DRAFT Dissertation Prospectus

Danton Noriega

November 21, 2016

Contents

1 Overview 3

2 An Evaluation of the Double Up Food Bucks program 5

Introduction 5

Financial Incentives to Encourage Healthy Food Purchases 9

The Healthy Incentives Pilot 12

The Double Up Food Bucks Program 14

Double Up Food Bucks vs the Healthy Incentives Pilot 16

Evaluating Double Up Food Bucks in Non-experimental Conditions 17

Data 19

Overview of Store Selection and Expansion 20

Selection and Expansion of Double Up Stores 20

Expansion on Observables 22

Selection of Control Stores 23

Chapter 1

Overview

Chapter 1 is an evaluation of the effectiveness of the Double Up Food Bucks (Double Up) program. "Effectiveness" will be defined by the change in total sales and volume of produce sold within a subset of grocery stores that implement Double Up (treatment group). The control group comprises 15 stores where Double Up was not implemented. A regression discontinuity design and difference-in-differences, between stores using Double Up (treatment) and those without (control), will be used to measure the size of the effect.

Improving health and food equity of SNAP participants is the broader policy concern. The mechanism is financial incentive designed to increase fruit and vegetables consumption. The focal point will be on the *Double Up Food Bucks*[1] program run by the non-profit *Fair Food Network* based out of Michigan. A comparison will be made with another financial incentive program called the *Healthy Incentives Pilot* (HIP). I will argue how and why an evaluation of the Double Up program is an important addition to the current literature.

Chapter 2 is an exploration of how the basket of products purchased by SNAP shoppers

[1] `http://www.doubleupfoodbucks.org/`

changes after the introduction of DUFB. This will use a few different sets of data where one is able to link purchases to individuals.

Chapter 3 is a descriptive paper on how Durham Social Services spends TANF dollars. There is no uniformity between how different states and local governments use TANF funds. There is also little published data research.

Chapter 2

An Evaluation of the Double Up Food Bucks program

Introduction

Chronic conditions like obesity, heart disease, and other metabolic risk factors (stroke, type II diabetes, etc.) are estimated to cost the US health care system between 200 to 400 billion dollars annually (Cawley and Meyerhoefer, 2012; Chatterjee et al., 2014). More importantly, these diseases account for hundreds of thousands of deaths each year. Heart disease alone is the leading cause of death for all persons in the US, with stroke fifth and diabetes seventh (National Center for Health Statistics, 2015). Diet is closely linked to these conditions, particularly obesity and cardiovascular disease. There is strong evidence that a diet high in (1) vegetables, fruits, nuts, unsaturated oils, fish, and poultry, but low in (2) red and processed meat and sugar-sweetened foods and drinks, helps lower body weight, blood pressure, and the risk of cardiovascular disease (Mente et al., 2009; Nutrition Evidence Library, 2014). Improving the diet of Americans has therefore become an increasing priority for the United States, especially for struggling families that participate in the Supplemental Nutrition Assistance (SNAP) program.

SNAP is a federal aid program administered by the Food and Nutrition Service (FNS), an agency of the U.S. Department of Agriculture (USDA). At 74 billion dollars in FY2015 with roughly 45.8 million participants, it is the largest food assistance program in the US (USDA FNS, 2016b). To be eligible for SNAP, a household must be sufficiently budget constrained that hunger is considered likely without assistance. Eligibility is a function of countable resources, vehicle ownership and value, household size, gross or net monthly income, household composition, and meeting certain work requirements.[1] Some eligibility requirements vary by state, but in general, a family with less than $2000 in countable resources, where the adults work at least part-time earning a gross (net) monthly income at or below 130% (100%) of the federal poverty line, is eligible to receive SNAP benefits. Aside from a few restrictions—no alcohol, tobacco, non-food items, read-to-eat meals, or hot foods—households can use SNAP benefits to purchase any foods that will be prepared and consumed at home. Unfortunately, the purchasing patterns of the average SNAP household are not conducive to a healthy diet.

Research on the dietary patterns of households receiving SNAP benefits has found that they are significantly *less* likely to meet USDA dietary guidelines than the average US household and much *more* likely to consume unhealthy foods (Andreyeva et al., 2015; Nguyen and Powell, 2015; Wolfson and Bleich, 2015). A smaller set of research has found that SNAP households, at best, consume same amount of unhealthy foods (e.g. sugar-sweetened beverages, baked goods, snacks, candy, etc) compared to SNAP-ineligible households (Todd and Ploeg, 2014; **?**). In other words, SNAP households consume foods that are less healthy or about the same as SNAP-ineligible households. This is a concerning result given that most US house-

[1] For more details, visit `http://www.fns.usda.gov/snap/eligibility`

holds, regardless of income, already purchase and consume far too much meat and foods rich in sugars and fats, and far too few fruits, vegetables and whole grains (USDA, 2015; Frazão, 1999). However, the purpose of SNAP is to keep struggling families from going hungry, not to ensure they consume the best possible diet. SNAP is designed to act like cash, helping families access more food than they could otherwise do so without assistance (Hoynes et al., 2015). It is therefore not a failing of the SNAP program if benefits are used to purchase unhealthy foods.

The SNAP program could be change such that it could continue satisfying its role as an anti-hunger program while simultaneously encouraging healthier purchases. Blumenthal et al. (2014) and Leung et al. (2013) both surveyed a field of stakeholders and policy experts in the SNAP program about what they would do to improve the dietary quality of purchases. In both studies, restricting the purchase of healthful foods (e.g. sugar-sweetened beverages) and promoting healthy purchases through monetary incentives were the two most popular improvements (i.e. ranked the highest or most often suggested). [2].

One common suggestion is to restrict the SNAP program to the same set of eligible foods as the Woman, Infants, and Children (WIC) program (Dinour et al., 2007). The WIC program provides food vouchers which limit households to a select group of food products. These food products are specifically selected to ensure women and their children receive nutritious, healthy foods. In other words, the WIC program, by design, places restrictions on food choices by defining a list of *eligible* items, as opposed to the SNAP program, which defines a list of *ineligible* items. Another common, and simpler, suggestion is to expand the existing list of ineligible items (i.e. alcohol) with products that are unambiguously lacking in nutrition and easy to identify, like soda or candy. New York City, for example, attempted to ban the

[2] It should be mentioned that there were other, less popular recommendations, such as modifying how SNAP benefits are distributed and improving nutrition education. For more details, see Blumenthal et al. (2014) and Leung et al. (2013)

purchase of sugar-sweetened beverages, and the state of Maine attempted to restrict the purchase of sodas, candy, and any other taxable food items (Gundersen, 2015). Both restrictions were overturned by the USDA.

There are problems with "improving" the SNAP program by implementing even greater purchasing restrictions. First, there is no reason to believe that such a restriction would work. The restriction assumes households will substitute unhealthy foods for healthy foods when using SNAP benefits. What would most likely happen is that households would shift to purchasing unhealthy foods with cash. Second, such restrictions would likely lead to a drop in SNAP participation (Gundersen, 2015). Restricting choice is a paternalistic policy that would further stigmatize SNAP participation. It would give the impression that SNAP beneficiaries are assumed to have worse diets and that they cannot be trusted to make healthy food purchases. Participation would also drop due to increased transactions costs of purchasing items with SNAP. Not all stores would clearly mark which items were SNAP eligible nor should participants be expected to remember. The result would be longer, more frustrating shopping trips. Lastly, it is important to remember that for many SNAP recipients, freedom of choice is what makes the SNAP program popular and easy to use (Edin et al., 2013).

The most popular "improvement" was providing a monetary incentive to SNAP participants for purchasing healthy foods (Blumenthal et al., 2014; Leung et al., 2013). Monetary (or financial) incentives, in this context, tend to be a rebate or voucher rewarded to SNAP households for using their benefits to buy certain healthful foods, generally mineral-rich and nutrient-dense fruits and vegetables (e.g. leafy greens but not white potatoes). These monetary incentives for buying "targeted" fruits and vegetables (aka TFVs) are exclusive to SNAP participants. Much like a grocery stores loyalty card or a student ID card, retailers can "discriminate on price" (aka "target the incentive") using SNAP Electronic Benefit Transfer (EBT) cards to identifying eligible participants. Monetary incentives in the food retail environment are popular for two main reason. First, the framing of the "improvement" is positive. Instead of "punishing" SNAP participants through paternal restriction or disincentives (not covered), monetary incentives reward participants for healthy shopping behavior (Gundersen, 2015). Retailers also prefer the positive framing of monetary incentives. For the moment, monetary incentives programs for SNAP participants are not wide spread. Therefore, taking up an incentive program, assuming the costs of implementation isn't too expensive, creates an opportunity for retailers to differentiate themselves from their competitors (Hartmann, 2011). The second reason is a strong theoretical framework established by neoclassical economics supporting incentives as an effective mechanism for changing human behavior. In practice, however, incentives have had mixed results, but there is building evidence that incentives may work in the food retail space.

How, why, and to what effect incentives may encourage SNAP participants to purchase more targeted fruits and vegetables will be discussed in detail below, and is the motivating question behind this paper.

Financial Incentives to Encourage Healthy Food Purchases

Encouraging healthy behavior through financial incentives has a long history. Results are mixed. For example, financial incentives have been shown to help individuals commit to regular exercise, improve dieting, increase weight loss, and to quit smoking, but the intended effect of the financial incentives were often only short-term (see Gneezy et al. (2011) and Cawley (2015) for an overview). Gneezy et al. (2011) also explain, through a review of the literature, that depending on the context and design, incentives can backfire, producing an effect known as "crowd out". Crowd out occurs when the incentives displaces the intrinsic reward of a behavior (originally defined for "prosocial" behavior, like volunteering or giving blood; see

Bénabou and Tirole (2006)). The behavior then becomes dependent on the extrinsic reward. As a result, having shifted from being intrinsically rewarding to extrinsically rewarding, the positive behavior continues only as long as the monetary incentive is provided. More significantly, the intrinsic reward of the behavior does not return once it has been "crowded out". Therefore, the long-term effect of a monetary incentive can be negative, despite showing positive effects in the short-term. That said, incentives can produce successful long-term results if they are instead used as a mechanism to build good habits. This requires that the incentives be salient and produce immediate feedback without neglecting behavioral findings such as loss aversion and mental accounting (John et al., 2011).

Given the research on incentives, it is reasonable to assume that a monetary incentive for SNAP participants to purchase healthy foods, like fruits and vegetables, may fail or even backfire. Should the act of purchasing healthy foods be intrinsically rewarding to SNAP shoppers, introducing an incentive may produce a "crowd out" effect. However, recent field experiments find that incentives can establish healthy food choice as a habit, possibly overriding any crowd out. Daily incentives encourage children to make healthier food choices in school lunchrooms, which in turn develop positive, long-term food habits (Loewenstein et al., 2016; List et al., 2015; Belot et al., 2014). Outside of the school lunchroom environment, List et al. (2015) found that similar habit formation is possible with incentives in a more traditional food retail environment. In their experiment, List et al. (2015) provided an incentive to 222 shoppers ($1) to use their rewards cards, but then randomly assigned each participant to the control group to one of three interventions: *information*, *incentive*, or *combination*. The *information* treatment was a flyer with tips on how to prepare fruit and vegetable dishes as well as the health benefits of eating more fruits and veggies. The *incentive* was an additional dollar for every 5 cups of targeted fruits and vegetables (TFVs) purchased. The *combination* treatment group included both. The intervention lasted for 5 months but each group continued to be observed for roughly 2 months after. The *information* intervention had no effect, but the *incentive* and *combination* interventions on average doubled their purchase of fruits and vegetables in comparison to the control group. Most promisingly, the gap persisted with minimal shrinkage for 6 weeks following the end of the intervention.

The design, food environment, and target of the incentive in each of these experiments is important. First, the incentive in these experiments is designed to be salient and immediate. In the school lunchroom experiments, the children are aware of the incentive and receive the reward (e.g. a small token) immediately after selecting the healthy food item. Likewise, the shoppers received their additional $1 reward for every 5 cups of TFVs at checkout. One distinction between the designs is frequency. The children are exposed to the incentive every school day in the lunchroom experiments. The shoppers, on the other hand, were exposed as frequently or as infrequently as they chose. This distinction is important, as the latter better reflects the experience of shoppers using the SNAP benefits. Second, the food environment is important because it determines what choices are available. The children have a finite set of options in the school lunchroom and they also have no outside option (besides not eating lunch). The children optimize on a relatively small set of choices and, for the duration of the intervention, the incentive always existed. Food retail environments are drastically different. There are numerous competing food choices and generally other outside options.[3] It is substantially more difficult for the shopper to optimize over such a large set of choices. Last, and most obviously, one set of studies targets children, the other adults. A priori, we would expect a monetary incentive to affect children differently than adults. The fact that habit formation through incentives appears possible for both target groups is promising.

Research where SNAP participants are the target group is nascent. The USDA's Food and

[3] It should be noted that this is not always the case. In List et al. (2015), for example, the store that was selected for the experiment was one of the few places local shoppers could find fresh produce. The store itself was located in one of Chicago's poorer neighborhoods.

Nutrition Services (FNS) ran the first large scale randomized control trial investigating the impact of a financial incentive for targeted fruits and vegetables in 2011. The experiment was called the Healthy Incentives Pilot (HIP). HIP is the precursor to every incentive program currently being funding by the USDA. It also provides the data for the few papers recently published on incentives for SNAP participants.

The Healthy Incentives Pilot

A brief overview of the Healthy Incentives Pilot is necessary to provide context to, and contrast with, more recent financial incentive programs.

The UDSA's Food and Nutrition Services designed the Healthy Incentives Pilot. The pilot was funded by the Food, Conservation, and Energy Act of 2008 to test whether financial incentives would increase consumption of targeted fruits and vegetables (TFVs). SNAP participants were the target group.

HIP was designed as a large scale randomized control trial (RCT). FNS partnered with the Massachusetts Department of Transitional Assistance (DTA) to implement HIP. The pilot lasted from early 2011 to the end of 2012. The population included all 55095 SNAP participants in Hampden County, MA. Hampden County is the poorest county in MA and has the highest rates of obesity and other diet-related chronic illness (e.g. type 2 diabetes).

Of the 55095 SNAP participants, 7500 were randomly assigned to the treatment group. The remainder fell into the control group. The treatment was a 30 cent (or 30%) rebate on every dollar spent on TFVs. The rebate was capped at $60 per month. To receive the rebate, selected SNAP participants had to use their EBT cards at participating retailers. The rebate, which was returned to their EBT account, could be then be used on any food item. That is, the rebate could only be earned buying TFVs, but could be redeemed buying any SNAP eligible food item. Most HIP participants spent about $12 a month on TFVs, earning an average of $3.65 per month in rebates—drastically lower than the $60 per month rebate cap.

The evaluation was conducted using 24-hour dietary recall surveys. A total of 5000 participants were selected to be surveyed, even split between treatment and control (2500 HIP, 2500 non-HIP). The first survey was conducted prior to the start of the pilot. This established a baseline. The second survey occurred 4 to 6 months in to the pilot and the third survey occurred 9 to 11 months in. (The variation, i.e. 4 to 6 months, was due to the treatment being implemented in 3 waves of 2500.)

The evaluation found that the 30% rebate lead to about a 26% increase in consumption of TFVs. This was equivalent to about 0.24 cups of TFVs. Roughly 60% of the increase was due to increased vegetable consumption and 40% due to increased fruit consumption. The effect, in absolute terms (0.24 cups), seems small. But a 0.87 price elasticity, relative to other results in the literature, is quite high—0.7 and 0.48, on average, for fruits and vegetables, respectively (Andreyeva et al., 2010).

Despite some limitations and technical problems—underreporting on the 24-hour recall survey and system glitches early in the pilot (see pages 60 and 208-210 of Bartlett et al. (2014))—HIP was considered to be an overall success (Klerman et al., 2014; Olsho et al., 2016). It implemented on of the largest, most complex RCTs to isolate how incentives can increase household consumption of TFVs. It also provided a feasible model for nationwide expansion (assuming cost reductions due to economies of scale; see An (2015)).

HIP also provides a framework for understanding how a financial incentive, expanded dramatically in one geographic area, could improve TFV consumption. But, as noted in the final HIP report, one of the most prominent retailers in Hampden County chose not to participate (page 61, Bartlett et al. (2014)). Its third-party processor decided it was too difficult and too costly to implement the financial incentive on its point-of-sale technology. This strategic behavior by the retailer, which had a significant presence in Hampden County, impacted where participants could use the incentive.

Most financial incentive programs work at the local level, expanding non-randomly. We should anticipate certain retailers (firms) to behave strategically when participating in any of these incentive programs. Likewise, we should anticipate voluntary (non-random) self-selection by SNAP beneficiaries into these financial incentives programs. To this end, more research is needed to understand the impact of incentive programs under *real-world* conditions. HIP provided evidence that incentive programs can work, but barring state-wide or nation-wide adoption of point-of-sale financial incentives, we should expect growth to occur organically under non-experimental conditions.

An example of such a financial incentive program for SNAP participants is the Double Up Food Bucks program (DUFB or Double Up). The non-random expansion and impact of this financial incentives program will remain the focus of this paper

The Double Up Food Bucks Program

The success of HIP paved the way for the Food Insecurity Nutrition Initiative (FINI), established by section 4208(b) of the Agricultural Act of 2014 (aka 2014 Farm Bill). FINI—a 100-million-dollar initiative—in turn piloted numerous non-profit financial incentive programs aimed at improving the diets of SNAP participants.

Of specific interest is Double Up Food Bucks (DUFB or Double Up), an incentives-based program funded by FINI. In 2009, the non-profit organization Fair Food Network (FFN) launched the Double Up Food Bucks program in Detroit, Michigan. The intention of the program was to get more low-income families visiting and participating in local Detroit farmer's markets. The mechanism for increasing participation was a dollar-for-dollar match of locally grown fruit and vegetable purchases. This subsidy was accessible only to low-income families receiving SNAP benefits, who could exchange up to $20 of their benefits for a wooden token that could be used on up to $40 worth of locally grown produce.

The DUFB program was considered successful given it had expanded to more than 150 farmer's markets in 2014 from just 5 farmer's markets in 2009. SNAP benefits have been used more than 200,000 times to purchase fresh produce, with more than 10,000 first time SNAP customers visiting farmer's markets in 2013 alone (Network, 2014). The program is considered by Fair Food Network to be a "three-fold" win given that the program helps local low-income families buy more fresh produce, provides new customers for local farmer's, and stimulates the local food economy. Relative to farmer's markets in other states, DUFB did seem to be bringing in substantially more SNAP dollars ($1.7 million in Michigan versus $307,000 in Illinois, the second largest).

A 5.17 million dollar FINI grant was awarded to Fair Food Network to help it pilot three adjustments to the Double Up Food Buck program (USDA NIFA, 2015). First, FFN needs to test DUFB as a year-round program in select locations instead of the current seasonal format. Second, shift away from the token system to providing DUFB electronically at point-of-sale. Third, the DUFB needs to expand from farmer's markets into other retail environments, like supermarkets and grocery stores.

Successful expansion into supermarkets and grocery stores is critical. Approximately 80% of all SNAP benefits in 2015 were used in supermarkets or super stores (USDA FNS, 2016a). Less than 1% percent of SNAP benefits were used at local farmer's markets. The amount of SNAP benefits used in local farmer's markets has increased since 2009, but no where near the growth necessary to reach the type of stores most frequented by low-income families. If localized financial incentive programs like DUFB are going to be considered as one of the USDA's many tools to increase food access and combat obesity, then they must be successfully implemented and scaled across supermarkets and grocery stores. Most importantly, incentive programs like DUFB must prove they are effective in changing purchasing habits within supermarket/grocery store food environments.

Double Up Food Bucks vs the Healthy Incentives Pilot

There are notable differences between DUFB and HIP that make the evaluation of DUFB more difficult. In short, HIP was implemented as an RCT. DUFB implementation is not. Let's explore in greater detail.

HIP had substantially more participating stores, all within the same county (Hampden County, MA). DUFB has fewer participating stores, spread across many different counties, and across many different grocery store chains. Therefore, the probability of a SNAP shopper in Hampden County having walked into a HIP participating store was therefore much higher than a SNAP shopper walking into any DUFB participating retailer.

The incentive delivery mechanisms also differ. First, all SNAP beneficiaries who shop at a DUFB participating store receive the benefit automatically. In other words, SNAP households that patron a store with DUFB receive the incentive regardless of their intentions or awareness of the DUFB incentive. Therefore, evaluating DUFB has the added difficulty of identifying which shoppers are optimizing in response to DUFB, as opposed to shopping normally. In contrast, SNAP households assigned to the HIP treatment group were made aware of incentive and were eligible to use it (even if they didn't quite understand how the incentive program worked — see Bartlett et al. (2014)). Households in the control group were not aware of the incentive and were not eligible to use it. And because participants were assigned, HIP evaluators could identify treated SNAP participants from control.

Second, the DUFB financial incentive is substantially larger but more restrictive. The DUFB incentive is a dollar-for-dollar match of locally grown produce purchases capped at $20 per day. The matched dollars are accrued as points on a store loyalty card. Existing points are then automatically redeemed as dollars on *any* fresh produce purchases, not just locally grown produce. In comparison, the HIP financial incentive was a return of 30 cents per dollar spent on TFVs which could be spent on *any* food item. That is, the DUFB incentive doubled the purchasing power of every dollars spent on TFVs *only for more TFVs*; the HIP incentive increased the purchasing power of every dollars spent on TFVs by 30% *for any SNAP eligible food item*.

Finally, the experimental designed of HIP allowed researchers to form a causal interpretation of their results; the average treatment effect is the same as the average treatment effect on the treated. Any difference in the purchase and consumption of TFV between the treatment and control groups could therefore be attributed to the incentive. This is not the case for DUFB. However, HIP implementation is the exception. How DUFB, and similar financial incentive programs are implemented, is the norm. The contribution of this paper will be evaluating and understanding the impact of DUFB, given that DUFB and similar programs are implemented in the "real-world" (non-experimental conditions)

Evaluating Double Up Food Bucks in Non-experimental Conditions

DUFB's expansion and implementation into supermarkets and grocery stores did not follow an experimental design. Fair Food Network searched for local partners in the Detroit area willing to participate in DUFB. Not all grocery stores, especially the smaller independent stores, had the capacity to implement the point-of-sale technology necessary for the incentive—even if FFN offered to help cover the upgrade costs. The result is a self-selected group of stores participating in DUFB. This, in some ways, parallels what occurred in HIP, where one of the largest retailers decided integrating their point-of-sale systems to include the incentive was too expensive. This type of strategic firm behavior is important to consider, even if complicates the evaluation of an incentive program like DUFB.

In the real world, stores seek to maximize profits and will opt to participate only if they expect to profit. Similarly, individuals will self-select into participating; participation is optional and more likely to occur with well-informed and motivated SNAP shoppers. Selection, in this case, is a feature, not a flaw, of such incentive programs when implemented by non-profits or policy makers. The evidence, thanks to HIP, exists that incentives can lead to an increase in consumption. The goal of this paper is therefore to accurately measure the effect of the DUFB on TFV purchases while taking the selection into account. That effect can then be extrapolated forward, albeit weakly, using the results of HIP, to measure changes in consumption.

Fair Food Network started testing and gathering data from grocery stores implementing DUFB in 2014. One of FFN's largest partners, a Michigan grocery retail and distribution company, piloted the program in 2 of its stores in 2014. The company has since expanded to 5 stores in 2015 and then to 17 of 62 stores in 2016. Rapid scaling was possible due to the point-of-sale technology used by the company to implement DUFB across its stores. It provides, to date, the best case study of a firm strategically scaling DUFB across numerous grocery stores that span different geographic areas and populations.

All transaction data from 2014 - 2016 will be provided for every store that has, at any point, participated in DUFB. These data are complete (i.e. no records have been removed) and at the item level. A complete set of data will also be provided from another 15 stores where DUFB was not implemented.

Currently, no research exists evaluating DUFB, or similar incentive programs, using a complete set of store transaction data. HIP, for example, only had transactions records for SNAP

EBT cards. Transactions, should a different tender be used by the same individual, could not be observed. Therefore, these data provide an unprecedented opportunity to analyze how the DUFB financial incentive performs under real-world conditions. This paper will be, to the best of my knowledge, the first to perform an evaluation of a financial incentive, targeted at SNAP participants, using a complete set of data, from multiple stores, across multiple years, and collected under non-experimental conditions.

Data

These data come from a large grocery distributor and retailer serving multiple grocery chains. Three years of data will be made available, 2014 through 2016. These data are transaction level and will include (at least) store number, register, transaction ID, date and time of purchase, payment type, item, dollars, and quantity.

Double Up implementation was considered for a single grocery chain. The chain has more than 60 stores, 17 of which were selected as "treatment" stores (with Double Up). Of the remaining stores, data is being made available from an addition 15 to serve as "controls". The quotes here signify that these terms will be used as shorthand, but the terminology is somewhat misleading. The use of "treatment" and "control" could lead one to think store assignment was random. It was not.

[TK real specific details about the data e.g. total transactions observed etc.]

One important variable that will not be made available is a variable for loyalty card numbers. The company's use of loyalty cards across its many chains was an exciting prospect. Previous transaction data from smaller independent grocery chains had no way linking purchases to a single unique identifier over time because these smaller chains did not have advanced point-of-sale systems.

In earlier conversations with the company, it was understood that loyalty cards would be made available. However, months into working with the company, we were informed that this was no longer possible. Per the company's legal department, the company cannot share any personal information about their customers. Unfortunately for us, in the loyalty card contract signed by customers, the loyalty card number itself is considered personal information, meaning loyalty card numbers fall under the same legal category as phone numbers and home addresses.

Overview of Store Selection and Expansion

How the 17 "treatment" stores and 15 "control" stores were selected in 2016 is important. First and foremost, selection was *not* random. Stores were either selected by the company (13 of 17) or self-selected into Double Up (4 of 17). Second, the 15 control stores were selected *after* the selection of the 17 treatment stores. Data from all remaining stores was requested but the request was denied; only 15 stores had been approved by the company's management. Finally, and most importantly, the selection criteria for the 17 treatment stores is *observable*. The implications of this will be covered in more detail in the Methods section.

Selection and Expansion of Double Up Stores

The first 2 stores were piloted with Double Up in 2014. Both were in geographically distinct areas (these will be referred to as "`Node 0`" and "`Node 1`"). There was a small expansion adding 3 more stores in 2015. The 3 stores were selected because they were geographically close to the 2 original pilot stores (2 close to `Node 0`, 1 close to `Node 1`). The 5 stores are referred to as the "core". The location of these 5 stores, separated in two clusters, established the geographic constraints that were then used to determine most of the additional stores in 2016.

Double Up was expanded to 12 more stores in 2016, totaling 17. Of those 12, 6 were selected due to their proximity to the 5 core stores, their SNAP EBT[4] sales figures, and similarity in surrounding demographics (high population density, more African-American). In other words, 9 of the 17 stores—excluding the initial 2 pilot stores—–were selected on a set of *observable* characteristics. The remaining 6 stores were not.

Of the remaining 6 stores, 4 asked if they could be included in the program. These stores *self-selected* into Double Up, making these stores fundamentally distinct. They were considered, and then included, only because they fell within the "Top 50". The final 2 stores were selected by the company for "strategic business decision". The best interpretation of this is that the company thought that Double Up would provide a competitive edge to the 2 included stores given some internal calculus. How the company came to this decision is *unknown* and therefore *unobserved*.

Table 2.1 helps understand the year by year expansion of Double Up. Stores are classified as either `assigned`, `self-selected`, or `unobserved`. To be `assigned` means a stores participation in Double Up was determined (assigned) by the company; `self-selected` means the store asked the company to participate; `unobserved` means that the company selected the store to participate in Double Up but for unknown and unobserved reasons. Numbers were assigned to each store for easy reference but otherwise have no meaningful interpretation.

[4]Electronic Benefit Transfer.

Table 2.1: Year by Year Store Selection. Stores 1 and 2 represent the initial 2014 pilot stores.

Store	2014	2015	2016
1	assigned	assigned	assigned
2	assigned	assigned	assigned
3		assigned	assigned
4		assigned	assigned
5		assigned	assigned
6			assigned
7			assigned
8			assigned
9			assigned
10			assigned
11			assigned
12			self-selected
13			self-selected
14			self-selected
15			self-selected
16			unobserved
17			unobserved

Expansion on Observables

An example expansion on *observables* (using fake data) can be seen in Figure 2.1. In the top frame, one can see two blue dots. These blue dots simulate the first two pilot stores in 2014. The left blue dot is `Node 0` and the right blue dot is `Node 1`. The gray zones represent areas of higher population density. Dark gray is considered *urban*, defined as having a population density of 1500 persons or more per square mile. The light gray are small towns and cities, more densely populated than very rural areas, but could not be considered *urban*. The expansion in 2015 (middle frame) proceeds to the stores closest to the original pilot stores. The expansion continues to 6 more stores in 2016 (bottom frame) away from the nodes but also along areas of higher population density.

Not conveyed in Figure 2.1 is that the 2015 and 2016 expansions also move through stores that happen to be "highly ranked"—that is, have relatively higher SNAP EBT sales.[5] Also not conveyed is the fact that there is a strong correlation between geography, population density, racial composition, and SNAP EBT sales. The 2015 expansion to the most nearby stores also meant that it was an expansion to stores with high SNAP EBT sales in densely populated, African-American neighborhoods. The 2016 Double Up expansion was more explicit given that set of feasible stores substantially increases as one moves away from each node. Double Up stores were thus specifically selected not just by geographic proximity, but also by SNAP EBT sales ranking and demographic compositions similar to the initial 2014 stores.

Selection of Control Stores

Ideally, all remaining stores would have been available to use as a control group but the company only approved that data be released for 15 stores. This left the added—and incredibly important—step of selecting the control stores since the company approved, but did not explicitly select, the 15 stores.

Selecting the control stores proceeded in two steps. First, stores that either self-selected or were selected using some unobservable criteria were matched using *Coarsened Exact Matching* (CEM) (Iacus et al., 2011). Second, stores assigned Double Up were pooled with nearby control stores and then scored using a linear probability model. Each step is explained in detail.

2.0.0.1 Step 1: Coarsened Exact Matching

The 6 stores classified as `self-selected` or `unobserved` (stores `12` through `17`; see Table 2.1) were compared against all possible control stores for matches. Matching was done across 5

[5] All stores within the chain were ranked by SNAP EBT sales as a percentage of total sales.

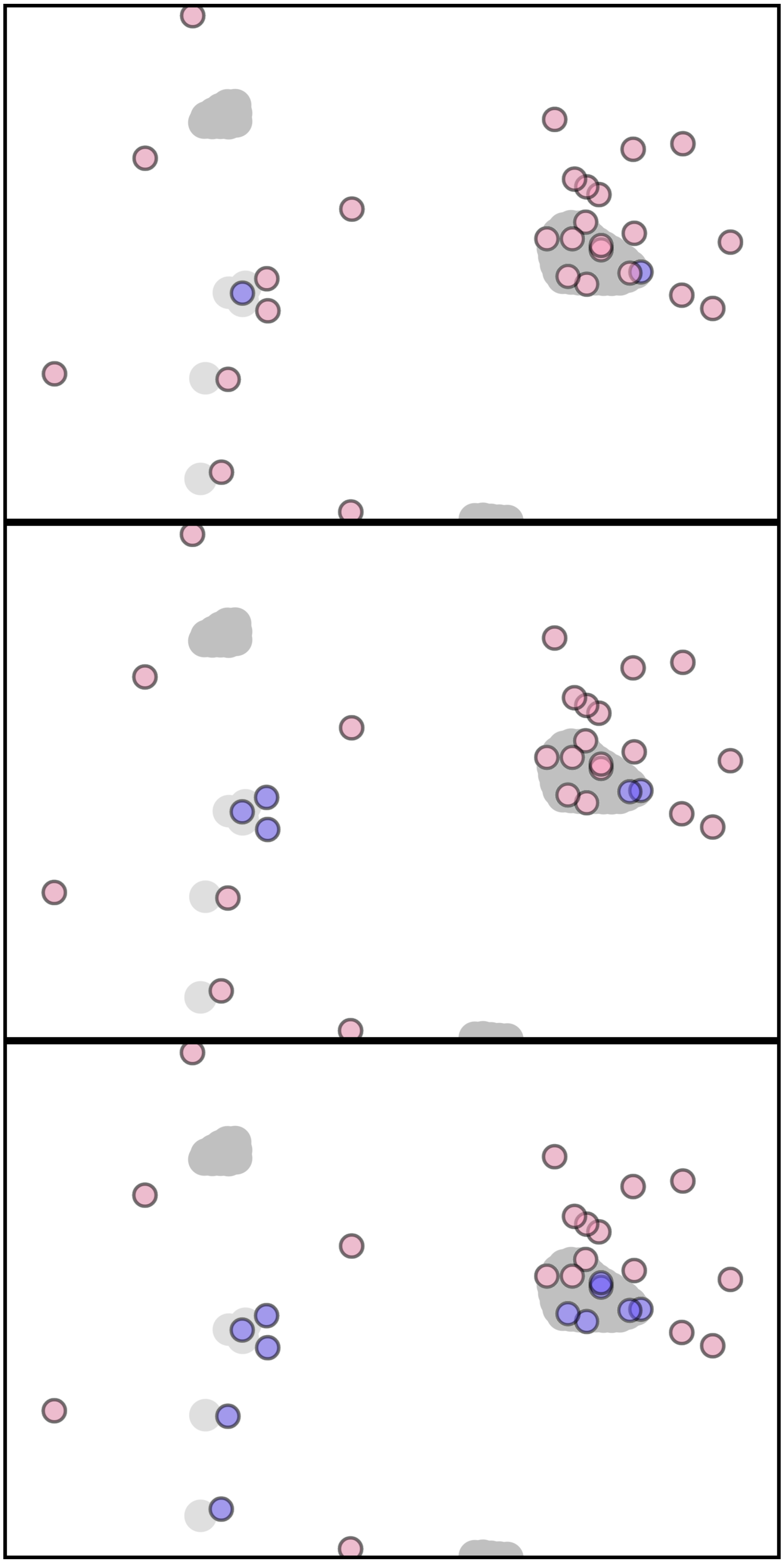

Figure 2.1: Example expansion over time from 2014 to 2016 (top to bottom) using fake data. Blue dots denote stores with Double Up, pink dots denote without. Gray sectors denote higher population density. The initial nodes can be seen in the top (2014) frame.

dimensions: race, income, population density, store attributes, store EBT sales. One variable per dimension was selected: percentage of population that is African-American (zip code level); people per square mile (zip code level); median income for people who have received SNAP or similar assistance (zip code level); the number of associates employed in each store; and the percentage of total stores sales attributed to EBT/SNAP.

Of the 6 stores (stores `12 - 17`), only 3 produced viable matches. However, each of the 3 matched stores had matched to more than one control stores. The closest stores, by driving distance, were selected as the tie-breaker for each matched store. Stores were sufficiently far apart, with very sparsely populated areas between, that "spill-over" was considered unlikely. That is, it is considered unlikely that a shopper near a store without Double Up would opt to drive 30 or more minutes to shop at the store *with* Double Up.

This left 12 stores to be allotted to the control group and 3 treatment stores to be effectively discarded.

2.0.0.2 Step 2: Scoring via Linear Probability Model

Assignment to treatment and control can be perfectly determined since we know and observe the criteria used for assignment: geographic distance from an initial store (node), SNAP EBT sales rank, and demographics—specifically population density and percentage African-American. A scoring function was created by fitting a linear probability model to all stores within 140 kilometers of the two initial pilot stores.

$$\begin{aligned} \boldsymbol{s} &= P(\mathbf{D} \widehat{= 1|\boldsymbol{X}}, \boldsymbol{N}) \\ &= \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\alpha}\mathbf{N} + (\mathbf{X} \odot \mathbf{N})\,\hat{\boldsymbol{\gamma}} \end{aligned}$$

$\boldsymbol{s}$ are the fitted values of the estimated linear probability model; $\mathbf{D} \in \{0, 1\}$ is a $n \times 1$ vector of

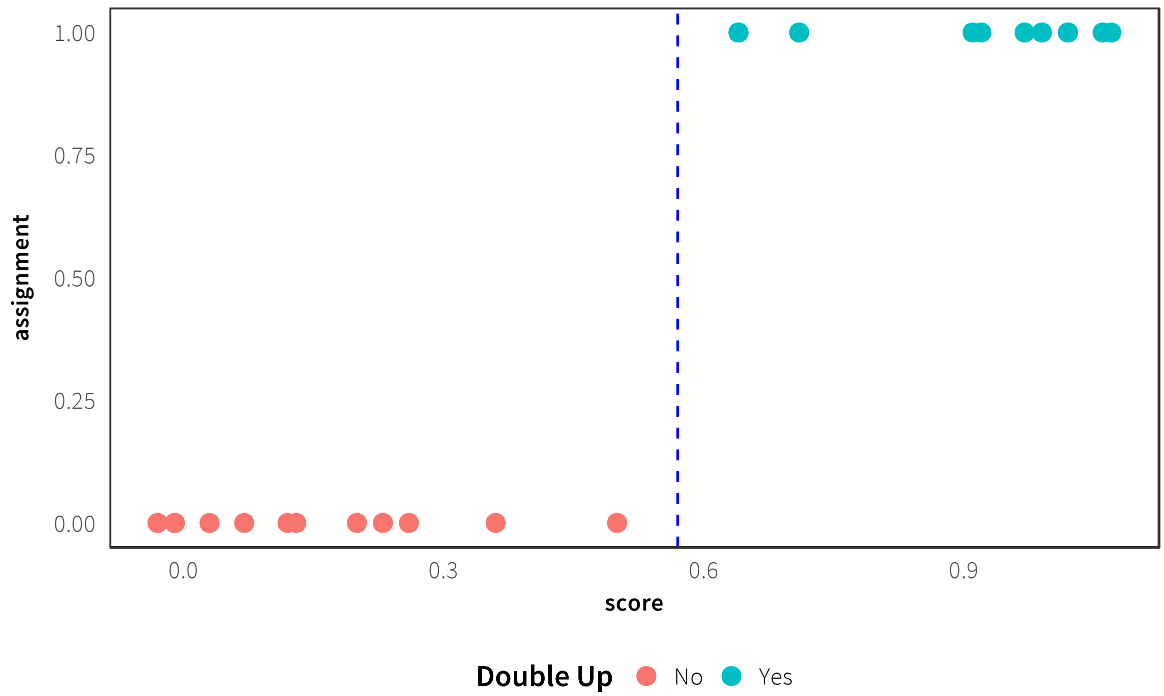

Figure 2.2: Store Score vs Double Up Assignment

store assignments to Double Up; $\mathbf{X}$ is an $n \times k$ matrix of normalized observable covariates that determine assignment; $\mathbf{N} \in \{0, 1\}$ is an $n \times 1$ dummy vector denoting the closest pilot store aka "Node", where 0 is `Node 0` and 1 is `Node 1`. $\odot$ represents element-wise multiplication aka "Hadamard product".

Stores were sorted by the fitted values of the model, $\boldsymbol{s}$. There is perfect separation between Double Up stores and those without (see Figure 2.2). Therefore, the top 11 stores by score value are all Double Up stores. The next 12 stores by score value are then allotted to the control group.

Bibliography

An, R. (2015). Nationwide expansion of a financial incentive program on fruit and vegetable purchases among Supplemental Nutrition Assistance Program participants: A cost-effectiveness analysis. *Social Science & Medicine*, 147:80–88.

Andreyeva, T., Long, M. W., and Brownell, K. D. (2010). The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food. *American Journal of Public Health*, 100(2):216–222.

Andreyeva, T., Tripp, A. S., and Schwartz, M. B. (2015). Dietary Quality of Americans by Supplemental Nutrition Assistance Program Participation Status: A Systematic Review. *American Journal of Preventive Medicine*, 49(4):594–604.

Bartlett, S., Klerman, J., and Olsho, L. (2014). Evaluation of the Healthy Incentives Pilot (HIP): Final Report. Technical report, Prepared by Abt Associates for the U.S. Department of Agriculture, Food and Nutrition Service.

Belot, M., James, J., and Nolen, P. J. (2014). Incentives and children's dietary choices: A field experiment in primary schools.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *The American economic review*, 96(5):1652–1678.

Blumenthal, S. J., Hoffnagle, E. E., Leung, C. W., Lofink, H., Jensen, H. H., Foerster, S. B., Cheung, L. W., Nestle, M., and Willett, W. C. (2014). Strategies to improve the dietary quality of Supplemental Nutrition Assistance Program (SNAP) beneficiaries: An assessment of stakeholder opinions. *Public Health Nutrition*, 17(12):2824–2833.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43:244–268.

Cawley, J. and Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics*, 31(1):219–230.

Chatterjee, A., Kubendran, S., King, J., and DeVol, R. (2014). Checkup Time: Chronic Disease and Wellness in America. Technical report, Milken Institute.

Dinour, L. M., Bergen, D., and Yeh, M.-C. (2007). The Food Insecurity–Obesity Paradox: A Review of the Literature and the Role Food Stamps May Play. *Journal of the American Dietetic Association*, 107(11):1952–1961.

Edin, K., Melody Boyd, James Mabli, Jim Ohls, Julie Worthington, Sara Greene, Nicholas Redel, and Swetha Sridharan (2013). SNAP Food Security In-Depth Interview Study. Technical report, USDA FNS, Office of Research and Analysis.

Frazão, E. (1999). High costs of poor eating patterns in the United States. 732:32–1.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *The Journal of Economic Perspectives*, 25(4):191–209.

Gundersen, C. (2015). SNAP and Obesity. *SNAP Matters: How Food Stamps Affect Health and Well Being*, pages 161–185.

Hartmann, M. (2011). Corporate social responsibility in the food sector. *European Review of Agricultural Economics*, 38(3):297–324.

Hoynes, H. W., McGranahan, L., and Schanzenbach, D. W. (2015). SNAP and Food Consumption. *SNAP Matters: How Food Stamps Affect Health and Well-Being*, page 107.

Iacus, S. M., King, G., and Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, page mpr013.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., and Volpp, K. G. (2011). Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial. *Journal of General Internal Medicine*, 26(6):621–626.

Klerman, J. A., Bartlett, S., Wilde, P., and Olsho, L. (2014). The Short-Run Impact of the Healthy Incentives Pilot Program on Fruit and Vegetable Intake. *American Journal of Agricultural Economics*, 96(5):1372–1382.

Leung, C. W., Hoffnagle, E. E., Lindsay, A. C., Lofink, H. E., Hoffman, V. A., Turrell, S., Willett, W. C., and Blumenthal, S. J. (2013). A Qualitative Study of Diverse Experts' Views about Barriers and Strategies to Improve the Diets and Health of Supplemental Nutrition Assistance Program (SNAP) Beneficiaries. *Journal of the Academy of Nutrition and Dietetics*, 113(1):70–76.

List, J. A., Samek, A. S., and Zhu, T. (2015). Incentives to Eat Healthy: Evidence from a Grocery Store Field Experiment. SSRN Scholarly Paper ID 2664818, Social Science Research Network, Rochester, NY.

Loewenstein, G., Price, J., and Volpp, K. (2016). Habit formation in children: Evidence from incentives for healthy eating. *Journal of Health Economics*, 45:47–54.

Mente, A., de Koning, L., Shannon, H. S., and Anand, S. S. (2009). A Systematic Review of the Evidence Supporting a Causal Link Between Dietary Factors and Coronary Heart Disease. *Archives of Internal Medicine*, 169(7):659–669.

National Center for Health Statistics (2015). *Health, United States, 2014: With Special Feature on Adults Aged 55–64*. Health, United States. National Center for Health Statistics (US), Hyattsville (MD).

Network, F. F. (2014). *Double Up Food Bucks: A Five-Year Success Story*. Fair Food Network.

Nguyen, B. T. and Powell, L. M. (2015). Supplemental nutrition assistance program participation and sugar-sweetened beverage consumption, overall and by source. *Preventive Medicine*, 81:82–86.

Nutrition Evidence Library (2014). A Series of Systematic Reviews on the Relationship Between Dietary Patterns and Health Outcomes. Technical report, U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, Alexandria, VA.

Olsho, L. E., Klerman, J. A., Wilde, P. E., and Bartlett, S. (2016). Financial incentives increase fruit and vegetable intake among Supplemental Nutrition Assistance Program participants: A randomized controlled trial of the USDA Healthy Incentives Pilot. *The American Journal of Clinical Nutrition*, 104(2):423–435.

Todd, J. E. and Ploeg, M. V. (2014). Caloric Beverage Intake Among Adult Supplemental Nutrition Assistance Program Participants. *American Journal of Public Health*, 104(9):e80–5.

USDA (2015). Scientific Report of the 2015 Dietary Guidelines Advisory Committee. Technical report, USDA.

USDA FNS (2016a). SNAP Retailer Data. `http://www.fns.usda.gov/snap-retailer-data`.

USDA FNS (2016b). Supplemental Nutrition Assistance Program (SNAP). `http://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap`.

USDA NIFA (2015). USDA Awards $31 Million in Grants to Help SNAP Participants Afford Healthy Foods. `https://nifa.usda.gov/resource/usda-awards-31-million-grants-help-snap-participants-afford-healthy-foods`.

Wolfson, J. A. and Bleich, S. N. (2015). Fruit and vegetable consumption and food values: National patterns in the United States by Supplemental Nutrition Assistance Program eligibility and cooking frequency. *Preventive Medicine*, 76:1–7.