Defining Statistically Significant Spatial Clusters of a Target Population using a Patient-Centered Approach within a GIS

Efforts to Improve Quality of Care

Stephen Jones, PhD

Bio-statistical Research Scientist

BlueCross BlueShield of Tennessee

Background – Issue #1

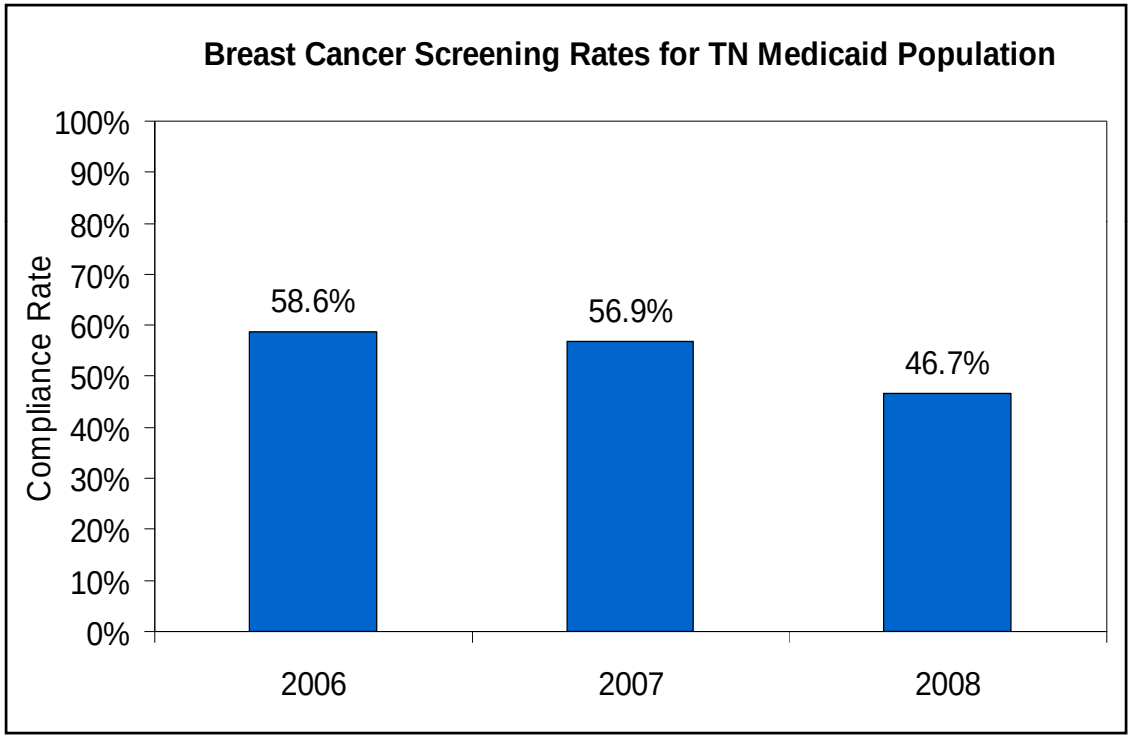
- Determining why someone is/is not compliant with evidence-based guidelines is complicated
- Mammogram screening rates are influenced by multiple factors:
  - Age, race, ethnicity, income status
  - Continuity of care with PCP
  - Logistic inconveniences
  - Variability in physicians and facilities
  - Positive views about initial screening
  - Practice of other preventive health behaviors
  - Knowledge of breast cancer and screening
  - Societal, familial and healthcare related influences

Background – Issue #2

- Traditional use of a GIS in outreach efforts involve isolating zip codes or even counties where the target population resides
  - Much information is lost using this granular approach; People's place of residence do not typically conform to zip code or county boundaries
  - Outreach efforts should be focused on where people actually live, not an arbitrary geographic entity

Breast Cancer Screening Rates over Time

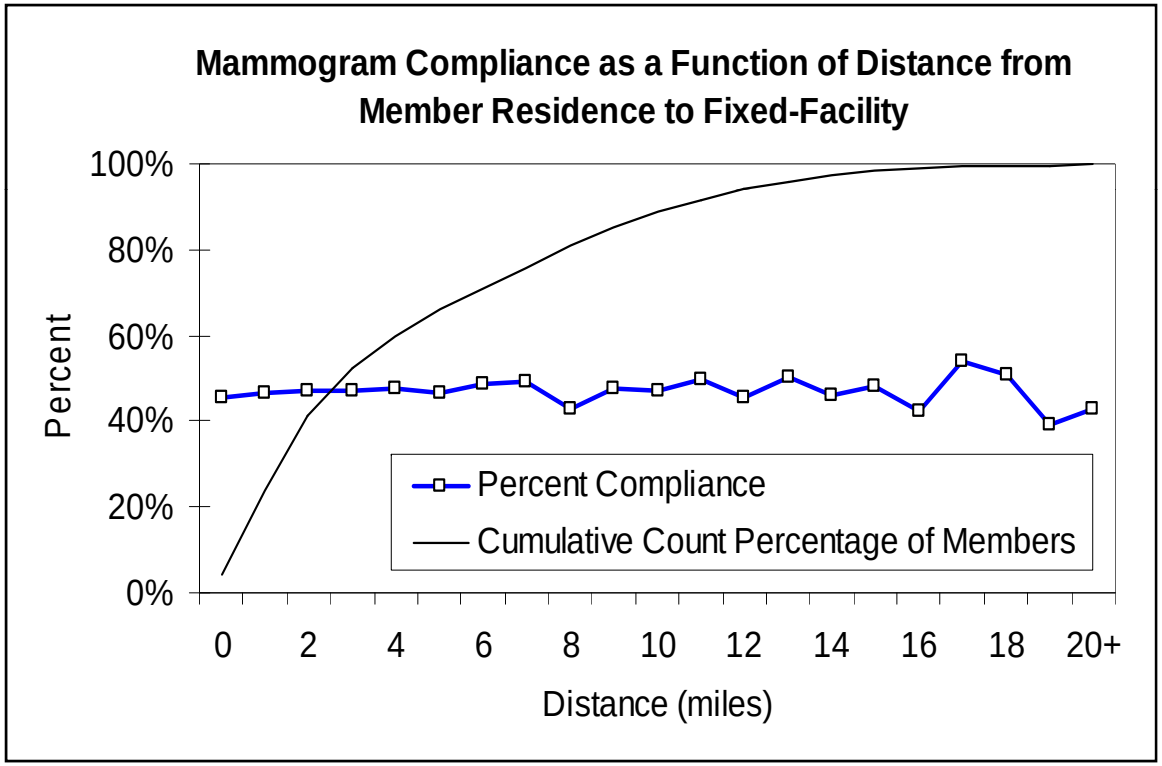
- Decline in breast cancer screening rates over time
- Multiple intervention programs have been implemented



of Tennessee

Compliance as a Function of Distance to Facility

- Compliance rates not influenced by distance to a fixed-facility
- 98% of the study population lived within 15 miles of a fixed facility
- Average distance from home to the nearest facility was not different (p=.32) for compliant members versus non-compliant members



of Tennessee

Study Objectives

- To develop a universal methodology utilizing data mining and spatial analytics to strategically target intervention efforts at the member level to improve compliance rates with evidence-based guidelines
- Specifically for this pilot, we wanted to:
  1. Build a predictive model to determine which members have a low probability of being compliant with their next mammogram screen
  2. Identify statistically significant spatial clusters of this target population

Data Mining Details

- Data mining enables the researcher to statistically model large amounts of data with many modeling techniques
- Increases efficiency
    - alleviates manual programming
    - build and compare multiple models
- Increases accuracy
    - advanced imputation, transformation, and standardization techniques available

Common Modeling Techniques

- Logistic Regression – predict the probability that a binary target (compliant vs. not compliant) will acquire the event of interest as a function of one or more independent covariates

- Decision Tree – segments data by applying a series of simple rules to make a “decision”
  - Useful when you have “missing” data

Data Details

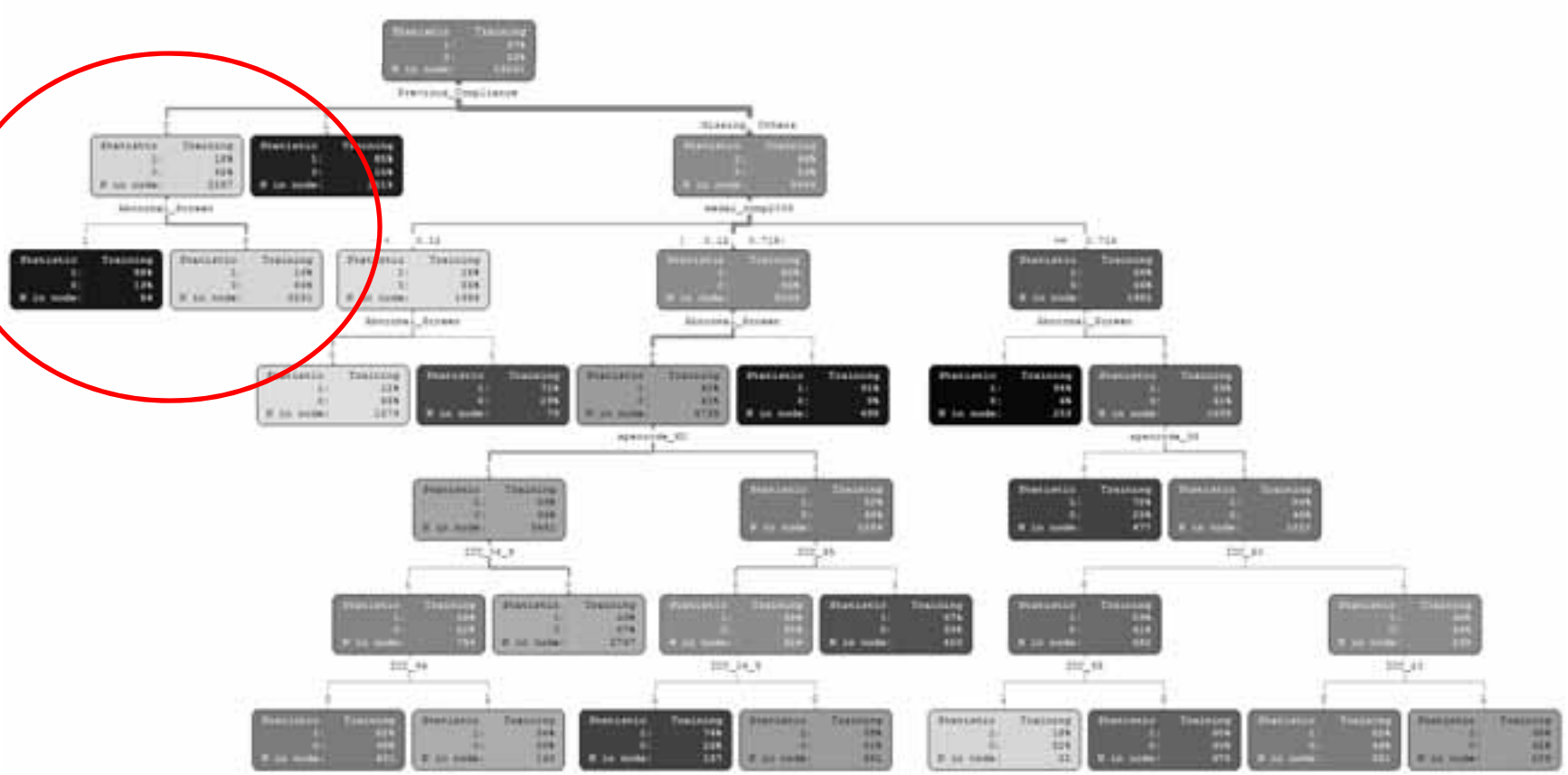
- Target variable:
  - Compliance with most recent mammography screening
- Explanatory Variables:
  - Claims Based
    - Clinical claims history
    - Demographic
    - Compliance with other evidence-based guidelines
  - Geo-Spatial
    - Distance to facility
    - Surrounding geographical information

Results

- 22,737 members (47% compliant)
  - 15% African-American (43% compliant)
  - 0.7% Hispanic (50% compliant)
  - 72% White (47% compliant)
  - 10% Unknown race (49% compliant)

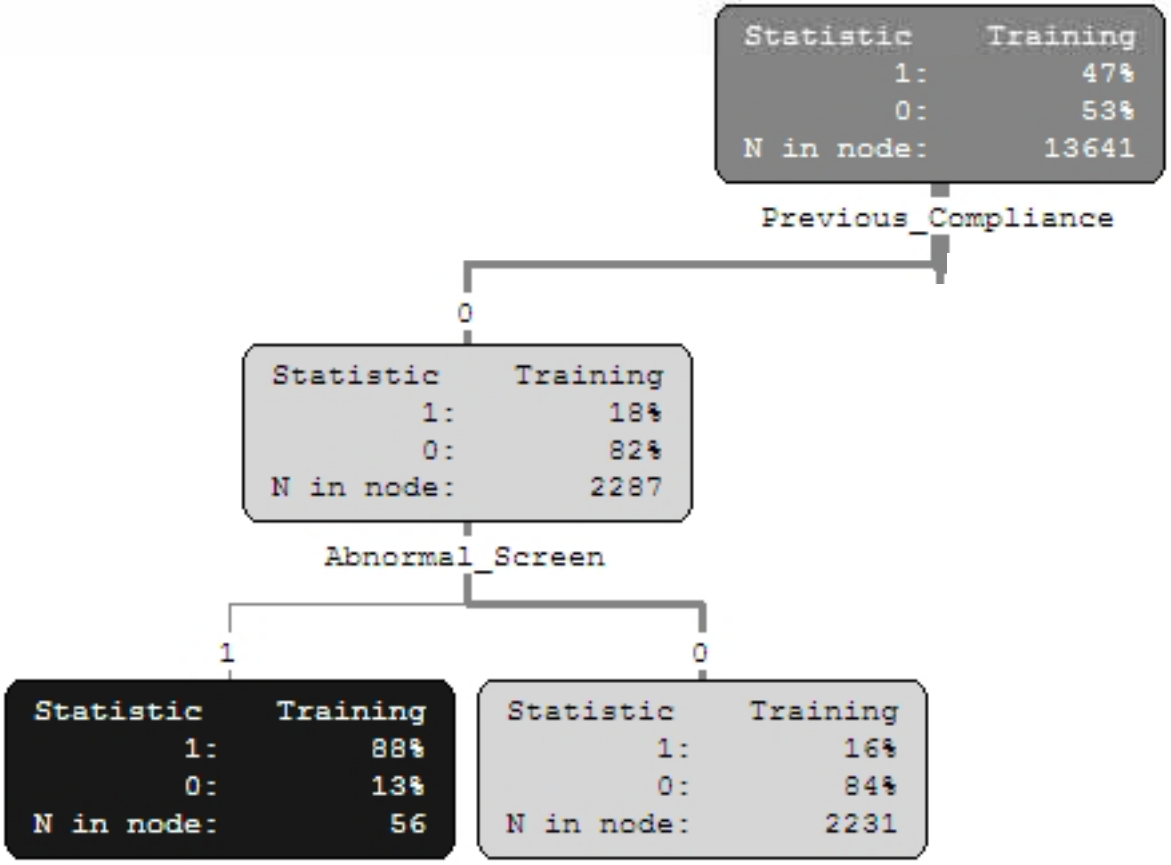
- Decision Tree selected as best model
  - ROC value = 0.81 (model fit measure ranges from 0.5 – 1.0 where 1.0 is best)
  - Misclassification rate = 25.6% across entire population

Results - Decision Tree Model



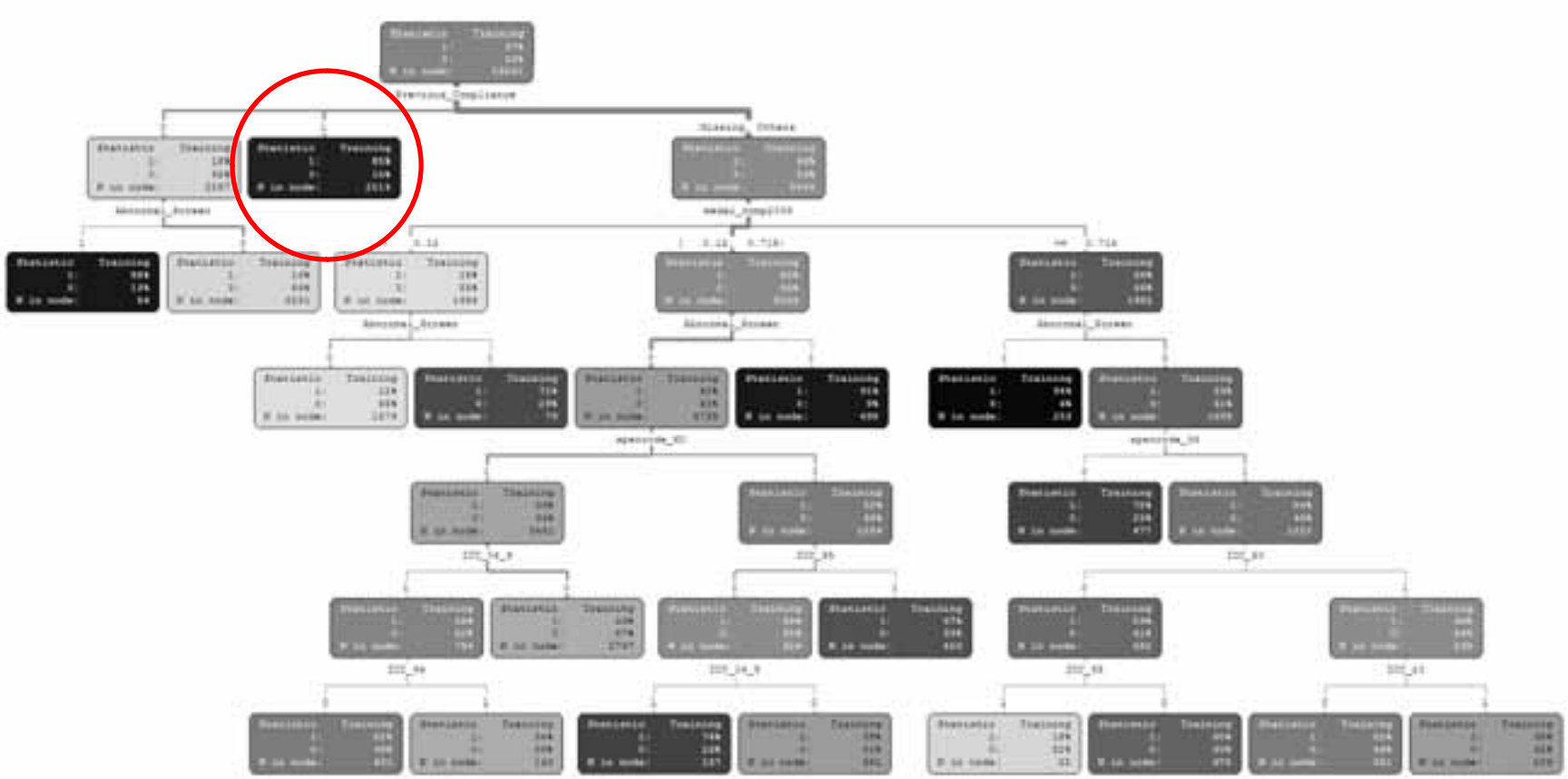
of Tennessee

Previously Non-Compliant



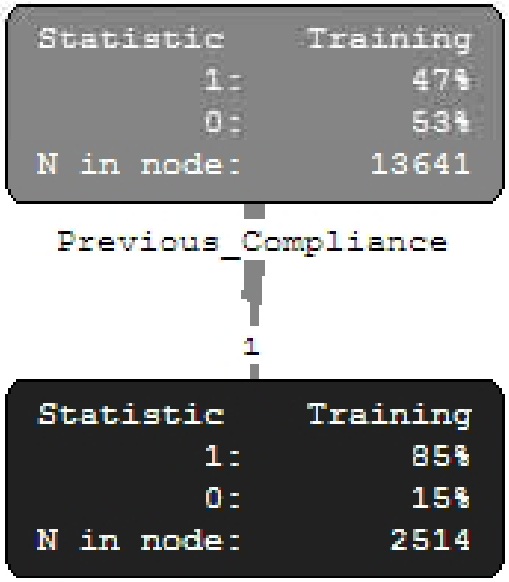
of Tennessee

Decision Tree - Compliant with prior screening

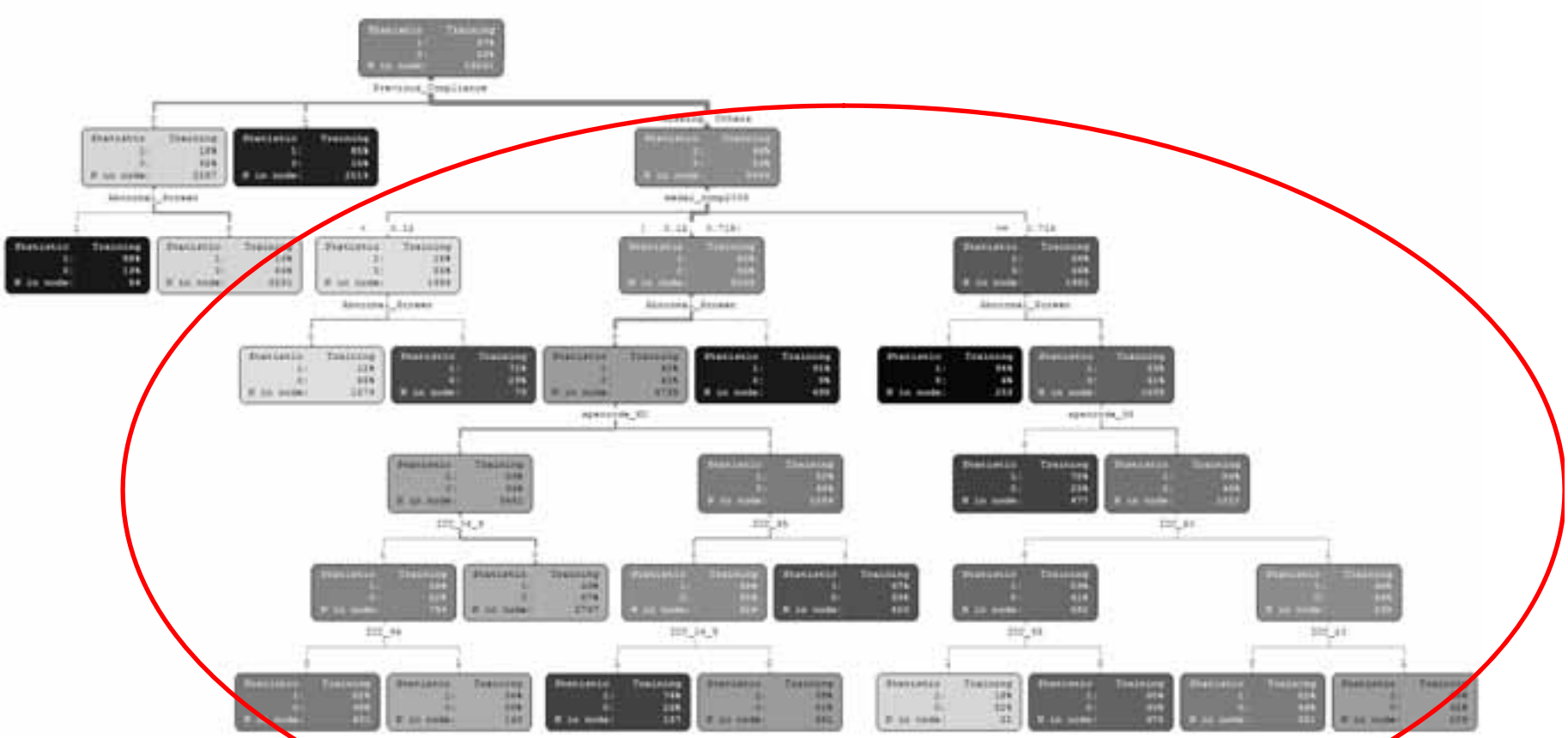


of Tennessee

Previously Compliant

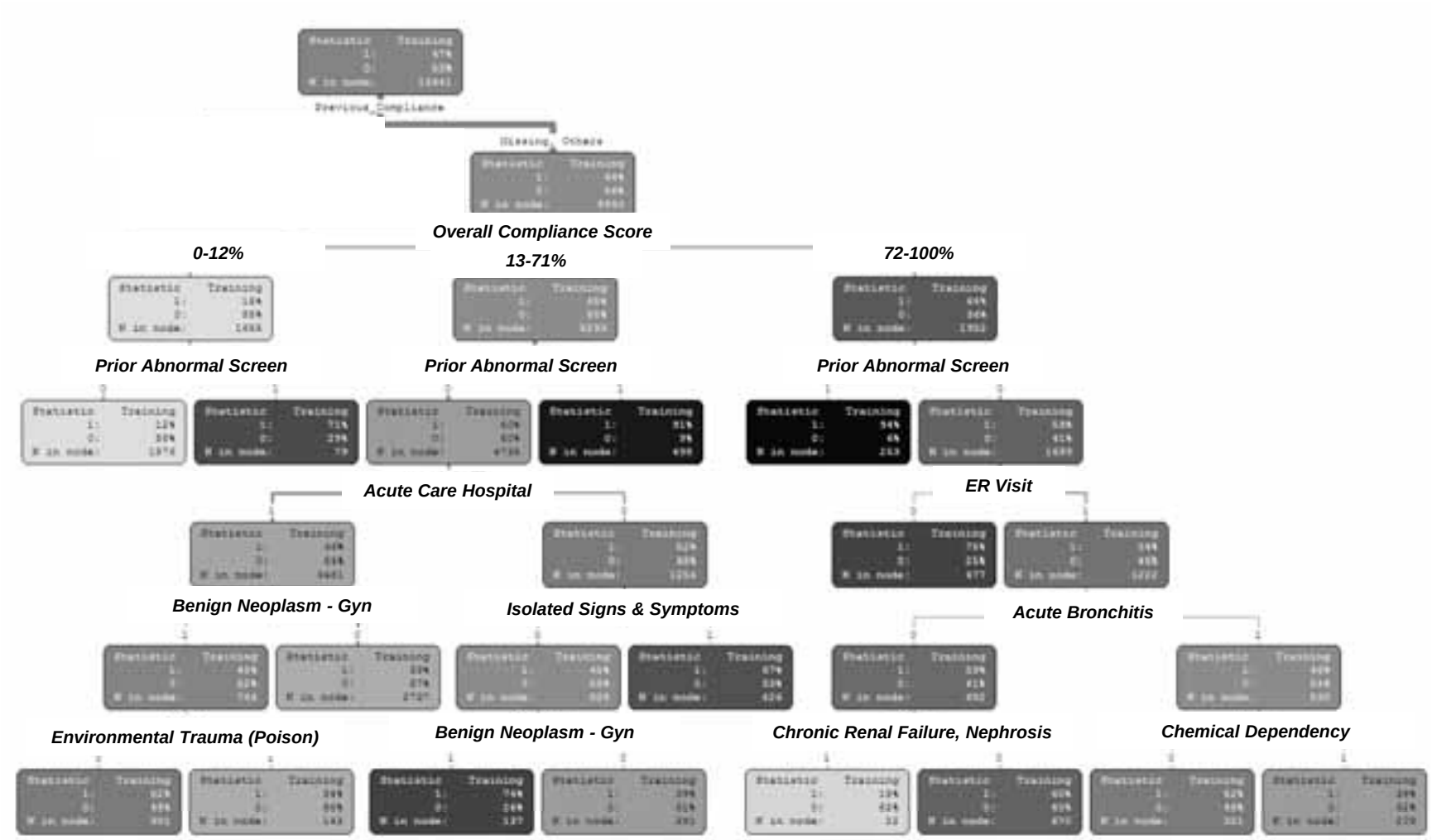


Decision Tree - No prior compliance history



of Tennessee

No Previous Compliance Data



of Tennessee

Deploying the Model

Predictive model provides 2 actionable items:

1. Outreach to the following people:
   - Members previously non-compliant
   - Members without a prior abnormal screen
   - Members non-compliant with other evidence-based guidelines
2. Select members with a low predicted probability to be compliant (e.g. those with scores 0% - 20%)

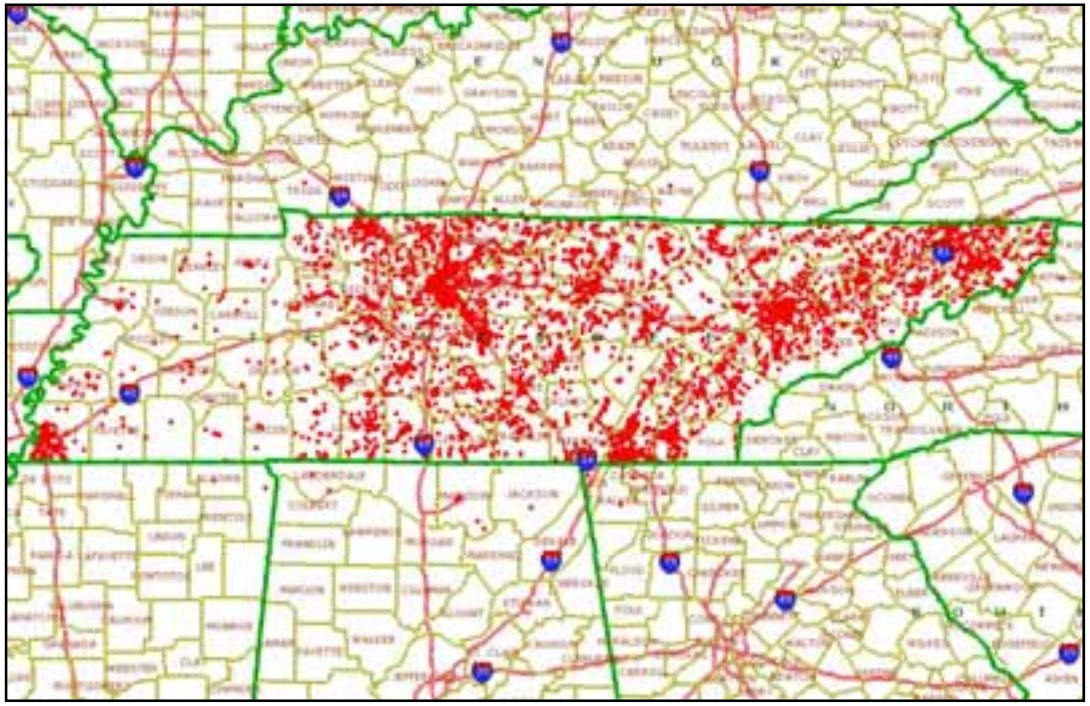
- Using this 20% threshold:
  - 28% (6,326) of the population predicted to be non-compliant
  - The model was 85% accurate in predicting a member's non-compliance

Deploying the Model – Using a GIS

- Using a Geographic Information System (GIS), locate members with a low probability of obtaining a mammography (n=6,326)

- Patient locations (latitude, longitude) were geocoded using residential street address and zip code information

Members with a low probability of obtaining a mammogram



of Tennessee

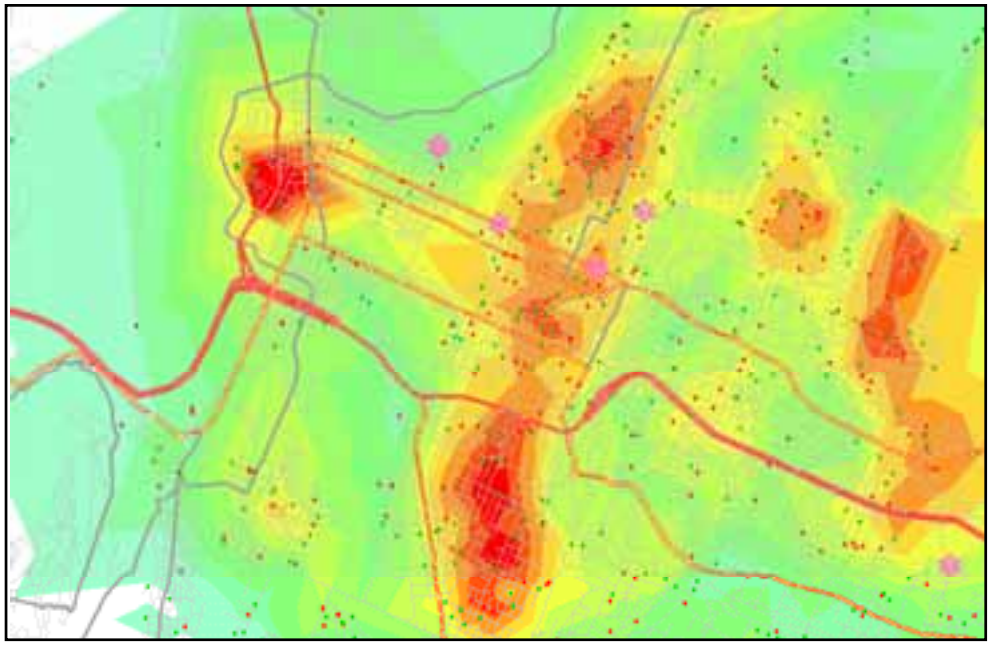
Spatial Clustering Tool

- Developed a spatial clustering tool that allows user to simultaneously cluster using:
  1. Any *n* number of variables and selection sets
  2. Counts or statistical metrics (sum, avg)
  3. Any aggregation distance
  4. Radial or drive-time polygon aggregations
  5. Differential weighting by variable
  6. Other predictive models as variables

Neighborhood Clustering

- Using a patient-centered approach, create surface contour maps (isopleth maps) that graphically detail higher densities (i.e. neighborhood clusters, or hotspots) of members with a low probability of obtaining a mammogram

Example of Patient-Centered hotspot clustering for members with a low probability of obtaining a mammogram (red dots)



of Tennessee

Statistical Testing of Spatial Clusters

- Incorporated the ability to test the clusters for statistical significance using the Getis-Ord local G-statistic
- Nearest neighbor analysis for spatially oriented data
- H_0 = no association exists between values found at one location versus neighboring points within a specified distance

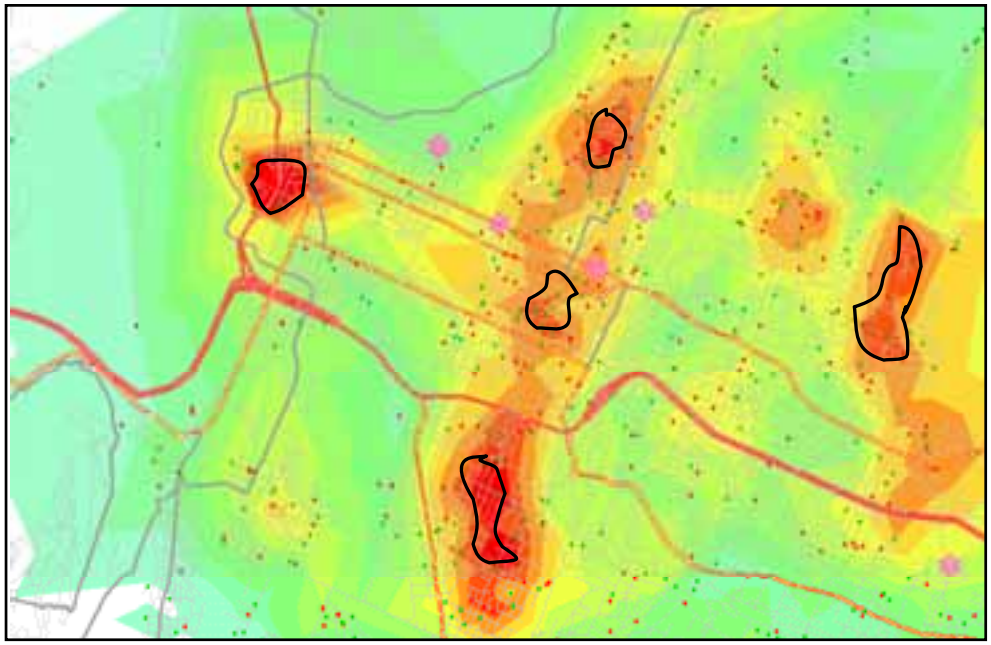
$$G_i^*(d) = \frac{\sum_{j=1}^{N} w_{ij}(d)x_j - \bar{x}\sum_{j=1}^{N} w_{ij}(d)}{S\sqrt{\frac{\left[N\sum_{j=1}^{N} w_{ij}^2(d) - \left(\sum_{j=1}^{N} w_{ij}(d)\right)^2\right]}{N-1}}}$$

- A significant spatial cluster is an isolated area where data points surrounding it are statistically different than the cluster itself

Statistically Testing of Clusters

- Retained only the members living within significant neighborhood clusters
- Reduced the number of members targeted for intervention from 6,326 to 1,018

Example of applying local G-statistic to hotspot clusters to determine statistically significant clusters (black outlines) of non-compliant members



of Tennessee

Conclusions

- Implementing data mining techniques and spatial analytics may significantly improve efficiency of outreach efforts by strategically targeting members
- Patient-centered clustering methods more clearly defined cluster boundaries and target areas versus traditional methods of using arbitrary geographic entities (i.e. counties, zip codes, census tracts, etc.)
- Techniques are transferable to other quality measures and healthcare research, including optimizing facility planning, increasing survey rates, accessing physician network adequacy, etc.