

# Data Science for Health Equity

As part of the internally funded RTI Innovation Award program, the **RTI Rarity™** team has curated a national database of **more than 200 SDoH measures within 10 domains** at the Census tract, ZIP code, and county level. The data come from 38 different publicly available federal, state, and nonprofit/academic resources, including the American Community Survey, USDA's Food Environment Atlas, CDC's Wide-ranging ONline Data for Epidemiologic Research (WONDER), HUD, Child Opportunity Index, Opportunity Insights, and more.

RTI Rarity uses supervised machine learning, including random forests and other state-of-the-art predictive methods, to create local social inequity (LSI) scores drawing on these measures of SDoH. The tool and its underlying data allow for the development of both **within-state** and **cross-state summary scores** and 10 domain-specific subscores informed by our conceptual framework (**Figure 1**). The scores yield meaningful insights into the neighborhood-level factors driving local health outcomes.

The RTI Rarity team's pilot phase involved an analysis within the [CMS State Innovation Models Round 2 Evaluation](#) project using LSI scores as a covariate in models testing Ohio's Comprehensive Primary Care demonstration. Within an all-Medicaid population, in both bivariate and fully adjusted triple-difference models, local social inequities were significantly associated with a number of health outcomes that are expensive, preventable, and potentially modifiable through policy mechanisms.

The scores have been benchmarked against 3 existing area-based composite measures related to SDoH: the Area Deprivation Index (ADI), the Social Deprivation Index (SDI), and the Social Vulnerability Index (SVI). In terms of life expectancy at birth, the SVI explains 26% of the variance, the SDI explains 29%, the ADI explains 43%, and LSI explains 67%. In other words, local social inequities account for two-thirds of the disparity between the neighborhoods with the highest and lowest life expectancies in the US.

Our LSI scores can be linked with individual-level data to improve predictive power of individual outcomes. In population-based analyses, these scores can help with identifying neighborhoods and areas for special focus and help, controlling for SDoH in analyses, tracking inequity over time, informing policies and interventions to address root causes, and simulating the effects of alternative policies and future scenarios.

For more information, please see: <https://www.themedicalcareblog.com/artificially-intelligent-social-risk-adjustment/> and <https://www.rti.org/insights/ai-local-social-risk-score-covid-19>.



Figure 1. RTI Rarity Conceptual Model

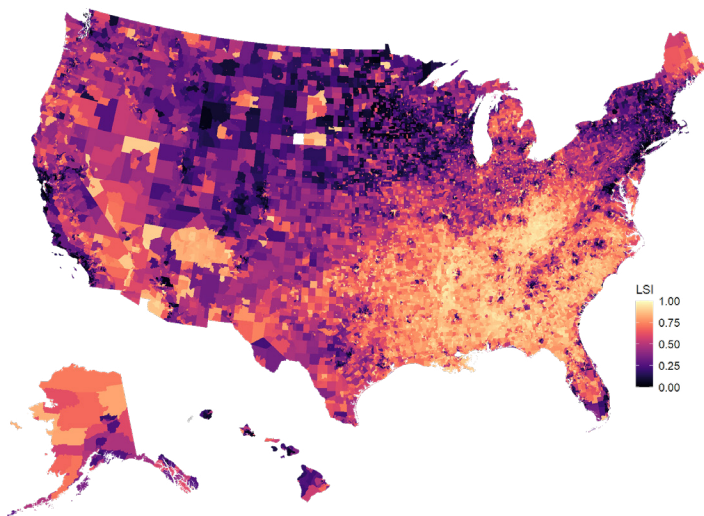


Figure 2. LSI Scores in US Census Tracts

**Figure 2** is a map of cross-state LSI scores at the Census tract level. Pale yellow indicates higher LSI scores, which correspond to neighborhoods with a higher social risk score. With the LSI, we can get a good, high-resolution picture of social and ecological context.

Across the US, we see distinct concentrations of high LSI in the Appalachian region, the Cotton Belt, the Four Corners, and sections of each coast. We also see that the urban cores of large cities have high social risk, surrounded by more wealthy and resourced suburban areas.

Because the LSI scores are output as percentile scores, we can easily compare areas on the same scale. This allows users to identify the top decile, percentile, or even 99.9th percentile to understand where the opportunities may be for interventions, adding resources, investing in those communities, and perhaps paying providers more to care for people in those communities.

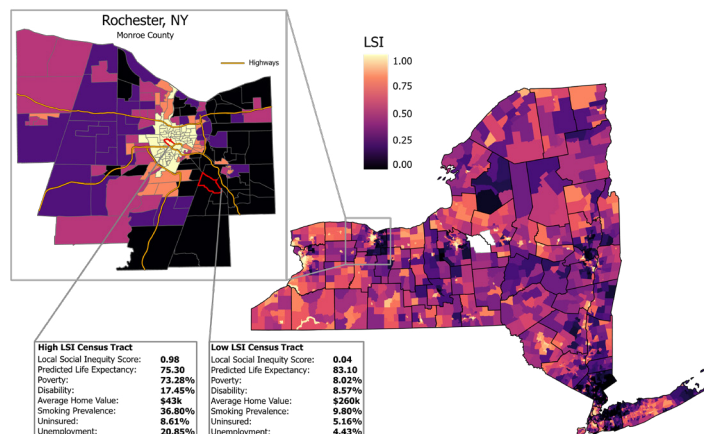


Figure 3. LSI Scores in New York Census Tracts

**Figure 3** is a map of the **within-state** LSI scores for New York Census tracts. The inset box on the left highlights two tracts in Rochester.

Rochester is the 5th poorest city in the US. It has a history of racial segregation and eminent domain that broke apart once-thriving Black communities. Rochester has also experienced economic challenges because of manufacturing jobs leaving the area.

The highlighted tracts are roughly 6 miles apart. They are defined by drastically different social contexts. The tract in the urban core of Rochester (pale yellow) has a LSI score of 0.98 (very high social risk). The suburban tract (dark purple) has a LSI score of 0.04.

Relative to the low LSI neighborhood, the high LSI neighborhood has 9x more residents living in poverty, 2x more residents with a disability, 4x higher rate of smoking, 5x higher rate of unemployment, and median home values that are 6x lower.

RTI Rarity's LSI scores provide a comprehensive health equity-focused index of neighborhood-level social inequity. These robust, cross-disciplinary, and spatially refined scores allow researchers, analysts, and policy makers across the research-practice spectrum to understand the co-occurrence of social and behavioral determinants of health and assess how best to intervene.



#### More Information

Lisa Lines, PhD, MPH  
Senior Health Services Researcher  
llines@rti.org

[www.rti.org](http://www.rti.org)

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. Clients rely on us to answer questions that demand an objective and multidisciplinary approach—one that integrates expertise across the social and laboratory sciences, engineering, and international development. We believe in the promise of science, and we are inspired every day to deliver on that promise for the good of people, communities, and businesses around the world. For more information, visit [www.rti.org](http://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.  
RTI CC 73984807 0522