Background

The fifteen wholly-owned health plans under WellPoint, Inc. (WellPoint) historically did not collect data in regard to the race/ethnicity of it members. In order to overcome this lack of data and meet a number of health plan quality improvement, operations and business planning needs, the Health Disparities Analytic Unit at WellPoint has developed a methodology to derive indirect race and ethnicity information for members. While WellPoint also has a long-term plan to move towards the collection of member self-reported race/ethnicity data on its over 35 million insured members, current indirect data methodologies offer a relatively inexpensive, yet reliably accurate strategy to bridge the data collection gap until primary source data is available for all members.

Traditionally, marketers have relied on surnames or geographically linked census data when attempting to predict race/ethnicity. Some of the first efforts to combine these two data sources were led by the RAND Corporation (RAND), which used a combination of absolute matches against surname lists and geocoded data linked to census blocks, then applied Bayesian statistical methods to determine the probability of each individual being Asian, African American, Hispanic, or White / other. In late 2006, WellPoint engaged RAND to access RAND's indirect methodologies to form an internal, dedicated Disparities Analytic Unit at WellPoint to support the quality and business needs of its plans.

Throughout 2007, leveraging the preliminary work done by RAND, the Health Disparities Analytic Unit continued independent research and development on indirect methodologies to develop its own model that combines available name and geographic Census data using a series of logistic regressions. This new logistic method eliminated the use of absolute surname matches and the calculation of specificities and sensitivities. In addition, updated surname lists and a proprietary African American first name list developed by the Health Disparities Analytic Unit were included in the model. Finally, logistic regressions replaced the use of Bayesian methods to combine surname and census data.

INTERVENTION AND IMPLEMENTATION

WellPoint operates a small but dedicated the Health Disparities Analytic Unit that conducts continuous refinement of its indirect data methodology and health disparities analyses. Using a combination of geocoding, name analyses, and logistic regressions, the analytic unit provides estimation of member race/ethnicity or preferred language needs based on membership demographics. The resulting indirect data estimates also allow WellPoint plans to examine differences between racial / ethnic groups in various health indicators, such as diabetes, colorectal and mammography screening rates by linking the proxy race/ethnicity data with member claims data and quality process measures. Along with the traditional graphs and charts, GIS tools have allowed further detailed examination of screening rates through mapping.

Application of Indirect Race/ Ethnicity Data in Quality Metric Analyses

Health disparities analyses may be performed by applying indirect race/ethnicity data to a clinical study factor. The most commonly requested study measures at WellPoint are clinical quality progress measurements such as the major Health Plan Employer Data and Information Set (HEDIS®) screening and process measures.

Major disadvantages of collecting and using primary source/member self-reported data are that race /ethnicity data is available for only a subset of the health plan's total insured population, and member self-selection in reporting the data that may impact and distort actual utilization patterns during summary-level comparisons of the clinical data.

Indirect race/ethnicity data estimates, coupled with study metrics, offer comprehensive performance comparisons on quality metrics at summary levels to aid in health plan business planning and decisions. The resulting output allows the identification of disparities among the various groups. Comparisons are most successful when complete data is used. To date, WellPoint has utilized summary reports that study and aggregate health disparities across multiple levels, including: state, regional, county, zip code, medical group/IPA, and provider practice levels for the following HEDIS® and other quality measure indicators.

Application of GIS mapping tools

As a positive side effect from the geocoding process completed as a part of the logistic model of estimating race/ethnicity, proxy race/ethnicity and quality metric data can easily be loaded into GIS software to produce maps illustrating items such as health disparities hotspots, network access needs, member dispersion, provider locations, and minority locations. These maps are visually impactful and may be used for business decision support in designing quality improvement programs, provider network management or contracting initiatives, and/or community collaboration projects with external health care stakeholders. (See Appendix – Case Study: Application of Indirect Race/ethnicity Data to Support Quality Improvement and Network Management Initiatives)

In Ohio, indirect race/ethnicity data was linked with mammography screening HEDIS data to identify geographic regions with high health disparities for increased reminder mailings. (Appendix – Sample Map A)

Indirect race/ethnicity data and subsequent health disparities analyses may also be used to support provider network management activities, such as contracting/service access initiatives, physician cultural competency training outreach, and/or physician-targeted quality improvement outreach initiatives. In Georgia, mapping of Hispanic diabetic members using indirect race/ethnicity data is used to identify potential addition of rural telemedicine facilities/locations that would offer members the ability to consult with Spanish-speaking/bi-cultural diabetes health educators from large metropolitan areas. (Appendix – Sample Map B)

In California, WellPoint has used the demographic overlay/screening density gradient maps with additional provider network specific data to meet its network management quality improvement and contract planning needs. To determine whether health care disparities might be a contributing factor in low HEDIS® performance in California's Inland Empire (San Bernardino and Riverside Counties), the Health Disparities Analytic Unit used indirect logistics modeling to apply racial/ethnic data to study geographic variation in the adherence to quality indicators. Member demographics and related quality process indicators were extracted based on 2006 HEDIS® guidelines for colorectal cancer screening (n=160,203), pharyngitis screening (n=10, 863), and the four diabetes-related HEDIS® screenings (n=40,375). (Appendix – Sample Map C)

EVALUATION METHODS

Unlike "intervention-based" programs, the logistic model of indirect race/ethnicity data methodology does not have quantifiable inherent performance measures that might be were used to evaluate overall success of the program. As a process/administrative best practice, however, the program's success may be assessed several qualitative criteria: the reliability of the data generated by the model, the impact of the initiative and the potential for replication by others.

The sensitivity, specificity and positive predictive value validation tests serve as a fair measure of the effectiveness and reliability of the logistic model of indirect race/ ethnicity data methodology as a diagnostic tool. Internal testing of WellPoint's current methodology has yielded excellent results to date. Model development and testing utilized a dataset from WellPoint's California and Connecticut State Sponsored business that contains race/ethnicity data on approximately 1.3 million members. Prior to running the regressions, 192,096 member records were extracted from the model development database and set aside to be used for validation testing. The resulting model predicted within 0.1% of the actual aggregate demographics for the validation sample (Chart 1). This is a marked improvement over the more traditional marketing approaches of using surname lists or census data separately.

When using the logistic model to predict the race/ethnicity of individual members, the model predicted the correct race/ethnicity for 86.2% of all members. Within ethnic minorities, however, indirect methodologies have proven to be highly reliable. The positive predictive value (PPV), or precision rate, is the proportion of people with specified demographic trait who are correctly diagnosed, and is the most important measure of a diagnostic method as it reflects the probability that a positive identification reflects the underlying trait that it is seeking. The logistic model's PPV for all racial / ethnic categories are over 90%, and in all categories except for Asians, the PPV is higher than those of more traditional indirect methods (Chart 2).

IMPACT OF INITIATIVE

As illustrated under the "Intervention" section, the logistic model of indirect race/ethnicity data methodology has proven itself to be flexible and applicable to address a wide range of project and business support needs. At WellPoint, indirect methodology and maps have been used to identify hotspots of unscreened members, study provider access for minority members, identify minority members for culturally/linguistically appropriate health screening reminders and health education materials, and determine member threshold language needs to meet regulatory requirements

Reports and maps generated using the indirect race/ethnicity data process have also allowed WellPoint to collaborate with its network medical groups, external health advocacy organization like the American Cancer Society, and elective government officials to encourage open dialogues on the issue of health disparities and jointly work towards collaborative quality initiatives to reduce the gap in care. It is our hope that expansion of such transparency and engagement tools will improve the health of our communities we serve and the quality of care overall.

APPENDIX

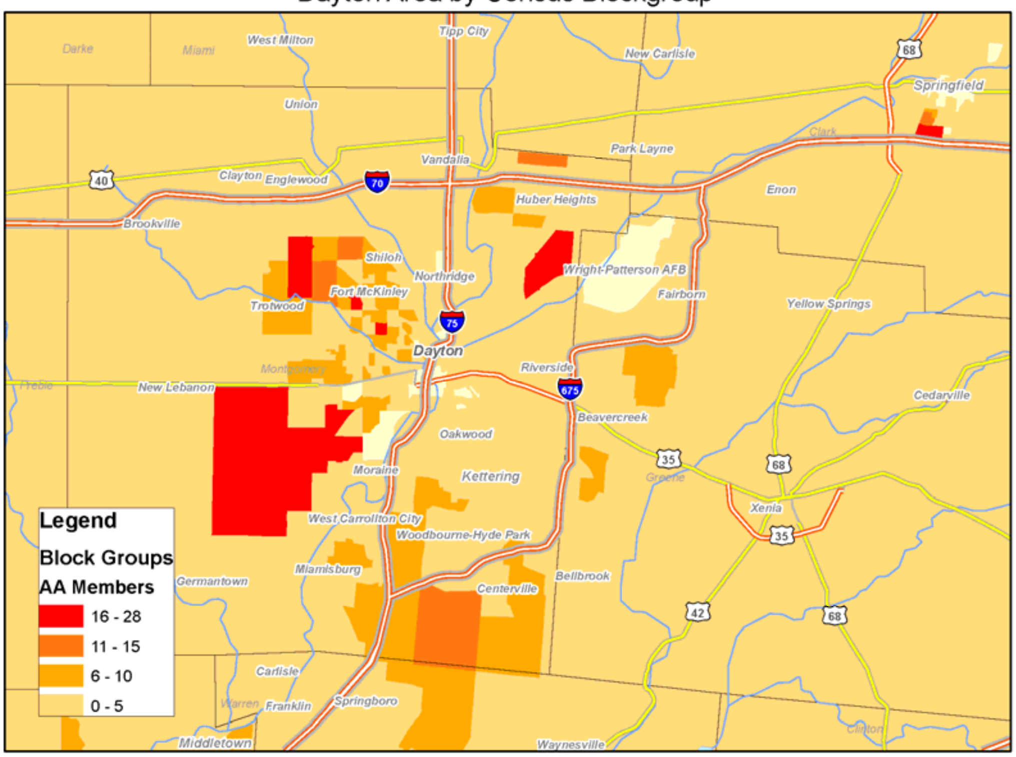
Case Study: Application of Indirect Race/ethnicity Data to Support Quality Improvement and Network Management Initiatives

More sophisticated maps could provide an overlay of zip codes with certain demographic characteristics (such as race, ethnicity, income or education level) with a density gradient map of members who have received a particular HEDIS® recommended screening.

Maps like sample Map A, titled "Eligible African American Members Not Receiving Mammography Screening Dayton Area by Census Blockgroup" were used in WellPoint's Ohio health plan to increase mammography screening rates among African American women. Indirect race/ethnicity data was linked with mammography screening HEDIS data to identify geographic regions with health disparities for increased reminder mailings.

Sample Map A

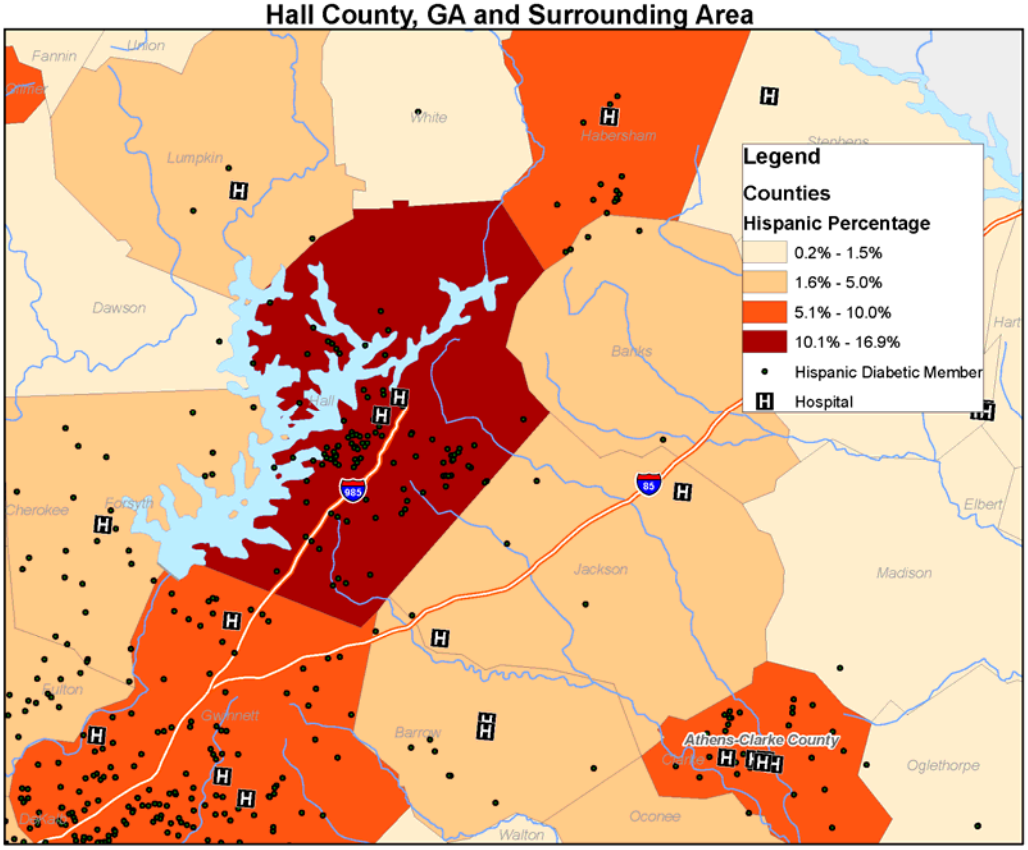
Eligible African American Members not Receiving Mammography Screening
Dayton Area by Census Blockgroup



In WellPoint's Georgia plan, maps such as sample Map B, titled "Hispanic Diabetic Members in Hall County, GA and Surrounding Area" were used for network access and provider contract planning. Hispanic diabetic members were identified and mapped using indirect race/ethnicity data. Hospital location data is layered onto the map to identify potential rural telemedicine sites at hospital locations that would offer rural Hispanic diabetic members the ability to consult with Spanish-speaking/bi-cultural diabetes health educators from large metropolitan areas.

Sample Map B

Hispanic Diabetic Members
Hall County, GA and Surrounding Area



Sample map C, titled, "Colorectal Screening Rate for Los Angeles" overlays the colorectal screening gradient density map with zip codes with Hispanic population majorities based on the 2000 U.S. Census. It is interesting to note that the more affluent coastal regions and areas like Beverly Hills show some of the highest screening rates, while many zip codes with Hispanic population majorities fall in areas with the lowest colorectal screening rates. At Anthem Blue Cross, these areas are called "disparities hot spots."

Sample Map C

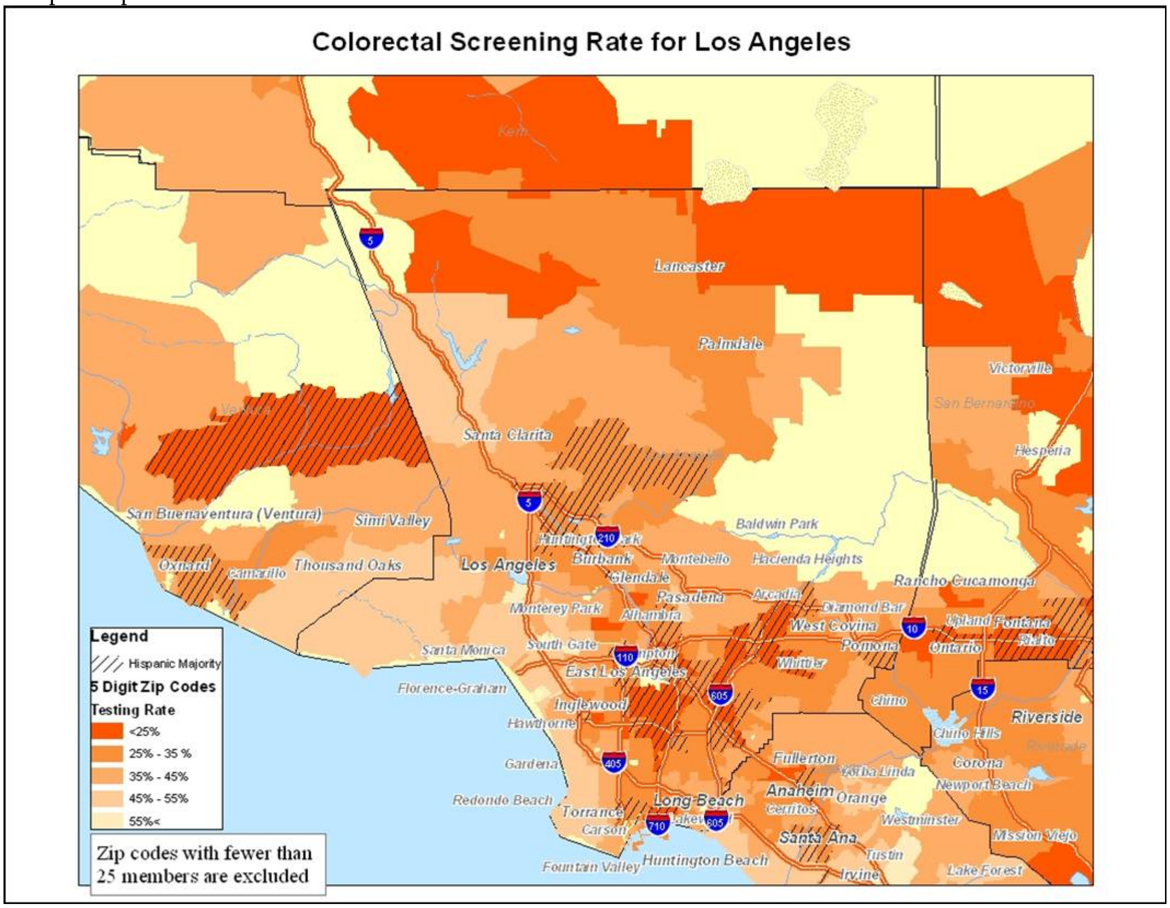


Chart 1: Aggregate Demographics – Predicted vs. Reported

Approach N = 192,096	Hispanic (%)	Asian (%)	Black (%)	White / Other (%)
Surname	46.1	6.6	7.1	40.2
Geocoding	41.3	7.9	11.9	39.0
Logistic Model	52.1	7.9	14.7	25.3
Self – report	52.0	8.0	14.8	25.2

Chart 2: Individual Predictions – Positive Predictive Value

Approach N = 192,096	Hispanic Positive Predictive Value	Asian Positive Predictive Value	Black Positive Predictive Value	White / Other Positive Predictive Value
Surname	94.9	96.0	85.5	76.8
Geocoding	93.3	87.1	83.9	76.8
Logistic Model	95.3	94.8	90.3	90.2