



# Transactions on Computational Modelling and Intelligent Systems

Volume 5, October 2026, Article 10088  
<https://doi.org/10.65112/tcmis.10088>



eISSN: 3108-5202  
[www.tcmis.org](http://www.tcmis.org)

## RESEARCH ARTICLE

### Data-driven decision support for community health outreach: Prioritizing senior health interventions using county-level public health indicators

Qi Xin\* 

Management Information Systems, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh, PA 15260, USA

\*Corresponding Author. Email: [qix29@pitt.edu](mailto:qix29@pitt.edu) (Q. Xin)

#### Article Information

Received: 22 May 2026  
Accepted: 15 July 2026  
Published: 20 October 2026

AMS 2020 Classification:  
68U35, 62H30

#### Abstract

Senior-serving nonprofit organizations and public health planners need transparent methods for deciding which counties should receive technology training, health education, chronic disease support, care navigation, and related outreach resources first. This study developed a county-level decision-support framework that converts CDC PLACES public health indicators into outreach priority scores, empirical risk clusters, rule-recovery audits, deterministic cluster-specific outreach plans, uncertainty checks, and a senior-focused resource-allocation scenario. The main analysis used the 2024 CDC PLACES county Open Data release (3,144 county equivalents after excluding the national aggregate), and external validation used the 2025 PLACES county GIS-friendly release (3,143 counties). Forty indicators were harmonized across health outcomes, preventive care, health risk behaviors, disability, health status, and health-related social needs. A 2024 Census county age-sex file supplied the estimated population aged 65 years and older for the optimization module. The 2024 priority score had a median of 49.50, a top-quintile threshold of 69.16, and a top-decile threshold of 76.17. The critical top-decile band contained 315 counties. A five-cluster operational solution separated critical multimorbidity and social-need counties, high chronic-disease and behavioral-risk counties, preventive-care gap counties, moderate mixed-risk counties, and lower-risk monitoring counties. KNN and MICE-style imputations produced ordered cluster mean-score correlations of 0.964 and 0.967. Logistic regression achieved five-fold cross-validation  $F1=0.979$ ; XGBoost achieved held-out  $F1=0.955$ ; and XGBoost achieved 2025 external  $F1=0.954$ . A PLACES confidence-interval simulation produced a median top-quintile threshold of 68.07 with a 2.5th-97.5th percentile interval of 67.77-68.34; the deterministic threshold was above this simulated interval, so threshold-level agreement was not assumed, although median county priority-band stability was 0.973. Under a 600-unit budget, the mixed-integer optimization selected 158 elevated-or-higher-priority counties and covered approximately 4.61 million residents aged 65 years or older. The framework provides a reproducible county-level planning workflow while preserving the limits of model-based local estimates.

**Keywords:** Decision support, community health outreach, senior communities, nonprofit planning, chronic disease intervention

## 1. Introduction

Senior communities often face overlapping needs: digital access and skills, chronic disease self-management, transportation barriers, preventive care gaps, disability support, and social isolation. Nonprofit organizations and local health planners must decide where to deploy limited staff time, mobile health education, telehealth training, screening events, and care navigation. A county can have high chronic disease burden but only moderate preventive-care gaps, while another county can have lower disease burden but high social-need stress. A practical decision-support system therefore has to transform many local health indicators into rankings, clusters, and action plans rather than leaving planners with disconnected tables.

CDC PLACES provides local estimates of health outcomes, preventive service use, health risk behaviors, disability, health status, and health-related social needs at geographic levels that include counties, places, census tracts, and ZIP Code tabulation areas [1–4]. Its methodology uses model-based small-area estimation, including multilevel regression and poststratification, to combine survey and population data into comparable local estimates [2, 6]. This makes PLACES useful for outreach planning because it provides consistent county indicators across the United States. At the same time, the estimates are not direct program impact measures, so the correct role is prioritization and planning rather than causal evaluation of an intervention [4, 6].

This paper builds a data-driven decision-support framework for county-level senior health outreach. The framework first computes an outreach priority score from risk-oriented public health indicators, then identifies empirical risk clusters, audits whether the high-priority decision rule can be recovered from the raw indicator matrix, and finally translates cluster profiles into deterministic outreach packages and a budget-constrained allocation scenario. The target user is a nonprofit, public health agency, or senior-serving coalition deciding where to prioritize technical training, health education, chronic disease programs, care navigation, transportation support, or community resource referral.

The paper makes five empirical contributions. First, it defines a transparent risk-scoring method that combines chronic disease, preventive care, behavior, disability, health status, and health-related social-need indicators. Second, it compares clustering solutions for operational risk-profile discovery. Third, it reframes supervised learning as a rule-recovery and feature-importance audit rather than as independent clinical prediction. Fourth, it adds uncertainty and imputation robustness checks for priority thresholds and cluster boundaries. Fifth, it implements a senior-specific resource-allocation module that combines county priority scores with Census estimates of residents aged 65 years and older.

## **2. Literature review**

Local public health decision making depends on geography-specific data. County Health Rankings showed that county-level indicators can make place-based health differences visible and can stimulate policy and community action [7, 8]. PLACES extends this local decision-support logic by providing standardized small-area estimates for counties and smaller geographic units [1, 6]. Greenlund et al. describe PLACES as a source of local estimates that can support public health planning, resource allocation, and community intervention targeting, while noting that the estimates should be interpreted with the limits of modeled local data [6].

The public health advocacy literature also emphasizes social and structural conditions. Healthy People 2030 frames social determinants of health as the conditions in which people are born, live, learn, work, play, worship, and age [9]. CDC PLACES includes health-related social needs such as food insecurity, housing insecurity, transportation barriers, lack of support, and social isolation [1, 4]. For senior communities, these indicators are directly relevant because outreach success often depends on transportation, technology access, social support, and linkage to care, not only on disease prevalence.

Composite scoring systems are common in applied public health planning, but their value depends on transparency. A score that hides weights or mixes favorable and unfavorable indicators can produce rankings that are hard to defend. The framework in this study therefore uses explicit category weights and risk-direction rules. Chronic outcomes receive the largest weight because senior outreach programs often focus on diabetes, blood pressure, heart disease, COPD, depression, obesity, and stroke risk. Disability and health status receive substantial weight because they affect whether residents can use telehealth, attend classes, or manage self-care. Preventive care, behavior, and social-need indicators remain in the score because they identify actionable outreach pathways.

Machine learning supports the framework in two ways. Clustering is used to identify different types of outreach need, not to rank counties by itself. K-means clustering is a long-standing least-squares partitioning method [15], and Ward hierarchical clustering is a variance-minimizing agglomerative method [16, 17]. Supervised models are used as an audit layer to test whether the decision rule can be recovered from the indicator matrix and to identify the variables that most strongly reproduce the high-priority flag. Random forests use ensembles of decision trees [13], and XGBoost implements gradient boosted trees for structured tabular data [14]. The implementation used scikit-learn as the primary modeling framework [18].

AI and machine-learning tools in health policy are most useful when they support human decision making rather than replacing it. Ramezani et al. describe AI as a decision-support tool for policy settings where transparency and governance are required [11]. The planning text in this study is placed after empirical scoring and clustering; it does not estimate disease burden. Evaluation of health-related text also requires more than fluency. Tam et al. emphasize human-relevant evaluation criteria for health-care language-model outputs [12]. The present study uses keyword consistency, readability, and risk-priority alignment as operational checks, with readability interpreted through classic Flesch, Flesch-Kincaid, and SMOG-related principles [20, 21, 22]. Resident-facing materials would require additional review using plain-language tools such as the CDC Clear Communication Index [10].

### 3. Methodology

#### 3.1. Data source and unit of analysis

The empirical pipeline used county-level CDC PLACES data. The 2024 county Open Data release provided the main modeling file, and the 2025 county GIS-friendly release provided an external validation file [3–5]. The Open Data file stores one geography-measure-data-value-type record per row, which preserved point estimates and confidence limits for the uncertainty analysis. The GIS-friendly file stores one county per row with measure-specific columns, which made it suitable for external validation. The unit of analysis was the county or county equivalent. The 2024 analysis included 3,144 county equivalents after excluding the national aggregate row. The 2025 validation included 3,143 counties, with complete all-measure validation available for 2,299 counties. The main score used crude prevalence because outreach allocation asks how large the local burden and service gap are in the county adult population. A sensitivity test repeated the scoring with age-adjusted prevalence.

To align the optimization layer with senior outreach, the county priority file was merged with the 2024 U.S. Census Bureau Population Estimates Program county age-sex file. Records were filtered to county/statistical equivalent rows (SUMLEV=50) and the July 1, 2024 estimate year (YEAR=6). The variable AGE65PLUS\_TOT supplied the estimated county population aged 65 years and older [23]. The Census 2024 National Counties Gazetteer file supplied county internal-point latitude and longitude for the centroid map [24].

#### 3.2. Indicator harmonization and risk scoring

Forty PLACES measures were grouped into six analytic categories: health outcomes, disability, health status, health risk behaviors, prevention, and health-related social needs. Each measure was assigned a risk direction. For disease, disability, poor health status, unhealthy behavior, insurance gap, and social-need measures, higher prevalence increased priority. For beneficial preventive-care measures such as checkups, dental visits, cholesterol screening, cancer screening, mammography, and blood pressure medication use among adults with high blood pressure, lower prevalence increased priority. This rule prevented beneficial care uptake from being incorrectly interpreted as risk.

For each measure, counties were ranked by risk-oriented percentile from 0 to 100 using average ranks. Let  $r_{ij}$  be the risk percentile for county  $i$  and measure  $j$ . For category  $c$ , the category score  $C_{ic}$  was the mean of available risk percentiles in that category. The overall priority score  $S_i$  was the weighted mean of available category scores:  $S_i = \sum_c w_c C_{ic} / \sum_c w_c$  over categories with observed values. The category weights were 0.35 for health outcomes, 0.15 for disability, 0.15 for health status, 0.15 for health risk behaviors, 0.10 for prevention, and 0.10 for health-related social needs. This weighting emphasizes chronic disease and functioning while preserving actionable care and social-need signals. Table 1 summarizes the decision-support modules, and Table 2 provides the full measure dictionary used for scoring.

Table 1. Decision-support modules, empirical methods, outputs, and planning use.

| Module                       | Method used                                                               | Primary output                                        | Decision use                                             |
|------------------------------|---------------------------------------------------------------------------|-------------------------------------------------------|----------------------------------------------------------|
| Risk scoring                 | Risk-oriented percentiles and weighted category aggregation               | County outreach priority score and priority band      | Rank counties for first-wave outreach                    |
| Clustering                   | K-means, Ward hierarchical clustering, PCA visualization                  | Five operational risk clusters                        | Match counties to intervention packages                  |
| Rule-recovery audit          | Logistic regression, decision tree, random forest, XGBoost                | Top-quintile flag recovery and feature importance     | Audit internal reproducibility and leading indicators    |
| Planning layer               | Deterministic cluster-to-plan template                                    | Cluster-specific outreach plan and intensity level    | Translate empirical profiles into nonprofit action plans |
| Uncertainty and optimization | PLACES CI simulation; binary budget allocation with Census 65+ population | Threshold stability and budget-constrained county set | Support stable prioritization and resource allocation    |

Table 2. Full PLACES measure dictionary, category weights, and risk-direction rules.

| Measure ID | Category        | Weight | Risk-direction rule                 | Measure                                                   |
|------------|-----------------|--------|-------------------------------------|-----------------------------------------------------------|
| ARTHRITIS  | Health Outcomes | 0.35   | Higher prevalence = higher priority | Arthritis among adults                                    |
| BPHIGH     | Health Outcomes | 0.35   | Higher prevalence = higher priority | High blood pressure among adults                          |
| CANCER     | Health Outcomes | 0.35   | Higher prevalence = higher priority | Cancer (non-skin) or melanoma among adults                |
| CASTHMA    | Health Outcomes | 0.35   | Higher prevalence = higher priority | Current asthma among adults                               |
| CHD        | Health Outcomes | 0.35   | Higher prevalence = higher priority | Coronary heart disease among adults                       |
| COPD       | Health Outcomes | 0.35   | Higher prevalence = higher priority | Chronic obstructive pulmonary disease among adults        |
| DEPRESSION | Health Outcomes | 0.35   | Higher prevalence = higher priority | Depression among adults                                   |
| DIABETES   | Health Outcomes | 0.35   | Higher prevalence = higher priority | Diagnosed diabetes among adults                           |
| HIGHCHOL   | Health Outcomes | 0.35   | Higher prevalence = higher priority | High cholesterol among adults who have ever been screened |
| OBESITY    | Health Outcomes | 0.35   | Higher prevalence = higher priority | Obesity among adults                                      |
| STROKE     | Health Outcomes | 0.35   | Higher prevalence = higher priority | Stroke among adults                                       |
| TEETHLOST  | Health Outcomes | 0.35   | Higher prevalence = higher priority | All teeth lost among adults aged $\geq 65$ years          |
| COGNITION  | Disability      | 0.15   | Higher prevalence = higher priority | Cognitive disability among adults                         |
| DISABILITY | Disability      | 0.15   | Higher prevalence = higher priority | Any disability among adults                               |
| HEARING    | Disability      | 0.15   | Higher prevalence = higher priority | Hearing disability among adults                           |
| INDEPLIVE  | Disability      | 0.15   | Higher prevalence = higher priority | Independent living disability among adults                |
| MOBILITY   | Disability      | 0.15   | Higher prevalence = higher priority | Mobility disability among adults                          |
| SELFCARE   | Disability      | 0.15   | Higher prevalence = higher priority | Self-care disability among adults                         |
| VISION     | Disability      | 0.15   | Higher prevalence = higher priority | Vision disability among adults                            |
| GHLTH      | Health Status   | 0.15   | Higher prevalence = higher priority | Fair or poor self-rated health status among adults        |
| MHLTH      | Health Status   | 0.15   | Higher prevalence = higher priority | Frequent mental distress among adults                     |
| PHLTH      | Health Status   | 0.15   | Higher prevalence = higher priority | Frequent physical distress among adults                   |

Table 2. Continued.

| Measure ID   | Category                    | Weight | Risk-direction rule                 | Measure                                                                              |
|--------------|-----------------------------|--------|-------------------------------------|--------------------------------------------------------------------------------------|
| BINGE        | Health Risk Behaviors       | 0.15   | Higher prevalence = higher priority | Binge drinking among adults                                                          |
| CSMOKING     | Health Risk Behaviors       | 0.15   | Higher prevalence = higher priority | Current cigarette smoking among adults                                               |
| LPA          | Health Risk Behaviors       | 0.15   | Higher prevalence = higher priority | No leisure-time physical activity among adults                                       |
| SLEEP        | Health Risk Behaviors       | 0.15   | Higher prevalence = higher priority | Short sleep duration among adults                                                    |
| ACCESS2      | Prevention                  | 0.10   | Higher prevalence = higher priority | Current lack of health insurance among adults aged 18-64 years                       |
| BPMED        | Prevention                  | 0.10   | Lower prevalence = higher priority  | Taking medicine to control high blood pressure among adults with high blood pressure |
| CHECKUP      | Prevention                  | 0.10   | Lower prevalence = higher priority  | Visits to doctor for routine checkup within the past year among adults               |
| CHOLSCREEN   | Prevention                  | 0.10   | Lower prevalence = higher priority  | Cholesterol screening among adults                                                   |
| COLON_SCREEN | Prevention                  | 0.10   | Lower prevalence = higher priority  | Colorectal cancer screening among adults aged 45–75 years                            |
| DENTAL       | Prevention                  | 0.10   | Lower prevalence = higher priority  | Visited dentist or dental clinic in the past year among adults                       |
| MAMMOUSE     | Prevention                  | 0.10   | Lower prevalence = higher priority  | Mammography use among women aged 50-74 years                                         |
| EMOTIONSPT   | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Lack of social and emotional support among adults                                    |
| FOODINSECU   | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Food insecurity in the past 12 months among adults                                   |
| FOODSTAMP    | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Received food stamps in the past 12 months among adults                              |
| HOUSINSECU   | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Housing insecurity in the past 12 months among adults                                |
| ISOLATION    | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Feeling socially isolated among adults                                               |
| LACKTRPT     | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Lack of reliable transportation in the past 12 months among adults                   |
| SHUTUTILITY  | Health-Related Social Needs | 0.10   | Higher prevalence = higher priority | Utility services shut-off threat in the past 12 months among adults                  |

### 3.3. Missingness, imputation, and clustering

The priority score was computed from observed category scores and renormalized across available category weights. This avoided assigning artificial HRSN values to counties without social-need measures while preserving their chronic disease, disability, health-status, behavior, and prevention evidence. For clustering and model auditing, missing values were handled with median imputation. Because HRSN coverage was incomplete, the clustering results were compared with KNN imputation ( $k=5$ ), MICE-style iterative imputation, and a complete-case analysis [25, 26].

Clustering used the 40 risk-oriented percentile indicators after imputation and standardization. K-means was evaluated for  $K=3$  through  $K=8$  using silhouette score, Calinski-Harabasz index, Davies-Bouldin index, and within-cluster sum of squared errors. The silhouette statistic was computed on a reproducible 500-county sample to avoid quadratic memory growth; other diagnostics used all counties. Ward hierarchical clustering was computed with Euclidean Ward linkage and compared at  $K=5$ .  $K=4$  produced the lowest Davies-Bouldin index in this run, but  $K=5$  was retained for the operational framework because it separated preventive-care and access-stressed counties from general moderate-risk counties, a distinction that directly changes outreach design.

Table 3 reports category-level data coverage in the 2024 county analysis.

Table 3. Category-level data coverage and missingness in the 2024 county analysis.

| Category                    | Measures | Counties with all category measures | Counties with partial category data | Counties with no category data | Mean missing cells (%) |
|-----------------------------|----------|-------------------------------------|-------------------------------------|--------------------------------|------------------------|
| Health Outcomes             | 12       | 3,077                               | 67                                  | 0                              | 0.355                  |
| Disability                  | 7        | 3,144                               | 0                                   | 0                              | 0.000                  |
| Health Status               | 3        | 3,144                               | 0                                   | 0                              | 0.000                  |
| Health Risk Behaviors       | 4        | 3,144                               | 0                                   | 0                              | 0.000                  |
| Prevention                  | 7        | 3,077                               | 67                                  | 0                              | 0.609                  |
| Health-Related Social Needs | 7        | 2,417                               | 0                                   | 727                            | 23.123                 |

### 3.4. Rule-recovery audit, planning layer, uncertainty, and optimization

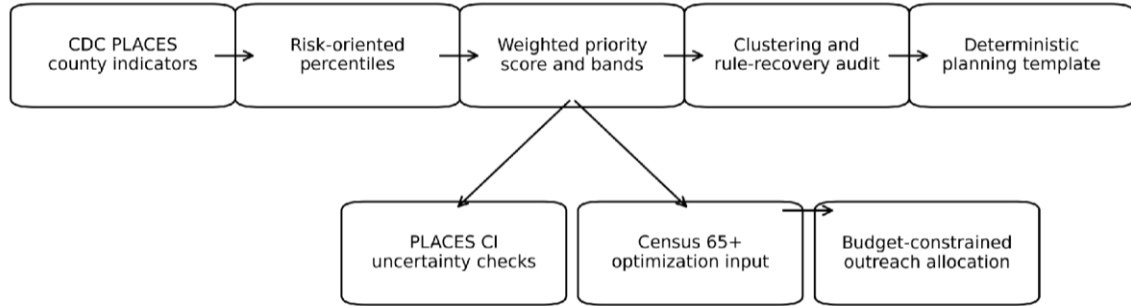
The supervised target was a high-priority county flag equal to one for counties in the top quintile of the 2024 priority score. This target is a decision-support label derived from the same indicator family used in scoring. The supervised models were therefore interpreted as a rule-recovery and feature-importance audit, not as independent clinical prediction or causal inference. Logistic regression, decision tree, random forest, and XGBoost models were evaluated using five-fold stratified cross-validation, a 70/30 held-out 2024 split, and external validation against the 2025 top-quintile flag. Metrics included accuracy, precision, recall, F1, ROC-AUC, PR-AUC, and confusion-matrix counts. Random seed 42 was fixed for stochastic procedures. The decision tree used `max_depth=5`, the random forest used 200 trees with `sqrt` feature sampling, and XGBoost used 100 trees, `max_depth=3`, `learning_rate=0.1`, `subsample=0.9`, and `colsample_bytree=0.9`.

The planning layer used a deterministic cluster-to-plan template rather than stochastic language generation. Each cluster was assigned an intensity from 1 to 5 based on its mean priority score and dominant risk categories. The template maps the three dominant category risks to concise outreach actions, so the reported empirical results do not depend on an external language-model engine, temperature, or top-p sampling parameter. Consistency was measured as the share of dominant risk domains reflected in the plan text. Readability was measured by Flesch Reading Ease and Flesch-Kincaid grade. Risk-priority alignment was measured as the Spearman correlation between county priority scores and the intervention intensity assigned by cluster.

Priority-threshold uncertainty was assessed with a Monte Carlo simulation based on the PLACES lower and upper confidence limits. For each observed county-measure estimate, the standard error was approximated as  $(\text{upper confidence limit} - \text{lower confidence limit}) / (2 \times 1.96)$ , simulated values were drawn from a normal approximation, values were truncated to the 0-100 prevalence range, and the full scoring rule was recomputed for 300 draws. The simulation assumed independent county-measure errors, approximate normality of PLACES confidence intervals, no covariance across measures or counties, and the same risk-direction and category-weight rules in every draw.

The resource-allocation module was formulated as a binary budget-constrained optimization problem. Let  $x_i$  indicate whether county  $i$  is selected,  $S_i$  be the 0-100 priority score,  $P_i$  be the estimated population aged 65 years and older, and  $c_i$  be the normalized cost unit assigned to the cluster intervention intensity. One cost unit represents one standardized outreach package unit combining staff time, logistics, materials, and care-navigation effort; a

level-5 cluster package consumes five units and a level-1 package consumes one unit. The objective was maximize  $\sum_i x_i (S_i/100) P_i$ , subject to  $\sum_i x_i c_i \leq B$ , a state-level cap of 25 counties to avoid geographic concentration, and  $x_i = 0$  for counties below the elevated-priority band. This population-weighted objective was used because senior-serving outreach must consider both severity and the number of older adults who could be reached. The mixed-integer linear program was solved with SciPy optimize.milp using the HiGHS solver, binary integrality, a 120-second time limit, and zero relative MIP gap tolerance [27, 28]. Budget sensitivity was assessed for  $B=100, 200, 300, 400, 500, 600, 800$ , and 1,000 cost units.



Transparent scoring links local health indicators to clusters, planning actions, uncertainty checks, and senior outreach allocation.

Figure 1. County outreach decision-support framework.

## 4. Experimental tests

### 4.1. Data coverage and priority score distribution

The experimental evaluation computed priority scores, clusters, rule-recovery models, planning-layer outputs, 2025 validation results, uncertainty checks, imputation robustness, and a senior-specific optimization scenario. The scoring method used category-weight renormalization, so counties without HRSN data remained eligible for ranking without imputing social needs into the priority score. For machine-learning matrices, missing values were handled separately as described in the methodology.

Table 4 reports selected empirical indicator distributions, and Table 5 compares release-level coverage for the 2024 scoring file and the 2025 validation file.

Table 4. Selected empirical indicator distributions in the 2024 county analysis.

| Measure ID  | Measure                     | Category                    | N observed | Mean  | Median | P90   |
|-------------|-----------------------------|-----------------------------|------------|-------|--------|-------|
| BPHIGH      | High Blood Pressure         | Health Outcomes             | 3,077      | 37.61 | 37.30  | 44.84 |
| COPD        | COPD                        | Health Outcomes             | 3,144      | 9.02  | 8.80   | 12.10 |
| DIABETES    | Diabetes                    | Health Outcomes             | 3,144      | 13.50 | 13.20  | 17.10 |
| OBESITY     | Obesity                     | Health Outcomes             | 3,144      | 37.63 | 38.20  | 42.90 |
| STROKE      | Stroke                      | Health Outcomes             | 3,144      | 4.26  | 4.20   | 5.50  |
| CSMOKING    | Current Cigarette Smoking   | Health Risk Behaviors       | 3,144      | 17.22 | 17.00  | 22.20 |
| LPA         | Physical Inactivity         | Health Risk Behaviors       | 3,144      | 27.71 | 27.40  | 34.60 |
| CHECKUP     | Annual Checkup              | Prevention                  | 3,144      | 76.56 | 77.20  | 80.50 |
| COLON_SCORE | Colorectal Cancer Screening | Prevention                  | 3,144      | 63.83 | 64.20  | 69.50 |
| DENTAL      | Dental Visit                | Prevention                  | 3,144      | 58.25 | 60.10  | 68.20 |
| DISABILITY  | Any Disability              | Disability                  | 3,144      | 34.82 | 34.50  | 42.57 |
| MOBILITY    | Mobility Disability         | Disability                  | 3,144      | 16.67 | 16.30  | 21.90 |
| GHLTH       | General Health              | Health Status               | 3,144      | 20.96 | 20.30  | 27.70 |
| PHLTH       | Frequent Physical Distress  | Health Status               | 3,144      | 14.73 | 14.50  | 18.20 |
| FOODINSECU  | Food Insecurity             | Health-Related Social Needs | 2,417      | 15.50 | 14.20  | 23.60 |
| ISOLATION   | Social Isolation            | Health-Related Social Needs | 2,417      | 32.88 | 32.80  | 36.40 |
| LACKTRPT    | Transportation Barriers     | Health-Related Social Needs | 2,417      | 9.08  | 8.50   | 12.70 |

Table 5. 2024 and 2025 county release coverage used in scoring and validation.

| Dataset                          | Counties | Complete all 40 measures | Complete core non-HRSN measures | Median priority score | Top-quintile threshold |
|----------------------------------|----------|--------------------------|---------------------------------|-----------------------|------------------------|
| 2024 county Open Data release    | 3,144    | 2,350                    | 3,077                           | 49.50                 | 69.16                  |
| 2025 county GIS-friendly release | 3,143    | 2,299                    | 2,956                           | 50.28                 | 70.40                  |

The 2024 priority score ranged across the operational spectrum, with a median of 49.50. The top-quintile threshold was 69.16 and the top-decile threshold was 76.17. The critical band contained 315 counties with a mean priority score of 80.76. Table 6 shows that these counties combined high disability scores with actionable prevention gaps.

Table 6. Priority bands, county counts, population, disability, and prevention gap scores.

| Priority band                    | Counties | Mean score | Median score | Mean population | Mean adult population | Mean disability score | Mean prevention gap score |
|----------------------------------|----------|------------|--------------|-----------------|-----------------------|-----------------------|---------------------------|
| Critical priority (top 10%)      | 315      | 80.76      | 80.17        | 18228.87        | 14110.42              | 90.76                 | 62.68                     |
| High priority (next 10%)         | 314      | 72.63      | 72.63        | 28601.02        | 21872.37              | 81.04                 | 58.72                     |
| Elevated priority (next 20%)     | 629      | 62.41      | 61.98        | 50082.36        | 38719.83              | 67.54                 | 55.27                     |
| Moderate priority (next 20%)     | 628      | 49.55      | 49.50        | 96216.70        | 74660.75              | 50.00                 | 50.72                     |
| Baseline monitoring (bottom 40%) | 1,258    | 30.51      | 31.55        | 180158.06       | 141718.84             | 23.32                 | 41.74                     |

Table 7 lists the 15 highest-scoring counties. Figure 2 maps county priority scores by geographic centroid, and Figure 3 shows the score distribution with the top-quintile and top-decile thresholds. The top-ranked counties were not simply the largest counties, which is important for nonprofit planning because rural and lower-population counties can have high per-resident need even when total population is small.

Table 7. Top 15 counties by 2024 outreach priority score.

| FIPS  | State | County   | Population | Score | Outcomes | Disability | Status | Behaviors | Prevention | HRSN  |
|-------|-------|----------|------------|-------|----------|------------|--------|-----------|------------|-------|
| 21237 | KY    | Wolfe    | 6,400      | 90.01 | 95.72    | 99.65      | 99.39  | 74.30     | 63.77      | 91.36 |
| 47067 | TN    | Hancock  | 6,845      | 89.02 | 95.59    | 98.45      | 99.36  | 73.19     | 66.62      | 82.52 |
| 54047 | WV    | McDowell | 17,850     | 88.54 | 96.96    | 99.30      | 99.69  | 74.89     | 49.16      | 86.06 |
| 38085 | ND    | Sioux    | 3,711      | 87.98 | 73.59    | 99.46      | 99.76  | 95.59     | 92.04      |       |
| 54059 | WV    | Mingo    | 22,573     | 87.95 | 94.75    | 95.80      | 99.06  | 76.26     | 49.91      | 91.26 |
| 48261 | TX    | Kenedy   | 358        | 87.90 | 80.36    | 99.96      | 99.67  | 74.90     | 87.01      | 98.94 |
| 21153 | KY    | Magoffin | 11,357     | 87.68 | 93.22    | 98.44      | 95.29  | 73.49     | 63.28      | 86.49 |
| 21129 | KY    | Lee      | 7,261      | 87.57 | 89.97    | 98.65      | 97.76  | 74.41     | 70.02      | 84.56 |
| 40001 | OK    | Adair    | 19,576     | 86.93 | 81.92    | 97.19      | 97.53  | 72.00     | 87.47      | 95.02 |
| 54109 | WV    | Wyoming  | 20,527     | 86.64 | 95.90    | 95.87      | 99.06  | 75.84     | 49.78      | 74.79 |
| 21095 | KY    | Harlan   | 25,662     | 86.30 | 90.96    | 97.59      | 97.14  | 73.61     | 59.28      | 82.89 |
| 54045 | WV    | Logan    | 31,316     | 86.28 | 94.07    | 95.21      | 98.81  | 75.33     | 52.11      | 77.37 |
| 21159 | KY    | Martin   | 11,095     | 86.26 | 84.60    | 97.73      | 98.16  | 75.02     | 77.06      | 83.06 |
| 48351 | TX    | Newton   | 12,052     | 86.22 | 84.32    | 94.82      | 94.82  | 77.73     | 74.13      | 91.87 |
| 05111 | AR    | Poinsett | 22,495     | 86.17 | 85.28    | 93.49      | 98.14  | 75.95     | 75.64      |       |

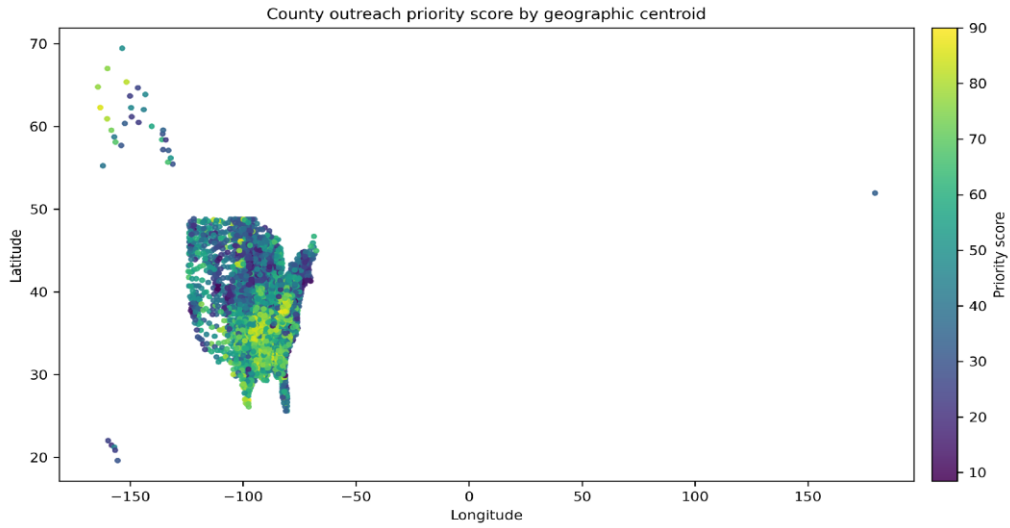


Figure 2. County outreach priority score by geographic centroid.

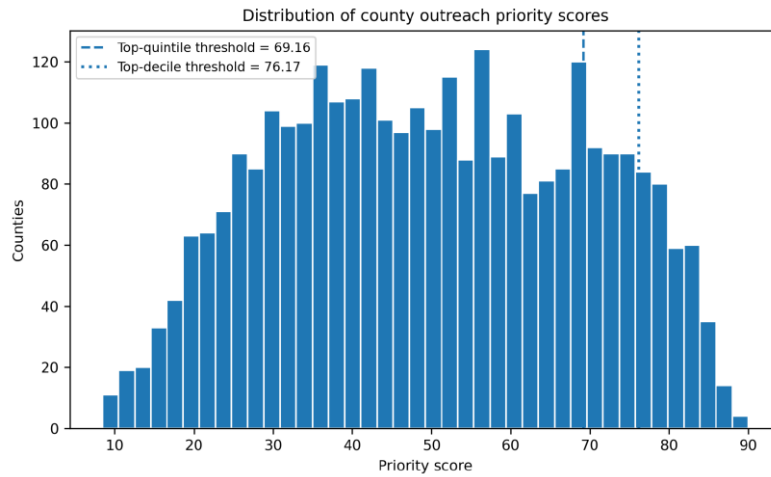


Figure 3. Distribution of the 2024 county outreach priority score with top-quintile and top-decile thresholds.

#### 4.2. Cluster selection and cluster profiles

Cluster diagnostics are shown in Table 8 and Figure 4.  $K=4$  produced the lowest Davies-Bouldin index, while  $K=5$  produced a more useful program structure by separating access-stressed preventive-care counties from general moderate-risk counties. Ward clustering at  $K=5$  was less aligned with the K-means solution, as shown in Table 9. The selected five-cluster K-means solution is visualized in Figure 5, and the cluster category-score heatmap is shown in Figure 6.

Table 8. K-means cluster selection diagnostics for  $K=3$  through  $K=8$ .

| K | Silhouette | Calinski-Harabasz | Davies-Bouldin | Within-cluster SSE |
|---|------------|-------------------|----------------|--------------------|
| 3 | 0.198      | 1353.15           | 1.67           | 67554.57           |
| 4 | 0.202      | 1168.21           | 1.51           | 59429.49           |
| 5 | 0.166      | 994.70            | 1.71           | 55461.01           |
| 6 | 0.154      | 887.67            | 1.64           | 52087.64           |
| 7 | 0.141      | 799.03            | 1.76           | 49741.51           |
| 8 | 0.135      | 728.36            | 1.83           | 47894.05           |

Table 9. K-means and Ward hierarchical clustering comparison at  $K=5$ .

| Method            | K | Silhouette | Calinski-Harabasz | Davies-Bouldin | Adjusted Rand vs K-means |
|-------------------|---|------------|-------------------|----------------|--------------------------|
| K-means           | 5 | 0.166      | 994.70            | 1.71           | 1.000                    |
| Ward hierarchical | 5 | 0.147      | 801.17            | 1.79           | 0.440                    |

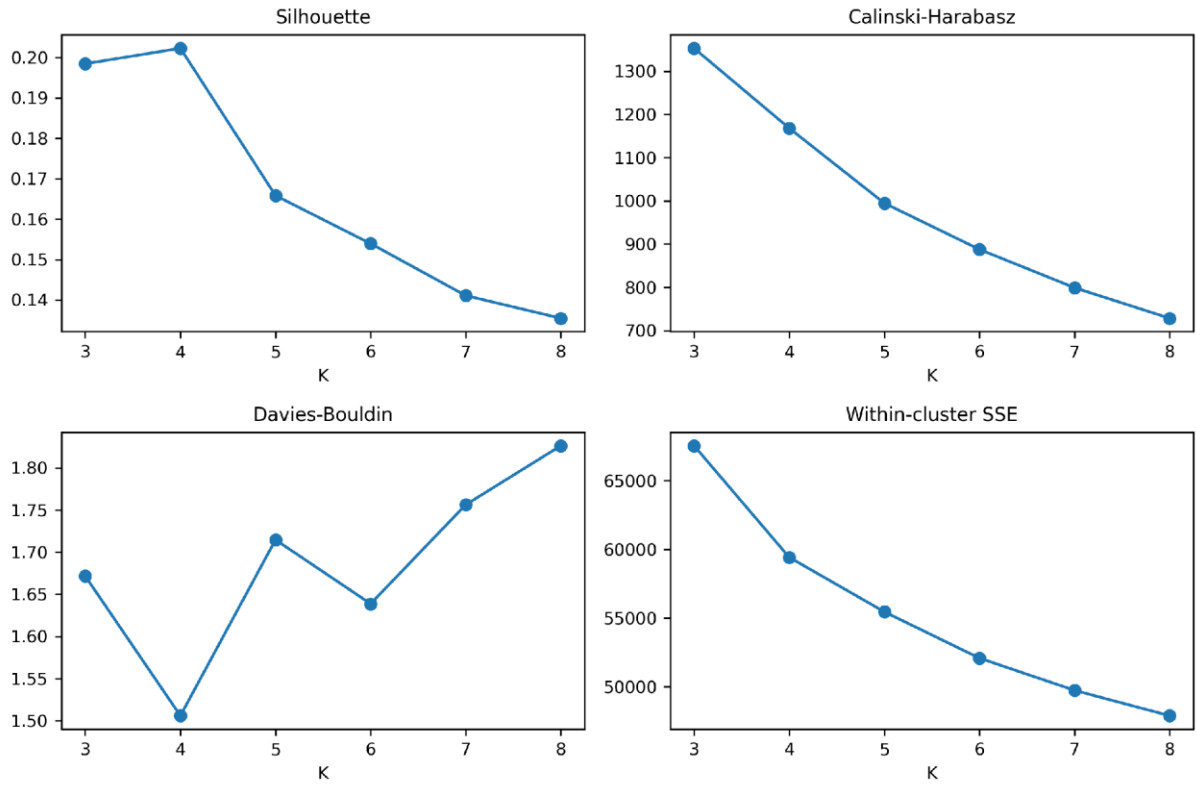


Figure 4. K-means cluster selection diagnostics.

Table 10. Empirical cluster profiles and mean category scores.

| Cluster | Label                                                         | Counties | Mean score | Top20 share | Health Outcomes | Disability | Health Status | Health Risk Behaviors | Prevention | Health-Related Social Needs |
|---------|---------------------------------------------------------------|----------|------------|-------------|-----------------|------------|---------------|-----------------------|------------|-----------------------------|
| 1       | Critical multimorbidity, disability, and social-need counties | 459      | 73.22      | 0.695       | 67.10           | 82.55      | 80.97         | 66.10                 | 67.76      | 85.35                       |
| 2       | High chronic-disease and behavioral-risk counties             | 681      | 68.83      | 0.455       | 73.52           | 75.02      | 76.17         | 62.55                 | 49.40      | 56.78                       |
| 3       | Preventive-care gap and access-stressed counties              | 532      | 48.01      | 0.000       | 36.15           | 46.84      | 52.28         | 51.86                 | 66.38      | 64.67                       |
| 4       | Moderate mixed-risk counties                                  | 748      | 43.52      | 0.000       | 52.47           | 42.77      | 37.00         | 42.44                 | 36.53      | 27.71                       |
| 5       | Lower-risk monitoring counties                                | 724      | 25.46      | 0.000       | 24.71           | 15.69      | 17.57         | 34.50                 | 41.34      | 25.37                       |

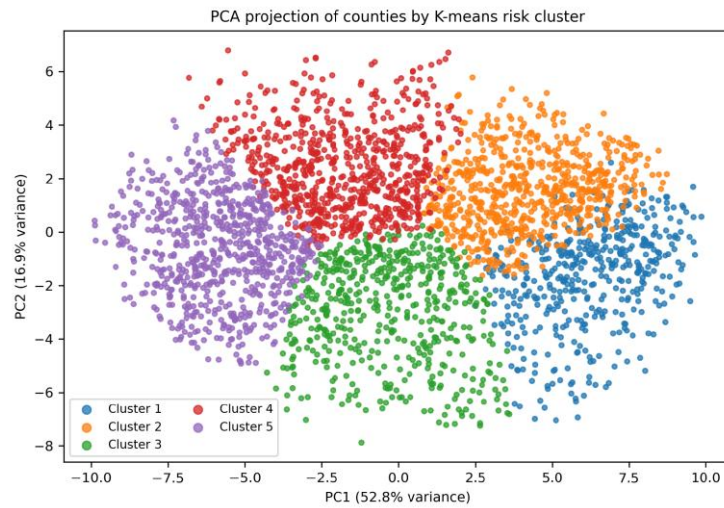


Figure 5. PCA projection of counties colored by K-means risk cluster.

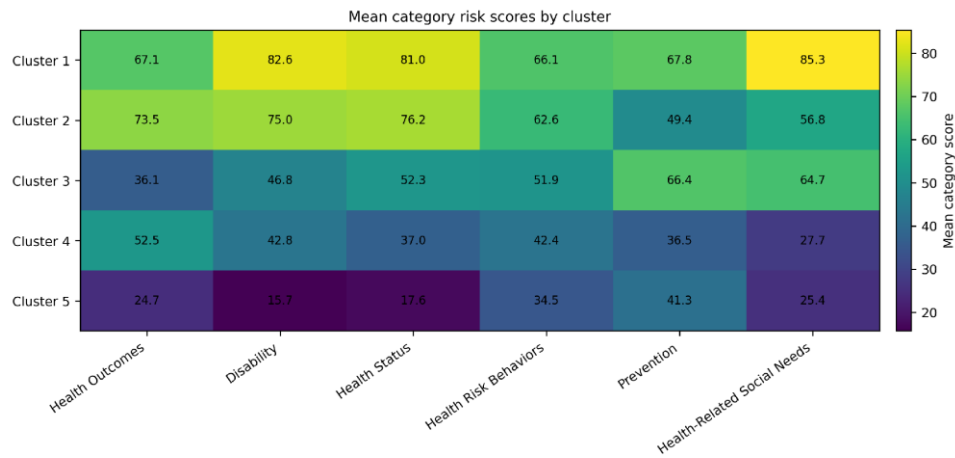


Figure 6. Cluster heatmap of mean category risk scores.

The five clusters were substantively different. Cluster 1 contained 459 counties and had the highest mean priority score, 73.22, with dominant health-related social needs, disability, and health-status distress. Cluster 2 contained 681 counties with high chronic disease, disability, health-status, and behavioral risk. Cluster 3 contained 532 counties with elevated preventive-care gaps and HRSN stress but lower outcome burden. Cluster 4 contained 748 moderate mixed-risk counties. Cluster 5 contained 724 lower-risk monitoring counties. These profiles support different outreach packages instead of a single uniform senior health intervention.

Table 11 addresses the concern that median imputation could distort cluster boundaries. KNN and MICE-style imputations changed some individual assignments, but ordered cluster mean scores remained highly correlated with the median-imputed solution. The complete-case analysis yielded an adjusted Rand index relative to the median-imputed solution on counties with all 40 measures. These results support the median-imputed matrix as a stable operational approximation while retaining the category-renormalized score as the primary ranking method.

Table 11. Cluster robustness under alternative missing-data strategies.

| Imputation strategy             | Adjusted Rand vs median K-means | Correlation of ordered cluster mean scores | Largest ordered cluster-size change | Top-quintile share range across ordered clusters |
|---------------------------------|---------------------------------|--------------------------------------------|-------------------------------------|--------------------------------------------------|
| Median imputation               | 1.000                           | 1.000                                      | 0                                   | 0.000-0.695                                      |
| KNN imputation (k=5)            | 0.570                           | 0.964                                      | 305                                 | 0.000-0.768                                      |
| MICE-style iterative imputation | 0.585                           | 0.967                                      | 280                                 | 0.000-0.786                                      |
| Complete-case analysis (n=2350) | 0.588                           | 0.968                                      | 103                                 | 0.000-0.822                                      |

### 4.3. Rule-recovery audit and external validation

The rule-recovery audit showed that the high-priority label was recoverable from the 40 risk-oriented indicators. In cross-validation, logistic regression produced the highest F1, followed by XGBoost and random forest, as shown in Table 12. These high values are expected because the target is derived from the same public health indicators that enter the score. They confirm that the decision rule is internally coherent and robust to model form, not that the models predict an independent clinical outcome.

Table 12. Five-fold cross-validation performance for high-priority rule recovery.

| Model               | Accuracy | Precision | Recall | F1    | ROC-AUC | PR-AUC | F1 SD |
|---------------------|----------|-----------|--------|-------|---------|--------|-------|
| Logistic regression | 0.992    | 0.981     | 0.978  | 0.979 | 1.000   | 0.998  | 0.009 |
| Decision tree       | 0.960    | 0.892     | 0.911  | 0.901 | 0.957   | 0.882  | 0.009 |
| Random forest       | 0.978    | 0.951     | 0.938  | 0.944 | 0.997   | 0.990  | 0.008 |
| XGBoost             | 0.981    | 0.952     | 0.951  | 0.951 | 0.998   | 0.992  | 0.008 |

Table 13 reports held-out 2024 test performance and confusion-matrix counts. Table 14 reports external validation on the 2025 GIS-friendly county release. Figure 7 displays the held-out ROC curves. For an outreach queue, logistic regression is attractive when transparency and recall are prioritized, while tree ensembles provide a useful feature-importance audit of the rule.

Table 13. Held-out 2024 test performance and confusion-matrix counts.

| Model               | Accuracy | Precision | Recall | F1    | ROC-AUC | PR-AUC | TN  | FP | FN | TP  |
|---------------------|----------|-----------|--------|-------|---------|--------|-----|----|----|-----|
| Logistic regression | 0.992    | 0.984     | 0.974  | 0.979 | 0.999   | 0.998  | 752 | 3  | 5  | 184 |
| Decision tree       | 0.968    | 0.925     | 0.915  | 0.920 | 0.976   | 0.899  | 741 | 14 | 16 | 173 |
| Random forest       | 0.981    | 0.952     | 0.952  | 0.952 | 0.998   | 0.991  | 746 | 9  | 9  | 180 |
| XGBoost             | 0.982    | 0.953     | 0.958  | 0.955 | 0.998   | 0.993  | 746 | 9  | 8  | 181 |

Table 14. External validation on the 2025 GIS-friendly county release.

| Model               | Counties | Accuracy | Precision | Recall | F1    | ROC-AUC | PR-AUC |
|---------------------|----------|----------|-----------|--------|-------|---------|--------|
| Logistic regression | 2,299    | 0.979    | 0.900     | 1.000  | 0.947 | 1.000   | 1.000  |
| Decision tree       | 2,299    | 0.968    | 0.918     | 0.910  | 0.914 | 0.967   | 0.885  |
| Random forest       | 2,299    | 0.976    | 0.902     | 0.979  | 0.939 | 0.998   | 0.993  |
| XGBoost             | 2,299    | 0.982    | 0.924     | 0.986  | 0.954 | 0.999   | 0.994  |

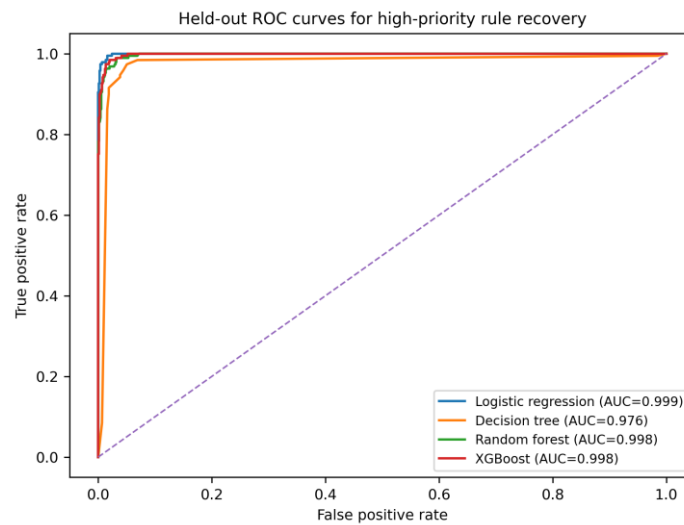


Figure 7. Held-out ROC curves for high-priority rule-recovery models.

Table 15 and Figure 8 show the leading feature importances. The leading variables indicate that disability and health-status distress are central to identifying counties where senior-serving outreach requires accessible delivery, telehealth setup, and care navigation. Future work could extend this audit with local explanation methods such as SHAP [19].

Table 15. Top feature importances for random forest and XGBoost.

| Model         | Measure ID | Measure                       | Category              | Importance |
|---------------|------------|-------------------------------|-----------------------|------------|
| Random forest | PHLTH      | Frequent Physical Distress    | Health Status         | 0.168      |
| Random forest | DISABILITY | Any Disability                | Disability            | 0.136      |
| Random forest | INDEPLIVE  | Independent Living Disability | Disability            | 0.117      |
| Random forest | MOBILITY   | Mobility Disability           | Disability            | 0.074      |
| Random forest | GHLTH      | General Health                | Health Status         | 0.068      |
| XGBoost       | DISABILITY | Any Disability                | Disability            | 0.194      |
| XGBoost       | PHLTH      | Frequent Physical Distress    | Health Status         | 0.189      |
| XGBoost       | INDEPLIVE  | Independent Living Disability | Disability            | 0.167      |
| XGBoost       | COGNITION  | Cognitive Disability          | Disability            | 0.065      |
| XGBoost       | LPA        | Physical Inactivity           | Health Risk Behaviors | 0.046      |

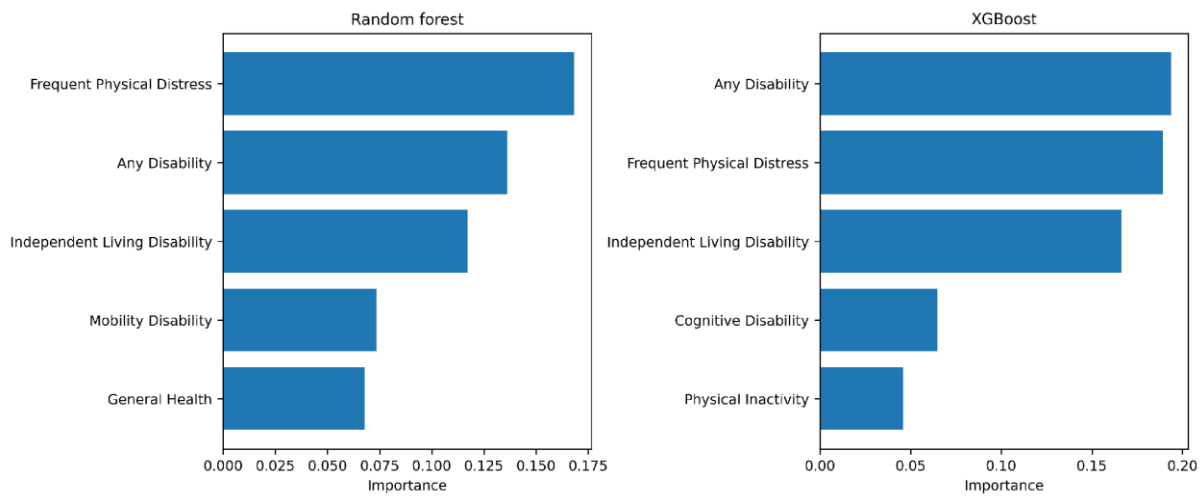


Figure 8. Model feature importances for the leading indicators of high-priority county status.

#### 4.4. Planning-layer checks, sensitivity, uncertainty, and optimization

The planning layer generated one intervention package for each empirical cluster. Table 16 reports the assigned intensity, dominant risks, consistency score, readability score, grade level, and word count. Consistency was high because the deterministic template was aligned with dominant empirical risks. The county-level Spearman correlation between priority score and assigned intervention intensity was 0.904, indicating that counties with higher risk scores received higher-intensity recommendation packages.

Table 16. Deterministic planning-layer evaluation by cluster.

| Cluster | Label                                                         | Intensity | Mean score | Dominant risks                                                 | Consistency | Flesch | Grade | Words |
|---------|---------------------------------------------------------------|-----------|------------|----------------------------------------------------------------|-------------|--------|-------|-------|
| 1       | Critical multimorbidity, disability, and social-need counties | 5         | 73.22      | Health-Related Social Needs; Disability; Health Status         | 0.667       | 65.01  | 4.89  | 14    |
| 2       | High chronic-disease and behavioral-risk counties             | 4         | 68.83      | Health Status; Disability; Health Outcomes                     | 0.667       | 46.17  | 7.69  | 14    |
| 3       | Preventive-care gap and access-stressed counties              | 3         | 48.01      | Prevention; Health-Related Social Needs; Health Status         | 0.667       | 79.89  | 2.92  | 13    |
| 4       | Moderate mixed-risk counties                                  | 2         | 43.52      | Health Outcomes; Disability; Health Risk Behaviors             | 1.000       | 41.64  | 8.20  | 12    |
| 5       | Lower-risk monitoring counties                                | 1         | 25.46      | Prevention; Health Risk Behaviors; Health-Related Social Needs | 1.000       | 50.23  | 6.94  | 11    |

The age-adjusted sensitivity check supported the stability of the risk ranking. Table 17 shows that crude and age-adjusted priority scores had a Spearman correlation of 0.982 across 3,144 county equivalents. The top-quintile overlap relative to the crude top quintile was 0.917. The crude score is appropriate for resource allocation because it reflects overall adult burden, while the age-adjusted score confirms that the main ranking is not an artifact of county age structure alone.

Table 17. Sensitivity of crude priority scoring against age-adjusted priority scoring.

| Comparison                                          | Counties | Spearman correlation | Top-quintile overlap relative to crude top quintile | Crude top-quintile counties | Age-adjusted top-quintile counties |
|-----------------------------------------------------|----------|----------------------|-----------------------------------------------------|-----------------------------|------------------------------------|
| Crude priority score vs age-adjusted priority score | 3,144    | 0.982                | 0.917                                               | 629                         | 629                                |

Table 18 and Figure 9 report the uncertainty analysis. Across 300 Monte Carlo draws using PLACES confidence limits, the simulated top-quintile threshold had a median of 68.07 and a 2.5th-97.5th percentile interval of 67.77-68.34. The deterministic threshold of 69.16 was above the simulated interval, so the simulation should not be interpreted as showing threshold-level agreement. Instead, it indicates that county band membership was relatively stable despite a downward shift in the simulated threshold distribution under the normal-approximation, truncation, and independence assumptions. Median county priority-band stability was 0.973, and the average stability of critical-band counties was 0.900. Among deterministic top-quintile counties, 548 retained top-quintile membership with probability at least 0.80.

Table 18. Priority-threshold uncertainty based on PLACES confidence limits.

| Deterministic top-quintile threshold | Monte Carlo draws | Median simulated top-quintile threshold | Threshold 2.5th percentile | Threshold 97.5th percentile | Median county priority-band stability | Critical-band stability mean | Top20 stable $p \geq 0.80$ |
|--------------------------------------|-------------------|-----------------------------------------|----------------------------|-----------------------------|---------------------------------------|------------------------------|----------------------------|
| 69.16                                | 300               | 68.07                                   | 67.77                      | 68.34                       | 0.973                                 | 0.900                        | 548                        |

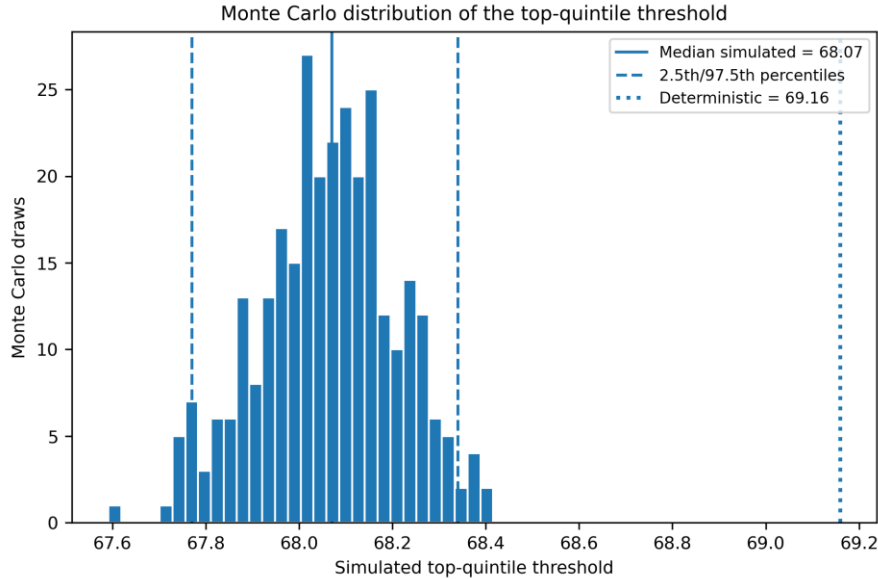


Figure 9. Monte Carlo distribution of the top-quintile threshold.

The optimization module demonstrates how the priority score can be converted into a resource-allocation decision. Candidate counties were limited to elevated, high, and critical priority bands. The objective weighted priority score by the estimated population aged 65 years and older. Under a 600-unit budget and a state cap of 25 counties, the mixed-integer optimization selected the county set summarized in Table 19. Table 20 shows the budget frontier, and Figure 10 plots the weighted senior-need units captured as the budget increases.

Table 19. Budget-constrained senior outreach allocation scenario.

| Budget units | Candidate counties | Selected counties | Total cost units used | Estimated older adults in selected counties | Mean selected priority score | Share critical/high priority | Optimization status |
|--------------|--------------------|-------------------|-----------------------|---------------------------------------------|------------------------------|------------------------------|---------------------|
| 600          | 1,258              | 158               | 600                   | 4,609,194                                   | 64.18                        | 0.241                        | optimal             |

Table 20. Budget frontier for the senior outreach optimization module.

| Budget units | Counties selected | Estimated older adults in selected counties | Weighted senior-need units captured |
|--------------|-------------------|---------------------------------------------|-------------------------------------|
| 100          | 27                | 2,232,009                                   | 1365471.36                          |
| 200          | 53                | 2,970,415                                   | 1817190.49                          |
| 300          | 81                | 3,499,029                                   | 2158180.35                          |
| 400          | 107               | 3,916,323                                   | 2437692.36                          |
| 500          | 133               | 4,287,277                                   | 2675884.11                          |
| 600          | 158               | 4,609,194                                   | 2890932.55                          |
| 800          | 207               | 5,156,841                                   | 3252813.83                          |
| 1,000        | 258               | 5,602,091                                   | 3552335.98                          |

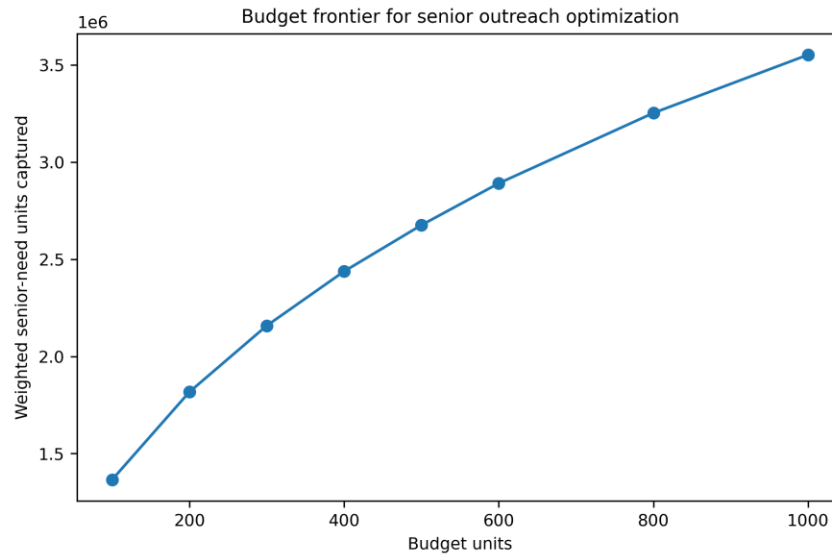


Figure 10. Budget frontier for weighted senior-need units captured.

## 5. Conclusions and future research

This study developed and empirically evaluated a data-driven decision-support framework for county-level senior health outreach. The framework converted local public health indicators into priority scores, five risk clusters, rule-recovery audits, cluster-specific intervention plans, threshold uncertainty estimates, and a budget-constrained senior allocation scenario. The 2024 county evaluation identified 315 critical-priority counties and a top-quintile score threshold of 69.16. The five-cluster solution separated counties with critical multimorbidity and social needs, high chronic and behavioral risk, preventive-care access gaps, moderate mixed risk, and lower-risk monitoring needs.

The results demonstrate that county-level PLACES indicators can support nonprofit and public health resource prioritization when they are transformed into explicit decision objects. The most important practical finding is that high-priority counties are not always high-population counties. Priority depends on the combination of chronic disease, disability, health status, behaviors, preventive-care gaps, and social needs. The cluster-based planning layer helps planners avoid a one-size-fits-all program by matching counties to outreach packages such as chronic disease education, telehealth and phone training, transportation support, care navigation, and baseline monitoring. The optimization module adds a managerial decision step by showing how a fixed budget can be allocated across eligible high-need counties while accounting for the number of older adults who could be reached.

The study has clear limits. PLACES measures are model-based local estimates, so they are suitable for planning and prioritization but not for measuring whether a local program caused a change in outcomes. The analysis used county-level adult indicators rather than direct health estimates for adults aged 65 years and older, although the

resource-allocation module incorporated county-level older-adult population counts. Health-related social-need coverage was incomplete in the 2024 county file, and the primary scoring method addressed this through category-weight renormalization rather than by inventing missing social-need values. The supervised high-priority target was derived from the priority score, so model performance evaluates internal reproducibility of the decision rule rather than independent disease prediction. The optimization cost unit is a normalized outreach-resource unit, not a dollar estimate; a program-specific implementation should replace it with local staff, travel, materials, and referral-navigation costs.

Future research should connect county scores with senior center locations, broadband access, public transit availability, local clinic capacity, and program cost. A prospective implementation study should test whether the priority score improves outreach uptake, screening completion, and referral success compared with population-only or first-come allocation. The planning layer should be evaluated by public health practitioners, nonprofit staff, and older adult advisory groups, then converted into resident-facing materials using formal plain-language review. Equity audits should test whether the scoring method systematically over- or under-prioritizes counties by rurality, race and ethnicity, disability, or socioeconomic status. These extensions would move the framework from reproducible decision support toward accountable community implementation.

### **Acknowledgments**

None.

### **Funding**

None.

### **Data Availability**

The public data files used in the analysis are available from the following sources.

CDC PLACES 2024 county Open Data release:

<https://data.cdc.gov/api/views/fu4u-a9bh/rows.csv?accessType=DOWNLOAD>

CDC PLACES 2025 county GIS-friendly release:

<https://data.cdc.gov/api/views/i46a-9kgh/rows.csv?accessType=DOWNLOAD>

U.S. Census Bureau 2024 county age-sex estimates:

<https://www2.census.gov/programs-surveys/popest/datasets/2020-2024/counties/asrh/cc-est2024-agesex-all.csv>

Census 2024 National Counties Gazetteer file:

[https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2024\\_Gazetteer/2024\\_Gaz\\_counties\\_national.zip](https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2024_Gazetteer/2024_Gaz_counties_national.zip)

### **Conflict of interest**

There is no conflict of interest to disclose.

### **Author contributions**

This is a single-author paper. The author is responsible for original article writing.

### **Declaration of using AI tools**

The author used AI-assisted language editing and structured drafting support during manuscript preparation.

### **References**

- [1] Centers for Disease Control and Prevention, "PLACES: Local Data for Better Health," 2024. [Online]. Available: <https://www.cdc.gov/places/>
- [2] Centers for Disease Control and Prevention, "Methodology: PLACES," 2025. [Online]. Available: <https://www.cdc.gov/places/methodology/>
- [3] Centers for Disease Control and Prevention, "PLACES Data Portal," 2024. [Online]. Available: <https://www.cdc.gov/places/tools/data-portal.html>
- [4] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, "PLACES: Local Data for Better Health, County Data 2024 release," Data.gov, 2024. [Online]. Available: <https://catalog.data.gov/dataset/places-local-data-for-better-health-county-data-2024-release-cb149>
- [5] Centers for Disease Control and Prevention, "PLACES County Data GIS Friendly Format, 2025 release," 2025. [Online]. Available: <https://data.cdc.gov/500-Cities-Places/PLACES-County-Data-GIS-Friendly-Format-2025-releas/i46a-9kgh>
- [6] K. J. Greenlund, H. Lu, Y. Wang, K. A. Matthews, J. M. LeClerc, B. Lee, et al., "PLACES: Local data for better health," *Prev. Chronic Dis.*, vol. 19, p. E31, Jun. 16 2022.

- [7] P. L. Remington, B. B. Catlin, and K. P. Gennuso, "The County Health Rankings: Rationale and methods," *Popul. Health Metr.*, vol. 13, no. 11, 2015.
- [8] C. M. Hood, K. P. Gennuso, G. R. Swain, and B. B. Catlin, "County Health Rankings: Relationships between determinant factors and health outcomes," *Am. J. Prev. Med.*, vol. 50, no. 2, pp. 129–135, 2016.
- [9] Office of Disease Prevention and Health Promotion, "Social determinants of health," *Healthy People 2030*. [Online]. Available: <https://health.gov/healthypeople/priority-areas/social-determinants-health>
- [10] Centers for Disease Control and Prevention, "The CDC Clear Communication Index," 2024. [Online]. Available: <https://www.cdc.gov/ccindex/>
- [11] M. Ramezani, et al., "Artificial intelligence and machine learning in health policy: A bibliometric analysis," *BMC Health Serv. Res.*, vol. 23, art. 1416, 2023.
- [12] T. Y. C. Tam, S. Sivarajkumar, S. Kapoor, A. V. Stolyar, K. Polanska, K. R. McCarthy, et al., "A framework for human evaluation of large language models in healthcare derived from literature review," *NPJ Digit. Med.*, vol. 7, no. 1, art. 258, 2024.
- [13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [14] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [15] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982., doi:
- [16] J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.
- [17] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?" *J. Classif.*, vol. 31, no. 3, pp. 274–295, 2014.
- [18] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [20] R. Flesch, "A new readability yardstick," *J. Appl. Psychol.*, vol. 32, no. 3, pp. 221–233, 1948.
- [21] J. P. Kincaid, R. P. Fishburne, L. Rogers, and B. Chissom, "Derivation of new readability formulas for Navy enlisted personnel," *Naval Technical Training Command, Research Branch Report 8-75*, 1975.
- [22] G. H. McLaughlin, "SMOG grading: A new readability formula," *J. Read.*, vol. 12, no. 8, pp. 639–646, 1969.
- [23] U.S. Census Bureau, "County Population by Characteristics: 2020–2024," 2025. [Online]. Available: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-detail.html>
- [24] U.S. Census Bureau, "2024 Gazetteer Files," 2025. [Online]. Available: <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html>
- [25] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, 2011.
- [26] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, et al., "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, Jun. 2001.
- [27] G. B. Dantzig, *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press, 1963.
- [28] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al.; SciPy 1.0 Contributors, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, 2020.



All open access articles published in *Transactions on Computational Modeling and Intelligent Systems* (<http://tcmis.org>) are distributed under the terms of the CC BY-NC 4.0 license (Creative Commons Attribution Non-Commercial 4.0 International Public License as currently displayed at <http://creativecommons.org/licenses/by-nc/4.0/legalcode>) which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work is properly cited.