NBER WORKING PAPER SERIES

MAKING NUTRITIONAL INFORMATION DIGESTIBLE: EFFECTS OF A RECEIPT-BASED INTERVENTION ON RESTAURANT PURCHASES

Kelly Bedard
Peter J. Kuhn

Working Paper 19654
http://www.nber.org/papers/w19654

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2013

Funded by NIH grants R21 DK075642 and 3R21DK075642-02S1. We thank Kyle Dean, Jay Ferro, Molly Chester and Nitin Pai for their patience and cooperation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Making Nutritional Information Digestible: Effects of a Receipt-Based Intervention on Restaurant Purchases
Kelly Bedard and Peter J. Kuhn
NBER Working Paper No. 19654
November 2013
JEL No. C33,D03,D12,I12

ABSTRACT

We study the effects of receipts that include personalized ordering suggestions designed to reduce fat and calorie consumption on purchasing behavior at a restaurant chain. We find that customers, in the aggregate, made most of the item substitutions that were encouraged by the messages, such as substituting ham for sausage in a breakfast sandwich, or substituting frozen yogurt for ice cream, though effects on overall calories and fat consumed were small. The results illustrate the potential of emerging information technologies, which allow retailers to tailor product marketing to individual consumers, to contribute in meaningful new ways to the battle against obesity.

Kelly Bedard
UCSB
kelly@econ.ucsb.edu

Peter J. Kuhn
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and NBER
pjkuhn@econ.ucsb.edu

1. Introduction

Countries throughout the developed world are concerned about rising obesity rates. By 2007-2008, 34 percent of American adults and 17 percent of American children were obese, double the shares in 1980.[1] While there is no shortage of proposed explanations for this trend—ranging from changes in food prices and increasingly sedentary lifestyles to technological change in food production—a popular candidate is consumption of restaurant food, particularly fast food.[2] Part of this explanation's appeal lies in the fact that the share of American food dollars spent on meals prepared outside the home rose from 26 percent to 46 percent from 1970 to 2003.[3]

In response to concerns about the link between obesity and restaurant food consumption, a number of U.S. jurisdictions have recently enacted mandatory calorie posting laws. In California, for example, restaurants with more than ten sites must now post calorie information on menus and/or menu boards. Despite this increase in regulations, there is still little evidence that providing such information changes purchasing patterns. The two best known studies find that calorie posting may have a small effect on calories purchased: Bollinger, Leslie and Sorenson (2011) find that calories per transaction at Starbuck stores fell by 6 percent after New York City mandated calorie posting at all chain restaurants. Wisdom, Downs, and Loewenstein (2010) found that randomly providing calorie information to a sample of Subway customers reduced calories per meal by about 8 percent. Both sets of authors, however, acknowledge that the interventions and settings they study are special, and that the effects identified may not be large or permanent enough to affect overall obesity levels. In consequence, additional, effective tools for improving food choices at restaurants may be needed.

The intervention studied in this paper is called the *Nutricate* receipt and was designed by *SmartReceipt* Corporation. In June 2009, it was implemented at a trial store in Burgerville, a restaurant chain in the Pacific Northwest. A distinct feature of this intervention is that it is printed on customers' receipt*s*; thus in principle it could be implemented either as a substitute for, or as a complement to calorie posting on menus. While at first glance a receipt-based intervention might not seem promising—it relies on consumer memory and can only affect consumption on future restaurant visits—receipt-based interventions also have some potential advantages. Of these, probably the most promising is the capability to deliver targeted purchase suggestions promoting healthier products that are close substitutes to what the consumer just bought.[4]

---

[1] Flegal et al (2010); CDC (2011).
[2] See Lakdawalla and Philipson (2002); Cutler, Glaeser, and Shapiro (2003); Lakdawalla, Philipson, and Bhattacharya (2005); Bleich et al. (2008) for assessments of various candidate hypotheses. Anderson and Matsa (2011) find little evidence for an effect of restaurants; as do Davis and Carpenter (2009) and Currie at al. (2010) who focus specifically on restaurants' effects on children.
[3] See USDA-ERS 2003 and NRA 2004. More importantly, the proportion of fat and calories coming from foods consumed away from home has risen substantially. In 1995, foods away-from-home provided 34 percent of total caloric intake and 38 percent of fat intake, nearly double the 18 percent caloric and fat intake in 1977-1978 (USDA-ERS 2000).
[4] In addition, restaurants and customers may consider receipt-based interventions less intrusive than menu-based ones.

Using store-level weekly purchase data from all 39 restaurants operated by Burgerville over a 125-week period, we estimate the effects of the *Nutricate* Receipt treatment on a variety of outcomes. Aside from the distinct nature of the *Nutricate* Receipt intervention, novel features of our study include a two-stage econometric approach that, in addition to accounting for store and week fixed effects, allows for persistent unobserved store-specific shocks. In addition, since we have 38 possible control stores, some of our estimates use the method proposed by Abadie, Diamond, and Hainmueller (2010) to construct an optimal synthetic control group, whose pre-treatment sales trends match the treatment store's more closely than those of an arbitrarily-selected set of stores.

Under our preferred specification, which applies both the above innovations, we do not detect a statistically significant effect of the *Nutricate* Receipt intervention on total calories per transaction at Burgerville. We do, however, estimate a reduction in average cholesterol per transaction of 2.1 percent. More significantly, we find robust evidence that consumers shifted their mix of purchases in the directions encouraged by the most common purchasing suggestions. For example, the share of adult main course items requesting "no sauce" increased by 6.8 percent, the share of kids' meals with apples (instead of fries) rose by 7.0 percent, and the share of breakfast sandwiches without sausage increased by 3.8 percent**.** Together, these patterns suggest that a significant share of fast food customers implemented their personalized ordering suggestion on future restaurant visits.

What mechanisms might account for the observed effects of ordering suggestions in our data? Since ordering suggestions on a receipt can only affect consumption on future purchases, we think it is unlikely that simple salience—i.e. reminding consumers of facts and consequences they already know—plays a major role. Instead, we propose that Nutricate receipts work because they provide information in a way that is personally relevant, and then combine this information with simple action suggestions. To see this point, it is worth noting—as the restaurant lobby has argued—that menus can be complex items, especially when customers can customize their orders. For example, even a simple deli sandwich can be ordered in over 4000 ways; selecting the optimal combination of taste and nutrition from this many choices can be cognitively demanding.[5] Faced with this complexity, consumers can respond in a number of suboptimal ways, including some heuristics that can be highly ineffective. For example, two recent experimental studies—Iyengar and Lepper (2000) and Bertrand et al (2010)—find that expanding consumers' choice sets led to choice *avoidance*: consumers were less likely to buy the product at all when offered a larger menu of possible product types. Relatedly, Hirshleifer, Lim, and Teoh (2009) provide highly suggestive evidence that information overload (in the form of a large number of earnings announcements) slows price adjustments in stock markets.[6]

[5] If a sandwich consists of one choice from each of 4 meats, 4 breads, 4 cheeses [including none], 4 spreads (e.g. mayo, mustard, butter, none) lettuce (yes/no), tomato (yes/no) and 4 condiments (olives, peppers, etc.) the total number of possible sandwiches is 4,096.

[6] Rubinstein and Salant (2006) review an extensive psychological literature demonstrating humans' reliance on simple, suboptimal heuristics when making choices from lists, such as restaurant menus. See also DellaVigna (2009) for an excellent recent review of the broader literature on psychological determinants of consumer choices.

In sum, we hypothesize that simple suggestions to make a health-improving switch from something the consumer recently chose to a close substitute may circumvent many of the cognitive difficulties associated with making choices from lists, such as menus. Further, since customized receipts are likely an early stage in the development of information technologies that allow retailers to tailor product marketing to individual consumers, our evidence that fast-food customers respond to tailored, health-improving suggestions suggests that these new information technologies may have the potential to contribute meaningfully to the battle against obesity.[7]

2. Food Labeling and its Effects

Long before the current wave of state and local mandates requiring restaurant menu labeling, the 1990 Nutrition Labeling and Education Act (NLEA) established mandatory nutrition labeling for packaged foods. The cornerstone of the NLEA is the Nutrition Facts Panel (NFP), which provides a variety of information including total calories, calories from fats, and amounts of other nutrients per serving and as a percentage of a recommended daily value. Until recently, restaurants were exempt from these labeling requirements. However, pressure to extend nutrition labeling to restaurants has increased in response to restaurants' expanding role in the American diet. Since New York introduced mandatory calorie posting in 2008, a number of other jurisdictions including California, Seattle, and Philadelphia followed suit. At the federal level, the Federal Drug Administration is currently drafting regulations mandated by the 2010 Patient Protection and Affordable Care Act. The proposed regulations would require restaurants with 20 or more locations to post calories on all menus and menu boards, clearly and prominently, adjacent to the number of calories.

Throughout this process of expanding legislation, the National Restaurant Association (NRA) has consistently opposed attempts to require posting of nutritional information on menus and menu boards.[8] The NRA argues that this method of nutritional disclosure presents many costly logistical and operational constraints. Among these is the fact that 70 percent of customers customize their restaurant meals; as noted this poses significant difficulties in providing accurate information in an easily-digestible form.[9] Other difficulties include the fact that restaurants have daily specials and that the FDA and other organizations periodically make changes to their nutrient recommendations, which would increase costs. Also, posting nutritional information poses special challenges for drive-through customers, who account for 60 percent of quick-service restaurant sales (NRA 2004; NRA letter to FDA 2003). Many of these costs associated with menus are less of a concern in the case of receipt-based interventions,

---

[7] Put a different way, menu-based interventions assist consumers by providing new information about calorie content, and by providing *salience*: before the customer orders, he/she is reminded of the calorie content in a meal he/she is considering. Receipt-based interventions provide little salience: the information they contain cannot be acted on till the next restaurant visit. On the other hand, they provide both information and *guidance*, i.e. assistance with the cognitively-demanding task of choosing from a large set of options. Guidance takes the form of suggesting a potentially welfare-improving perturbation to the consumer's previous choice.

[8] That said, there is some support by the restaurant association for a uniform federal law that would replace the expanding patchwork of state and local regulations.

[9] In the case of the deli sandwich consisting of 7 items in 4096 combinations, providing complete information would require the restaurant either to list values for all 4096 combinations, or to post the information for each of the 24 possible components (4+4+4+4+2+2+4=24) and rely on the consumer to add up seven two- or three-digit numbers.

which do not occupy new physical space and can be easily reprogrammed. Thus, receipt-based interventions may be a useful complement *or* substitute for menu-based ones.[10]

As noted, the key existing studies of calorie posting are Bollinger, Leslie and Sorenson (2011); and Wisdom, Downs and Loewenstein (2010).[11] Bollinger, Leslie, and Sorenson (2011) use internal company data from Starbucks to study the reaction of Starbucks' customers to a mid-2008 law that required *all* chain restaurants in New York City to post calories on menus or menu boards. While average calories per transaction fell from 247 to 236, this effect was entirely driven by the small fraction of consumers purchasing food—there was no decline in purchased drink calories. Whether or not the reduction in calories at Starbucks constitutes a large or a small impact is to some extent in eye of the beholder. The 11 calorie decline is statistically significant and corresponds to a 6 percent decrease, but it constitutes a small fraction of daily calories.[12]

Wisdom, Downs, and Loewenstein (2010) designed a pair of field experiments where a relatively small number of Subway customers were randomly assigned to different types of menus.[13] The context was one in which no restaurants operating in the market were required to post nutrition information. Pooling their two studies, Wisdom et al.'s results suggest that calorie information reduces calories by approximately 7 percent, although many of their point estimates are imprecisely estimated. Their results also suggest that a different intervention that made healthy choices more convenient (by making them a 'featured option') could reduce ordered calories, depending on the format.

Other survey and receipt collection studies come to similar conclusions. Elbel et al. (2009) collected receipts from guests outside of fast-food chain restaurants, before and after calorie posting in New York City, using Newark NJ stores as controls. They could detect no change in calories purchased.[14] Finally, in a study design similar to ours, Finkelstein et al. (2011) studied the effects of mandatory calorie posting in King County, WA using monthly sales data from 28 TacoTime restaurants. Seven of these restaurants were near but not inside King

---

[10] For more detail on the *Nutricate* receipt and its possible advantages for restaurants, see http://receipt.com/nutrition_solutions.php

[11] There is also a small literature that studies the effect of packaged food labeling under the NLEA. For an example and review, see Variyam and Cawley (2006). Using National Health Interview Survey data, they study the effects of packaged food labeling by comparing the change before and after the implementation of NLEA in body weight among those who use labels when food shopping to that among those who do not use labels. They report a relative decrease in body weight among the label-users relative to non-users.

[12] Bollinger et al.'s regression tables do not indicate whether their standard errors are clustered or whether other adjustments were made for within-group error correlations. In footnote 27, they report that their results are robust when they account for serial correlation by aggregating all transaction data before calorie posting and all transaction data after calorie posting, then testing for a before-after difference in calories per transaction. Assuming the aggregation was done by store, these tests would then require the 316 store-level pre-post differences in their data to be statistically independent across stores. Most of our estimated standard errors do not rely on this assumption.

[13] The inducement to participate was a free meal; according to the authors "more than half" of the customers who were approached agreed to participate.

[14] Dumanovsky et al. (2011) conduct a similar study, but using data from New York restaurants only; they found modest reductions in calories purchased in some specifications, but interpretation of these differences as causal is problematic due to the absence of a control group. Bassett et al. (2008) show that Subway consumers who reported seeing posted calorie information purchased few calories than other Subway consumers; inferring causality is difficult here as well.

County, and served as controls. Their econometric approach does not appear to include store fixed effects, or to adjust standard errors and optimize the control group in the ways we do here. They find no effect of menu labeling on calories purchased.

To our knowledge, ours is the only study to estimate the effect of *receipt-based* interventions on restaurant food purchases. In its general form, the *Nutricate* receipt consists of three components: information, motivation and recommendations. The information component displays each item's relevant nutritional data, as ordered. Thus, if a consumer asked to "hold the mayo," on her order, the nutritional information provided on the receipt will reflect this. Receipts can also contain *motivational* statements, aimed to remind consumers of the benefits of a healthier diet, such as "A diet low in total fat may reduce the risk of heart disease." Finally, and of greatest interest to us in this study, are customized *recommendations* for how a customer can improve his/her nutrition, by making small changes to the items the customer just ordered. For example, one message used at Burgerville says, "Did you know if you held the mayonnaise on a burger you would cut your fat intake by 57%, or 15 grams?" In our analysis, we pay special attention to the effects of these recommendations, both because they are unique to receipt-based interventions and because they might offer an additional policy tool in improving customers' food choices. Appendix A shows an example receipt from a Burgerville restaurant.

3. Data and Descriptive Statistics

Our data consist of weekly purchase information for all 39 restaurants operated by Burgerville for the 125-week period running from January 2008 to mid-May 2010. Beginning on June 4, 2009, the receipts at a single store (henceforth the "treatment" store) were changed from a conventional sales receipt to the *Nutricate* receipt. During our study period, no form of mandatory calorie posting was required in any jurisdiction containing a Burgerville store.[15] Overall, we therefore have a difference-in-differences, or "comparative case study" design with pre- and post-treatment information on one treated store and 38 potential control stores.

In the Burgerville implementation, aside from promoting healthier choices more generally, the recommendations in the *Nutricate* receipts focused on encouraging nine main changes in ordering patterns. These were, (1) to substitute to a side salad for any adult-size fries[16]; (2) to substitute grilled chicken for fried chicken in sandwiches; (3) to substitute apples for fries in kids' meals; (4) to substitute milk for soda or juice in kids' meals; (5) to substitute frozen yogurt for ice cream; (6) to substitute any other kind of breakfast sandwich (ham, bacon, cheese, or egg only) for the breakfast sandwich with sausage; (7) and (8) to 'hold the sauce' or 'hold the cheese' on any sandwich; and (9) to substitute a meal-sized (entrée) salad for other main dishes. These changes were motivated by the notion that the proposed replacement item was both a reasonable substitute for the purchased item and significantly lower in calories and

[15] Burgerville has stores in Washington State, but none in King County, which mandated fast-food calorie posting during our study period.

[16] Throughout this paper "fries" includes both french fries and onion rings. Onion rings are only available seasonally; when they are introduced purchases of fries fall precipitously but the total of fries plus rings remains essentially unchanged. The opposite occurs when onion rings are removed from the menu.

fat.[17] While most customers likely know that salads are healthier than fries, some of the information provided by these suggestions (e.g. the fact that bacon—or at least the amount of bacon in a Burgerville breakfast sandwich—has less fat and calories than sausage, or that frozen yogurt has many fewer calories than ice cream) could be news to some customers. Clearly, the estimated effects of messages in our study will incorporate the effects of new information of this nature in addition to their role in mitigating the cognitive challenges associated with menu choice.

Throughout this paper, we report results for three types of outcomes. First, following the existing literature, we focus on the total nutritional content --specifically calories and fat-- of a typical transaction at Burgerville. Next, to shed some light on the effects of targeted suggestions, we study the share of transactions that included an item that was (a) encouraged, or (b) discouraged in the above messages. Finally, with the same goal in mind we focus on specific item classes (e.g. frozen desserts or breakfast sandwiches) within which the messages encouraged consumers to substitute one item for another and ask whether those substitutions occurred.

Table 1 reports summary statistics for nutrition per transaction in the treatment store, separately for the pre- and post-intervention periods (January 2008 through June 3, 2009, and June 4 2009 through mid-May 2010 respectively). Also reported is the equally weighted average across control stores for the same periods. In order to protect confidential aspects of Burgerville transaction data, all data in the paper are normalized by the mean for the treatment store in the pre-treatment period. The one exception is column 1 in Table 1, which reports some pre-treatment means to provide context. All summary measures are based on weekly sales data.[18] The average transaction at the treatment store in the pre-treatment period included 1657 calories and 80 grams of fat.[19] In comparison, the non-treated stores sold 3.7 (4.4) percent more calories (fat) per transaction. In the post-treatment period, average calories and fat fell by about 2 percent in both treatment and control stores. According to the simple difference-in-difference estimates in column 6 (which simply compare the change in the treated and untreated stores), the treatment seems to have had essentially no effect on calories or total fat per transaction, though it appears to be associated with a 2.7 percent reduction in cholesterol per transaction. If anything, the treatment is associated with a small increase in store revenue.[20]

In Table 2 we turn our attention to the menu items that were encouraged or discouraged in the intervention's overall messaging strategy. Since the receipts contained a variety of messages, Table 2 defines "discouraged" items as items that were systematically discouraged in a number of messages, but were never encouraged in any message. Encouraged items are defined

---

[17] Some of the item substitutions encouraged in the Burgerville messages change the cost of the consumer's order; for example substituting a side salad for fries costs 40 cents. Importantly, in any given week this price differential is the same across all stores because Burgerville menus and pricing are set at the corporate level.

[18] As both the Donald and Lang (2007) and Abadie, Diamond, and Hainmueller (2010) methods require a balanced panel, we linearly interpolated a small number of missing data points.

[19] Note that this is more than six times the pre-treatment average of 247 calories at Starbucks: Interventions at traditional fast-food establishments have much more potential to affect daily calorie intake than at stores specializing in coffee and other drinks.

[20] Assessing whether the differences in differences in Tables 1-3 are statistically significant raises a number of issues which we treat in detail in the following section. Accordingly we do not report significance levels here.

analogously; examples of messages corresponding to each of the discouraged and encouraged items are provided in Appendix B.[21] The numbers in Table 2 are the share of weekly transactions that include the (encouraged or discouraged) item in question. Any apparent treatment effects in Table 2 thus refer to the impact of the *overall* messaging strategy in the Burgerville receipts on the share of transactions that include these encouraged or discouraged items. As in Table 1 and in all subsequent tables, the numbers in Table 2 are normalized to equal one in the treatment store in the pre-treatment period. While purchases of some discouraged items (notably breakfasts including sausage) fell precipitously in the treatment store relative to the control store, and some encouraged items (such as side salads, kids' apples, and kids' milk) rose sharply.

Finally, in Table 3 we focus on substitution between encouraged and discouraged items within six item classes: adult sides (which can be either fries or a side salad); chicken sandwiches (which can be either grilled or fried); kids' meals (which can be ordered with fries or apples, and with soda/juice or milk); desserts (which can contain either ice cream or yogurt); breakfast sandwiches (which can include sausage or not); and adult main dishes (which can include or exclude cheese and sauce, and can be either a sandwich or a salad). In contrast to Tables 1 and 2, here the difference-in difference estimates are all in the expected direction and generally substantial in magnitude. For example, the share of salads in adult sides rose by 7.4 percent, with similar increases in the share of kids' meals that included apples and milk. The share of frozen yogurt in desserts rose by 8.4 percent. Consumers also shifted away from sausage in their breakfast sandwiches, and were more likely to choose main dishes without cheese or sauce. The share of main dishes that were salads rose by eight percent.

4. Econometric Framework

As noted, our data consist of 125 weekly observations on 39 stores, one of which changes treatment status during our sample period. A common way to estimate the effects of the treatment in cases like this is to estimate the following fixed-effects regression on the full sample of 125*39 = 4875 observations:

$$Y_{sw} = \varphi_s + \gamma_w + \delta T_{sw} + \varepsilon_{sw}, \tag{1}$$

where Y is the outcome of interest, and φ_s and γ_w denote store and week fixed effects respectively. T is a treatment indicator which equals one in the treatment store starting in week 76, and zero in all other store-week cells. Together, the fixed effects in this specification absorb an arbitrary pattern of time-invariant store characteristics and an arbitrary pattern of time (and season) effects that is common across stores. Following Bertrand, Duflo, and Mullainathan (2004), it is standard practice to cluster the standard errors at the store level. We refer to estimates of (1) with store-clustered errors as one-stage fixed-effects (FE) estimates, and use them as a starting point for our analysis.

[21] Some items were encouraged in some messages, but discouraged in others, as part of a 'chain' intended to move consumers towards lower-calorie items in small steps. We do not study consumption of these items because it is theoretically unclear whether aggregate consumption of these items should rise or fall as a result of the treatment.

a) Alternative standard errors

Donald and Lang (2007) argue that the above one-stage FE approach can seriously underestimate standard errors when the number of groups is small. While superficially we have 39 stores, in another sense we have one treated entity and a single, imperfectly defined, control group. To address this concern we use the two-stage approach suggested by Donald and Lang, as applied in their paper to two well known studies (Card 1990, and Gruber and Poterba 1994). Specifically, Card's Mariel boatlift study has annual labor market outcome data from one treatment city (Miami) and four control cities over a period of seven years. In that context, the first step of Donald and Lang's approach is simply to calculate seven cross-sectional differences between the outcome (say, the unemployment rate) in Miami and the mean of the other four cities. In the second stage, Donald and Lang then regress this difference on an indicator variable for the post-treatment period. Under the assumption that the difference between the annual unemployment rate in Miami and the comparison cities is subject to an *iid* shock, the significance of the resulting coefficient can then be assessed using a t-statistic with four degrees of freedom.[22]

Applying this procedure to our context, we first collapse the 38 control stores into a single unweighted average, then calculate 125 weekly differences between the treatment store's outcome and the mean outcome in the control stores.[23] Denoting these 125 differences by D_w, the second stage of the DL procedure regresses them on a dummy variable for the 50 post-treatment weeks, P_w:

$$D_w = \alpha + \delta P_w + \mu_w. \tag{2}$$

By construction, estimates of δ from (2) are numerically equivalent to those obtained from estimating (1) on the full sample of 4875 observations; they therefore control for all the same confounding influences (i.e. common store and week fixed effects). However in at least one important sense the 125 observations in (2) reflect the true number of degrees of freedom in the estimation more accurately. In particular, if unobserved shocks that are idiosyncratic to a store are *iid*, then (with 125 observations) we can appeal to standard asymptotics in interpreting the standard errors from this regression. In what follows, we refer to estimates and standard errors computed this way as DL estimates.

It seems highly likely, however, that an individual store's sales in one week may be correlated with its own sales in recent weeks, for example due to local events and conditions that last longer than a week. Thus treating the 125 differenced observations in (2) as *iid* is still likely to underestimate our standard errors; indeed because we have weekly data the problem is likely more severe than in Card's or Gruber-Poterba's annual data. To address this issue, we note that the regression in (2) is just a single time series. Therefore, we can allow for unobserved store-specific shocks to have some persistence by calculating Newey-West (1987) autocorrelation-consistent standard errors for this time-series regression (we denote these by NW); this allows

[22] Two degrees of freedom are used to estimate the constant and slope term in the regression, and the year in which the boatlift occurred (1980) is excluded from the sample.

[23] Equivalently, the first stage could be specified as a regression of the differenced outcome on a full set of week effects.

unobserved local conditions that affect stores differently to last for more than one period. In all of our DL-NW estimates, we allow for autocorrelation among observations up to five weeks apart, though the standard errors are not highly sensitive to alternative values of the window allowed. Of course, like the DL estimates described above, the regression coefficients from this DL-NW approach are also numerically equivalent to the one-stage FE coefficients.

b) Constructing a control group

A common question affecting many non-experimental studies using a difference-in-difference design is which of the untreated units should be used as controls. For example, while Card and Krueger's (1994) well known minimum-wage study used Pennsylvania as a control for New Jersey, a number of subsequent studies (such as, for example, Dube et al 2010) have attempted to construct more comparable control groups based on criteria such as geographical distance. It has also become much more common to provide evidence that the treatment and control groups have similar observables, and that their outcomes evolve similarly over time during the pre-intervention period. Still, the question of how to select an appropriate or convincing control group remains subject to considerable discretion. The most widely-used approach uses plausible but arbitrary criteria such as geographical distance to select a set of control sites, and uses an unweighted mean of the control sites to represent the counterfactual for the treated site.

Recently, however, Abadie, Diamond, and Hainmueller (ADH, 2010) have proposed a method for constructing an optimal control group in a comparative case study context when a number of possible control groups are available. The technique seems especially useful in situations such as ours where there is a long time series of pre-intervention data for a substantial number of potential control groups. Essentially, the investigator specifies a set of pre-intervention characteristics he/she wishes to match between the treatment store and the synthetic control; call this vector X_i. X_i can include non-time-varying characteristics of the store, as well as time-varying store characteristics (values of a characteristic at two different pre-intervention dates enter X_i as separate variables). In this way X_i can also include pre-treatment values of the outcome variable Y_{it}; the only restriction on X_i is that it cannot include any variable that might be affected by the treatment. The optimal synthetic control is then defined by a vector of (non-time-varying) store weights w that minimizes the discrepancy between the treated store's X and the weighted mean X of the synthetic control.[24] Typically (and in our application) the vector of weights is restricted to be nonnegative and to sum to one.[25] Informally, then, our synthetic control store is constructed as a weighted average of potential control stores, with weights chosen so that the resulting synthetic store best reproduces the values of a set of predictors of the outcome of interest in the treated store before the implementation of the treatment.

[24] The discrepancy between the treated group's characteristics, X_1, and the weighted mean characteristics of the synthetic control, X_0W, is specified as $\sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$, where V is a positive semidefinite matrix. While the technique can be implemented with any V, a natural criterion is to pick the V that minimizes the mean squared prediction error of the outcome variable during the pre-intervention periods. In this paper we choose V to be the positive definite, diagonal matrix with this property.

[25] The ability to restrict the weights in this way is an attractive feature of the synthetic control method, because it provides a built-in safeguard against unwittingly using linearity assumptions to extrapolate to conditions where observed data are sparse or nonexistent..

In the baseline results reported here, *X* includes only the means of the outcome variable in the pre-treatment period, grouped into fifteen five-week windows.[26] Thus, a different synthetic control is designed for each outcome we study. Finally, note that one can calculate DL and DL-NW standard errors for *any* pair of treatment and control groups, including one constructed by the ADH procedure. Thus, our preferred estimates in this paper –denoted ADH-DL-NW estimates-- first construct an optimal synthetic control using the methods just described separately for each outcome under consideration, then implement the DL-NW procedure on the two resulting time series. Later in the paper, we illustrate the ability of our synthetic control groups to track the pre-treatment outcomes in the treated store. Section 6 also explores the effects of alternative ways to construct control groups, such as geographical distance.

5. Regression Results

The regression results reported in Tables 4-7 all follow the progression described in Section 4. Row 1 shows the one-stage fixed effects estimates; row 2 uses the two-stage Donald-Lang procedure to adjust standard errors, and row 3 implements the Newey-West autocorrelation-consistent standard errors in the DL regressions. As noted, the reported estimates allow for autocorrelation among observations up to five weeks apart, though the standard errors are not very sensitive to alternative values of this window. Rows 4 and 5 repeat the above exercises, replacing the simple mean of the outcome in all the non-treated stores by a weighted mean, where these (non-time-varying) weights are derived using the ADH procedure detailed in the last section.[27] Finally, the bottom panel of the Table replicates these ADH-DL-NW regressions, splitting the post-treatment period into two parts to assess the permanence of the treatment effects.

a) Nutrition and sales:

Table 4 reports the aforementioned results for nutrition and sales. Columns 1-4 estimate treatment effects on calories, fat, and cholesterol per transaction; columns 5 and 6 look at the number of items and revenues per transaction. Columns 7-9 estimate treatment effects for three measures of total sales volume at a store. Comparing point estimates and standard errors across rows 1-3 reveals the effects of aggressively adjusting the standard errors. To illustrate, consider the effects on calories per transaction. Using the standard one-stage FE approach, we estimate a 0.56 percent decrease in calories per transaction as a result of the treatment; this appears to be statistically significant at the 1 percent level. But using the DL difference approach more than doubles the standard error, and further adjusting for first-order autocorrelation using DL-NW increases the standard error to over three times its original level. In consequence, we cannot conclude that the treatment reduced calories per transaction at any conventional level of statistical significance. Similar patterns across specifications are found for all outcomes:

[26] We chose five-week windows in order to match longer-term trends rather than week-to-week within-store variation (which is substantial); the theoretical motivation is that treatment effects should take a few weeks to appear (since the messages can only work on a repeat visit). Results from shorter and longer windows are reported in Section 6. We can think of no theoretical reason to match stores on other pre-treatment observables. Also, matching only on the outcome variable eliminates discretion in choosing which observables (store size, location, mix of items sold, etc.) to match on.

[27] Since there is no obvious counterpart to the FE approach with the synthetic control group, we only present DL and DL-NW results here.

standard errors increase substantially as we move from the conventional FE approach to DL, and again when we move on to DL-NW. Importantly, these adjustments do not reduce all the coefficients to insignificance. In Table 4, only a 2.7 percent decline in cholesterol survives both adjustments, remaining significant at one percent. In subsequent tables, adjusting standard errors in this way plays an important role in distinguishing what are likely genuine effects of the *Nutricate* receipt from chance patterns.

Rows 4 and 5 show that the ADH optimal synthetic cohort approach leads to some changes in the point estimates, though not for the one outcome (cholesterol) that was robustly significant in rows 1-3. Standard errors on the other hand tend to fall slightly relative to the comparable specification in rows 2 and 3. Overall, based on the point estimates reported in Row 5, the introduction of the *Nutricate* receipt reduced cholesterol per transaction by 2.1 percent and had no significant discernible impact on any other outcome. Notably, only one of our measures of sales (items per transaction) was significantly affected by the treatment, and this effect was (small and) positive. Thus, it does not appear that introducing the *Nutricate* receipt had detrimental effects on the treatment store's business.[28]

The effects of constructing a synthetic control group on the control group's ability to track the treatment store before treatment is introduced and to isolate divergences thereafter is illustrated in Figures 1 and 2, which compare time trends in the treatment store to an average of all other stores (Figure 1) or to our synthetic control groups (Figure 2) for the four nutritional outcomes in Table 4.[29] Clearly, the synthetic controls in Figure 2 track the treatment store much more accurately during the pretreatment period, and illustrate a divergence in total fat and cholesterol that seems to emerge after the treatment, consistent with Table 4's estimates. The temporary nature of the total fat effect is also evident. These effects are less visually evident in Figure 1.

Finally, rows 6 and 7 of Table 4 break the post-treatment period into two parts to distinguish the short- and longer-term effects of the *Nutricate* receipt treatment. Interestingly, if we restrict our attention to effects that occur within the first 25 weeks after the introduction of *Nutricate* receipts, we see statistically significant reductions in both cholesterol and total fat per transaction, with little change in the remaining results. Also, the reduction in cholesterol appears to be relatively permanent. A statistically significant positive effect on revenue per transaction emerges, but only 25 weeks after treatment begins.

b) Purchases of Discouraged and Encouraged Items:

We now turn our attention to the impact of the nutritional messages printed on the *Nutricate* receipt. In particular, Tables 5 and 6 report the impact of *Nutricate* receipt introduction for consistently "discouraged" and "encouraged" items, respectively. Beginning with the

---

[28] In additional analysis available from the authors, we also asked whether the five stores located closest to the treatment store experienced any changes in business when *Nutricate* receipts were introduced at the treatment store, since the receipts could conceivably attract or alienate customers and change where they eat. No effect was found.
[29] Since high-frequency (week-to-week) noise makes it hard to visually discern longer-term trends, all our time series graphs in the paper, including Figures 1 and 2, show five-week moving averages (the current week plus four lags). This smoothing is applied only to the graphs; all our statistical analysis is performed on the raw, unsmoothed data.

discouraged items in Table 5 and focusing on the synthetic control estimates reported in Row 5, the only precisely estimated effect is a 13.2 percent decline in sausage breakfast sandwiches. Looking at the timing of these effects in more detail reveals that this effect is relatively long-lasting (i.e. up to 50 weeks). According to row 7, a significant reduction in ice cream sales emerges 25 weeks after the treatment, as does a small and marginally significant increase in fried chicken sandwiches. As we show below these and other similar results may be artifacts of demand shocks to narrower item categories, such as dessert items or kids' meals as a group. Finally, to measure the extent to which the treatment was associated with a broad-based decline in discouraged item consumption, column (6) of Table 5 reports results for an index that combines all five items. To construct this index, we normalized each store's purchases of the item in question to have a mean of zero and a standard deviation of one, then added the five normalized items together. Thus, column (6) shows some evidence of a broad-based decline in discouraged item purchases, though only in the first twenty-five weeks after treatment commenced.[30]

Table 6 reports the companion results for encouraged items. Again, focusing on the synthetic-control estimates (row 5) that are statistically significant at conventional levels, the introduction of the *Nutricate* receipt increased requests for grilled chicken sandwiches by 6.3 percent, for 'no sauce' by 7.0 percent, raised side salad purchases by 13.8 percent, kids' milk by 9.5 percent, but (contrary to expectations) reduced non-sausage breakfast sandwiches by 11.7 percent. Combining all the estimated effects in column 10's index reveals a positive, insignificant overall effect.

Looking at the persistence of effects on individual items in rows 6 and 7 reveals a variety of patterns. Some effects (main and side salads) either fall in magnitude or disappear after the first 25 weeks, while others (grilled chicken, no sauce, no cheese, and kids' milk) seem to strengthen over time (recall that this is not implausible--for the receipt to have an effect, customers need first to be exposed to it on one visit, then act on it in a repeat visit). The frozen yogurt effect actually reverses sign after 25 weeks. Row 7 also reveals that the unexpected negative treatment effect on non-sausage breakfast sandwiches only emerges 25 weeks after treatment; also, combining all the encouraged items into an index reveals a highly significant overall increase in encouraged item purchases during the first 25 weeks after the treatment was introduced. As in Table 5, some of these unexpected patterns could be artifacts of category-specific demand shocks, which we address in Table 7 below.

c) Within-category Item Substitutions:

One possible concern with the previous section's items-per-transaction estimates is that they do not purge the effects of unobserved shocks that are common to items in a category. For example, a store may be affected by a demand shock (such as road construction or the end of the school year) that temporarily reduces breakfast traffic, or a weather shock that raises the demand

[30] In terms of magnitudes, the coefficient indicates that purchases of a 'typical' item fell by .65/5 = .13 standard deviations, since the index is a sum of five items.

for frozen desserts.[31] If this shock is, by accident, correlated with the introduction of our treatment, an items-per-transaction approach would mistakenly conclude that (consistent with the messaging strategy) the treatment reduced the consumption of breakfast sausage when in fact it reduced the consumption of all breakfast items.

To address the possible contaminating effect of category-specific shocks such as these, and to provide more precise tests of whether customers actually made the substitutions recommended in the messages, Table 7 estimates the effects of the treatment on the share of encouraged items within more narrowly-defined item categories. Table 7 considers within-category substitutions towards the nine encouraged items in Table 6. For four of these items (side salads, grilled chicken sandwiches, frozen yogurt, non-sausage breakfast sandwiches) the relevant category includes only that item and the associated discouraged item in Table 5. In the remaining cases data constraints require us to use somewhat broader categories. For example, children's apples and milk are both measured as a share of children's meals, and main course salads, as well as orders with 'no cheese' and 'no sauce' and are modeled as a share of adult main items.

Focusing again on the synthetic control results with Newey-West standard errors in Row 5 of Table 7, the point estimates now all have the expected sign, and all but one are statistically significant. This suggests that some of the unexpected results in Tables 5 and 6 may be driven by unobserved category-specific shocks that are correlated with our treatment. Estimated effect sizes range from a 3.8 percent increase in the share of breakfast sandwiches without sausage to a 14.5 percent increase in the share of adult side dishes that are salads.[32] Column 10 confirms the impression of a broad-based shift from discouraged to encouraged items, with the within-category share of a 'typical' encouraged item rising by 1.74/9 = about 0.19 standard deviations after the treatment is introduced.

Breaking the post-treatment period in half in the last two rows of Table 7 reveals heterogeneous patterns in the duration of these effects, though the dominant pattern seems to be a reduction in the magnitude of the coefficient, in several cases leading to a loss of statistical significance. Indeed, only four of the nine encouraged items exhibit statistically significant increases more than 25 weeks after the treatment: side salads, kids' apples, non-sausage breakfast sandwiches and 'no sauce' requests. Reflecting this, the index of encouraged items shows no evidence of a treatment effect after 25 weeks.

The results of Table 7 are illustrated visually in Figure 3, which shows time trends for the treatment store versus synthetic controls for all ten outcomes studied in the Table. Consistent with row 6 of the Table, the treatment store's sales of side salad, grilled chicken, kid's milk, frozen yogurt, non-sausage bagels, meal salads, and 'no sauce' requests all show a short-term rise relative to the controls in the 25 weeks after the treatment. Also consistent with Table 7, a number of the above effects either vanish or significantly attenuate after 25 weeks—specifically,

---

[31] Note also that, for example, shocks to non-breakfast business can also affect estimates for breakfast items in the items-per-transaction approach: a rise in afternoon and evening business would reduce breakfast sausage per transaction without implying a treatment effect of the receipt.

[32] Note that the latter increase is from a small base, since –as in most hamburger chains-- the vast majority of adult sides are fries.

those for grilled chicken, kid's milk, frozen yogurt, non-sausage bagels and meal salads. Altogether, Table 3 and Figure 7 together provide strong evidence that consumers initially tried out the substitution suggestions in the *Nutricate* receipts, but evidence of longer-term changes in consumption patterns is more mixed.

6. Robustness Checks

a) Alternative Control Stores

So far, we have presented estimates using all non-treated stores as the control group, and estimates for a synthetic control group, selected to match the entire pre-treatment evolution of the outcome variable as well as possible. In Table 8 we explore the effects of using other definitions of the control group, focusing on the results for within-category shares in Table 7. Rows 1-3 explore the effects of different constructions of the synthetic control group, by matching on 3- and 7-week averages of the pre-treatment outcome instead of a 5-week average. For the most part, this has very little effect on the estimated coefficients, though the levels of statistical significance are highest in our baseline case of five-week averages.

Rows 4-6 of Table 8 show the effects of two alternatives to the approach of using a simple average of all the nontreated stores as a control group. Row 5 just uses the five stores closest in size (measured by total revenues) to the treatment store, while Row 6 uses the more common criterion of geographical distance, selecting the five closest stores. (Row 4 reproduces our baseline estimates from row 3 of Table 7, for comparison.) The coefficient estimates using these alternative control groups are very similar in both magnitude and significance to the baseline case.

b) Other Item Substitutions

It is also possible that the *Nutricate* receipt treatment led customers to make item substitutions other than the nine within-category substitutions that were repeatedly suggested in the printed messages. One possible cause of such an effect is the calorie counts and other quantitative nutritional information printed on the receipts: if this information was new to consumers (or simply provided some extra salience) customers may have made other calorie-saving item substitutions that were not explicitly recommended. Relatedly, the introduction of the new receipts may have created an overall atmosphere of nutrition-consciousness that encouraged healthy substitutions beyond those that were recommended. In both these cases, customers would make other item substitutions that reinforce the calorie-reducing effects of the *Nutricate* receipt messages. An alternative possibility is that customers made other substitutions that cancelled out the calorie-saving impact of the recommended substitutions. For example, on learning that she could save a significant number of calories from substituting bacon for sausage on a breakfast sandwich, a consumer with a target breakfast calorie count might add hash browns to her breakfast order.

To check for these types of effects, we tried to identify possible item substitutions that (a) would result in a significant change in fat or calories, (b) involved items that were purchased frequently enough to allow an effect to be detected if one existed, and (c) were *not* explicitly encouraged in the *Nutricate* receipt messages as implemented at Burgerville. We came up with three possibilities: forgoing hash browns with a breakfast sandwich, avoiding the large size of fries, and picking a main item that was under 500 calories. The effects of introducing the *Nutricate* receipt treatment on the incidence of these three choices (as shares of their respective item categories) are shown in Table 9. The regression and control group specifications are identical to those in Table 7.

The dependent variable in column one of Table 9 is the share of breakfast sandwich orders that did not include hash browns.[33] Interestingly, the results in this column are consistent with the canceling-out scenario above: the introduction of the *Nutricate* receipt system is associated with a 3.8 percent decline in the share of breakfast sandwich orders that 'held the hash browns'. A similar effect appears to hold for the share of main items that contained under 500 calories, though here the effect is both smaller and magnitude and borderline in statistical significance.[34] In contrast, however, customers seem to have downsized their fries by a small but statistically significant amount when *Nutricate* receipt was introduced, even though this change was never suggested in the receipts.[35] As noted, possible explanations for this include either a positive 'atmosphere' effect or a direct effect of the quantitative nutrition information in the receipts. Another possibility is that customers reacted to the suggestion to substitute a side salad for fries by downsizing their fries instead.

c) Placebo Tests and Spillovers

If the estimated effects in Tables 7 and 8 are genuine causal effects of the *Nutricate* receipt and not artifacts of our statistical approach, then applying our approach to policy interventions that did not occur should yield estimates of a zero effect. Rows 1-3 of Table 10 address this question by estimating the effects of placebo treatments that occurred at different dates than the actual treatment. Specifically, rows 2 and 3 of Table 10 estimate the effects of placebo treatments that occurred 25 and 50 weeks after the start of our data, respectively. In both cases, the sample is restricted to the 75-week-long actual pre-treatment period. In all cases the specification is identical to row 5 of Table 7, which is reproduced in row 1 for comparison. Of the eighteen estimated coefficients in rows 2 and 3, none are statistically significant at 10 percent or better. Thus our statistical approach does not estimate a treatment effect at times when no treatment occurred.

A distinct type of placebo test asks how often our estimation approach generates a statistically significant treatment effect in stores that were not treated. To answer this question, we replicated the entire analysis in row 5 of Table 7 39 times: once for each store in our sample. Each time, we designated the store in question as the treatment store and assumed treatment

[33] As in many fast food restaurants, Burgerville customers can order breakfast items either individually (*a la carte)*, or bundled into a 'basket' consisting of the sandwich, drink and hash browns. Our measure of of breakfast sandwiches without hash browns in column 1 of Table 9 is the share of breakfast sandwiches sold that were not part of a basket.

[34] Unlike the other two effects in Table 9, the effect on low-calorie mains also becomes insignificant after 25 weeks of treatment (see rows 6 and 7).

[35] The decline is only about half a percent, but is precisely measured because a very large share of Burgerville transactions included an order of fries.

commenced in week 76. We then constructed an ADH control group from the remaining 38 stores using the same methods as before, then re-estimated columns 1-9 of Table 7. Table 11 then reports the number of item substitutions (out of a total of nine possible substitutions for each placebo treatment store) for which this procedure estimates a statistically significant positive coefficient. For example, in four such stores, none of the nine estimated coefficients were positive and significant at the ten percent level; in 16 stores, one significant positive effect was estimated. Thirty-two of the 39 stores yielded two or fewer positive, significant coefficients. Overall, Table 11 strongly suggests there is something special about the true treatment store—it is the only store that shows a consistent and significant pattern of item substitutions in the direction encouraged by the *Nutricate* receipt messages.

A final robustness check concerns the possibility of spillovers in consumption patterns between Burgerville stores. This could occur, for example, if customers read a *Nutricate* receipt at the treatment store, then subsequently visited another Burgerville store. If this induces similar within-category item substitutions at the subsequent stores, it will lead to attenuation bias in our estimated coefficients, leading us to underestimate the effects of the *Nutricate* receipt intervention. To assess this possibility, we ask whether customers at the five stores that are closest (geographically) to the treated store made similar within-category item substitutions as occurred at the treatment store after *Nutricate* receipts were introduced at the treatment store. Specifically, in row 4 of Table 10 we exclude the treatment store from our sample, use the 5 nearest stores to it as our "treatment store", and use the 5 stores furthest from the treatment store as controls. Only one of the nine possible item-substitution effects is statistically significant, and the encouraged share index shows no significant change either (though the estimated effect is positive). We conclude that evidence of cross-store spillover effects of the *Nutricate* treatment is weak at best.

7. Discussion

We have shown that the introduction of receipts containing customized ordering suggestions designed to reduce fat and calories had a detectable effect on purchasing behavior at a chain of fast-food restaurants. While we find no statistically significant effect on total calories purchased, we do find some effects for fat intake. In particular, the combination of messages in the Burgerville implementation of the *Nutricate* receipt induces a robust and long-lasting (at least up to 50 weeks after treatment) two- percent reduction in cholesterol per transaction. Our confidence that this behavioral change is causally linked to the new receipts is increased by the fact that customers in the aggregate made most of the specific item substitutions that were promoted by the receipts, such as substituting grilled chicken for fried chicken in sandwiches, frozen yogurt for ice cream, and non-sausage for sausage breakfast sandwiches. While these specific substitutions vary in their permanence, the estimated substitutions are robust to the use of different control groups, both in terms of whether and how a synthetic control is calculated and which alternative criteria (geographical distance, similar store size, or all nontreated stores) are used to construct a non-synthetic control group. We also find no evidence that introducing the *Nutricate* receipt was harmful to our treated store for any measure of total sales.

How do we reconcile our relatively strong findings about item substitutions with the relatively muted overall effects of this intervention on nutrition and calories? One possible explanation is that customers made other item substitutions that counteract the substitutions detected in Table 7. Indeed, Table 9 presents evidence of at least two such changes: an increase in hash browns, and in main dishes containing more than 500 calories. Second, several of the encouraged items (whose consumption rose substantially in percentage terms when *Nutricate* receipts were introduced) constitute a very small share of fast-food purchases. Examples include side salads, kids' apples and grilled chicken sandwiches.[36] Even large substitutions towards items like these will have only minimal effects on total fat and calories purchased. And as in all studies of restaurant consumption, we caution that our estimates refer only to purchases at a particular restaurant; thus it is possible that nutritional improvements we detect may be cancelled out by changes in food consumption at home and at other restaurants. On the other hand, our estimates may understate the potential of restaurant-based interventions because the treatment store in our study could not change its menu offerings in response to the treatment. Anecdotal evidence suggests that several well known chains reworked their menus in response to actual and anticipated calorie labeling requirements (Bernstein 2011).

Overall, we think that our results send a qualified but positive message about the health-improving potential of interventions like the *Nutricate* receipt, which has three key features that distinguish it from existing nutritional interventions: (1) this intervention is receipt-based (not menu-based); (2) it is personalized (in the sense that the message is triggered by the customer's current purchase) and (3) it goes beyond the provision of information (and beyond reminding consumers of consequences they already know) by suggesting specific changes in behavior. In addition to providing a vehicle for customized recommendations, a receipt-based approach may be useful as a convenient and unobtrusive substitute for menu-based interventions in cases where those interventions are too costly, intrusive, uninformative, or simply avoided by restaurants. Receipt-based approaches can, of course, also be used together with menu-based ones with possibly complementary effects.

The *Nutricate* receipt's novel approach of making personalized recommendations for future purchases mirrors rapidly-spreading practices by Internet retailers like Amazon and Netflix that recommend items a customer might like based on his or her recent purchase history. As such, it is worth noting that the particular system we tested at Burgerville is a relatively early implementation of a concept that may have significant potential to improve nutritional choices in the future. Suppose that a customer's favorite restaurant has developed a healthier version of a sandwich he orders frequently—why not inform him of this option in a personalized way? (He might not even look at the menu any more.) Future systems might remember that a customer always prefers 'no sauce' on her burger or salad instead of fries, and could make this her default choice for future orders. Although we focus on an early example of this type of technology, our study suggests that personalized ordering suggestions may have the potential to facilitate healthier food choices by serving restaurant customers better.

We conclude by highlighting a number methodological lessons that emerge from our analysis. The first concerns standard errors. Our results suggest that a failure to make aggressive adjustments to standard errors in difference-in-difference research designs (in our case implementing both Donald and Lang-style differencing and applying Newey-West standard

[36] Our agreement with Burgerville prohibits us from reporting these means directly, but we can say,for example, that side salads and apples are less than 10 percent of adult and child side dishes respectively.

errors to the resulting time series) could lead to a large number of false positives. In other words, clustering by group is not enough, and Donald-Lang differencing alone may not be enough because it does not allow for persistent group-specific shocks. A second methodological note is the overriding importance in these studies of choosing or constructing appropriate control groups, or of showing that one's key results are robust to the choice of control groups. In our case, we show that our results are robust to several methods of constructing control groups, including the Abadie, Diamond, and Hainmueller (2010) procedure to construct 'optimal' control groups.

A third methodological point addresses the challenge of detecting the effects of marketing interventions like the Nutricate receipt's ordering suggestions. In this regard, our paper illustrates the importance of focusing on substitutions within detailed item categories, such as desserts, chicken sandwiches, or kids' meals in our case. Both because of the confounding role of category-specific demand shocks (affecting, for example, all breakfast or dessert items together) and because of the wide range of other substitutions consumers can make, detecting the effects of these types of interventions difficult unless one conditions on these detailed categories.[37]

Finally, our results underscore the fact that simply providing nutritional information, or repeatedly reminding people that, say, fries are unhealthy may not be the most effective ways to improve nutritional choices. While lack of information and willpower may both be important determinants of choices, our results suggest that other cognitive constraints, such as information processing costs when choosing from long lists, may also play a role. To that extent, the literature on health behaviors can perhaps take a lesson from the recent literature on financial literacy, where simplifying complex choices by reducing the number of options and suggesting specific behaviors (for example, as a default option) have been shown to have beneficial effects.[38]

[37] Incidentally, these analytical challenges illustrate the tremendous difficulty in designing effective interventions to improve diet, compared for example to anti-smoking interventions: compared to tobacco (a fairly unique product with no close substitutes) there are thousands of possible food choices even within the same restaurant, and reducing the consumption of any one unhealthy item does not translate easily into an overall improvement in nutrition.
[38] See for example Bertrand et al., 2010.

References

Abadie, Alberto , Alexis Diamond, and Jens Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program *Journal of the American Statistical Association* 105(490): 493 -505.

Anderson, Michael L. and David A. Matsa. 2011. Are Restaurants Really Supersizing America? *American Economic Journal: Applied Economics* 3:1, 152-188.

Bassett, Mary T., Tamara Dumanovsky, Christina Huang, Lynn D. Silver, Candace Young, Cathy Nonas, Thomas D. Matte, Sekai Chideya, and Thomas R. Frieden. 2008. Purchasing Behavior and Calorie Information at Fast-Food Chains in New York City, 2007. *American Journal of Public Health*, 98(8): 1457–59.

Bernstein, Sharon. Making Every Calorie Count: Upcoming FDA disclosure rules prompt many chains to overhaul their offerings *Los Angeles Times*, Weds. June 11, 2011, pp.B1 and B4.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How Much Should We Trust Differences-in- Differences Estimates? *The Quarterly Journal of Economics*, Vol. 119, No. 1 (Feb), pp. 249-275.

Bertrand, Marianne, Dean Karlin, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment.*Quarterly Journal of Economics*. vol. 125, no. 1, (February): 263-306.

Bleich, Sara, David Cutler, Christopher Murray, and Alyce Adams. 2008. Why Is the Developed World Obese? *Annual Review of Public Health*, 29: 273–95.

Bollinger, B., P. Leslie and A. Sorensen. 2011. Calorie Posting in Chain Restaurants, *American Economic Journal: Economic Policy*, 3 (1), 91-128.

Card, David. 1990. The Impact of the Mariel Boatlift on the Miami Labor Market, *Industrial and Labor Relations Review* 43 (January 1990), 245–257.

Card, D. and A. Krueger. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, *American Economic Review* 84 (Sept): 772-793.

Center for Disease Control. 2011. Obesity Halting the Epidemic by making Health Easier, At A Glance http://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2011/Obesity_AAG_WEB_508.pdf, Accessed on January 17, 2011.

Currie, Janet, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania. 2010. The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy*, 2(3): 32–63.

Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro. 2003. Why Have Americans Become More Obese? *Journal of Economic Perspectives*, 17(3): 93–118.

Davis, Brennan, and Christopher Carpenter. 2009. Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity. *American Journal of Public Health*, 99(3): 505–10.

DellaVigna, Stefano. 2009. Psychology and Economics:Evidence from the Field *Journal of Economic Literature* 2009, 47:2, 315–372

Donald, Stephen G. and Kevin Lang. 2007. Inference with Differences-in-Differences and other Panel Data. *The Review of Economics and Statistics* 89(2) (May): 221–233.

Downs, Julie S., George Loewenstein, and Jessica Wisdom. 2009. Strategies for Promoting Healthier Food Choices. *American Economic Review* 99(2): 159–64.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties *Review of Economics and Statistics* 92(4) (November): 945-64

Dumanovsky, Tamara, Christina H. Yuang, Cathy A. Nonas, Thomas D. Matte, Mary To. Bassett, and Lynn D. Silver. 2011. Changes in energy content of lunchtime purchases from fast food restaurants after introduction of calorie labeling: cross sectional customer surveys. *British Medical Journal* 343: d4464. Published online July 26. doi: 10.1136/bmj.d4464

Elbel, B., Kersh, R., Brescoll, V.L., Dixon, L.B. 2009. Calorie labeling and food choices: A first look at the effects on low-income people in New York City, *Health Affairs*, 28(6), 1110-21

Finkelstein, E.A., K.L. Strombotne, N.L. Chan and J. Krieger. 2011. The Impact of Mandatory Menu Labeling In One Fast Food Chain in King County, Washington. *American Journal of Preventive Medicine* 40(2) (Feb): 122-127.

Flegal, Katherine M., Margaret D. Carroll, Cynthia L. Ogden, and Lester R. Curtin. 2010. Prevalence and Trends in Obesity Among US Adults, 1999-2008, *Journal of the American Medical Association*, 303(3): 235-241.

Gruber, Jonathan, and James Poterba, Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-Employed, *Quarterly Journal of Economics* 109 (August 1994), 701–734.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh. 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *Journal of Finance* 64(5) (October): 2289-2325

Iyengar, Sheena S. and Mark R. Lepper. 2000. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology*, 2000, Vol. 79, No. 6, 995-1006.

Lakdawalla, Darius, and Tomas Philipson. 2002. The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination. National Bureau of Economic Research Working Paper 8946.

Lakdawalla, Darius, Tomas Philipson, and Jay Bhattacharya. 2005. Welfare-Enhancing Technological Change and the Growth of Obesity. *American Economic Review*, 95(2): 253–57.

Newey, Whitney K, and Kenneth D. West. 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix *Econometrica* 55(3) (May):703-08

Rubinstein, Ariel and Yuval Salant. 2006. A model of choice from lists *Theoretical Economics* 1: 3–17

Vadiveloo, Maya K, L Beth Dixon and Brian Elbel. 2011. Consumer purchasing patterns in response to calorie labeling legislation in New York City *International Journal of Behavioral Nutrition and Physical Activity*, 8:51

Variyam, Jayachandran N and John Cawley. 2006. Nutrition Labels and Obesity National Bureau of Economic Research working paper no. 11956.

Wisdom, Jessica , Julie S. Downs and George Loewenstein. 2010. Promoting Healthy Choices: Information vs. Convenience *American Economic Journal: Applied Economics* 2:164-178

Appendix A: Sample Receipt

Order #F-0008
11/05/11 10:28am Server: ERIKA H
18350 Willamette Dr. West Linn, OR 97068
(503) 635-7339 www.burgerville.com

BURGERVILLE

Fresh ▸ Local ▸ Sustainable

NUTRICATE™ receipt (nu-tri-kAt) v. To nutritionally educate

Qty	Item	Price	Calories	Fiber(g)	Fat(g)	Carbs(g)
1	ENG MFN EGG/SAUS BASKET	$4.89	-	-	-	-
	Eng Mfn w/ Egg & Saus		550	1	42	23
	Add Tillamook Cheese		110	0	9	0
	Hashbrown		140	2	9	12
	Large Coffee		8	0	0	2
	NUTRITION TOTALS		808	3	60	36
	% DAILY VALUE - 2000 CALORIES		40%	10%	92%	12%
	% DAILY VALUE - 2500 CALORIES		32%	8%	73%	10%

Sub Total	$4.89
TOTAL	$4.89
CASH	$5.00
CHANGE	$0.11
PAID	$4.89

Did You Know ?

Are you watching your diet, but don't want to sacrifice taste? Try ham instead of sausage on your English Muffin Sandwich and save about 160 calories and 20g fat.

TURN OVER FOR SPECIAL OFFERS!

Appendix B: Sample Messages

Discouraged Items with an Encouraged Substitution:

Fries to Side Salad: "Our side salad with fresh greens, Tillamook cheddar cheese, carrots, red cabbage, grape tomatoes topped with Bleu Cheese dressing contains 160 fewer calories than a regular fry."

Kids' Fries to Apples: "If you're looking for a refreshing complement to your kids meal, try ordering apples instead of fries and save 165 calories and 8g of fat."

Fried to Grilled Chicken Sandwich: "If you're trying to live a healthier lifestyle, consider our Low Fat Grilled Chicken Sandwich instead of the Crispy Chicken Sandwich and save 90 calories and 7g of fat."

Ice Cream to Yogurt: "Trying to eat healthier? Try our YoCream Frozen Yogurt Caramel Sundae with 32% fewer calories and 93% fewer grams of fat than our Caramel Ice Cream Sundae. See if you can taste the difference."

Sausage to Other Breakfast Sandwich: "For the same great taste but fewer calories, next time try a bagel with egg and bacon (instead of sausage) and save approximately 150 calories!"

Other Encouraged Choices:

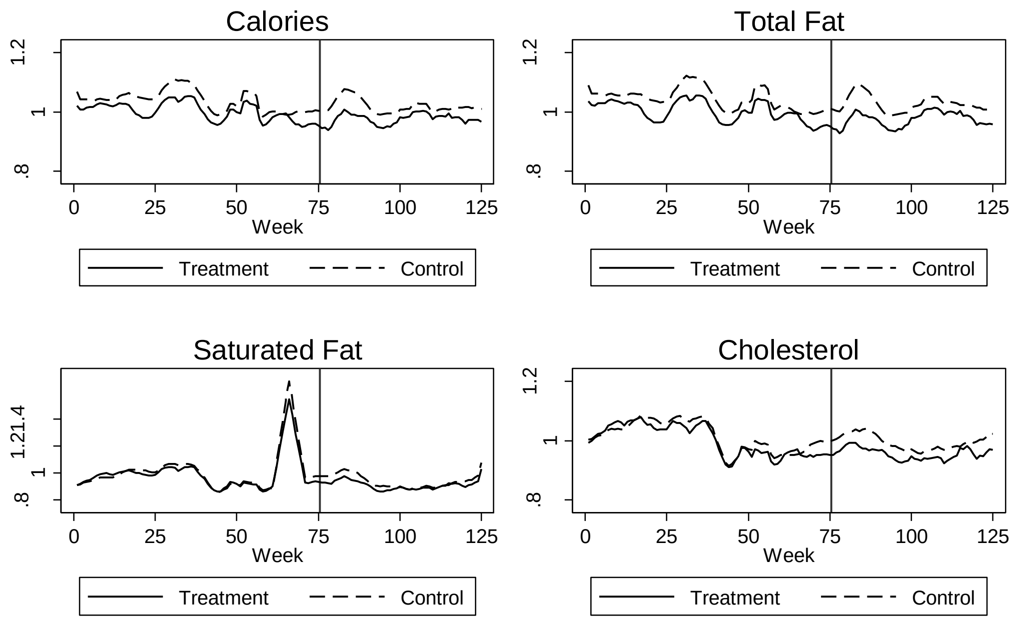
"Hold the sauce": "If you're looking to eat healthier, ask us to hold the Burgerville Spread and save about 70 calories and 8g of fat."

"Hold the cheese: "While cheese is a good source of calcium, if you are watching your calorie or fat intake, next time consider skipping the cheese to save up to 110 calories and 9g of fat."

Kid's milk: "Sized for smaller appetites and featuring the same exceptional quality, Kids' Meals offer a variety of healthful choices, like apple slices and local, hormone-free milk."

Main salad: "Our entre [sic] salads feature fresh greens and sustainable, local ingredients, such as smoked salmon and Oregon hazelnuts." Or, "Salmon is rich in omega-3 fatty acids, which are good for heart health and brain function."

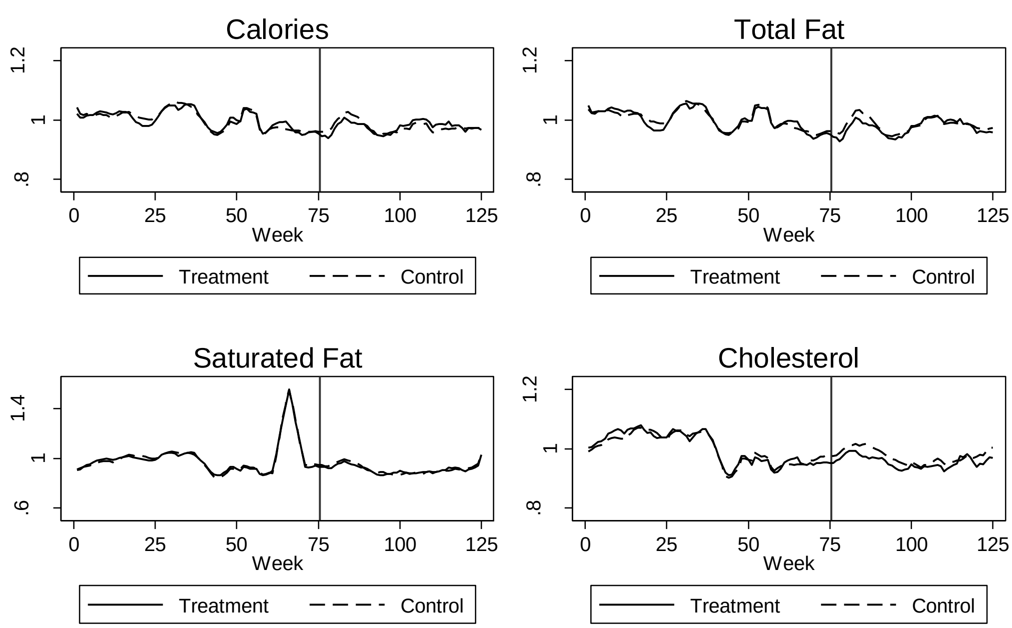
Figure 1. Weekly Nutrition for Treatment and Control Stores



Control is the unweighted average of all non-treatment stores

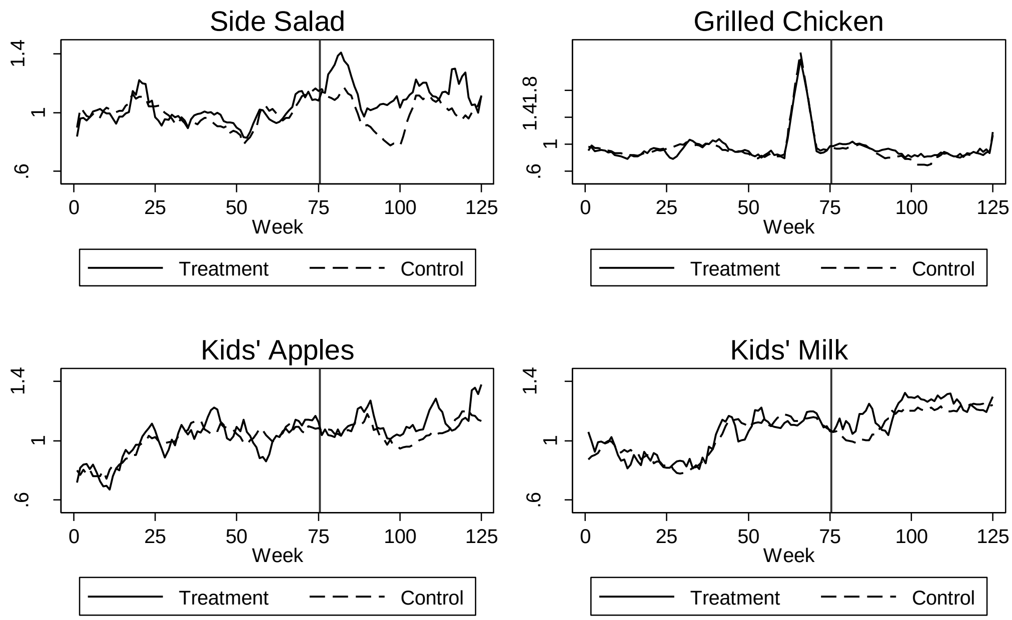
Note: All figures show five-week moving averages.

Figure 2. Weekly Nutrition for Treatment and Control Stores



Control is the synthetic control store

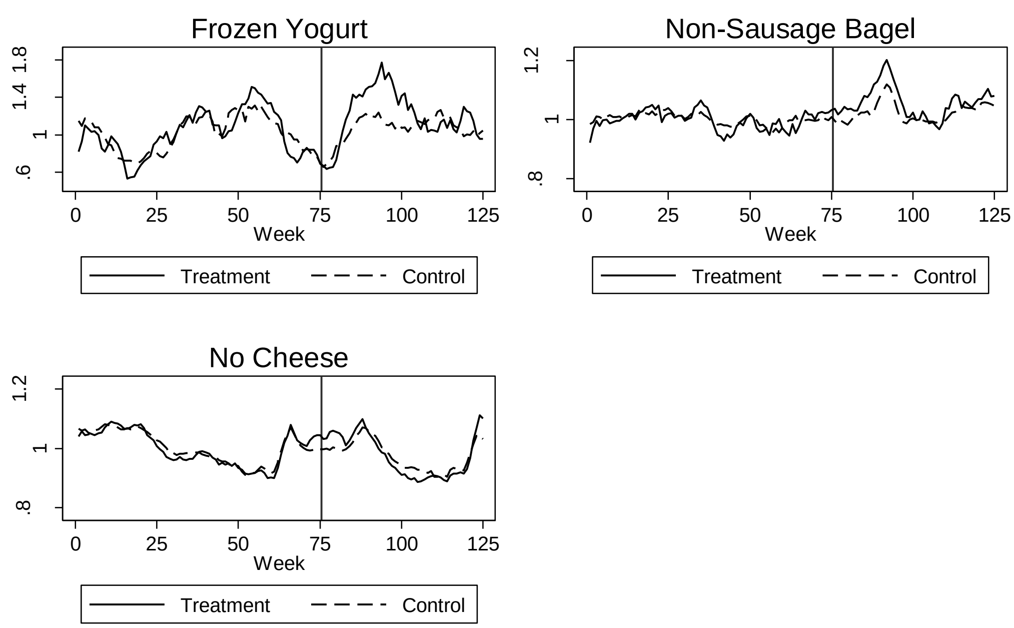
Figure 3a. Category Shares for Treatment and Control Stores



Control is the synthetic control store

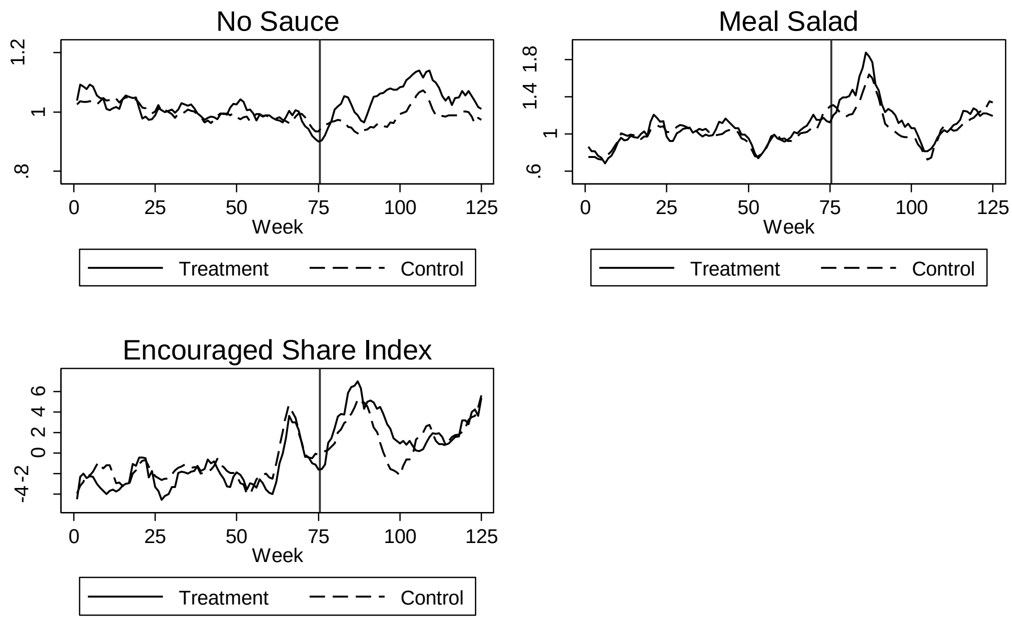
Note: All figures show five-week moving averages.

Figure 3b. Category Shares for Treatment and Control Stores



Control is the synthetic control store

Figure 3c. Category Shares for Treatment and Control Stores



Control is the synthetic control store

Note: All figures show five-week moving averages

Table 1: Nutrition per Transaction and Sales, By Store and Period

	Treatment Store			All Other Stores		Difference-in-Difference
	Baseline Means (1)	Period 1 (2)	Period 2 (3)	Period 1 (4)	Period 2 (5)	[(3)-(2)] - [(5)-(4)]
Nutrition per transaction						
Calories	1657	1.000	0.977	1.037	1.019	-0.006
Total fat (grams)	80	1.000	0.975	1.044	1.027	-0.008
Saturated fat (grams)	23	1.000	0.919	1.017	0.947	-0.010
Cholesterol	153	1.000	0.956	1.014	0.996	-0.027
Sales						
Revenue per transaction	--	1.000	1.044	1.024	1.066	0.003
Total transactions	--	1.000	0.976	0.577	0.543	0.010
Total revenue	--	1.000	1.022	0.594	0.584	0.032

Treatment begins at the treatment store starting in week 76. Period 1 includes weeks 1-75 and period 2includes weeks 76-125. The overall sample is 4875 store*week cells.

Table 2: Purchases of Selected Items per Transaction, By Store and Period

	Treatment Store		All Other Stores		Difference-in-Difference
	Period 1	Period 2	Period 1	Period 2	
	(1)	(2)	(3)	(4)	[(2)-(1)] - [(4)-(3)]
Discouraged Items:					
Fries (adult)	1.000	0.985	1.011	0.993	0.002
Fries (kids)	1.000	1.045	1.123	1.125	0.043
Fried chicken sandwich	1.000	1.012	0.844	0.848	0.007
Ice cream	1.000	0.783	1.444	1.239	-0.012
Breakfast sandwich w. Sausage	1.000	0.867	0.970	0.985	-0.148
Encouraged Items:					
Main salad	1.000	1.226	0.841	1.012	0.056
Grilled chicken sandwich	1.000	0.836	0.938	0.702	0.071
No sauce	1.000	1.053	0.853	0.846	0.059
No cheese	1.000	0.975	1.138	1.103	0.010
Side salad	1.000	1.143	0.778	0.849	0.072
Frozen Yogurt	1.000	1.007	0.841	0.950	-0.102
Kids' apples	1.000	1.224	0.956	1.070	0.110
Kids' milk	1.000	1.307	1.187	1.365	0.129
All other breakfast sandwiches	1.000	1.005	0.999	1.057	-0.052

Treatment begins at the treatment store starting in week 76. Period 1 includes weeks 1-75 and period 2includes weeks 76-125. The overall sample is 4875 store*week cells.

Table 3: Shares of Encouraged Items within Selected Item Categories, by Store and Period

	Treatment Store		All Other Stores		Difference-in-Difference
	Period 1 (1)	Period 2 (2)	Period 1 (3)	Period 2 (4)	[(2)-(1)] - [(4)-(3)]
Share of Adult Sides:					
Salad	1.000	1.155	0.776	0.857	0.074
Share of Chicken Sandwiches:					
Grilled	1.000	0.918	1.037	0.911	0.044
Share of Kids' Meals:					
With Apples	1.000	1.141	0.855	0.931	0.065
With Milk	1.000	1.223	1.063	1.205	0.081
Share of Desserts:					
Frozen Yogurt	1.000	1.222	0.627	0.765	0.084
Share of Breakfast Sandwiches:					
Without Sausage	1.000	1.061	1.017	1.035	0.043
Share of Mains:					
With No Cheese	1.000	0.978	1.023	0.991	0.010
With No Sauce	1.000	1.053	0.770	0.763	0.060
With Main Salad	1.000	1.230	0.760	0.910	0.080

Treatment begins at the treatment store starting in week 76. Period 1 includes weeks 1-75 and period 2includes weeks 76-125. The overall sample is 4875 store*week cells.

Table 4: Estimated Treatment Effects for Nutrition and Sales per Transaction

	Nutrition				Sales				
	Calories	Total Fat	Saturated Fat	Cholesterol	Items per Transaction	Revenue per Transaction	Total Items	Total Transactions	Total Revenue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<u>All other stores as controls</u>									
One-stage FE	-0.0056***	-0.0082***	-0.0102***	-0.0267***	0.0031	0.0028	0.0273***	0.0097**	0.0316***
	(0.0020)	(0.0020)	(0.0022)	(0.0025)	(0.0026)	(0.0021)	(0.0053)	(0.0037)	(0.0047)
DL	-0.0056	-0.0082*	-0.0102	-0.0267***	0.0031	0.0028	0.0273	0.0097	0.0316
	(0.0045)	(0.0047)	(0.0071)	(0.0048)	(0.0040)	(0.0045)	(0.0228)	(0.0222)	(0.0237)
DL-NW	-0.0056	-0.0082	-0.0102	-0.0267***	0.0031	0.0028	0.0273	0.0097	0.0316
	(0.0073)	(0.0076)	(0.0110)	(0.0067)	(0.0052)	(0.0077)	(0.0310)	(0.0298)	(0.0329)
<u>Synthetic control group</u>									
ADH-DL	0.0005	-0.0066	-0.0042	-0.0213***	0.0099***	0.0024	-0.0074	-0.0164	-0.0028
	(0.0041)	(0.0040)	(0.0055)	(0.0044)	(0.0037)	(0.0040)	(0.0239)	(0.0221)	(0.0250)
ADH-DL-NW	0.0005	-0.0066	-0.0042	-0.0213***	0.0099***	0.0024	-0.0074	-0.0164	-0.0028
	(0.0055)	(0.0050)	(0.0055)	(0.0054)	(0.0036)	(0.0062)	(0.0364)	(0.0323)	(0.0391)
<u>Synthetic control group (ADH-DL-NW)</u>									
Weeks 76-100	-0.0071	-0.0129**	-0.0054	-0.0248***	0.0094***	-0.0071	-0.0084	-0.0111	-0.011
	(0.0064)	(0.0058)	(0.0086)	(0.0054)	(0.0034)	(0.0073)	(0.0391)	(0.0327)	(0.0427)
Weeks 101-125	0.008	-0.0003	-0.0029	-0.0179**	0.0103*	0.0118**	-0.0065	-0.0217	0.0054
	(0.0057)	(0.0053)	(0.0063)	(0.0072)	(0.0054)	(0.0058)	(0.0447)	(0.0408)	(0.0472)

<u>Notes</u>

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

One-stage FE: The dependent variable is the level of the outcome in the store*week cell. Treatment equals one in treatment store starting in week 76, zero otherwise. Sample size is 4875 store*week cells. Regression includes a full set of 125 week and 39 store fixed effects. Standard errors are clustered by store.

ADH-DL: The dependent variable is the difference between the outcome in treatment store and the mean for all other stores, by week. Sample size is 125 weeks. Estimation is by OLS.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Table 5: Estimated Treatment Effects for Purchases of Discouraged Items, per Transaction

	Fries (adult) (1)	Fries (child) (2)	Fried Chicken Sandwich (3)	Ice Cream (4)	Sausage Breakfast Sandwich (5)	Discouraged Index (6)
All other stores as controls						
One-stage FE	0.0022	0.0426***	0.0074	-0.0116	-0.1477***	-0.6439***
	(0.0026)	(0.0088)	(0.0055)	(0.0165)	(0.0171)	(0.1207)
DL	0.0022	0.0426**	0.0074	-0.0116	-0.1477***	-0.6439**
	(0.0060)	(0.0186)	(0.0163)	(0.0552)	(0.0264)	(0.2759)
DL-NW	0.0022	0.0426*	0.0074	-0.0116	-0.1477***	-0.6439*
	(0.0099)	(0.0251)	(0.0214)	(0.1125)	(0.0268)	(0.3593)
Synthetic control group						
ADH-DL	-0.0036	-0.0007	0.0146	-0.0580**	-0.1320***	-0.3212
	(0.0055)	(0.0181)	(0.0171)	(0.0242)	(0.0266)	(0.2664)
ADH-DL-NW	-0.0036	-0.0007	0.0146	-0.0580	-0.1320***	-0.3212
	(0.0071)	(0.0195)	(0.0205)	(0.0403)	(0.0239)	(0.3532)
Synthetic control group (ADH-DL-NW)						
Weeks 76-100	-0.0124	-0.024	-0.0162	0.0199	-0.1312***	-0.6506***
	(0.0084)	(0.0221)	(0.0166)	(0.0314)	(0.0342)	(0.2442)
Weeks 101-125	0.0052	0.0226	0.0454*	-0.1359***	-0.1327***	0.0081
	(0.0084)	(0.0226)	(0.0266)	(0.0467)	(0.0292)	(0.5950)

Notes

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

One-stage FE: The dependent variable is the level of the outcome in the store*week cell. Treatment equals one in treatment store starting in week 76, zero otherwise. Sample size is 4875 store*week cells. Regression includes a full set of 125 week and 39 store fixed effects. Standard errors are clustered by store.

ADH-DL: The dependent variable is the difference between the outcome in treatment store and the mean for all other stores, by week. Sample size is 125 weeks. Estimation is by OLS.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Table 6: Estimated Treatment Effects for Purchases of Encouraged Items, per Transaction

	Main Salad	Grilled Chicken Sandwich	No Sauce	No Cheese	Side Salad	Frozen Yogurt	Kids' Apples	Kids' Milk	Non-Sausage Breakfast Sandwiches	Encouraged Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<u>All other stores as controls</u>										
One-stage FE	0.0556***	0.0715***	0.0590***	0.0100***	0.0722***	-0.1020***	0.1104***	0.1295***	-0.0522***	1.7566***
	(0.0100)	(0.0103)	(0.0055)	(0.0033)	(0.0112)	(0.0362)	(0.0132)	(0.0145)	(0.0175)	(0.1646)
DL	0.0556**	0.0715**	0.0590***	0.01	0.0722***	-0.102	0.1104**	0.1295***	-0.0522*	1.7566***
	(0.0278)	(0.0322)	(0.0127)	(0.0075)	(0.0249)	(0.0622)	(0.0440)	(0.0292)	(0.0288)	(0.3865)
DL-NW	0.0556	0.0715*	0.0590***	0.01	0.0722**	-0.102	0.1104**	0.1295***	-0.0522	1.7566***
	(0.0354)	(0.0388)	(0.0207)	(0.0116)	(0.0286)	(0.1178)	(0.0515)	(0.0352)	(0.0375)	(0.3215)
<u>Synthetic control group</u>										
ADH-DL	0.0357	0.0626**	0.0700***	0.0157**	0.1377***	0.0141	0.0523	0.0954***	-0.1172***	0.5484
	(0.0303)	(0.0274)	(0.0122)	(0.0077)	(0.0258)	(0.0616)	(0.0435)	(0.0306)	(0.0304)	(0.4463)
ADH-DL-NW	0.0357	0.0626**	0.0700***	0.0157	0.1377***	0.0141	0.0523	0.0954**	-0.1172***	0.5484
	(0.0390)	(0.0252)	(0.0142)	(0.0096)	(0.0313)	(0.1028)	(0.0467)	(0.0378)	(0.0412)	(0.5003)
<u>Synthetic control group (ADH-DL-NW)</u>										
Weeks 76-100	0.1235***	0.0467**	0.0592***	0.0127	0.1784***	0.2420***	-0.0085	0.0605	-0.0774	1.6567***
	(0.0259)	(0.0215)	(0.0213)	(0.0156)	(0.0310)	(0.0880)	(0.0423)	(0.0392)	(0.0486)	(0.3279)
Weeks 101-125	-0.0521	0.0785**	0.0807***	0.0188*	0.0970**	-0.2139***	0.1131	0.1304**	-0.1571***	-0.5599
	(0.0386)	(0.0375)	(0.0161)	(0.0100)	(0.0444)	(0.0763)	(0.0685)	(0.0571)	(0.0531)	(0.5185)

<u>Notes</u>

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

One-stage FE: The dependent variable is the level of the outcome in the store*week cell. Treatment equals one in treatment store starting in week 76, zero otherwise. Sample size is 4875 store*week cells. Regression includes a full set of 125 week and 39 store fixed effects. Standard errors are clustered by store.

ADH-DL: The dependent variable is the difference between the outcome in treatment store and the mean for all other stores, by week. Sample size is 125 weeks. Estimation is by OLS.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Table 7: Estimated Treatment Effects for Purchases of Encouraged Items, as a Share of Category

	Side Salads as a Share of Adult Sides	Grilled as a Share of Chicken Sandwiches	Apples as a as a Share of Child Meals	Milk as a as a Share of Child Meals	Frozen Yogurt as a Share of Desserts	Non-Sausage as a Share of Breakfast Sandwiches	No Cheese as a Share of Adult Mains	No Sauce as a Share of Adult Mains	Main Salad as a Share of Adult Mains	Encouraged Share Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<u>All other stores as controls</u>										
One-stage FE	0.0741***	0.0437***	0.0650***	0.0811***	0.0841***	0.0433***	0.0098***	0.0598***	0.0800***	3.1391***
	(0.0115)	(0.0080)	(0.0105)	(0.0096)	(0.0276)	(0.0064)	(0.0025)	(0.0050)	(0.0082)	(0.1609)
DL	0.0741***	0.0437*	0.0650*	0.0811***	0.0841	0.0433***	0.0098	0.0598***	0.0800**	3.1391***
	(0.0263)	(0.0229)	(0.0333)	(0.0275)	(0.0605)	(0.0127)	(0.0077)	(0.0108)	(0.0310)	(0.4560)
DL-NW	0.0741**	0.0437*	0.0650*	0.0811***	0.0841	0.0433***	0.0098	0.0598***	0.0800*	3.1391***
	(0.0308)	(0.0258)	(0.0329)	(0.0305)	(0.0930)	(0.0113)	(0.0126)	(0.0157)	(0.0437)	(0.4824)
<u>Synthetic control group</u>										
ADH-DL	0.1453***	0.0422*	0.0697*	0.0654**	0.1266**	0.0382***	0.0032	0.0677***	0.0803**	1.7425***
	(0.0278)	(0.0227)	(0.0361)	(0.0292)	(0.0607)	(0.0134)	(0.0073)	(0.0106)	(0.0323)	(0.4912)
ADH-DL-NW	0.1453***	0.0422*	0.0697**	0.0654**	0.1266*	0.0382***	0.0032	0.0677***	0.0803**	1.7425***
	(0.0335)	(0.0222)	(0.0319)	(0.0284)	(0.0751)	(0.0109)	(0.0112)	(0.0105)	(0.0400)	(0.5598)
<u>Two-Time Periods: Synthetic control group (ADH-DL-NW)</u>										
Weeks 76-100	0.1904***	0.0574**	0.0447	0.0918**	0.2647***	0.0550***	0.0071	0.0712***	0.1618***	2.9804***
	(0.0358)	(0.0228)	(0.0272)	(0.0415)	(0.0871)	(0.0130)	(0.0147)	(0.0146)	(0.0337)	(0.5484)
Weeks 101-125	0.1002**	0.0271	0.0948*	0.039	-0.0115	0.0215**	-0.0007	0.0642***	-0.0012	0.5046
	(0.0450)	(0.0293)	(0.0499)	(0.0280)	(0.0627)	(0.0108)	(0.0152)	(0.0124)	(0.0412)	(0.4216)

<u>Notes</u>

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

One-stage FE: The dependent variable is the level of the outcome in the store*week cell. Treatment equals one in treatment store starting in week 76, zero otherwise. Sample size is 4875 store*week cells. Regression includes a full set of 125 week and 39 store fixed effects. Standard errors are clustered by store.

ADH-DL: The dependent variable is the difference between the outcome in treatment store and the mean for all other stores, by week. Sample size is 125 weeks. Estimation is by OLS.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Table 8: Estimated Treatment Effects for Purchases of Encouraged Items, as a Share of Category - Alternate Control Groups

	Side Salads as a Share of Adult Sides	Grilled as a Share of Chicken Sandwiches	Apples as a as a Share of Child Meals	Milk as a as a Share of Child Meals	Frozen Yogurt as a Share of Desserts	Non-Sausage as a Share of Breakfast Sandwiches	No Cheese as a Share of Adult Mains	No Sauce as a Share of Adult Mains	Main Salad as a Share of Adult Mains	Encouraged Share Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Synthetic control group (ADH-DL-NW)										
3-week	0.1651***	0.0425**	0.0790***	0.0695**	0.1058	0.0519***	0.0045	0.0791***	0.0599	1.6935***
	(0.0341)	(0.0213)	(0.0300)	(0.0327)	(0.0740)	(0.0122)	(0.0106)	(0.0116)	(0.0369)	(0.5640)
5-week (Baseline)	0.1453***	0.0422*	0.0697**	0.0654**	0.1266*	0.0382***	0.0032	0.0677***	0.0803**	1.7425***
	(0.0335)	(0.0222)	(0.0319)	(0.0284)	(0.0751)	(0.0109)	(0.0112)	(0.0105)	(0.0400)	(0.5598)
7-week averages	0.1439***	0.0289	0.0787**	0.0762**	0.0787	0.0323**	0.0065	0.0753***	0.0681*	1.7321***
	(0.0323)	(0.0225)	(0.0341)	(0.0309)	(0.0817)	(0.0142)	(0.0107)	(0.0115)	(0.0380)	(0.5533)
Other control groups										
DL-NW (Baseline)	0.0741**	0.0437*	0.0650*	0.0811***	0.0841	0.0433***	0.0098	0.0598***	0.0800*	3.1391***
	(0.0308)	(0.0258)	(0.0329)	(0.0305)	(0.0930)	(0.0113)	(0.0126)	(0.0157)	(0.0437)	(0.4824)
5 most similar sized stores	0.0810**	0.0711***	0.0627*	0.1029***	0.0668	0.0539***	0.0088	0.0726***	0.0795*	3.2703***
	(0.0330)	(0.0252)	(0.0361)	(0.0251)	(0.0905)	(0.0135)	(0.0117)	(0.0132)	(0.0414)	(0.5083)
5 closest stores	0.0747**	0.0642**	0.1064***	0.0727**	0.1075	0.0621***	0.0143	0.0585***	0.0718*	3.5972***
	(0.0338)	(0.0247)	(0.0360)	(0.0318)	(0.1004)	(0.0122)	(0.012)	(0.013)	(0.041)	(0.543)

Notes

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Other control groups: As defined in each column. Estimation is ADH-DL-NW.

Table 9: Estimated Treatment Effects for Purchases of Encouraged Items, as a Share of Category - Non-Messaged Items

	Breakfast Sandwich Only as a Share of All Sandwiches (not in a basket)	Small or Regular Fries as a Share of All Fries	Mains 500 or Fewer Calories as a Share of All Mains
	(1)	(2)	(3)
All other stores as controls			
One-stage FE	-0.0344***	0.0031***	-0.0226***
	(0.0063)	(0.0007)	(0.0026)
DL	-0.0344***	0.0031**	-0.0226***
	(0.0130)	(0.0014)	(0.0052)
DL-NW	-0.0344**	0.0031	-0.0226***
	(0.0144)	(0.0022)	(0.0078)
Synthetic control group			
ADH-DL	-0.0388***	0.0050***	-0.0113**
	(0.0136)	(0.0013)	(0.0051)
ADH-DL-NW	-0.0388***	0.0050***	-0.0113*
	(0.0139)	(0.0017)	(0.0067)
Synthetic control group (ADH-DL-NW)			
Weeks 76-100	-0.0359**	0.0063***	-0.0135*
	(0.0154)	(0.0023)	(0.0081)
Weeks 101-125	-0.0418*	0.0036*	-0.0092
	(0.0220)	(0.0020)	(0.0091)

Notes

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

One-stage FE: The dependent variable is the level of the outcome in the store*week cell. Treatment equals one in treatment store starting in week 76, zero otherwise. Sample size is 4875 store*week cells. Regression includes a full set of 125 week and 39 store fixed effects. Standard errors are clustered by store.

ADH-DL: The dependent variable is the difference between the outcome in treatment store and the mean for all other stores, by week. Sample size is 125 weeks. Estimation is by OLS.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Table 10: Estimated Treatment Effects for Purchases of Encouraged Items, as a Share of Category - Placebo Dates and Spillovers

	Side Salads as a Share of Adult Sides	Grilled as a Share of Chicken Sandwiches	Apples as a as a Share of Child Meals	Milk as a as a Share of Child Meals	Frozen Yogurt as a Share of Desserts	Non-Sausage as a Share of Breakfast Sandwiches	No Cheese as a Share of Adult Mains	No Sauce as a Share of Adult Mains	Main Salad as a Share of Adult Mains	Encouraged Share Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Synthetic control group (ADH-DL-NW)										
Baseline	0.1453***	0.0422*	0.0697**	0.0654**	0.1266*	0.0382***	0.0032	0.0677***	0.0803**	1.7425***
	(0.0335)	(0.0222)	(0.0319)	(0.0284)	(0.0751)	(0.0109)	(0.0112)	(0.0105)	(0.0400)	(0.5598)
Placebo Treatment Dates										
Treat date = week 25	0.0124	0.0136	-0.0222	0.0153	0.0491	0.0027	-0.0006	0.0053	-0.003	-0.1724
	(0.0272)	(0.0212)	(0.0373)	(0.0278)	(0.0569)	(0.0113)	(0.0075)	(0.0145)	(0.0311)	(0.5824)
Treat date = week 50	-0.0186	-0.0054	-0.0151	0.003	0.0439	0.0054	0.01	-0.0068	-0.0135	-0.2521
	(0.0271)	(0.0268)	(0.0393)	(0.0336)	(0.0723)	(0.0130)	(0.0107)	(0.0129)	(0.0365)	(0.5246)
Spillovers										
Neighboring Stores	0.0612***	-0.0211	-0.0148	0.0065	-0.102	0.0042	-0.0009	0.0199	0.0053	5.2417
	(0.0168)	(0.0158)	(0.0288)	(0.0177)	(0.0680)	(0.0115)	(0.0074)	(0.0204)	(0.0209)	(3.7108)

Notes

*, **, and *** denote p<0.10, p<0.05, and p<0.01, respectively.

ADH-DL-NW: The dependent variable is the difference between the outcome in treatement store and the mean for all other stores, by week. Sample size is 125 weeks. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.

Placebo Treatment Dates: Estimation is ADH-DL-NW. Sample includes all stores and all weeks.

Spillovers: Estimation is ADH-DL-NW. Sample is restricted to stores more than 15 miles from the center of Portland (controls stores) and a single weighted average of the five closest stores ("treatment store").

Table 11: Placebo Stores: Number of Positive Statistically Significant Encouraged Share Coefficients (Max 9)

	Store Count
Zero	4
One	16
Two	12
Three	2
Four	3
Five	1
Six	0
Seven	0
Eight	1 (Treatment Store)
Nine	0

Notes

Count of statistically significant (at the 10% level or better) coefficients for the first nine columns of Table 7 regressions.

Synthetic control group: The dependent variable is the difference between the outcome at the treatment store and the outcome for the synthetic control group, by week. Synthetic controls are constructed separately for each outcome. Stores are matched using five week averages in the pre-treatment period.