ALGOREX
HEALTHCARE TECHNOLOGIES

Take-Home Data Challenge

Medicare Food Access Program Report

Nour Elkhalawy

Introduction

A national Medicare Advantage plan is seeking to address the food access challenges for its members. So, the focus of this report is to find the food access challenges that face the members of the plan, the locations where these challenges are most significant and need to be addressed first, the population segments of the members who will benefit the most from addressing the food access challenges in the chosen locations, and the effect of the food access challenges reduction on the cost of the Medicare services in these locations.

Data

The data chosen for solving this challenge are:

1. CMS – State/County Medicare Utilization Summary: This dataset represents the members of the Medicare Advantage plan and the costs of the services they use across the US counties.

2. FDA – Food Atlas: This dataset represents the food access challenges across the US counties.

There are some important decisions and assumptions made about the datasets:

1. The CDC dataset was dropped because its locations didn't match the locations inferred from the merged two datasets CMS and FDA. More about that in Question 2 of the Results section.

2. The dates in the FDA datasets were ignored since sometimes a variable from 2010 was used because it was the most recent in the dataset and another variable was used from 2015 for the same reason. Note that the variables from 2010 and 2015 correlated well with correlation coefficients ~0.98.

3. None of the columns for the different variable dates in the FDA dataset were dropped, but the recent ones were used when shown in the results.

Methods

First, both the CMS and the FDA datasets were cleaned and their FIPS code was matched in type and number of digits.

Finding a Connection Between Datasets. In order to identify what are the food access challenges that face the members of the Medicare Advantage plan, the FDA and the CMS datasets were merged on the FIPS code because it serves as the unique identifier of both of the rows of the dataset. That is because each row of each of the datasets represents the data in one US county. After merging both of these datasets, a connection needed to be found between the members of the Medicare Advantage plan and the food access data in order for the analysis to make sense. If the correlation between the two different datasets were high without such justification for a connection, the datasets cannot be assumed to be sampled from the same population. That is why in this analysis such a connection was tested by correlating the characteristics of the two datasets populations. Since the only characteristics present about the populations of the datasets, especially about the members of the Medicare Advantage plan, are the ethnicity segments in every county, the two datasets were correlated on the percentage of the ethnicity segment. Such correlation shows how much the FDA dataset ethnicity segments represent the CMS ethnicity segments of the Medicare Advantage plan members and vice versa. Trying to understand how the characteristics of the Medicare Advantage plan members match with the characteristics of the Food Access Data population segments would give us a better understanding of whether the food access dataset addresses the true needs of the members or not.

The following table shows the correlations of the ethnicities' segments of both datasets populations.

ALGOREX
HEALTHCARE TECHNOLOGIES

Table 1.

Correlation Coefficients between CMS and FDA Datasets Ethnicity Segments

CMS Data Ethnicity	FDA Data Ethnicity	Correlation Coefficient
White	White	-0.040500
African American	Black	0.631613
Hispanic	Hispanic	0.669346
Other	Asian	0.429844
Other	American Indian or Alaska Native	0.413076
Other	Hawaiian or Pacific Islander	0.185470

As shown in Table 1, there is a strong correlation between the same segments of the two datasets, except for the White ethnicity segment. That is why the food access data will be analyzed based on all the FDA ethnicity segments except for the White ethnicity segment, which will be analyzed based on the CMS Data White ethnicity segment because it represents the White segment of the Medicare Advantage plan members better than that of the FDA dataset.

Finding the Main Food Access Challenges. The White ethnicity segment in the CMS dataset showed positive correlation coefficients with almost the same factors as the other ethnicities segments from the FDA dataset. The top 10 FDA food access variables that correlate with the members of the Medicare Advantage plan are listed below in descending order from the most important to the least important for each category.

Table 2.

Top Food Access Factors Correlated with Ethnicity Segments of the Members

White	Black	Hispanic	Asian	American Indian	Hawaiian
White low access to store	Black low access to store	Multiracial low access to store	Asian low access to store	American Indian low access to store	Students eligible for reduced-price lunch
Population low access to store	Convenience Stores	Hispanic low access to store	Full-service restaurants	FDPIR Sites	
Seniors low access to store	SNAP households, low access to store	Low income & low access to store	Fast-food restaurants		
Children low access to store	SNAP-authorized stores	Fast-food restaurants	Specialized food stores		
Supercenters & club stores	Low income & low access to store	Children low access to store	Population low access to store		
Convenience stores	Households no car & low access to store	SNAP-authorized stores	Hawaiian low access to store		
Food Banks	Population low access to store	Full-service restaurants	Children low access to store		
Low income & low access to store	WIC-authorized stores	Population low access to store	Grocery stores		
SNAP households, low access to store	Children low access to store	Specialized food stores	SNAP-authorized stores		
SNAP-authorized stores	Fast-food restaurants	Grocery stores	Convenience stores		

Table 2 shows the food access problems for every ethnic segment of the members. For the White segment, the main problem is the low access to stores as well as the low income. The low-income factor is more significant in the other population segments because they correlate highly with the food access programs in addition to the low income and the absence of a household vehicle. As for the Hispanic and the Asian segments, the specialized food stores became a significant factor that wasn't highlighted in the White and the Black ethnicity segments. Overall, the low access to stores and low income as well as the population's low access to stores are the unification factors for all of the members' segments. The locations where the food access program should take place would be the top 10 locations that have the lowest access to store and income average, and the population lowest access to store.

Note that in Table 2, some of the workplaces that included convenience stores and fast-food restaurants were correlated with some of the ethnicity segments. These workplaces were included in the top 10 factors while other workplaces like farms were excluded because such stores can serve as the locations to reach the members with the correlated segments when implementing the food access program. For example, if full-service restaurants are highly correlated with the Asian segment of the members, this means that a food access program in cooperation with these restaurants would reach the Asian segments of the members easier than convenience stores. More on that topic in the discussion section.

Results

Question 1. The top 10 low access to store and low income and the population low access to store counties are listed in Table 3.

Table 3.

Top 10 Counties on Variables Low Income & Low Access to Store and Population Low Access to Store.

Population Low Access to Store Top 10 Counties	Low Income & Low Access to Store Top 10 Counties
Harris	Harris
Los Angeles	Bexar
Maricopa	Maricopa
Tarrant	Dallas
Bexar	Hidalgo
Riverside	San Bernardino
San Bernardino	Los Angeles
San Diego	Riverside
Dallas	Tarrant
Suffolk	Shelby

Table 3 shows that the two variables match on 8/10 of the counties. That is why only the Low Income & Low Access to Store top 10 counties data will be used for further analysis. Figures 1 and 2 take a closer look at the location of these top 10 counties for each variable on the US map.
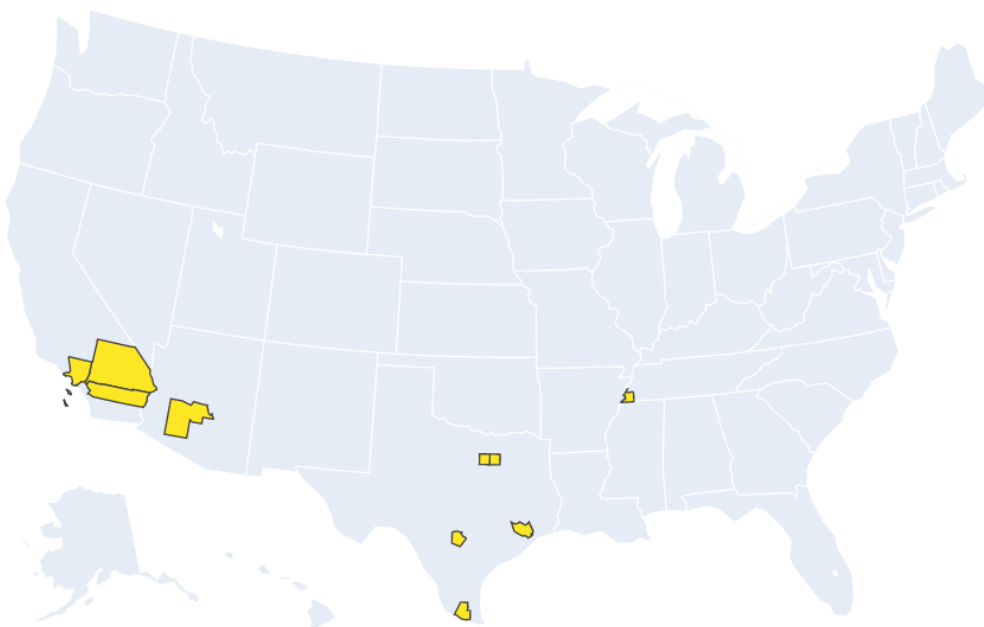
ALGOREX
HEALTHCARE TECHNOLOGIES

Figure 1.

Top 10 Population Low Access to Store Counties
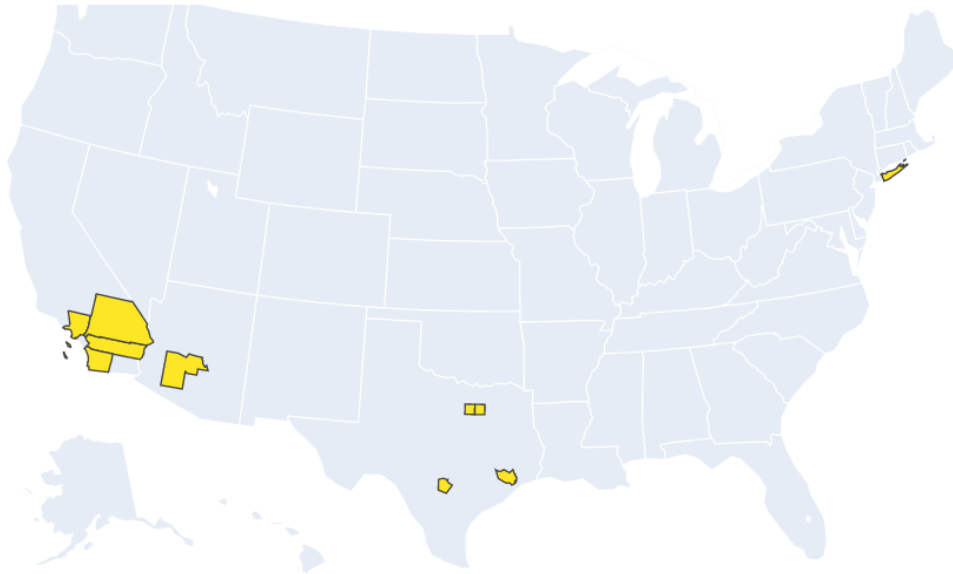
Figure 2.

Top 10 Low Income & Low Access to Store Counties

Both Figures 1 and 2 show the concentration of the counties in the border states between Mexico and the US, which suggests a hypothesis about the characteristics of the members in these counties. The hypothesis is that most of the members located in these counties would be of a Hispanic ethnicity based on the immigration routes and numbers summarized in Figure 3. The hypothesis will be tested when answering the following question.

Figure 3.

Popular Migration Routes from Central America through Mexico

* Figure Source

Question 2. To test the hypothesis about the Hispanic ethnicity of the majority of the Medicare Advantage plan members in the 10 counties from Question 1 and to discover the other top characteristics of the members' population, I correlated the low access to store segments of the FDA with the Low Income & Low Access to Store variable and found the top characteristics about the populations of the counties listed in Table 4.

Table 4.

Correlation Coefficients of the Top 5 Variables Correlated with Low Income & Low Access to Store

Variable Correlated with Low Income & Low Access to Store	Correlation Coefficient
Children low access to store	0.873221
Population low access to store	0.753215
Hispanic low access to store	0.740205
SNAP low access to store	0.591847
Households no car & low access to store	0.546029

Table 4 shows the top characteristics of the members residing in the top 10 counties. The correlation of the Hispanic segment verifies the hypothesis from Question 1. Note that the Multiracial segment was dropped since it didn't add value to the specific characteristics of the members in these counties.

Note also that the CDC dataset was used to find if there were other characteristics correlated with the Low Income & Low Access to Store variable in the given 10 counties but the CDC dataset only matched with the 10 counties on only 2 of them, while the rest of the 8 counties didn't exist. That is why the CDC dataset was dropped from the analysis because obtaining any correlation on only 2 counties would be meaningless. There would be an unjustified correlation due to the presence of only two rows of data.

Question 3. To predict the future impact of addressing the low food access problems associated with the Variables in Table 4 in the top 10 counties, a linear regression model was used to predict the Total Actual Medicare Costs based on the Variables in Table 4 and the Low Income & Low Access to Store Variable. The coefficient of determination of the model, which is commonly known as the accuracy score, turned out to be 92%, which is high for a model based only on 6 independent variables. Figure 4 shows how each of the 5 variables correlates with the Total Actual Costs, excluding the SNAP variable since it indicates low income, which is already correlated with the Actual Costs in Figure 3.
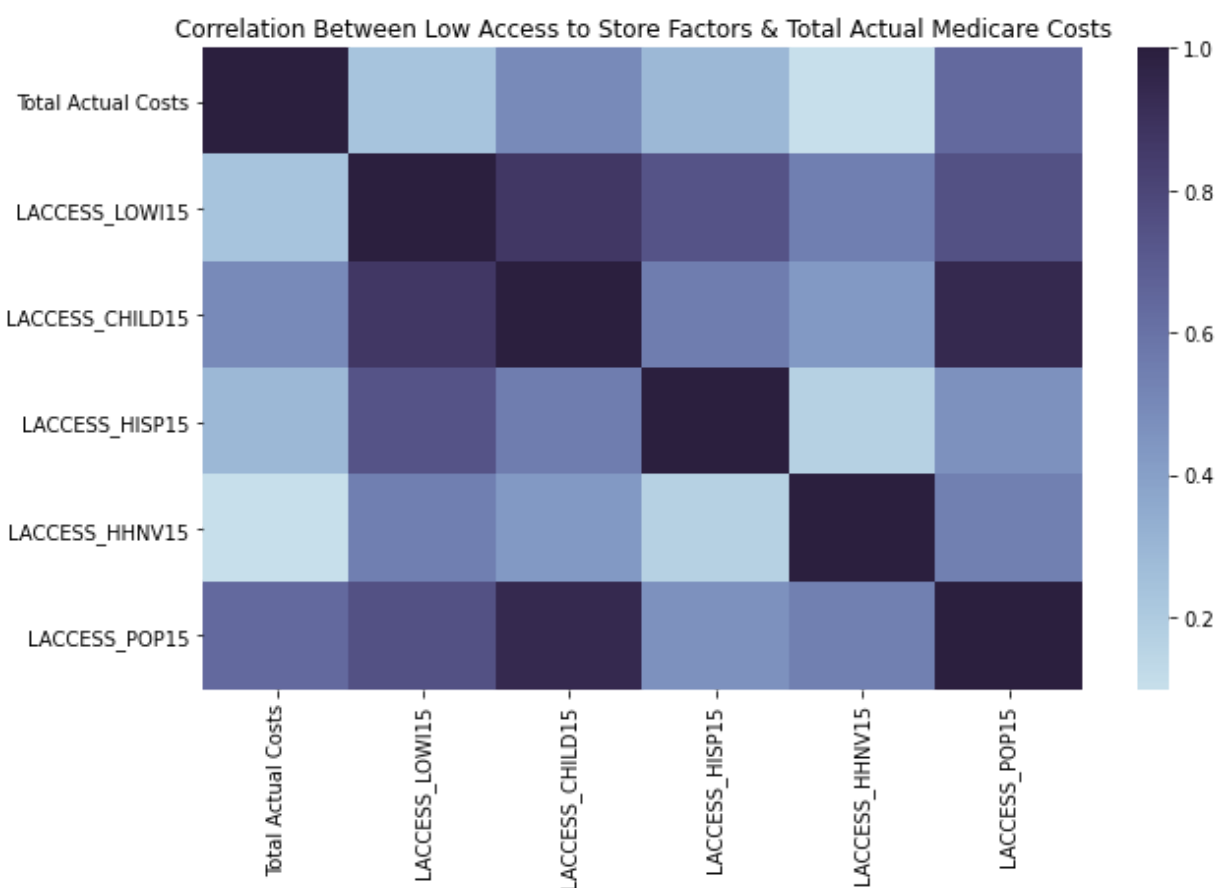
Figure 3.

Figure 3 variables LACCESS_LOW15 is Low income & low access to store, LACCESS_CHILD15 is Children low access to store, LACCESS_HISP15 is Hispanic low access to store, LACCESS_HHNV15 is Household with no vehicle and low access to store, LACCESS_POP15 is Population low access to store.

Discussion

The food access challenges faced by the members of the Medicare Advantage plan seem to center around the low access to stores and the low-income factors. Such challenges are particularly highlighted in the top 10 counties with the lowest income and access to stores. The characteristics of the members who live in these counties and would benefit the most from a food access program are the Hispanic and Multiracial ethnicity members with families that have children, low income, and no vehicle. Most likely these members are also members of other food access programs like SNAP. Addressing the challenges in these 10 counties can reduce the total costs of the Medicare services by 92%.

In order to address the food accessibility issues and provide the Hispanic majority of members with their food needs, the Medicare Advantage plan can cooperate with the Fast Food Restaurants that the Hispanic members correlate with, as shown in Table 2, to provide healthy meals or open specific distribution points for healthy food near the locations of fast-food restaurants. The Medicare Advantage plan can provide services for the children of their members at the local schools to address the low children food accessibility.

Conclusion

The members of the Medicare Advantage plan would benefit from a program that addresses the store proximity and the price points of healthy food. Hispanic members with children and low income located in the Southern US counties near the US-Mexican border will benefit the most from the program. The program will also benefit the Medicare Advantage plan by potentially reducing the total cost of the services utilized in these counties by 92%.