Title
Micro-marketing healthier choices: Effects of personalized ordering suggestions on restaurant purchases

Authors
Bedard, Kelly
Kuhn, Peter

Publication Date
2015

DOI
10.1016/j.jhealeco.2014.10.006

Peer reviewed

Published in final edited form as:
J Health Econ. 2015 January ; 39: 106–122. doi:10.1016/j.jhealeco.2014.10.006.

Micro-Marketing Healthier Choices: Effects of Personalized Ordering Suggestions on Restaurant Purchases

Kelly Bedard[*] and **Peter Kuhn**[†]

[*]Department of Economics, University of California, Santa Barbara, 93106 USA. kelly.bedard@ucsb.edu, (805) 893-5571.

Abstract

We study the effects of the *Nutricate* receipt, which makes personalized recommendations to switch from unhealthy to healthier items at a restaurant chain. We find that the receipts shifted the mix of items purchased towards the healthier alternatives. For example, the share of adult main dishes 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. The results illustrate the potential of emerging information technologies, which allow retailers to tailor product marketing to individual consumers, to generate healthier choices.

Keywords

restaurant; health; fast food; marketing; obesity; calories; fat

1. Introduction

Over the past several decades, social scientists have studied the effects of policy tools designed to reduce the consumption of products that are harmful to health, such as cigarettes, drugs, and unprotected sex. Interventions that have been studied include outright prohibitions, taxes, publicity campaigns, commitment contracts, mandated disclosure of adverse consequences, and changes in the way choices are presented to consumers. More recently, prompted in part by the rise in obesity across the developed world, researchers have focused on reducing the consumption of sugary and fatty foods, using tools like labeling of packaged foods (Variyam and Cawley 2006), mandatory calorie posting on menus (Bollinger, Leslie and Sorenson 2011), and added convenience of healthier choices (Wisdom, Downs, and Loewenstein 2010). While some studies of these interventions report

[†]Corresponding author, Department of Economics, University of California, Santa Barbara, 93106 USA. peter.kuhn@ucsb.edu (805) 893 3666..

statistically significant effects, policy tools that induce sizable long-term improvements in nutritional choices remain an elusive goal.

A parallel development over the past couple of decades is the growth of information technologies that allow retailers not only to track their customers' purchasing behavior but to design individualized marketing strategies based on that information. For example, commercial software packages like *Adobe Target* provide automated behavioral targeting algorithms that adaptively learn what individual consumers want, and test models of each consumer's preferences against alternatives. New information is continuously gathered from a variety of sources including detailed purchase histories. While systems like these are now widely used to increase firms' sales revenues, their potential to induce health-improving changes in consumer behavior remains largely unexplored.

To that end, this paper studies the effects of an intervention called the *Nutricate* receipt. Designed by *SmartReceipt* Corporation, the receipt technology was implemented at a trial store of *Burgerville*, a restaurant chain in the Pacific Northwest in June 2009. A novel feature of this intervention is the fact that --in addition to providing tabular information on the calories and fat contained in the items the customer just ordered-- it delivers personalized purchase suggestions promoting healthier products that are close substitutes to an item the consumer just bought. While the *Nutricate* receipt is an early and simple example of the use of individual purchase history data to market healthier choices, it may provide some indication of this approach's potential.

Using store-level weekly purchase data from all 39 restaurants operated by *Burgerville* over a 125-week period, we find that the *Nutricate* receipt did, in fact, shift the mix of items purchased in directions encouraged by the most common ordering 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. While the implications of these changes for overall calories and fat consumed at *Burgerville* stores are modest, the results suggest that the next generation of targeted, adaptive interventions might have additional potential. For example, the *Nutricate* system bases its recommendations on the consumer's most recent purchase only, and –because it is printed on the receipt—is not accessible electronically and can only be acted on at the consumer's next purchase. None of these are necessary features of adaptive micro-marketing systems.

Also of interest for the direction of future interventions are the mechanisms that appear to account for the effects of the *Nutricate* receipts in our data. While it is possible that customers are responding primarily to the tabular information on fat and calories printed on those receipts, in the paper we argue this is unlikely because –rather than being broadly based—consumers' item substitutions are quite focused on the items targeted by the receipts' ordering suggestions. Further, because most customers will not be able to act on these ordering suggestions until their next restaurant visit, it seems unlikely that the suggestions are mitigating problems of impulse control (Laibson 1997; O'Donoghue and Rabin 1999) or otherwise affecting the immediate decision environment at the time of purchase via framing effects or jogging a consumer's memory. Instead, we suggest that the

Nutricate receipts work primarily because the individualized ordering suggestions provide new, possibly restaurant-specific information in a form that mitigates well known cognitive constraints associated with choices from lists (Rubinstein and Salant, 2006). Choice from lists characterizes many consumer decision problems; with lists becoming ever longer due to the expansion of internet commerce, mechanisms that improve the effectiveness of such choices may have significant social value.

2. Previous Studies

While some other interventions designed to reduce caloric intake in restaurants have been considered, most of the research to date studies the effects of calorie-posting on menus.[1] Since New York introduced mandatory calorie posting in 2008, a number of other jurisdictions including California, Seattle, and Philadelphia followed suit.[2] The best known of the calorie-posting studies are probably Bollinger, Leslie and Sorenson (2011) and Wisdom, Downs and Loewenstein (2010). 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. The 11 calorie decline is statistically significant, but constitutes less than half a percent of recommended daily calories.[3]

Wisdom, Downs, and Loewenstein (2010) designed a pair of field experiments where a small number of Subway customers were randomly assigned to different types of menus. 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 imprecise. 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. Dumanovsky et al. (2011) conduct a similar study, but using data from New York restaurants only; they found modest reductions in calories purchased in

[1]An earlier literature studies the effects of the 1990 *Nutrition Labeling and Education Act* (*NLEA*), which mandated nutritional labeling of packaged foods (see for example Variyam and Cawley 2006). Most studies find small impacts. Other researchers have asked whether *access* to fast food has increased obesity, with decidedly mixed results; see for example Davis and Carpenter (2009), Currie et al. (2010) and Anderson and Matsa (2011).

[2]At the time of writing, the Federal Drug Administration was still reviewing national calorie posting regulations mandated by Section 4205 of the 2010 Patient Protection and Affordable Care Act. We discuss the significance of our findings in view of the ACA mandate in Section 8.

[3]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.

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. 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 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 using micro-marketing methods based on a customer's purchase history to encourage health-improving choices in any commercial context, including restaurants.

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 December 27, 2007 to May 19, 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. 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.

While our confidentiality agreement with *Burgerville* limits the amount of information we can provide about *Burgerville*'s stores and customer base, Table 1 of the online Appendix provides some contextual information on Multnomah County. Multnomah County includes central Portland, more than one-quarter of *Burgerville*'s stores, and is by far the most populous county in Oregon. Two comparison groups are provided: the most central counties for the four most similarly-sized cities in the United States (Milwaukee WI, Jefferson KY, Oklahoma OK and Baltimore City MD) and the national figures for the United States. Compared to similar central counties and the nation as a whole, Multnomah County has the highest percent white, the highest education level, and the lowest share of children in poverty. It also has relatively low rates of smoking and obesity. Overall, *Burgerville* stores thus serve a relatively healthy, well-educated and prosperous population.[4]

Nutritional information provided by *Burgerville* about their treated store reveals that during the 75-week pre-treatment period, mean nutritional content per transaction was 1657 calories, which 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. An average transaction also contained 80 grams of fat, 23 grams of saturated fat, and 153 mg of cholesterol, providing considerable scope for reductions in these nutrients. Since means of all other variables in our analysis are confidential, all our reported results use normalized data, where

[4]Some summary demographic information on *Burgerville* customers is available to researchers on request from the authors. In brief, it corroborates the notion that *Burgerville* customers are likely better educated than a typical U.S, fast food customer.

the mean of every outcome variable is set equal to 1 during the pretreatment period at the treated store.

How was *Burgerville*'s treatment store chosen, and how does it compare to *Burgerville*'s other stores? Conversations with *Burgerville*'s management indicate that in addition to logistical considerations (ease of rollout and a willing manager), this store's relatively large size and less suburban location were seen as advantageous for the rollout, as it would allow them to judge the response of a large number and wider variety of guests to the new receipt. Thus, the non-treated stores in our sample had on average 58% of the transactions and 59% of revenue of the treated store. On a per-transaction basis, however, the non-treated stores were quite similar to the treated store, with similar revenues and 3.7 (4.4) percent more calories (fat) per transaction.

Appendix A shows an example of a *Nutricate* receipt from a *Burgerville* restaurant.[5] In addition to the standard price and quantity information, the *Nutricate* receipt has two other components. One, the nutrition information table, displays the calorie, fat, fiber and carbohydrate content of the customer's order, separately by item. This information is provided both in absolute amounts and (after aggregation across items) as a percentage of a daily recommended amount. Second, the receipt also contains a message that is customized based on the client's current order. In transactions where the customer's current order contains one of the ‘discouraged’ items, this message is a customized ordering suggestion, proposing that the customer consider a similar substitute item that is lower in fat and/or calories (usually both). For example, one message used at *Burgerville* (triggered by the purchase of a large strawberry milkshake) says “Looking for a simple way to cut calories and fat? Ask for our large Strawberry Milkshake to be made with non-fat frozen yogurt instead of ice cream and cut approximately 36% of the calories and 100% of the fat.”[6] Like this message, most ordering suggestions also provide information on the amount of fat or calories that would be saved by making the substitution.

While our estimates of the *Nutricate* receipt's effect represent the combined impact of its nutrition table and ordering suggestions, we attempt to distinguish the effects of these channels by asking whether the receipt was associated with a broad-based substitution away from all high-fat and high-calorie items towards all lower fat- and calorie items, or with a more targeted shift from the specific discouraged items to their encouraged substitutes. Since fat and calorie information was provided for all items, the former, broad-based pattern is what we would expect if nutrition-conscious customers paid attention only to the nutrition table in the receipts. On the other hand, if customers focused only on the ordering suggestions, only the ‘messaged’ items should be directly affected, and it is even possible that customers could increase their consumption of *Burgerville*'s high-calorie and –fat items that were not specifically discouraged. In this regard, we find strong evidence that customers made the specific item substitutions recommended in the messages, and very limited evidence that they made calorie- or fat-saving substitutions that were not encouraged by the

[5]For more detail on the *Nutricate* receipt, see http://receipt.com/nutrition_solutions.php

[6]Not all messages recommended an altenative item. For example, some consumers who bought grilled chicken sandwiches received the message “By choosing grilled chicken instead of crispy chicken in your sandwich, you saved 90 calories and 8g of fat. Great choice!”

messages. Indeed we find that consumers partially compensated for the decline in discouraged items by increasing their consumption of some other, caloric items that were not discouraged by the messages. Overall, this suggests that the recommended item substitutions played at least some role in addition to the quantitative nutritional information that was provided.

In the *Burgerville* implementation, 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;[7] (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 healthier than and a reasonable substitute for the purchased item.[8]

Throughout the paper, we report results for three main types of outcomes. First, following the existing literature, we report on the total nutritional content --specifically calories and fat-- of a typical transaction at *Burgerville*. Second, we estimate whether the *Nutricate* receipt had any detectable effect on total revenues and sales at the treated store, for example by deterring customers who wanted to avoid the messages. Finally, to shed some light on the effects of targeted suggestions, we study the mix of items purchased. While we present some results for the share of all transactions that included an encouraged or discouraged item, our main analysis focuses on purchase shares of encouraged items within item classes (e.g. the share of frozen yogurt in frozen desserts) within which the messages encouraged consumers to substitute one item for another.

Simple difference-in-difference estimates of the effect of the *Burgerville* treatment on the outcomes studied in this paper are presented in Figures 1-3. The numbers in all these tables represent the difference between the (post- minus pre-treatment) change in our normalized outcome variable in the treatment store and the (mean of) the same change in all the non-treated stores. When multiplied by 100, these are just the relative percent changes in the treated versus non-treated stores.[9] The pre- and post-intervention periods were December 27, 2007 through June 3, 2009, and June 4, 2009 through May 19, 2010 respectively. According to Figure 1, calories, total fat, saturated fat and cholesterol per transaction all fell in the treated store relative to the non-treated stores, though the effect was considerably larger (at 2.67 percent) for cholesterol than the other aggregate nutritional measures.[10] There is also no indication that the *Nutricate* treatment harmed the treatment store's sales; if

[7]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.

[8]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.

[9]The levels of the pre-and post-treatment means from which Figures 1-3 were calculated are provided in Online Appendix Tables 2-4.

[10]Assessing whether the differences in differences in Figures 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.

anything, the treatment is associated with about a one percent increase in the number of transactions per store and a three percent increase in revenues per store.

In Figure 2 we turn our attention to the menu items that were encouraged or discouraged in the intervention's overall messaging strategy. The first five items in Figure 2 –adult and child-size fries, fried chicken sandwiches, ice cream and breakfast sandwiches with sausage—are unambiguously *discouraged* items, which are defined as items that were systematically discouraged in a number of messages, but were never encouraged in any message. The remaining nine *encouraged* items are defined analogously; examples of messages corresponding to each of the discouraged and encouraged items are provided in Appendix B.[11] The numbers in Figure 2 are the relative (post- minus pre-treatment) percent changes in the share of weekly transactions that include the item in the treatment versus the non-treated stores. Any apparent treatment effects in Figure 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. Taken together, the patterns in Figure 2 are not clear cut. While purchases of seven of the nine encouraged items increased in the treated store relative to the non-treated stores, the other two items (frozen yogurt and all other breakfast sandwiches) experienced sizable drops. And while breakfast sandwiches with sausage declined considerably, three of the discouraged items actually increased in the treated store, though these increases were small.

One possible concern with Figure 2'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 for frozen desserts.[12] If this shock is, by accident, correlated with the introduction of our treatment, an items-per-transaction approach would mistakenly conclude that the treatment reduced the purchases of breakfast sausage when in fact it reduced the purchases of all breakfast items. To address this issue, Figure 3 focuses 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); frozen 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).[13]

In sharp contrast to Figure 2, Figure 3's 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 7.41 percent, with slightly larger increases in the share of kids' meals

[11] To give a broader sense of the set of messages, summary statistics for the entire population of messages in the system on a particular post-treatment date are provided in Part B of the online Appendix.

[12] 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.

[13] Another advantage of the within-category shares approach relates to the fact that *Burgerville* transactions can serve multiple customers. While the average order size suggests this is relatively infrequent, and average order size did not respond to the *Nutricate* treatment, a per-transaction approach could still be vulnerable to changes in the allocations of transactions to customers that happen to coincide with the treatment.

that included apples and milk. The share of frozen yogurt in desserts rose by 8.41 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. Given the apparent confounding effects of category-specific shocks, the remaining econometric analyses will focus on the outcomes in Figure 1 (store-level nutrition and sales) and Figure 3 (shares of encouraged and discouraged items within categories).

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:[14]

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

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.

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.[15]

[14]As both the Donald and Lang (2007) and Abadie, Diamond, and Hainmueller (2010) methods used below require a balanced panel, we linearly interpolated three of these 4875 observations. According to *Burgerville*, these store-week cells were affected by technical difficulties with their point-of-sale (POS) software.

[15]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.

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.[16] 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:

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

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 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.[17]

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

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

[17]In additional analysis (available on request) we take an arguably even less restrictive approach by simply conducting *t*-tests (allowing for unequal variances) for differences between two means for each outcome in Table 2: the difference between the treated store and the synthetic cohort store before the treatment (75 observations) and the same difference after the treatment (50 observations). The resulting differences are –by definition-- identical to the corresponding regression coefficients Table 2. As it turns out, the standard errors are very similar to the DL-NW standard errors in that Table.

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.[18] Typically (and in our application) the vector of weights is restricted to be nonnegative and to sum to one.[19] 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 means of the outcome variable in the pre-treatment period, grouped into fifteen five-week windows.[20] 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.

[18]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.

[19]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..

[20]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.

5. Regression Results

The regression results reported in Tables 1 and 2 follow the progression described in Section 4. Panel A shows the (identical) coefficient estimates from the one-stage fixed effects procedure, the two-stage Donald-Lang procedure (DL), and DL with Newey-West standard errors (DL-NW), plus the standard errors corresponding to each of these three approaches. As noted, the DL-NW 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. Panel B repeats the above exercise, 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. Since there is no obvious counterpart to the FE approach with the synthetic control group, we only present DL and DL-NW standard errors in Panel B. Finally, Panel C 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 1 reports the aforementioned results for nutrition and sales. Columns 1-4 estimate treatment effects on calories, fat, saturated 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 the alternative standard errors in Panel A shows substantial differences that affect judgments of statistical significance. 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: 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 Panel A, 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.

Panel B shows that the ADH optimal synthetic cohort approach leads to some changes in the point estimates, though the estimated cholesterol effect is roughly similar at 2.1 percent. Standard errors on the other hand tend to fall slightly relative to the comparable specification Panel A. As in Panel A, the only statistically significant nutritional effect is on cholesterol per transaction. 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.[21]

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 Figure 4, which compare time trends in the treatment store to an average of all other stores (in part a) or to our synthetic control groups (part b) for the four nutritional outcomes in Table 1.[22] Clearly, the synthetic controls in part (b) track the treatment store much more accurately during the pre-treatment period, and illustrate a divergence in total fat and cholesterol that seems to emerge after the treatment, consistent with Table 1's estimates. The temporary nature of the total fat effect is also evident. These effects are less visually evident in part (a).

Finally, Panel C of Table 1 breaks 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. The treatment had no statistically significant effects on total items, total transactions, or total revenue. Thus it does not appear that the *Nutricate* receipt was either a net deterrent to customers –who might find it intrusive—or a net attractor of customers –who might start coming because they find the new receipts helpful.

b) Within-category item substitutions

In Table 2 we ask whether changes in the mix of items bought at *Burgerville* correspond to the item substitutions recommended in the *Nutricate* receipts. As argued earlier, our approach focuses on changes in the share of the nine encouraged items within narrow item categories, i.e. on the outcomes summarized in Figure 3.[23] 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 Figure 2. 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, the point estimates now all have the expected sign, and all but one are statistically significant. 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.[24]

---

[21] 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.

[22] 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.

[23] Regression results for the share-of-transaction outcomes in Figure 2 are provided in Online Appendix Tables 5 and 6. They show short-term (20 week) decreases (increases) in an index of discouraged (encouraged) item purchases, but less consistent patterns for individual items than Table 2.

To measure the extent to which the treatment was associated with a broad-based within-category shift towards the encouraged items, column (10) of Table 1 reports results for an index that combines all nine items. To construct this index, we normalized each store's category share for the item in question to have a mean of zero and a standard deviation of one, then added the nine normalized shares together. Thus, column (10) shows strong evidence 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 2 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, while seven of the nine only encouraged items exhibit statistically significant increases during the first 25 weeks after treatment, this share falls to four of nine in the following 25 weeks. Overall, the treatment's effects appear to be stronger in the six months or more so after the intervention than after that.

The results of Table 2 are illustrated visually in Figure 5, which shows time trends for the treatment store versus synthetic controls for all ten outcomes studied in the Table. Consistent with Panel C 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 2, a number of the above effects either vanish or significantly attenuate after 25 weeks—specifically, those for grilled chicken, kid's milk, frozen yogurt, non-sausage bagels and meal salads. Altogether, Table 2 and Figure 5 provide strong evidence that consumers initially tried out the substitution suggestions in the *Nutricate* receipts, but evidence of longer-term changes in purchase patterns is more mixed.

6. Robustness Tests

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 each outcome variable as well as possible. In Table 3 we explore the effects of using other definitions of the control group, focusing on the results for within-category shares in Table 2. Panel A explores 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 5-week averages. 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.

Panel B of Table 3 shows the effects of two alternatives to using a simple average of all the non-treated stores as a control group. The first uses the just five stores closest in size (measured by total revenues) to the treatment store, while the second uses the more common

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

criterion of geographical distance, selecting the five closest stores. (Panel B also reproduces our baseline estimate Table 2, for easy 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. As already noted, this is something we would expect if customers were reacting primarily to the nutritional information tables on the receipts: these tables may have induced customers to make calorie- or fat-saving item substitutions that were not explicitly recommended. The receipts may also have created an overall atmosphere of health-consciousness that spilled over to non-messaged items. An alternative possibility is that customers who complied with the ordering suggestions (but ignored the nutrition table) made other item substitutions that cancelled out their calorie-saving impact. 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.[25]

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.[26] 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 4. The regression and control group specifications are identical to those in Tables 1 and 2.

The dependent variable in column one of Table 4 is the share of breakfast sandwich orders that did not include hash browns.[27] 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 3.9 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.[28] 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.[29] As noted, possible explanations

[25]Unexpected effects of nutritional messages have been documented for the case of packaged food labeling, by Wansink and Chandon (2006) and Kiesel and Villas Boas (2013), among others.

[26]Unfortunately, customers serve their own sodas from a fountain in *Burgerville* stores, so information on diet versus regular sodas is not available in their transaction data.

[27]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 4 is the share of breakfast sandwiches sold that were not part of a basket.

[28]Unlike the other two effects in Table 4, the effect on low-calorie mains also becomes insignificant after 25 weeks of treatment (see Panel C).

for this include a positive ‘atmosphere’ effect and 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 2 and 3 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. Panel B of Table 5 address this question by estimating the effects of placebo treatments that occurred at different dates than the actual treatment. Specifically, Panel B estimates 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 our preferred baseline specification in Table 2, which is reproduced in Panel A for easy comparison. Of the twenty estimated coefficients in Panel B, 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 Panel B of Table 2 (i.e. the ADH-DL-NW specification) 39 times: once for each store in our sample. Each time, we designated the store in question as the treatment store and assumed treatment 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 2. Table 6 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 6 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.

Another robustness check concerns the possibility of spillovers in purchase 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. Another important possibility is that the *Nutricate* treatment caused some customers to switch stores. For example, while there was no detectable change in total sales at the treated store, it is conceivable that the treatment attracted some health-conscious consumers while repelling an equal number of persons who were annoyed by the *Nutricate* receipts. To the extent that these customers came and went from neighboring stores, this

[29]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.

behavior would appear as a negative spillover in our data, i.e. a decline in the share of encouraged items at neighboring stores.

To assess these possibilities, we asked 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 Panel C of Table 5 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. This suggests that changes in customer mix are probably not the main explanation for our estimated treatment effects.

d) Shorter observation windows

A well-known challenge for difference-in-difference estimates like the ones in this paper is controlling for time-varying shocks that differentially impact the treated and non-treated groups. So far, we have attempted to confront this challenge by constructing a variety of control groups and by adjusting standard errors for persistent unobserved store-specific shocks. Still, in most cases the ability of these methods to provide valid counterfactuals tends to deteriorate as the temporal distance from the introduction of the treatment rises. To address this problem, Appendix Tables 1 and 2 estimate the effects of the treatment using successively narrower windows around the intervention. The specification is the preferred one in our main estimates (ADH-DL-NW in Table 1), but using data from only 50, 40, 30, 20 or 10 weeks before and after the treatment was introduced (week 75). In addition to vitiating the need to control for unobserved store-specific trends far from the introduction of treatment, another advantage of this approach is that --to the extent that it takes customers some time to learn about the new receipt and start to switch stores-- this method may also remove the effect of any changes in customer mix in reaction to the receipts. A drawback is the fact that it can only estimate short run effects of the treatment; this is noteworthy because, as noted, *Nutricate* receipts can only affect behavior on the first purchase after they are seen.[30]

Bearing these caveats in mind, Appendix Table 1 shows that the treatment effect on cholesterol remains statistically significant with roughly the same magnitude as our main specification, even when using data from only the 20 weeks before and after the intervention. Interestingly, the estimated effects on calories and total fat are negative and significant in this specification also. Using the same two twenty-week windows, Appendix Table 2 shows that purchases of seven of the nine encouraged items increased as a share of their category (compared to all nine in the baseline specification). The encouraged share index increases significantly using only ten weeks of data before and after the intervention. Taken together, the appendix tables show that customer reaction to the new receipts was

[30]While a confidentiality agreement does not allow us to publish numerical estimates, individual *Burgerville* stores do get a substantial amount of repeat business. Additional information on repeat purchase frequencies is available to researchers on request from the authors.

relatively rapid, suggesting that changes in customer mix are probably not the main reason why the treatments affected the mix of purchases at *Burgerville*. They also increase our confidence in our main estimates, and suggest that the *Nutricate* receipt may have had at least short-term (20 week) effects on total calories and fat purchased at *Burgerville*, not just on cholesterol.

7. Mechanisms

In attempting to understand the mechanisms via which the *Nutricate* receipt may have operated in our context, recall again that in most cases, information printed on any receipt can only affect choices on a customer's next visit to a store or restaurant. Thus, it seems unlikely that the receipts operate by changing the immediate decision environment when the customer is placing his or her order, for example by jogging the customer's memory, by re-framing the decision, by priming the customer to pay attention to nutrition, or by ameliorating problems associated with impulse control or present bias (Laibson 1997; O'Donoghue and Rabin 1999). We also note that the *Nutricate* receipt was introduced into a context where calorie posting on menus was not mandated. Thus, both the nutrition information table and the customized ordering suggestions may have given consumers some information they did not have before. Accordingly, it seems likely that the receipts work either by giving customers new information (some of which may be restaurant-specific), or perhaps by helping customers process information they already have access to.

Having ruled out some possible mechanisms we now ask which component of the *Nutricate* receipt –the nutrition table or ordering suggestions— best explains the receipt's observed effects. To that end, suppose first that customers looked only at the nutrition table printed on their receipt, which only provides information on the items the customer just ordered. If some of the information provided there is new to consumers, and if at least some customers prefer a *ceteris paribus* reduction in fat and calories, the introduction of the *Nutricate* receipt should lead to a broadly-based and relatively gradual reduction in those nutrients. The shift should be broadly based, because the nutrition table is provided for every item purchased (not just those specifically discouraged in the messages) and because no specific substitute item is suggested. The shift should be more gradual because nutritional information for substitute items is not provided; thus it might take some time for customers to identify acceptable lower fat and calorie substitutes.

Now, in contrast, imagine that customers only paid attention to the ordering suggestions in the *Nutricate* receipt (thus ignoring the nutrition table). Compared to the previous case, we should expect much more targeted changes in purchase behavior. Only the 'messaged' items should be directly affected, and it is even possible that customers could increase their consumption of high-calorie and –fat items that were not specifically discouraged. And while no receipt-based intervention is likely to have immediate effects, we might expect the observed item substitutions to be more rapid, since consumers who follow the suggestion will transition directly to the recommended item. Since our findings are more in line with this pattern, we conclude that the receipts' customized ordering suggestions probably played at least some role in explaining customers' reactions to the *Nutricate* receipt.

Finally, given that customized ordering suggestions 'worked' in our context, what economic or psychological processes might account for their effect? For reasons already discussed, we think that explanations based on changes in the immediate decision environment at the time of purchase, such as impulse control, are unlikely. Remaining possibilities are (a) new information that is conveyed *by the ordering suggestions*, and (b) by alleviating cognitive constraints associated with choice from lists (Rubinstein and Salant, 2006). Concerning the former, consumers may not have been aware that *Burgerville*'s breakfast sausage contains much more fat and calories *than any of their other breakfast meats*, or that the *difference between* ice cream and yogurt is so large. The suggestions may also communicate the information that *Burgerville* servers will be happy to accommodate customized orders (such as holding the sauce or cheese) which may not be the case in other restaurants.

To illustrate the cognitive constraints affecting choice from lists, consider the list of books available on Amazon.com as a 'menu' of books from which customers can choose. *Even with* (and perhaps because of) the extensive information and reviews of each book easily available on the site, making an effective choice from this list is a daunting task, and probably benefits tremendously from the recommendations Amazon makes based on one's past purchases. While restaurant menus are smaller, they can still be highly complex, especially when patrons can combine and customize items. For example, even a seven-item deli sandwich can be ordered in over 4000 ways.[31] Faced with the task of selecting the optimal combination of taste, cost and nutrition from that many choices, consumers could resort to simple habits and 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.[32] In our case, simply informing the consumer that his last purchase was highly caloric still leaves her with the problem of finding an appealing, cost-effective substitute from a large menu. Suggesting such a substitute alleviates this problem.

8. Discussion

We have shown that the introduction of receipts containing customized ordering suggestions promoting healthier menu items had a detectable effect on purchasing behavior at a chain of fast-food restaurants. While we detect only short run effects on total calories and fat purchased, 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. For frequent fast-food customers, this effect could be similar in magnitude to the estimated effects of statins, though the magnitude of the effect is sensitive to how much consumers compensate by increasing cholesterol intake at other meals.[33] Our confidence that this behavioral change is causally linked to the personalized ordering suggestions in the new receipts is increased by the fact that customers in the aggregate made most of the

[31]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.

[32]Relatedly, Hirshleifer, Lim, and Teoh (2009) provide suggestive evidence that information overload (in the form of a large number of earnings announcements) slows price adjustments in stock markets.

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, such as whether and how a synthetic control is calculated and which alternative criteria (geographical distance, similar store size, or all non-treated 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 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 that were encouraged. Indeed, we found 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 purchases 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.[34] Even large substitutions towards items like these will have only minimal effects on total fat and calories purchased.

One noteworthy *caveat* to our results is the fact that --as in all studies of restaurant purchases-- our estimates refer only to purchases at a particular restaurant; thus it is possible that the nutritional improvements we detect may be cancelled out by changes in food consumption at home and at other restaurants. Another caution is that *Burgerville*'s relatively educated Portland-based clientele may be more health-conscious than a typical fast-food restaurant, and correspondingly more responsive to interventions like ours. Related, with our aggregate data we cannot distinguish *which Burgerville* customers changed their behavior most in response to the receipts, and whether these customers are those who would experience the greatest or least health benefits from that behavioral change. 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 newly-available technologies that make individualized purchase suggestions based on a consumer's purchase history. Refinements of the *Nutricate* system might offer additional potential. Suppose, for example, that a customer's favorite restaurant has developed a healthier version of a sandwich he orders frequently—why not inform him

[33]Suppose that every milligram of consumed cholesterol increases LDL by 0.15 mg (an assumption based on the fact that approximately 15 percent of blood cholesterol comes from food-- see Gordon, 2014). If an average fast food meal contains approximately 125 mg of cholesterol, a consumer who eats at *Burgerville* twice per week and does not change their food intake elsewhere would then reduce weekly cholesterol intake by 5 mg, and LDL by 0.15*5 = 0.75 mg/dL. In comparison, a consumer with an LDL reading of 200 who takes a statin that reduces their LDL by 30 percent (Rosenson 2014) would experience a 60 mg/dL reduction.

[34]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.

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. Migrating these systems to the cloud and to consumers’ mobile devices would untie them from paper receipts, and hence from the delay between the *Nutricate* message and the re-purchase decision. Other health-improving options may remain to be discovered.[35]

Another important question about our findings is what they mean in the context of the calorie-posting requirements mandated by Section 4205 of the 2010 Affordable Care Act (ACA), which require U.S. restaurants with 20 or more locations to list calorie content information on their menus.[36] Since no form of mandatory calorie posting was required in any jurisdiction containing a *Burgerville* store during our study period, our results are most directly relevant to a context where mandated calorie posting is absent. Indeed, if the only thing the *Nutricate* receipt did was to provide information that will be made available by menu boards, receipts of this type should have no effect once the ACA menu-posting requirements are in effect. However, the fact that customers seem to respond mostly to the ordering suggestions in the receipt (and not the nutrition tables) suggests that the receipt is doing more than providing quantitative information that was not previously available. If that is the case, technologies like the receipt may have effects even in the presence of menu calorie posting.

One possible reason for this is that information in *Nutricate*-type receipts address some gaps that are likely to exist in the new labeling requirements. These gaps apply to customizable orders (such as combination meals with a variety of sides or sizes), where the FDA's recommended rule is to require only ranges to be displayed; some of these ranges could be so great as to be uninformative. Another gap is that *Nutricate*-type receipts automatically total up the calories in a consumer's order, which requires extra effort in the case of menu posting. Indeed, compared to menu posting --which attempts to display the nutritional consequences of *all possible choices*—the *Nutricate* receipt presents a limited amount of personally relevant information in an easily-digestible form. For example, in the case of the deli sandwich consisting of 7 items in 4096 combinations, providing complete *ex ante* 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 for each nutritional outcome of interest.

Finally, ‘smart’ recommendations such as those in the *Nutricate* receipt could be complementary with the additional menu information mandated by the ACA. For example, returning to the Amazon case, imagine that the government mandated that all books on the site have at least ten consumer reviews before being offered for sale. On its own, this might be marginally helpful (though it could even be counterproductive if the additional reviews

[35]Of course, many of these possibilities –like all aspects of firms’ increasing use of ‘big data’ to track individual consumer behavior —raise issues of privacy and consent, which would need to be addressed via opt-in agreements or other means. Repeated (and badly calibrated) messages to non-motivated consumers could conceivably be counterproductive for both nutrition and sales.

[36]As of April 2014, the regulations implementing Section 4205 were under Office of Management and Budget (OMB) review, which is the final stage before publication.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

were difficult to ignore). In combination with customized purchase recommendations, however, the new information could be more helpful since the consumer can now choose more effectively from among a set of recommended items. For similar reasons, recommendations based on a consumer's purchase history might be more effective in improving nutritional choices in the presence than the absence of the ACA's mandated calorie posting requirements.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was 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.

References

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

Anderson, Michael L.; Matsa, David A. Are Restaurants Really Supersizing America? American Economic Journal: Applied Economics. 2011; 3(1):152–188.

Bassett, Mary T.; Dumanovsky, Tamara; Huang, Christina; Silver, Lynn D.; Young, Candace; Nonas, Cathy; Matte, Thomas D.; Chideya, Sekai; Frieden, Thomas R. Purchasing Behavior and Calorie Information at Fast-Food Chains in New York City, 2007. American Journal of Public Health. 2008; 98(8):1457–59. [PubMed: 18556597]

Bernstein, Sharon. Making Every Calorie Count: Upcoming FDA disclosure rules prompt many chains to overhaul their offerings. Los Angeles Times. Jun 11.2011 :B1, B4. Weds.

Bertrand, Marianne; Duflo, Esther; Mullainathan, Sendhil. How Much Should We Trust Differences-in-Differences Estimates? The Quarterly Journal of Economics. Feb; 2004 119(1):249–275.

Bertrand, Marianne; Karlin, Dean; Mullainathan, Sendhil; Shafir, Eldar; Zinman, Jonathan. What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. Quarterly Journal of Economics. Feb; 2010 125(1):263–306.

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

Card, David. The Impact of the Mariel Boatlift on the Miami Labor Market. Industrial and Labor Relations Review. Jan.1990 43:245–257. 1990.

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

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

Davis, Brennan; Carpenter, Christopher. Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity. American Journal of Public Health. 2009; 99(3):505–10. [PubMed: 19106421]

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

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

Dumanovsky, Tamara; Yuang, Christina H.; Nonas, Cathy A.; Matte, Thomas D.; Bassett, Mary To.; Silver, Lynn D. Changes in energy content of lunchtime purchases from fast food restaurants after introduction of calorie labeling: cross sectional customer surveys. British Medical Journal. 2011; 343:d4464. Published online July 26. doi: 10.1136/bmj.d4464. [PubMed: 21791497]

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

Finkelstein EA, Strombotne KL, Chan NL, Krieger J. The Impact of Mandatory Menu Labeling In One Fast Food Chain in King County. Washington. American Journal of Preventive Medicine. Feb; 2011 40(2):122–127.

Gordon, Jerry. [October 17, 2014] Blood Cholesterol versus Dietary Cholesterol. 2014. HowStuffWorks.com. Available at: http://health.howstuffworks.com/diseases-conditions/cardiovascular/cholesterol/cholesterol2.htm

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

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

Iyengar, Sheena S.; Lepper, Mark R. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? Journal of Personality and Social Psychology. 2000; 2000; 79(6):995–1006. [PubMed: 11138768]

Kiesel K, Villas-Boas SB. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment. International Journal of Industrial Organization. 2013; 31(2):153–163.

Laibson D. Golden Eggs and Hyperbolic Discounting. Quarterly Journal of Economics. 1997; 112(2): 443–478.

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

O'Donoghue T, Rabin M. Doing It Now or Later? American Economic Review. 1999; 89(1):103–124.

Rosenson, Robert R. [October 17, 2014] Patient information: High cholesterol treatment options (Beyond the Basics). 2014. www.uptodate.com, available at:http://www.uptodate.com/contents/high-cholesterol-treatment-options-beyond-the-basics.

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

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

Wansink, Brian; Chandon, Pierre. Can "Low-Fat" Nutrition Labels Lead to Obesity? Journal of Marketing Research. Vol. 2006; 43:605–617.

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

Burgerville Highlights

- We study the effects of the *Nutricate* receipt, introduced at a fast food chain
- The receipt makes personalized recommendations to switch to healthier items
- The receipts shifted the mix of items purchased towards the healthier alternatives
- The results illustrate the health-improving potential of ‘smart’ micro-marketing

Figure 1.

Raw Difference-in-Difference Estimates, Nutrition and Sales

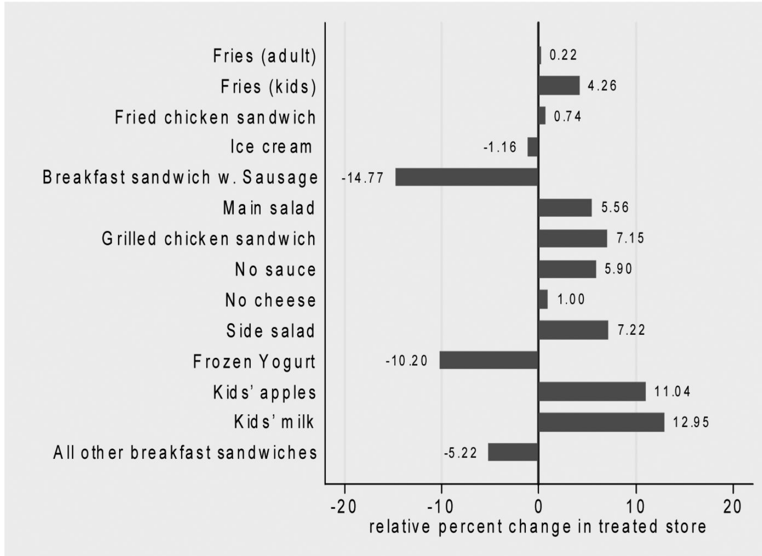

Figure 2.

Raw Difference-in-Difference Estimates, Items per Transaction

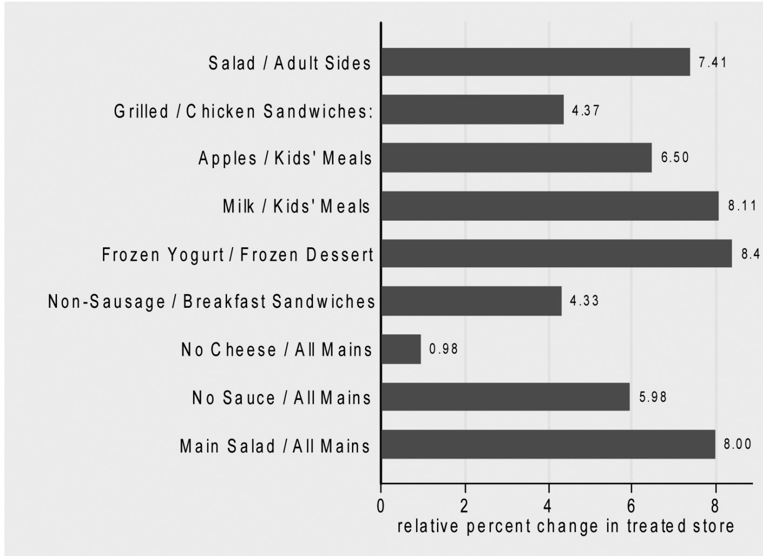

Figure 3.

Raw Difference-in-Difference Estimates, Category Shares

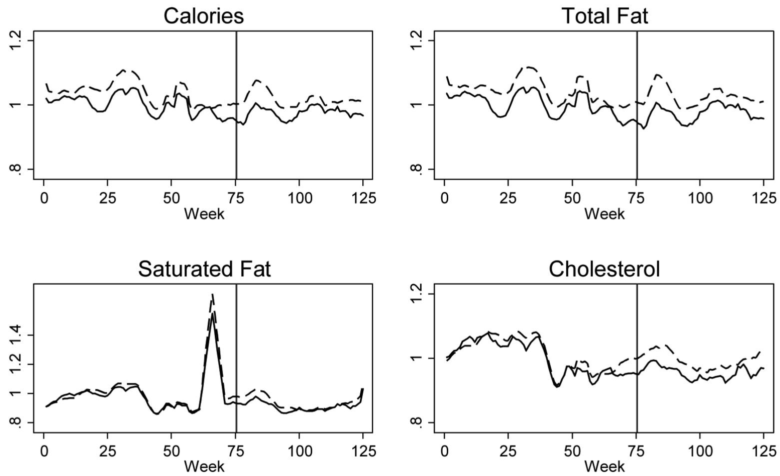

Figure 4a.
Weekly Nutrition for Treatment Store versus All Other Stores

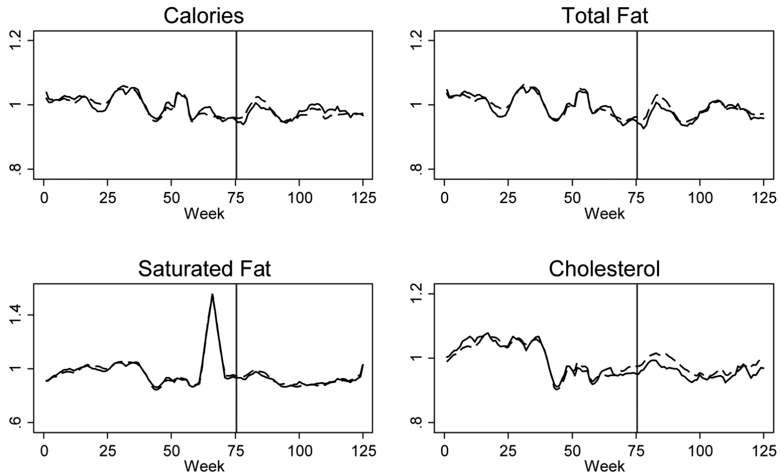

Figure 4b.
Weekly Nutrition for Treatment Store versus Synthetic Control

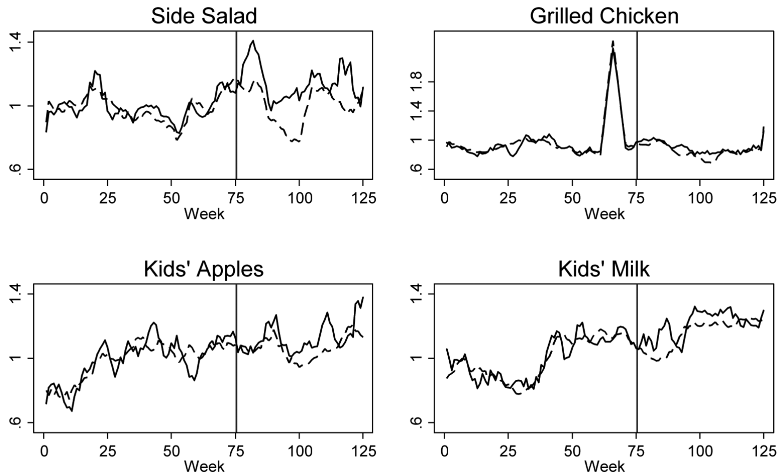

Figure 5a.
Category Shares for Treatment Store versus Synthetic Control

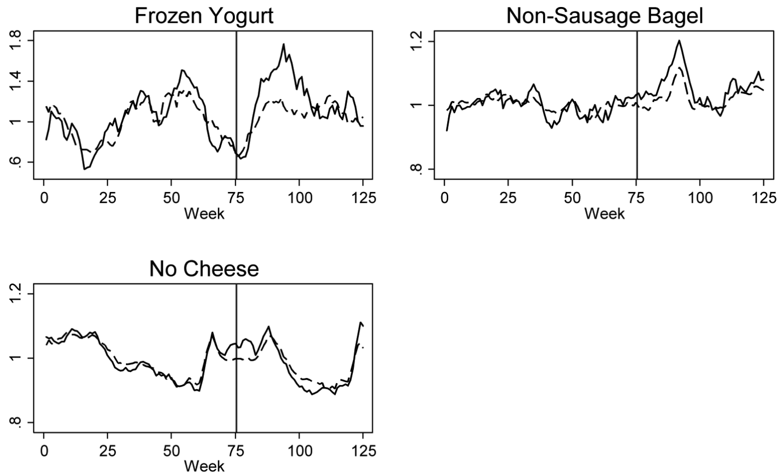

Figure 5b.
Category Shares for Treatment Store versus Synthetic Control

Figure 5c.
Category Shares for Treatment Store versus Synthetic Control