Food Deserts and Health Outcomes

Using Difference-in-difference Design to Determine the Effect of Food Access on Obesity in the US

Hannah Jones

5/11/2021

Abstract

The term 'food desert' is used to characterize communities with low access to affordable, fresh food. These communities are likely to be low income and predominantly minority. When specific people groups are being consistently affected by the same issue across the country, it is important to examine the impact of the issue. The continued impact on historically disenfranchised groups suggests that many gaps in income, opportunity and other outcomes may be attributed to systemic issues such as lack of access to food (2).

It has been well documented that limited food access negatively impacts the health of the affected community. Mapping food deserts reveals that these areas also suffer from higher rates of obesity, as well as other diet-related diseases (3). In this paper, I use food desert data from the USDA (5), as well as health outcome data from the CDC (4) to study the impact of changing food desert status on affected communities. Specifically, I look at communities where there were grocery store closures between the observation years (2010 and 2015), and see if there is a corresponding rise in obesity, controlling for other variables. In order to isolate the causal effect of the store closures on obesity, I utilize a weighted difference-in-difference technique.

This analysis finds that there is an increase in obesity of .5% in communities that experience grocery store closures between 2010 and 2015. This finding suggests that policy to increase healthy food options as well as education would impact health positively in these communities. Notably, this result is driven by propensity weighting. The treatment group is found to be about twice as likely to experience grocery score closure over the control group. Considering the systemic forces around health and quality of food access for disenfranchised populations, this finding emphasizes the need for further research and policy action.

Introduction

In order to study the effect of food insecurity on impacted populations, I used USDA food access data from 2010 and 2015, and CDC health data from within the time period of change (2013) and after (2017). I then examine health outcomes for these treated to groups to see if increased distance from grocerie stores or supermarkets causes poor health outcomes compared to those groups with no change in food access status.

The primary goal of this paper is to measure the causal effect between access to food and health outcomes. The identifying assumption underlying this differences-in-difference strategy is the changes in health outcomes for counties with small changes in access to food provide a good counterfactual for the changes in health outcomes that would have been observed for counties with larger changes in access if their access to food had changed similarly (1). Notably, the treatment group will be census tracts that are not considered low food access in 2010, and become low food access in 2015. The control group will be the census tracts that do not experience low access in 2010 or 2015.

Methodology:

To assess the assumption that lack of grocery store access causes increased obesity, I focus on the distance to the closest grocery store, plotting health outcomes over time for groups sorted according to their changes in distance to the closest store between 2010 and 2015. The data contains flags to indicate which census tracts contain more than 33% of the population farther than 0.5 miles or 1 mile of a grocery store for urban areas, and further than 10 or 20 miles for rural areas. The movement of these flags between time periods will serve indicate grocery store openings if distance reduces, and grocery store closings if the distance increases for any tract. For the outcome variable, we will look at health data collected in 2013 and 2017, and specifically at the prevalence of obesity. The markers of interest that are generally associated with low food access are obesity, diabetes, and high blood pressure (2). The data from the CDC surveys indicate the prevalence of obesity, diabetes, and high blood pressure as a percentage of the population.

Using the distance indicators, I designate the treatment group as any rural or urban census tract that becomes a food desert between the two observation years. This is to say that in the first set of data from 2010, the census tract observed has a grocery store within half or 1 mile if it is urban, and within 10 or 20 miles if it is rural, but then in the second set of data, the census tract no longer has a grocery store within the appropriate distance. This “treatment” indicates the closure of grocery stores since the distance to the closest store has now increased outside of the studied threshold.

To run a differences-in-difference model, we use the emergence of a food desert (grocery store closures) as the treatment, and we can use the prevalence obesity as the outcome. We will compare the prevalence of obesity in areas that retained their proximity to grocery stores to these areas that emerged as food deserts between 2010 and 2015. I only use census tracts that were not considered low access in either dataset as the control group in order to assume parallel trends in this analysis. This way, both the treated and untreated groups are not considered low access in 2010, and then the treatment group becomes low access in 2015, while the control group stays the same. This assumption requires that nothing unobserved changes within a census tract that would also determine obesity prevalence.

In order to account for a difference in the likelihood that an area becomes a food desert between sample periods based on the existing health outcomes in the area, we look at the propensity score weighting. This can be understood as perhaps a certain tract does not value access to fresh food or does not engage with access to fresh food, resulting in worse health outcomes and less demand for fresh food, which would increase the likelihood of grocery store closures.

I use propensity scores in order to capture the idea that the expected outcome of non-treatment for the treatment group will not equal the expected outcome of non-treatment for the non-treatment group. In other words: $E(y^0 \mid D = 1) \neq E(y^0 \mid D = 0)$

Due to Selection bias from systemic issues, it is more likely that the expected health outcome for the treated group is worse than the expected health outcome for the non treated group. In this case, we would expect the outcome (obesity prevalence) to be higher in the treatment group in the case of non-treatment. This can be shown as: $E(y^0 \mid D = 1) > E(y^0 \mid D = 0)$

This phenomena will bias the average treatment effect upwards because it will capture some of the difference in the groups due to the difference in their propensity for treatment.

Analysis:

To analyze the impact of food desert status on health, I use a weighted dif-in-dif technique. I predicted a propensity score for each Census Tract based on population, urban status, income status, and a variety of health factors. I chose these covariates to capture the propensity of already low-income and low-health populations to lose grocery store access. Creating this propensity score captures the tendency for systemically low-income, low-health communities are less accounted for and advocated for on the policy front.

I create dummy variables for food deserts for both 2010 and 2015. I determine a census tract is a food desert in either given year based on the "Low Access Tract" flags in the data. There is a flag for tracts that are not within half a mile of a grocery store, within 1 mile, and within 20 miles. If a tract is flagged at a certain level in 2010 or 2015, then I consider it a food desert for that year. In order to build a treatment group, I look at the change in flags between 2010 and 2015. For example, if a census tract is not flagged in 2010 at the 1/2 mile level, and then it is flagged in 2015 and the 1/2 mile level, I consider this an indication that the tract has experienced grocery store closures, and therefore should be considered part of the treatment group. I run this comparison for the 1/2 mi., 1 mi., 10 mi., and 20 mi. access levels, and then create a single treatment variable that captures every census tract that experienced and increase in distance from grocery stores due to closures.

Table 1: Food Desert Totals 2010 and 2015

Years	Low Access 0.5 mi	Low Access 1 mi	Low Access 10 mi	Low Access 20 mi
2010	64638	42364	3346	422
2015	46520	24322	3205	388

Table 1 shows the raw total of each type of low access tract in the two data years. As you can see, in general the trend between the two years was a reduction in food deserts. Overall this reflects a greater awareness and movement towards equitable access. However, in 2015 there were still ~46500 census tracts in the US in which more than a third of the residents were not within a half mile of a grocery store or supermarket. It is also particularly interesting to note that while urban low access at the 0.5 mile and 1 mile levels decreased by about 1/3 and 1/2, respectively, rural low access did not see the same improvements. It seems that rural census tracts who experienced low access at the 10 and 20 mile levels saw very little improvement in access between 2010 and 2015.

Now I will use this data to create a treatment variable that targets areas that were not food deserts in the first set of data, and became a food desert in the second set of data. This will serve as a proxy for grocery store closures.

Table 2: Food Desert Treatment and Control Groups

	Treated	Untreated (Control)
Urban	2783	7020
Rural	342	13946

Table 2 reflects the size of the treatment group, separated into urban and rural groups. The treatment group is defined as a census track that experienced an increase in grocery store distance (as signified by a flag appearing between data rounds) between the data rounds. About 3000 total census tracts experienced this. I also created an untreated group that was not experiencing any sort of low access in either 2010 or 2015. Therefore, for this analysis I will be comparing the treated to the untreated, or in other words the non-desert to desert group versus the never desert group. In order to account for the selection bias present in which areas become food deserts, I will calculate a propensity score for the treatment group based on health outcomes, population, and income level. This propensity score should help account for any selection bias in areas that experience grocery store closures.

First, we can look at the raw rates of obesity and other health issue prevalence in 2013 and 2017.

Table 3: Average Prevalence of Health Outcomes 2013 and 2017 (in percent of population)

Years	Obesity Prevalence	Diabetes Prevalence	High Blood Pressure Prevalence
2013	29.758	10.770	30.949
2017	30.473	10.811	30.653

Table 3 shows that on a population average level, the prevalence of obesity and diabetes increased slightly between 2013 and 2017, while the prevalence of high blood pressure decreased slightly. Already we can observe some dissonance at at least an aggregate level around health outcomes and food access. While food access got better between 2010 and 2015, health outcomes at an aggregate level got worse. This disconnect suggests that perhaps some populations are not experiencing improved health and improved food access in conjunction. One possible explanation might be that on the whole food access is improving but due to cultural or lifestyle factors, health outcomes are not improving. This may be due to specific affected populations, or a larger cultural movement in the US around convenience foods. In order to determine if certain populations have a higher propensity for losing food access, we will construct propensity scores base on demographic and health information. This propensity score will take the form:

$Pr(D = 1 \mid X) = F(\beta_0 + \gamma Treat + \alpha X)$

Where the covariates captured by X are urban status, population, low income status, state, black, prevalence of SNAP benefits, access to health care, obesity prevalence, high blood pressure prevalence, and diabetes prevalence. I have chosen these covariates as they would traditionally ne correlated systemic issues that may increase the likelihood of losing grocery access.

Table 4: Propensity Scores based on Treatment

Treatment	Mean P-score	Min	Max
Treated	0.435	0.101	0.897
Untreated	0.258	0.100	0.863

Table 4 shows the mean propensity scores for the treated and untreated census tracts. I dropped the pscores over 0.9 and under 0.1 to remove extreme cases. As you can see, the treated group has a higher mean propensity for treatment based on demographics and health characteristics. For the untreated group, the average propensity score is greater than 1, but still is highly weighted on the low end, while the propensity for the treated group is much higher and more distributed down the right tail.

Figure 1: Propensity Scores for Untreated

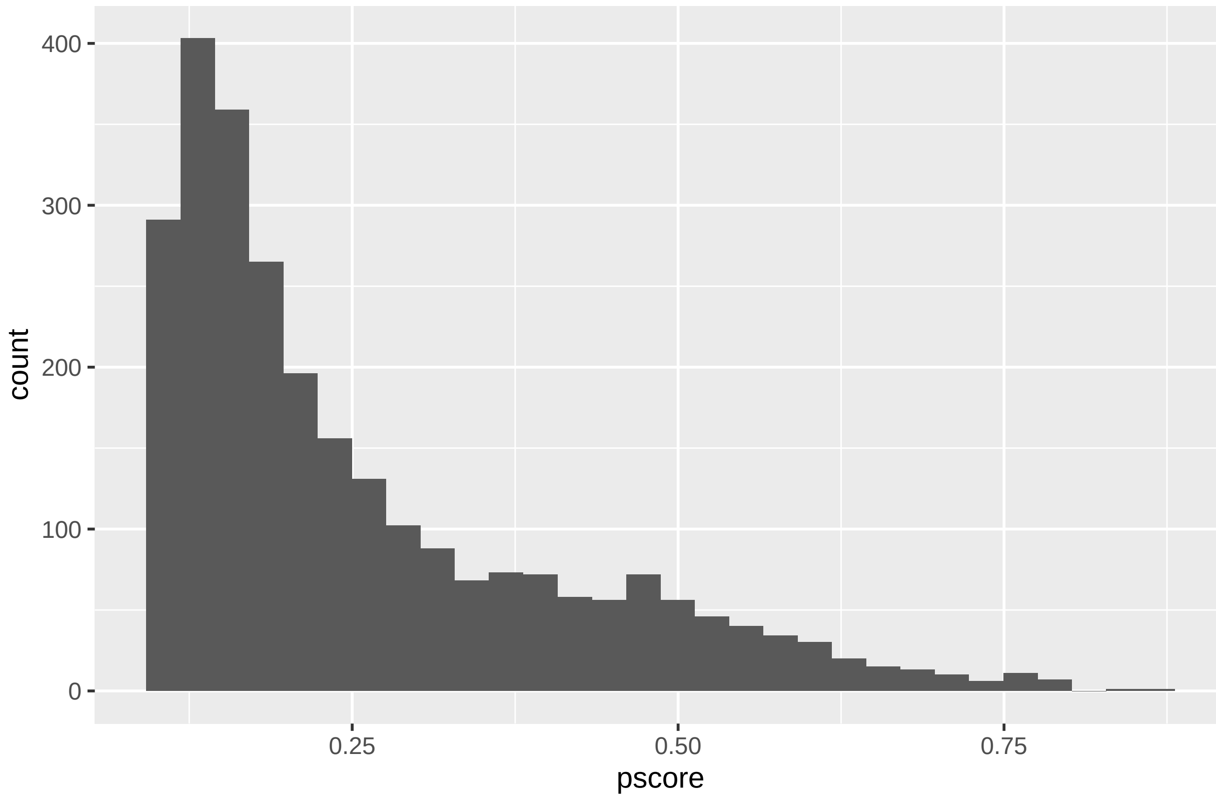
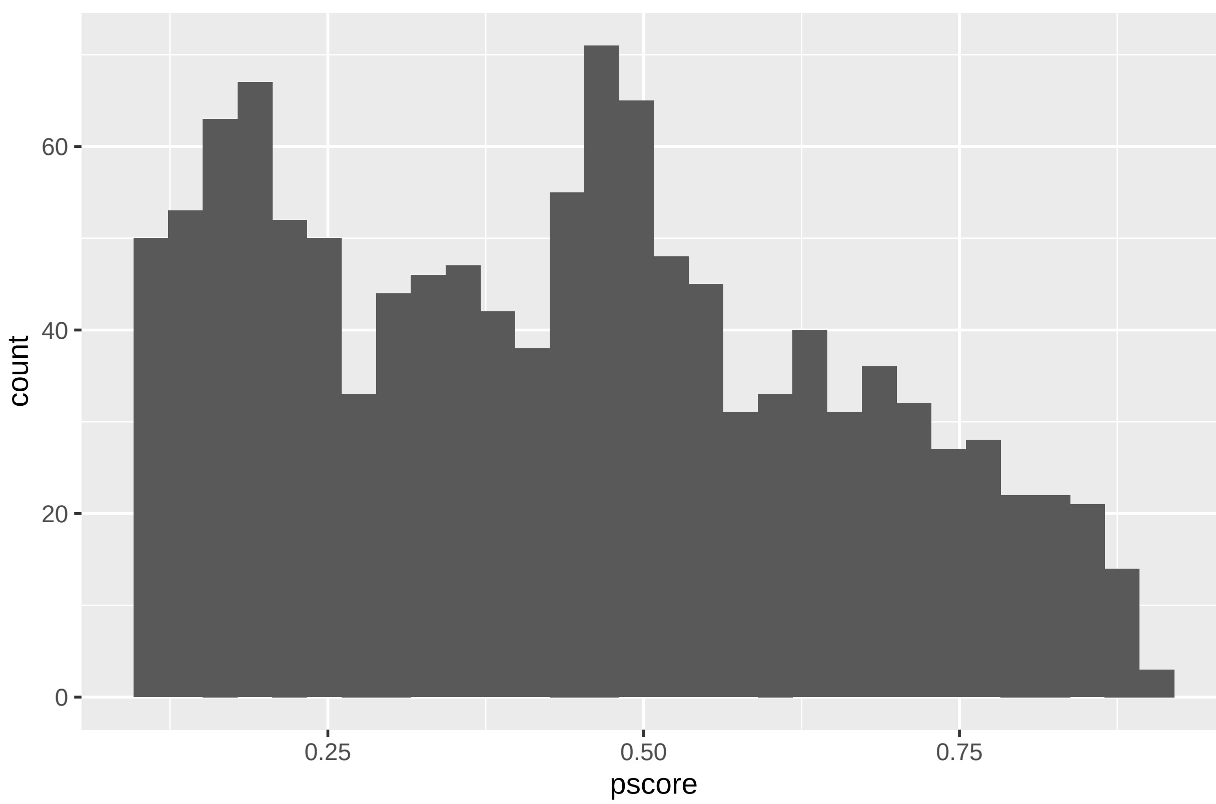


Figure 2: Propensity Scores for Treated



Figures 1 and 2 show the distribution of propensity scores for the treated and untreated Census Tracts.

The summary shows that the mean propensity score is higher for the treatment group. We will use these propensity scores to weight the treatment effect. First, we can look at the average treatment effect of grocery store closures between 2010 and 2015 on obesity prevalence as a percentage of the US population between 2013 and 2017. As shown below, the unweighted average treatment effect of grocery store closure is an increase in obesity prevalence of ~2.4%.

Table 5:

	Prevalence of Obesity (% of Population)
Untreated	30.665
Treated	33.079
Average Unweighted Treatment Effect	2.414

Now I will look at the weighted difference using the propensity scores, using both non-normalized and normalized weights. Due to the high propensity score of the treated group, I expect the average treatment effect to be smaller than the 2.4% shown in Table 5. Since the treated group has about double the propensity for treatment based on characteristics stated previously, the impact on obesity, controlled for propensity, will be lower.

Table 6:

	Prevalence of Obesity (% of Population)
Untreated	31.078
Treated	31.192
Average Weighted Treatment Effect	0.114

Table 6 shows the non-weighted difference in obesity prevalence. This suggests that grocery store closures result in a .11% increase in the prevalence of obesity, but once we normalize the weights, as shown in table 7, we find a difference of .5%. This suggests that the difference in the prevalence of obesity in a census tract that has experienced grocery store closures is about a .5% increase over a tract that has not experienced closures.

Table 7:

	Prevalence of Obesity (% of Population)
Untreated	31.283
Treated	31.805
Average Normalized Treatment Effect	0.522

Conclusion

It is important to consider the impact of socio-economics and race on health outcomes as they connect to opportunity and access. Though a small sample size of only about 5% of the total census tracts experienced grocery store closures, this small sample tended to be more susceptible to grocery store closures than the whole group. This susceptibility has the ability to impact health outcomes, that may in turn limit opportunities for advancement economically and socially.

This study of health outcomes related to food access found that areas that experience grocery store closures saw an causal increase in obesity by about .5% over areas whose food desert status was not jeopardized. In this study, I only examined areas where food desert status changed from high access to low access between 2010 and 2015. However, plenty of census tracts were food deserts across both observation points, and therefore did not make it into the treatment group, but are likely experiencing ongoing poor health outcomes due to consistent food desert status over time.

Further Research

When considering the propensity of a census tract to lose grocery store access based on race, health and low income status, we find a propensity score twice as large for the treatment group than the control group. This suggests that certain populations are more prone to losing access and to having poor health outcomes. In further research, it would be interesting to dive deeper into exactly what systems and policies make it such that these populations are more prone to losing access to a basic human right, fresh food. Advocacy and education are likely instrumental in improving health outcomes for these populations.

References

1. Causal Inference: the Mixtape, by Scott Cunningham, Yale University Press, 2021.
2. National Research Council (US). The Public Health Effects of Food Deserts: Workshop Summary. Washington (DC): National Academies Press (US); 2009. Summary. Available from: https://www.ncbi.nlm.nih.gov/books/NBK208018/
3. Kelli, Heval M., et al. "Living in Food Deserts and Adverse Cardiovascular Outcomes in Patients With Cardiovascular Disease." Journal of the American Heart Association, 11 Feb. 2019, www.ahajournals.org/doi/10.1161/JAHA.118.010694.
4. "PLACES: Local Data for Better Health." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 8 Dec. 2020, www.cdc.gov/places/index.html.
5. Rhone, Alana. "USDA Food Access Data." USDA ERS - Go to the Atlas, 27 Apr. 2021, www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas/.