Making Nutritional Information Digestible:
Effects of a Receipt-Based Intervention on Fast Food Purchases

Preliminary: please do not quote without authors' permission

Kelly Bedard
Peter Kuhn
University of California, Santa Barbara

March 9, 2013

Abstract

We find that the introduction of receipts that include both nutrition information and ordering suggestions had a measurable effect on purchasing behavior at a fast-food chain. The strongest and most persistent effects were on within-category item substitutions that were encouraged by the messages, such as substituting ham or bacon for sausage in a breakfast sandwich, or switching from fries to a side salad. The results suggest that providing easily-understood information about healthier product substitutions that are based on a consumer's current purchases can affect purchasing behavior on future restaurant visits.

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 that consumption of restaurant food, particularly fast food, contributes to obesity.[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 are needed.

We refer to the intervention studied in this paper as the *Health Receipt*. In June 2009, it was implemented at a trial store within a fast food chain we shall Hamburg Chain. A distinct feature of this intervention is that it is printed on customers' *receipts*; thus in principle it could be implemented either as a substitute for, or in addition to calorie posting on menu boards. 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 Hamburg Chain over a 125-week period, we estimate the effects of the *Health Receipt* treatment on a variety of outcomes. Aside from the distinct nature of the *Health Receipt* intervention, novel features of our study include a two-stage econometric approach that, in addition to accounting for store and week fixed effects in a difference-in-differences design, allows for persistent unobserved store-specific shocks. In addition, since we have 38 possible control stores, we also 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 find do not detect a statistically significant effect of the *Health Receipt* intervention on total calories per transaction at Hamburg Chain. We do, however, estimate a reduction in average cholesterol per transaction of 3.0 percent. More significantly, we find robust evidence that consumers shifted their mix of purchases in the direction encouraged by the most common purchasing suggestions. For example, the share of adult main courses items requesting "no sauce" increased by 6.1 percent, side salads as share of adult sides increased by 12.6 percent, and the share of breakfast sandwiches without sausage increased by 5.4 percent. Together, these patterns suggest that Hamburg Chain consumers respond not only to the calorie and fat counts printed on all *Health Receipt*s but to the customized ordering suggestions as well.

Why might proposed item substitutions affect consumer choices in a way that is distinct from the effects of nutritional information alone? One important consideration is that –as the restaurant lobby has argued—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 set 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] 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. Thus, receipt-based suggestions may provide a useful

---

[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.

additional tool in the battle against obesity, either in addition to or as a substitute for menu-based interventions.[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, 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.

---

[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.

As noted, the key existing studies of calorie posting are Bollinger, Leslie and Sorenson (2011); and Wisdom, Downs and Loewenstein (2010).[10] 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 a 6 percent decrease, but it constitutes a small fraction of daily calories.[11]

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.[12] The context was one in which no restaurants operating in the market were required to post nutrition information. Pooling across 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.[13] 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 County, and served as controls. Their econometric approach does not appear to include store

---

[10] 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.

[11] Bollinger et al.'s regression tables do not indicate whether their standard errors are clustered or that 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.

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

[13] 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.

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 *Health 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 Hamburg Chain 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 Hamburg Chain restaurant.

3. Data and Descriptive Statistics

Our data consist of weekly purchase information for all 39 restaurants operated by Hamburg Chain 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 switched from a receipt containing price information only to the *Health Receipt*. During our study period, no form of mandatory calorie posting was required in any jurisdiction containing a Hamburg Chain store. 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 Hamburg Chain implementation, aside from promoting healthier choices more generally, the recommendations in the *Health Receipt*s focused on encouraging nine main changes in ordering patterns. These were, (1) to substitute to a side salad for any adult-size fries[14]; (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 just egg) 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

---

[14] 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.[15] 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 Hamburg Chain 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 Hamburg Chain. 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 Hamburg Chain 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.[16] The average transaction at the treatment store in the pre-treatment period included 1651 calories and 79.96 grams of fat. In comparison, the non-treated stores sold 4 (5) 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 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.[17]

In Table 2 we turn our attention to the consumption of menu items that were encouraged or discouraged in the intervention's overall messaging strategy. Specifically, the numbers in this Table are the mean weekly share of transactions that include the (encouraged or discouraged)

---

[15] Some of the item substitutions encouraged in the Hamburg Chain 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 Hamburg Chain menus and pricing are set at the corporate level.

[16] 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.

[17] 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.

item in question. As in Table 1 and in all subsequent tables, these numbers are normalized to equal one in the treatment store in the pre-treatment period. Since the receipts included a large number of messages that were triggered by the customer's choices, the five "discouraged" items in Table 2 refer only to items that were systematically discouraged in a number of messages, and *never* encouraged in any message.[18] Similarly, the nine encouraged items were systematically encouraged in a number of messages, but never discouraged in any message. Any apparent treatment effects in Table 2 thus refer to the impact of the *overall* messaging strategy in the Hamburg Chain receipts on the share of transactions that include these encouraged or discouraged items. Examples of messages corresponding to each of the discouraged and encouraged items are provided in Appendix B.

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, Table 2 shows no strong overall movement out of discouraged and into encouraged items. In fact, the simple difference-in-difference estimator suggests that the treatment did not change the number of discouraged items purchased, though it may have raised the number both encouraged items slightly.[19]

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 the previous two tables, here the difference-in difference estimates are almost all in the expected direction and substantial in magnitude. For example, the share of salads in adult sides rose by 7.3 percent, as did the share of kids' meals that included apples and milk. The share of frozen yogurt in desserts rose by 8.5 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 8 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:

---

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

[19] Trends in the total number of discouraged items are dominated by fries, which are purchased in a very high share of transactions.

$$(1) \qquad Y_{sw} = \varphi_s + \gamma_w + \delta T_{sw} + \varepsilon_{sw},$$

where Y is the outcome of interest, s denotes store, and w denotes week. 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.

a) Alternative standard errors

Donald and Lang (2007) argue quite convincingly 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 a very real 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.[20]

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.[21] Denoting these 125 differences by D_w, the second stage of the DL procedure regresses them on a dummy for the 50 post-treatment weeks, P_w,:

$$(2) \qquad D_w = \alpha + + \delta P_w + \mu_w.$$

Note that, 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

---

[20] 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.

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

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; this allows 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, 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 available 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.[22] Typically (and in our application) the vector of weights is restricted to be nonnegative and to sum to one.[23] 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.

In the baseline results reported here, X includes only the 75 values of the outcome variable in the pre-treatment period.[24] 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. 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 (FE) 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. 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

---

[22] The discrepancy between the treated group's characteristics, X_1, and the weighted mean characteristics of the synthetic control, $X_0 W$, is specified as $\sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$, 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.

[23] This restriction to positive weights that sum to one 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. By construction, the synthetic control cannot have observable characteristics that fall outside the range of actual comparison units found in the data.

[24] Thus, in each case we simply use the bundle of control stores that best matches the temporal evolution of the treatment store's outcome before the treatment is implemented. 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.

detailed in the last section.[25] 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 total calories and fat 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 0.58 percent decrease in calories per transaction as a result of the treatment; this appears to be statistically significant at the 1% 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 almost four 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: 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 *Health Receipt* from chance patterns.

Rows 4 and 5 show that the ADH optimal synthetic cohort approach in most cases leads to only modest changes in the point estimates; standard errors on the other hand tend to fall slightly relative to the comparable specification in rows 2 and 3. Specifically, the estimated effect on cholesterol remains essentially unchanged. Overall, based on the point estimates reported in Row 5, the introduction of the *Health Receipt* reduced cholesterol per transaction by 3.0 percent and had no significant discernible impact on any other outcome. Notably, none of our measures of sales at the store level (transactions, revenue, or revenue per transaction) was significantly affected by the treatment, and all the point estimates are positive in sign. Thus, it does not appear that introducing the *Health Receipt* had detrimental effects on the treatment store's business.[26]

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 *Health Receipt* treatment. Interestingly, if

---

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

[26] 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 *Health Receipt*s were introduced at the treatment store, since the receipts could conceivably attract or alienate customers and change where they eat. No effect was found.

we restrict our attention to effects that occur within the first 25 weeks after the introduction of *Health Receipt*s, we see statistically significant reductions in all three of our measures of fat intake: total fat, saturated fat, and cholesterol. Of these, only the cholesterol effect continues into the next 25 weeks. 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 *Health Receipt*. In particular, Tables 5 and 6 report the impact of *Health Receipt* introduction for consistently "discouraged" and "encouraged" items, respectively. Beginning with the discouraged items in Table 5 and focusing on the synthetic control estimates reported in Row 5, the only precisely estimated effects are a 7.5 percent reduction in ice cream orders and an 11.8 percent decline in sausage breakfast sandwiches. Looking at the timing of these effects in more detail reveals, perhaps surprisingly, that the response appears to be greater more than 25 weeks after the treatment than immediately afterwards. Row 7 of Table 5 also shows marginally significant *increases* in purchases of adult and kids' fries emerging more than 25 weeks after the treatment; as we show below these and other similar results may be artifacts of demand shocks to narrower item categories, such as kids' meals as a group.

Table 6 reports the companion results for encouraged items. Again, focusing on the synthetic-control estimates that are statistically significant at conventional levels, the introduction of the *Health Receipt* increased requests for grilled chicken sandwiches by 5.1 percent, for 'no sauce' by 7.4 percent, raised side salad purchases by 13 percent, kids' apples by 11.1 percent, and kids' milk by 14.7 percent. The introduction of the receipt did not have a detectable impact on requests for main salads, no cheese, frozen yogurt, or non-sausage breakfast sandwiches. According to the final column of Table 6, the total number of encouraged items per transaction increased by 1.4 percent.

Looking at the persistence of these effects in rows 6 and 7 reveals a variety of patterns. Some effects (main and side salad) fade or disappear after the first 25 weeks, while others (no sauce, kids' milk and kids' apples) 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. As in Table 5, we think that some of these unexpected patterns are artifacts of category-specific demand shocks, which we address in Table 7 below.

c) Within-category Item Substitutions:

One possible concern about the previous section's items-per-transaction approach 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

for frozen desserts.[27] 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. Altogether, 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, suggesting 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. While the standard errors in row 5 are too large to reject a zero overall treatment effect for four of the outcomes (chicken sandwiches, kids' milk, frozen yogurt, and 'no cheese'), the introduction of the receipt increased the share of salads in adult side dishes by 12.6 percent, and the share of kids meals with apples by 9.8 percent. Similarly, the treatment reduced the use of sausage in breakfast sandwiches by 5.4 percent, the share of adult main courses requesting no sauce by 6.1 percent and the share of adult main courses that were a main salad by 7.5 percent.

Breaking the post-treatment period in half in the last two rows of Table 7 again reveal heterogeneous patterns in the duration of these effects. For example, the impact fades for side salads, breakfast sandwiches and main salads. It persists for 'no sauce', and in the case of kids' apples actually takes some time to emerge after the treatment begins. A priori, it is not clear which of these patterns we should expect. Whether an effect fades or not depends in part on whether customers enjoyed the proposed substitute after trying it. Delayed effects are also possible, especially for infrequently-purchased items, since the customer has to return to the restaurant at least once after receiving a message to implement the substitution it proposes.[28]

---

[27] 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.

[28] Unless, of course, customers share information on the printed receipts.

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. Rows 2 and 3 change the way we construct the synthetic control group as follows: instead of matching the entire pre-treatment output pattern (75 weekly observations) we match a smoothed version of that pattern. Specifically, row 2 matches three-week averages of the pre-treatment outcome (75/3 = 25 observations), and row 3 matches 75/5 = 15 five-week averages. The (informal) motivation is that high-frequency purchase variation may consist largely of noise, and matching this noise between the treatment and control stores may compromise our goal of picking control stores with the same low-frequency trends as the treatment store.[29] According to Table 8, if anything these changes slightly strengthen our results: in Row 3, we now measure statistically significant, positive treatment effects on eight of the nine encouraged items, compared with only five in our original specification (reproduced in Row 1). Overall, the estimated coefficients are similar in magnitude, with the exception of the grilled chicken and yogurt shares, which increase substantially.

Rows 5 and 6 of Table 8 consider two alternatives to the simple 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. 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, of course, possible that the *Health 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 mechanism for such an effect are 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, even though this could only apply to the next restaurant visit), 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 *Health 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

---

[29] Matching at these lower frequencies also more closely mirrors ADH's original application of their methodology, which was to a relatively short panel of annual data on U.S. states.

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 save a significant number of 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 *Health Receipt* messages as implemented at Hamburg Chain. 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 *Health Receipt* treatment on the incidence of these three choices 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.[30] Interestingly, the results in this column are more consistent with the *canceling-out* scenario above: the introduction of the *Health Receipt* system is associated with a three-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 had under 500 calories, though here the effect is both smaller and magnitude and borderline in statistical significance. In contrast, however, customers seem to have downsized their fries by a small but highly statistically significant amount when *Health Receipt* was introduced, even though this change was never suggested in the receipts.[31] 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.

7. Discussion

We show that the introduction of receipts that include both quantitative nutritional information and customized ordering suggestions designed to reduce fat and calories had a measurable effect on purchasing behavior at Hamburg Chain restaurants. While we find no statistically significant effect on calories purchased, we do find some effects for fat intake. In particular, we measure a robust and long-lasting (at least up to 50 weeks after treatment) three-percent reduction in cholesterol per transaction, and weaker effects on other measures of fat. While these behavioral changes could have been caused both by the nutritional information and the ordering suggestions on the receipts, we find strong evidence that customers in the aggregate made most of the item substitutions that were promoted by the receipts. Specifically, customers

---

[30] As in many fast food restaurants, Hamburg Chain 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.

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

substituted side salads for fries, 'main dish' salads for sandwiches, apples for fries in children's meals, non-sausage for sausage breakfast sandwiches, and requested 'no sauce' on their sandwiches more frequently. In some specifications, we also detect a shift from fried to grilled chicken sandwiches, substitution of milk for soda in children's meals, and substitution of yogurt for ice cream in frozen desserts. These effects however vary in their permanence, with some being only temporary changes. The estimated effects are highly 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. Placebo tests for alternative treatment dates and stores show no consistent effects.

Overall, we think that our results send a qualified, positive message about the possible health impacts of receipt-based interventions like the *Health Receipt*. One the one hand, the receipts seem to have been (perhaps surprisingly) effective at promoting specific item substitutions in a healthy direction, leading to large percentage increases in purchases of items like side salads for adults and side apple slices for children. On the other hand, we detect no overall effects on calories purchased, and only a modest (three percent) reduction in cholesterol, with weaker effects on total fats. We can think of at least two reasons behind these muted overall effects. First, customers may have made other item substitutions that 'cancel out' the substitutions that were made in response to the suggestions. Indeed, we find evidence of at least one such change: the decline in high-fat (i.e. sausage) breakfast sandwiches that was encouraged by the messages was accompanied by an increase in hash browns ordered. Second, several of the encouraged items (whose consumption rose substantially in percentage terms when *Health Receipt*s were introduced) constitute a very small share of fast-food purchases. Examples include side salads, kids' apples and grilled chicken sandwiches.[32] Even large substitutions towards items like these will have only minimal effects on total fat and calories purchased.

All of this being said, we also note that there is no evidence that introducing the *Health Receipt* was harmful to our treated store for any measure of total sales; indeed our preferred point estimates for sales effects are all positive, though statistically insignificant. And the cost of installing and operating the *Health Receipt* system are minimal. Thus, the system may still be a cost effective way of generating healthier restaurant food choices. As in all studies of restaurant consumption, however, we caution that customers may 'undo' any nutritional improvements at the restaurant being studied with changes in food intake elsewhere, not only at home but at other restaurants. On the other hand it is also possible that our estimates understate the potential of restaurant-based interventions, since 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 have reworked their menus in response to actual and anticipated calorie labeling requirements (Bernstein 2011).

---

[32] Our agreement with Hamburg Chain prohibits us from reporting these item 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.

Methodologically, 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 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 deal with *persistent* group-specific shocks. Our results also highlight the importance of assessing the sensitivity of difference-in-difference estimates to the choice of control group. In our case, we show that our results are robust to several methods of constructing those groups, including the Abadie, Diamond, and Hainmueller (2010) procedure to construct 'optimal' control groups.

Finally, our results may also contain a lesson for the design of restaurant-based interventions aimed at generating healthier food choices: *salience* may not be the only, or the most important mechanism via which such mechanisms can work. Interventions (such as the *Health Receipt*) that attempt to mitigate the cognitive demands associated with choices from lists may offer additional potential in the battle against obesity.

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_50 8.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
(excluding identifying restaurant information)

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 XXX 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 Hamburg Chain 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 XXX hazelnuts." Or, "Salmon is rich in omega-3 fatty acids, which are good for heart health and brain function."

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)] (6)
Nutrition per transaction						
Calories	1650.917	1.000	0.977	1.040	1.023	-0.006
Total fat (grams)	79.959	1.000	0.975	1.047	1.031	-0.008
Saturated fat (grams)	22.183	1.000	0.919	1.021	0.951	-0.011
Cholesterol	153.122	1.000	0.956	1.017	0.999	-0.027
Sales						
Revenue per transaction	--	1.000	1.044	1.027	1.069	0.002
Total transactions	--	1.000	0.976	1.167	1.101	0.042
Total revenue	--	1.000	1.022	1.203	1.184	0.040

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 (1)	Period 2 (2)	Period 1 (3)	Period 2 (4)	[(2)-(1)] - [(4)-(3)]
Discouraged Items:					
Fries (adult)	1.000	0.985	1.015	0.997	0.002
Fries (kids)	1.000	1.045	1.127	1.130	0.041
Fried chicken sandwich	1.000	1.012	0.849	0.854	0.007
Ice cream	1.000	0.783	1.433	1.231	-0.014
Breakfast sandwich w. Sausage	1.000	0.867	0.965	0.979	-0.147
Total Discouraged Items	1.000	0.980	1.031	1.010	0.001
Encouraged Items:					
Main salad	1.000	1.226	0.842	1.014	0.055
Grilled chicken sandwich	1.000	0.836	0.939	0.705	0.070
No sauce	1.000	1.053	0.856	0.851	0.058
No cheese	1.000	0.975	1.139	1.104	0.010
Side salad	1.000	1.143	0.779	0.851	0.071
Frozen Yogurt	1.000	1.007	0.824	0.933	-0.102
Kids' apples	1.000	1.224	0.965	1.082	0.107
Kids' milk	1.000	1.307	1.193	1.375	0.125
All other breakfast sandwiches	1.000	1.005	0.996	1.055	-0.054
Total Encouraged Items	1.000	1.019	1.047	1.044	0.023

Treatment begins at the treatment store starting in week 76. Period 1 includes weeks 1-75 and period 2 includes 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	0.774	0.856	0.073
Share of Chicken Sandwiches:					
Grilled	1.000	0.918	1.035	0.909	0.043
Share of Kids' Meals:					
With Apples	1.000	1.141	0.861	0.938	0.063
With Milk	1.000	1.223	1.064	1.208	0.079
Share of Desserts:					
Frozen Yogurt	1.000	1.222	0.621	0.758	0.085
Share of Breakfast Sandwiches:					
Without Sausage	1.000	1.061	1.018	1.036	0.043
Share of Mains:					
With No Cheese	1.000	0.978	1.022	0.989	0.010
With No Sauce	1.000	1.053	0.771	0.766	0.059
With Main Salad	1.000	1.230	0.759	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				Revenue per Transaction	Total Transactions	Total Revenue
	Calories (1)	Total Fat (2)	Saturated Fat (3)	Cholesterol (4)	(5)	(6)	(7)
All other stores as controls							
1. One-stage FE	-0.0058***	-0.0084***	-0.0107***	-0.0272***	0.0023	0.0425	0.0398***
	(0.0020)	(0.0020)	(0.0021)	(0.0025)	(0.0021)	(0.0338)	(0.0101)
2. DL	-0.0058	-0.0084*	-0.0107	-0.0272***	0.0023	0.0425*	0.0398
	(0.0045)	(0.0047)	(0.0070)	(0.0048)	(0.0045)	(0.0242)	(0.0282)
3. DL-NW	-0.0058	-0.0084	-0.0107	-0.0272***	0.0023	0.0425	0.0398
	(0.0074)	(0.0077)	(0.0110)	(0.0068)	(0.0079)	(0.0387)	(0.0484)
Synthetic control group							
4. ADH-DL	-0.0021	-0.0068*	-0.0093*	-0.0300***	0.0032	0.0054	0.0331
	(0.0040)	(0.0040)	(0.0056)	(0.0048)	(0.0041)	(0.0208)	(0.0234)
5. ADH-DL-NW	-0.0021	-0.0068	-0.0093	-0.0300***	0.0032	0.0054	0.0331
	(0.0059)	(0.0055)	(0.0081)	(0.0061)	(0.0068)	(0.0284)	(0.0335)
Synthetic control group (ADH-DL-NW)							
6. Weeks 76-100	-0.0103	-0.0144**	-0.0188*	-0.0269***	-0.0071	0.0196	0.0422
	(0.0071)	(0.0065)	(0.0106)	(0.0058)	(0.0085)	(0.0266)	(0.0332)
7. Weeks 101-125	0.0061	0.0008	0.0001	-0.0332***	0.0136**	-0.0087	0.024
	(0.0058)	(0.0056)	(0.0081)	(0.0083)	(0.0058)	(0.0370)	(0.0418)

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.

ADH-DL-NW: 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. The standard errors adjusted for first-order autocorrelation via the Newey-West method, with maximum lag set at 5 weeks.

Synthetic controls are constructed separately for each outcome. Stores are matched according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.

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)	Total Discouraged Items (6)
All other stores as controls						
1. One-stage FE	0.0024	0.0409***	0.0067	-0.0142	-0.1472***	0.0015
	(0.0026)	(0.0087)	(0.0055)	(0.0162)	(0.0171)	(0.0023)
2. DL	0.0024	0.0409**	0.0067	-0.0142	-0.1472***	0.0015
	(0.0061)	(0.0186)	(0.0163)	(0.0548)	(0.0262)	(0.0061)
3. DL-NW	0.0024	0.0409	0.0067	-0.0142	-0.1472***	0.0015
	(0.0100)	(0.0251)	(0.0215)	(0.1116)	(0.0263)	(0.0106)
Synthetic control group						
4. ADH-DL	0.0036	0.0121	-0.0057	-0.0745***	-0.1182***	-0.0006
	(0.0054)	(0.0186)	(0.0161)	(0.0234)	(0.0263)	(0.0050)
5. ADH-DL-NW	0.0036	0.0121	-0.0057	-0.0745*	-0.1182***	-0.0006
	(0.0080)	(0.0251)	(0.0179)	(0.0392)	(0.0280)	(0.0081)
Synthetic control group (ADH-DL-NW)						
6. Weeks 76-100	-0.0085	-0.0279	-0.0228	-0.0124	-0.0897***	-0.0124
	(0.0092)	(0.0317)	(0.0149)	(0.0233)	(0.0330)	(0.0086)
7. Weeks 101-125	0.0156*	0.0521**	0.0114	-0.1367**	-0.1467***	0.0111
	(0.0083)	(0.0231)	(0.0248)	(0.0556)	(0.0349)	(0.0087)

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 treatment 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 according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.

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	Total Encouraged Items
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All other stores as controls										
1. One-stage FE	0.0545***	0.0701***	0.0579***	0.0103***	0.0708***	-0.1021***	0.1070***	0.1251***	-0.0538***	0.0243***
	(0.0100)	(0.0102)	(0.0054)	(0.0033)	(0.0111)	(0.0362)	(0.0128)	(0.0138)	(0.0174)	(0.0023)
2. DL	0.0545*	0.0701**	0.0579***	0.0103	0.0708***	-0.1021	0.1070**	0.1251***	-0.0538*	0.0243***
	(0.0278)	(0.0318)	(0.0127)	(0.0075)	(0.0249)	(0.0623)	(0.0439)	(0.0290)	(0.0287)	(0.0055)
3. DL-NW	0.0545	0.0701*	0.0579***	0.0103	0.0708**	-0.1021	0.1070**	0.1251***	-0.0538	0.0243***
	(0.0353)	(0.0383)	(0.0206)	(0.0115)	(0.0285)	(0.1186)	(0.0517)	(0.0350)	(0.0376)	(0.0056)
Synthetic control group										
4. ADH-DL	0.0434	0.0508*	0.0738***	-0.0036	0.1300***	0.0041	0.1109**	0.1473***	-0.0292	0.0139***
	(0.0306)	(0.0270)	(0.0123)	(0.0070)	(0.0257)	(0.0617)	(0.0461)	(0.0303)	(0.0277)	(0.0052)
5. ADH-DL-NW	0.0434	0.0508*	0.0738***	-0.0036	0.1300***	0.0041	0.1109*	0.1473***	-0.0292	0.0139***
	(0.0380)	(0.0284)	(0.0169)	(0.0095)	(0.0301)	(0.1172)	(0.0578)	(0.0403)	(0.0349)	(0.0053)
Synthetic control group (ADH-DL-NW)										
6. Weeks 76-100	0.1199***	0.036	0.0495**	-0.003	0.1652***	0.2711***	0.0144	0.0952**	0.0283	0.0152*
	(0.0263)	(0.0256)	(0.0227)	(0.0149)	(0.0312)	(0.1018)	(0.0543)	(0.0390)	(0.0409)	(0.0089)
7. Weeks 101-125	-0.0331	0.0656*	0.0981***	-0.0043	0.0949**	-0.2629***	0.2074***	0.1995***	-0.0867***	0.0127***
	(0.0451)	(0.0384)	(0.0169)	(0.0090)	(0.0422)	(0.0700)	(0.0699)	(0.0573)	(0.0328)	(0.0044)

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 treatment 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 according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.

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

	Side Salads as a Share of Adult Sides	Share of Grilled 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	Share of Non-Sausage 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
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All other stores as controls									
1. One-stage FE	0.0727***	0.0434***	0.0632***	0.0791***	0.0855***	0.0425***	0.0102***	0.0588***	0.0796***
	(0.0114)	(0.0080)	(0.0104)	(0.0093)	(0.0275)	(0.0063)	(0.0025)	(0.0049)	(0.0082)
2. DL	0.0727***	0.0434*	0.0632*	0.0791***	0.0855	0.0425***	0.0102	0.0588***	0.0796**
	(0.0263)	(0.0228)	(0.0332)	(0.0274)	(0.0604)	(0.0127)	(0.0076)	(0.0108)	(0.0310)
3. DL-NW	0.0727**	0.0434*	0.0632*	0.0791**	0.0855	0.0425***	0.0102	0.0588***	0.0796*
	(0.0308)	(0.0256)	(0.0331)	(0.0304)	(0.0933)	(0.0113)	(0.0125)	(0.0156)	(0.0437)
Synthetic control group									
4. ADH-DL	0.1264***	0.0182	0.0975***	0.0516*	0.0454	0.0540***	0.0072	0.0608***	0.0751**
	(0.0276)	(0.0217)	(0.0334)	(0.0276)	(0.0679)	(0.0125)	(0.0072)	(0.0104)	(0.0323)
5. ADH-DL-NW	0.1264***	0.0182	0.0975***	0.0516	0.0454	0.0540***	0.0072	0.0608***	0.0751*
	(0.0353)	(0.0211)	(0.0366)	(0.0319)	(0.0995)	(0.0111)	(0.0111)	(0.0104)	(0.0394)
Two-Time Periods: Synthetic control group (ADH-DL-NW)									
6. Weeks 76-100	0.1848***	0.0237	0.0350	0.0746	0.2395**	0.0696***	0.0141	0.0559***	0.1571***
	(0.0340)	(0.0213)	(0.0249)	(0.0492)	(0.0963)	(0.0122)	(0.0146)	(0.0128)	(0.0298)
7. Weeks 101-125	0.0680	0.0127	0.1601***	0.0286	-0.1487*	0.0384***	0.0002	0.0658***	-0.0068
	(0.0430)	(0.0253)	(0.0473)	(0.0291)	(0.0851)	(0.0119)	(0.0144)	(0.0133)	(0.0424)

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 treatment 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 according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.

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 (1)	Share of Grilled Chicken (2)	Apples as a as a Share of Child Meals (3)	Milk as a as a Share of Child Meals (4)	Frozen Yogurt as a Share of Desserts (5)	Share of Non-Sausage Breakfast (6)	No Cheese as a Share of Adult Mains (7)	No Sauce as a Share of Adult Mains (8)	Main Salad as a Share of Adult Mains (9)
Synthetic control group (ADH-DL-NW)									
1. 1-week (Baseline)	0.1264***	0.0182	0.0975***	0.0516	0.0454	0.0540***	0.0072	0.0608***	0.0751*
	(0.0353)	(0.0211)	(0.0366)	(0.0319)	(0.0995)	(0.0111)	(0.0111)	(0.0104)	(0.0394)
2. 3-week averages	0.1576***	0.0416*	0.0787**	0.0686**	0.106	0.0531***	0.0046	0.0799***	0.0621*
	(0.0358)	(0.0213)	(0.0301)	(0.0329)	(0.0740)	(0.0122)	(0.0106)	(0.0119)	(0.0369)
3. 5-week averages	0.1497***	0.0427*	0.0699**	0.0635**	0.1260*	0.0385***	0.0032	0.0669***	0.0803**
	(0.0380)	(0.0221)	(0.0319)	(0.0283)	(0.0751)	(0.0109)	(0.0112)	(0.0105)	(0.0399)
Other control groups									
4. DL-NW (Baseline)	0.0727**	0.0434*	0.0632*	0.0791**	0.0855	0.0425***	0.0102	0.0588***	0.0796*
	(0.0308)	(0.0256)	(0.0331)	(0.0304)	(0.0933)	(0.0113)	(0.0125)	(0.0156)	(0.0437)
5. 5 most similar sized stores	0.0810**	0.0711***	0.0627*	0.1029***	0.0668	0.0539***	0.0088	0.0726***	0.0795*
	(0.0330)	(0.0252)	(0.0361)	(0.0251)	(0.0905)	(0.0135)	(0.0117)	(0.0132)	(0.0414)
6. 5 closest stores	0.0747**	0.0642**	0.1064***	0.0727**	0.1075	0.0621***	0.0143	0.0585***	0.0718*
	(0.0338)	(0.0247)	(0.0360)	(0.0318)	(0.1004)	(0.0122)	(0.012)	(0.013)	(0.041)

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 treatment 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 according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.

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			
1. One-stage FE	-0.0352***	0.0031***	-0.0216***
	(0.0063)	(0.0007)	(0.0025)
2. DL	-0.0352***	0.0031**	-0.0216***
	(0.0130)	(0.0014)	(0.0052)
3. DL-NW	-0.0352**	0.0031	-0.0216***
	(0.0144)	(0.0022)	(0.0078)
Synthetic control group			
4. ADH-DL	-0.0267**	0.0041***	-0.0100*
	(0.0131)	(0.0012)	(0.0051)
5. ADH-DL-NW	-0.0267*	0.0041**	-0.0100
	(0.0152)	(0.0019)	(0.0065)
Synthetic control group (ADH-DL-NW)			
6. Weeks 76-100	-0.0329**	0.0065***	-0.0116
	(0.0150)	(0.0024)	(0.0080)
7. Weeks 101-125	-0.0205	0.0017	-0.0083
	(0.0236)	(0.0018)	(0.0089)

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 treatment 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 according to pre-treatment week-by-week specific pre-displacement levels of the dependent variable.