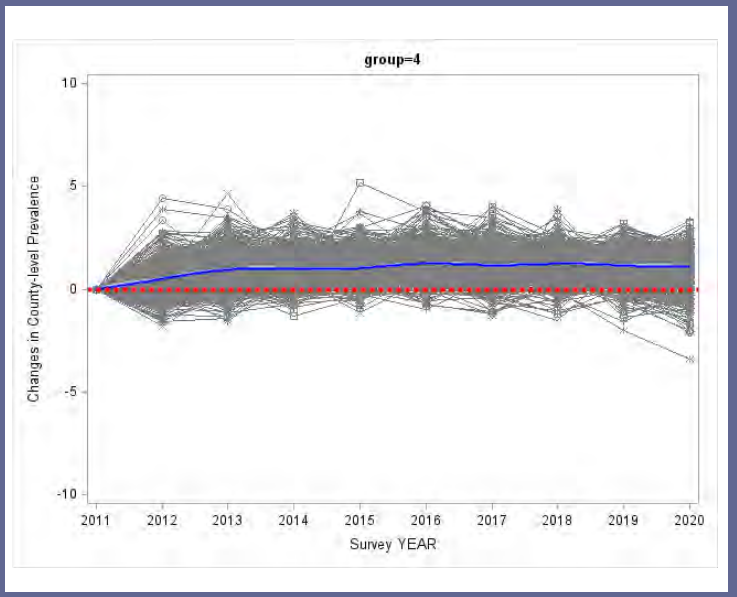
Results I : Temporal Analysis

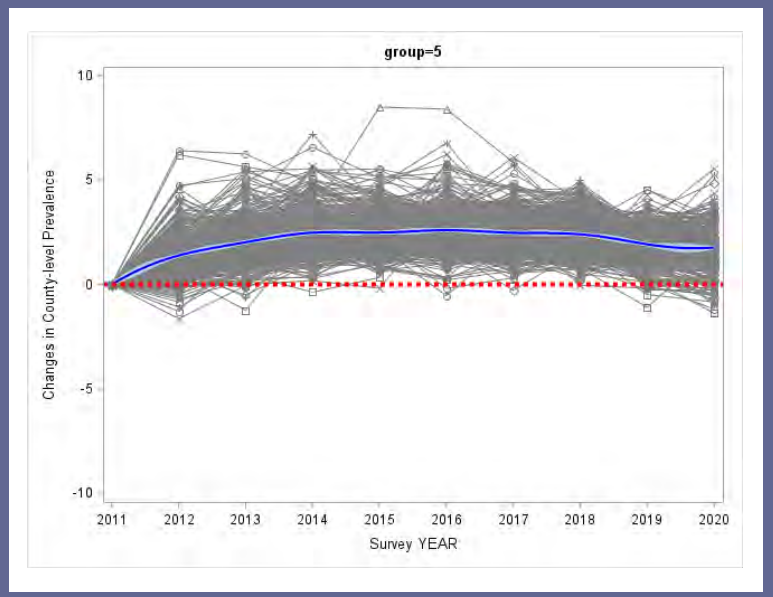

Prevalence Averaged by Counties Groups and Years
group
1
2
3
4
5
6
Average Prevalence
4
3
2
1
0
−1
2012
2014
2016
2018
2020
Year

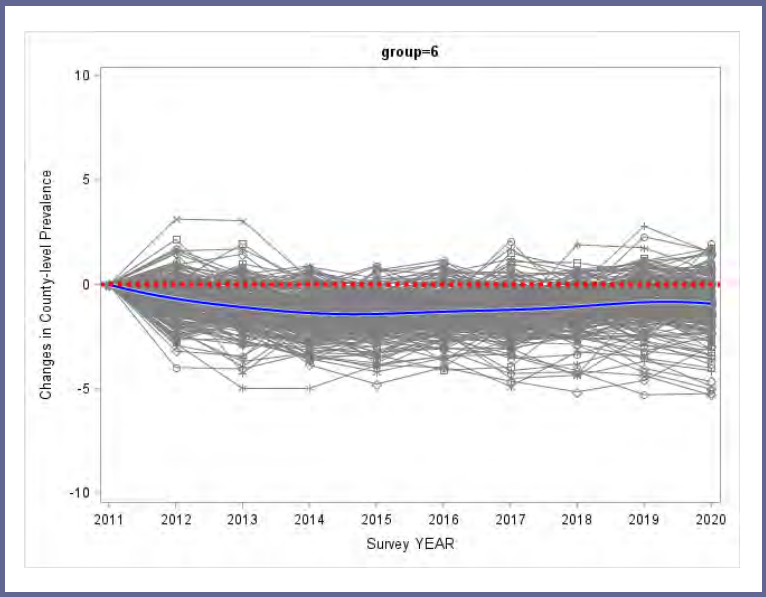

group=1
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR


group=2
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR


group=3
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR

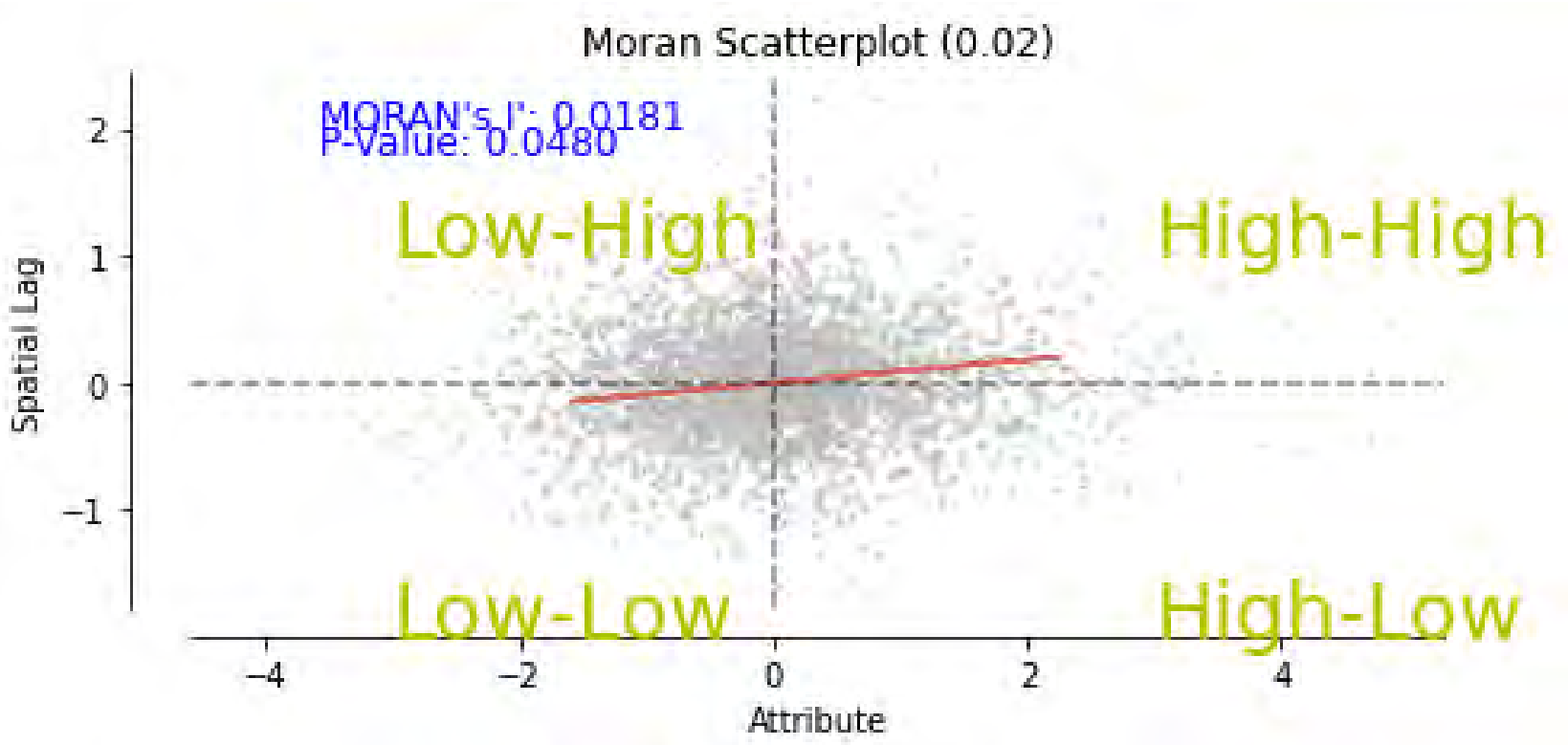

group=4
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR


group=5
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR

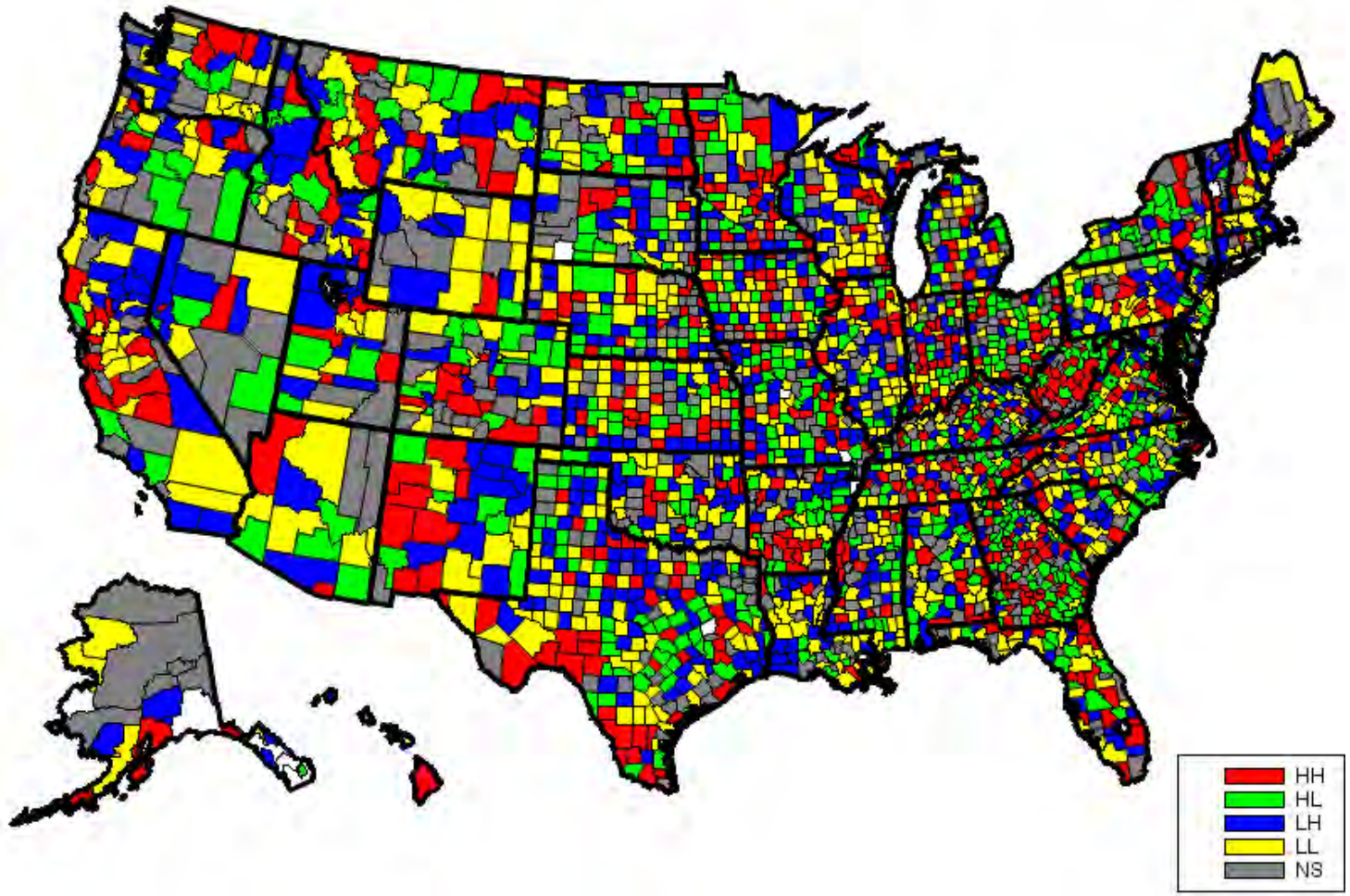

group=6
Changes in County-level Prevalence
10
5
0
-5
-10
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Survey YEAR

Results II. Spatial Analysis

Global Moran’s I



Spatio Patterns of County-level Diabetes Prevalence Trends: LISA Cluster Map



Discussions

- We identified distinct temporal patterns in the trends of county-level diagnosed diabetes prevalence, with 37.62% of counties having a continuously increasing trend, while others displayed more intricate patterns.
- Meanwhile, we observed notable spatial variations, with 22.89% of counties (mostly located in the southeast) having consistently higher changes of prevalence than others.
- The findings provide insights into the temporal and spatial dynamics of diabetes prevalence and could inform targeted interventions to help reduce diabetes prevalence in the most affected areas.
- Two limitations: (1) The analysis covered a ten-year period. A more extended time frame might provide a broader perspective on trends; (2) the uncertainties of county-level estimates by SAE were not counted, which will be considered in future studies

Thank You