NBER WORKING PAPER SERIES

DIABETES AND DIET:
BEHAVIORAL RESPONSE AND THE VALUE OF HEALTH

Emily Oster

Working Paper 21600
http://www.nber.org/papers/w21600

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2015

I thank Leemore Dafny, Ben Keys, Jim Salee, Steve Cicala, Aviv Nevo, Matt Notowidigdo, Amy Finkelstein, Andrei Shleifer, Jerome Adda and participants in seminars at Brown University, Columbia University, Harvard University, University of Chicago and Kellogg School of Management for helpful comments. Kejia Ren, Angela Li and David Birke provided excellent research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Emily Oster. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Diabetes and Diet: Behavioral Response and the Value of Health
Emily Oster
NBER Working Paper No. 21600
October 2015
JEL No. I12,J17

ABSTRACT

Individuals with obesity often appear reluctant to undertake dietary changes. Evaluating the reasons for this reluctance, as well as appropriate policy responses, is hampered by a lack of data on behavioral response to dietary advice. I use household scanner data to estimate food purchase response to a diagnosis of diabetes, a common complication of obesity. I infer diabetes diagnosis within the scanner data from purchases of glucose testing products. Households engage in statistically significant but small calorie reductions following diagnosis. The changes are sufficient to lose 4 to 8 pounds in the first year, but are only about 10% of what would be suggested by a doctor. The scanner data allows detailed analysis of changes by food type. In the first month after diagnosis, healthy foods increase and unhealthy foods decrease. However, only the decreases in unhealthy food persist. Changes are most pronounced on large, unhealthy, food categories. Those individuals whose pre-diagnosis diet is concentrated in one or a few foods groups show bigger subsequent calorie reductions, with these reductions occurring primarily occurring in these largest food groups. I suggest the facts may be consistent with a psychological framework in which rule-based behavior change is more successful. I compare the results to a policy of taxes or subsidies.

Emily Oster
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
and NBER
emily_oster@brown.edu

1 Introduction

In many health contexts, individuals appear resistant to undertaking costly behaviors with health benefits. Examples include resistance to sexual behavior change in the face of HIV (Caldwell et al, 1999; Oster, 2012) and lack of regular cancer screening (DeSantis et al, 2011; Cummings and Cooper, 2011). Among the most common examples of this phenomenon is resistance to dietary improvement among obese individuals, or those with conditions associated with obesity (Ogden et al, 2007). Encouraging behavior change in this context is of significant policy importance: estimates suggest that the morbidity and mortality costs of obesity were $75 billion per year in the US in 2003 and rising (Wang et al, 2011).

Dietary changes are a significant component of prevention and treatment. Despite this, individuals generally find it difficult to change their diet and lose weight, even in the face of strong health incentives to do so. There are many explanations for this fact. Individuals may lack information about the importance of weight loss; this explanation underlies much of the prevailing policy approach to this problem, which focuses on information campaigns.[1] Individuals may have good information but lack motivation to lose weight, either because they put limited value on their long-term health or because they have high discount rates. It is also possible that individuals are informed and motivated but face external constraints on behavior change.

The basic observation that dietary changes are limited on average does not provide much insight into why, and hampers our ability to design effective policy. In this paper I use data on dietary changes after health news - specifically, news about Type 2 diabetes status - to present a set of facts about dietary response. This setting is a useful laboratory since a diagnosis of Type 2 diabetes has a limited impact on the *benefit* of weight loss, but does come with a large change in information and monitoring of diet. I suggest that we may use this approach to evaluate how individuals change their dietary behavior when given salient information on benefits and guidelines for what to do. I am able to look at overall changes as well as detailed information on changes by food type.

The analysis described above requires observing panel data with detailed diet information among a sample of individuals with a diabetes diagnosis. Standard health data sources do not allow for this. The key data innovation in the paper is the use of the Nielsen HomeScan panel, a dataset which is commonly used in industrial organization and marketing applications. Household participants in the panel are asked to scan the UPC codes of purchases, including all grocery and drug store item purchases.[2] I use purchases of glucose testing products, following a period of exclusion, as a marker for diabetes diagnosis or news.[3] I observe

[1] See, for example, http://ndep.nih.gov/partners-community-organization/campaigns/ for diabetics in particular, and Michelle Obama's "Let's Move!" campaign (http://www.letsmove.gov/).

[2] Panelists participate in the panel for varying periods, but typically for at least a year, and are incentivized for their participation. Other validation exercises have supported the quality of these data (Einav et al, 2010). Throughout the paper I will discuss various issues with the data which will need to be addressed in the empirical work.

[3] A small survey of diabetics confirms that nearly all newly diagnosed diabetics acquire these products within a month of diagnosis, and most of them do so through direct purchase. Glucose monitoring is not a recommended treatment for conditions other than diabetes, so it is unlikely this procedure identifies non-diabetics.

UPC-level evidence on food purchasing behavior before and after this event. I merge these data with a second dataset which provides calorie and nutrient information for foods, so I am able to observe an estimate of calories purchased as well as quantities and prices.

The two significant limitations of the data, discussed in more detail below, are (1) it is available only at the household level and (2) it does not measure food away from home. I limit the primary analysis to two person households, with some robustness checks using the (small) sample of single-person households. I discuss bounding the estimates given this approach. I address the second issue by looking at changes in a subset of foods (example: breakfast cereals) which are rarely purchased away from home.

Given the data, the methodology in this paper is straightforward. Using a household fixed effects framework, I estimate the evolution of food purchase behavior after diagnosis. I argue this provides a causal effect of a diagnosis in the household on diet.

I begin by using these data to confirm the observation that behavior change is limited. I find statistically significant reductions in calories purchased after diagnosis, but the changes are small. A simple illustration, which also illustrates the identification strategy, can be seen in Figure 2a. Calorie purchases are flat leading up to and including the month of (estimated) diagnosis, and then decline. This decline persists, but is only about 2% of total calories. I discuss additional assumptions to convert this change into a weight loss magnitude; I predict a weight loss in the range of 4.3 to 8.3 pounds in the first year. This is consistent with weight loss estimates after diabetes diagnosis from other sources (Feldstein et al (2008); the Health and Retirement Survey). I subject these results to a large set of robustness test with varying specifications, exclusion periods, time trends and sub-samples.

Following this I turn to analysis which dis-aggregates changes by food group. In the first month, changes by foods are broadly consistent with doctor dietary advice. To precisely measure this advice, I fielded a small survey of doctors who treat diabetics and asked them to rank food modules as a "good source of calories," a "bad source of calories" or "neither good nor bad." I group foods as "All Good" (indicating that all doctors surveyed felt this was a good source of calories), "All Bad" (all doctors felt it was bad), "Majority Good" and "Majority Bad". In the first month, households purchase more calories (and quantities) of the foods doctors say are good and fewer of those doctors say are bad. By the second month following, the decline in the bad food group persists, but the increase in good foods fade.

In addition to heterogeneity by food quality, I also find there is heterogeneity across food groups based on their importance in the pre-period diet. Calorie declines occur disproportionately among the largest unhealthy food categories.

I then explore heterogeneity across households in calorie reductions. There is limited variation across demographic groups, despite strong demographic differences in diagnosis. However, there is strong evidence of heterogeneity by *ex ante* dietary characteristics. Individuals whose pre-diagnosis diet is more concentrated

(i.e. a larger share of their calories are accounted for by a small number of food groups) have much larger dietary changes than those whose diet is less concentrated. This is true holding constant total number of calories pre-diagnosis and the total number of food categories purchased.[4] This effect is driven by differential changes across concentration groups in the largest pre-period food categories.

Together, the evidence suggests that overall changes in behavior are small, but that changes in the largest unhealthy food categories play a crucial role in behavioral response. Further, more successful dieters seem to be characterized by having a diet concentrated in one or a few food categories, and their success is driven by larger reductions on these particular foods. In Section 5 of the paper I argue this may reflect findings in psychology which suggests that identifying specific behavior changes (such as "don't eat a particular food") may produce greater overall change than more general advice ("eat a healthier diet"). To the extent this helps explain the patterns in the data, it suggests greater behavior change could be achieved on average by encouraging everyone – even those with less concentrated diets – to focus their efforts on a small number of food groups.

The small effects estimated here suggest approaches other than individual-motivated behavior change may be more productive in combating obesity. I make this concrete by comparing the behavior change here to what we would expect to obtain with a program of taxes and subsidies on foods, using external data on price elasticity to infer the impact of taxes. I find that moderate taxes (in the 10% range) would achieve similar reductions in unhealthy foods and, in terms of increases spending on healthy foods, subsidies are likely to have a larger impact than the changes here.

The primary contribution of this paper is to better understand this important health behavior and to speak to how obesity policy may be improved. A secondary contribution, however, is to illustrate a new way that household scanner data might be used by health researchers. Although these data are commonly used in industrial organization and marketing applications, they have not been used to evaluate questions in health.

2 Background on Diabetes and Diabetes Management

Diabetes is a medical condition in which the pancreas cannot create enough insulin. There are two types. In Type 1 diabetes, the pancreas cannot make any insulin; this disease typically manifests in childhood and individuals with the illness must manage it with insulin injections to replace pancreatic function. In Type 2 diabetes the pancreas produces some insulin, but not enough to process all glucose consumed. This illness more commonly manifests in adulthood and is very often a complication of obesity. Medical treatment of Type 2 diabetes includes oral medication and, if the disease progresses, injected insulin. This paper will focus on Type 2 diabetes, which is more common and more responsive to behavior modification.

[4]This relationship is not mechanical; it does not appear in an analysis of non-diabetics.

The health consequences of Type 2 diabetes relate to the possible buildup of glucose in the blood. This buildup can damage blood vessels, leading to a variety of problems. Complications from poorly managed diabetes include blindness, kidney failure, amputation of extremities (feet in particular), heart attack and stroke. Even with treatment, Type 2 diabetics have significantly elevated mortality risk compared to non-diabetics (Taylor et al, 2013). Similar to other complications of obesity, Type 2 diabetes is on the rise in the US. An estimated 29 million Americans live with the disease, and 1.7 million new cases are diagnosed each year (CDC, 2014). The vast majority of these are Type 2 diabetes. Estimates from 2012 put the annual cost of diabetes to the US health care system at $176 billion, with $69 billion in further costs from reduced productivity (American Diabetes Association, 2013).

A central component of diabetes treatment is changes in diet and exercise behavior. Diet recommendations are made by the American Diabetic Association (Franz et al, 2002) and have several components. First and foremost is weight loss. A very large majority of Type 2 diabetics are overweight or obese, and the ADA recommends weight loss through a deficit of 500 to 1000 calories per day relative to what would be required for weight maintenance. The ADA also makes recommendations on the makeup of these calories: roughly 60-70% should be from carbohydrates, 15-20% from protein and less than 10% from saturated fat. Although in general a diet rich in whole grains and vegetables is recommended, the ADA has in recent periods noted that the total calorie intake is more important than the source.

The observation that weight loss is an important component of diabetes treatment is reasonably well accepted (Wilding, 2014). Williamson et al (2000), for example, shows individuals who lose weight after diagnosis have approximately a 25% decreased mortality rate compared to those who do not lose or who gain weight. Intensive lifestyle intervention has been shown to produce disease remission in a limited share of individuals (Gregg et al, 2012). The evidence is not uniform: a recent large-scale randomized trial has demonstrated limited benefits of a weight loss intervention on overall mortality, although many intermediate outcomes were affected (Wing et al, 2013). In Appendix A I discuss in more detail evidence of the impacts of weight loss on various health outcomes among diabetics.

It is quite important to note that the benefits to weight loss are also very large *prior* to diagnosis. At least two randomized controlled trials (Lindstrom et al, 2006; Diabetes Prevention Program et al, 2002) have shown that weight loss programs for individuals at risk for (but not yet diagnosed with) diabetes can reduce the chance of diabetes onset. Given the large impact of diabetes on mortality, these changes have significant mortality impacts. Progression to diabetes entails changes in pancreatic function that are difficult or impossible to reverse; avoiding those in the first place is of value.

Given this, the change in the medical benefit to weight loss upon diagnosis is likely quite small (it could even be negative). A major change at diagnosis, however, is the frequency of interaction with the medical system and the severity of the advice given. I argue it is therefore appropriate to think of diagnosis as largely

an information treatment. Before and after the individual feels physically similar, and has a similar objective benefit to weight loss. The difference is they are provided with a much more specific and directed set of dietary advice and more frequent feedback on progress.

3 Data and Empirical Strategy

The primary data used in this paper cover consumer purchases and are collected by Nielsen through its HomeScan panel. In addition, I make use of data from a small survey of doctors on dietary advice and data on calorie contents of foods. These sources are described in the first subsection below. The second subsection discusses data limitations. The third subsection describes the empirical strategy used.

3.1 Consumer Purchase Data

3.1.1 Nielsen HomeScan

The primary dataset used in this paper is the Nielsen HomeScan panel. This dataset tracks consumer purchases using at-home scanner technology. Individuals who are part of the HomeScan panel are asked to scan their purchases after all shopping trips; this includes grocery and pharmacy purchases, large retailer and super-center purchases, as well as purchases made online and at smaller retailers. The Nielsen data records the UPC of items purchased and panelists provide information on the quantities, as well as information on the store. Prices are recorded by the panelists or drawn from Nielsen store-level data, where available. Einav, Leibtag and Nevo (2010) have a validation of the reliability of the HomeScan panel. I use Nielsen data available through the Kilts Center at the University of Chicago Booth School of Business. This data covers purchases from 2004 through 2013.

I construct measures of quantity of food purchased in ounces and total expenditures. Where necessary, I convert non-ounce measurements (i.e. pounds) into ounces. In the case of products which are recorded in counts (i.e. eggs) I use external evidence on the weight of the item to convert to ounces.

All Nielsen household are asked to scan all items with UPC codes; this will exclude items like loose coffee, loose vegetables or butcher-counter meats, among others. A subset of households, called Magnet Households, are asked to record these items as well. These records are typically limited to prices. Throughout the paper I will show results on expenditures for the whole sample as well as for Magnet households alone, which will give a sense of the importance of the exclusion of these items.

In addition to purchase data, Nielsen records demographic information on individuals. This includes household size, structure, income, education of the household heads and age of household heads and children. The data also include information on individual zip codes. I merge in data from the USDA on "food deserts"

by zip code; these are defined as low income census tracts more than 1 (10) miles away from a supermarket in urban (rural) areas.

The analysis will rely on the subset of two-person households for whom I infer a diabetes diagnosis during the panel (this inference is described in detail in Section 3.3). This includes roughly 4000 households; summary statistics for these individuals appear in Table 1.[5] Panel A summarizes the demographics of these households, and Panel B summarizes characteristics of their trips and purchases.

3.1.2 Gladson Product Information Data

I merge the Nielsen data with nutrient information purchased from Gladson.[6] Gladson maintains a database of information on consumer products, including virtually all information available on the packaging. The primary objects of interest are total calories and the nutrient breakdown. I use a single pull of the Gladson data as of 2010.

The Gladson data does not contain a UPC match for every code in HomeScan. I undertake a sequential match procedure similar to what is used in Dubois, Griffith and Nevo (2014). For 61% of purchases there is a direct UPC match to Gladson. For products which do not have a match in the Gladson data, I impute nutrition values based on product module, brand, description and size. I calculate average nutrition per size from the matched products and multiply it with the product sizes of the unmatched products to obtain the imputed values.[7]

Calorie and nutrient summary statistics appear in the final rows of Panel B of Table 1. The average household records purchases of 1460 calories per person per day, with 11% of calories from protein, 13% from saturated fat and 53% from carbohydrates.

3.1.3 Doctor Survey Data

The discussion in Section 2 provides a sense of the general dietary advice for diabetics. To get more specific information, I fielded a small survey of doctors. Seventeen primary care doctors who treat individuals with diabetes were surveyed about food choices for diabetic patients. They were provided with a list of food items designed to correspond to categories in the Nielsen HomeScan data (examples: applesauce, shrimp, frozen vegetables). For each one, the doctors were asked to indicate if the item is a "Good Source of Calories", a "Bad Source of Calories" or "Neither Good nor Bad". In the analysis below I classify foods into four groups: "All Good" (all 17 of the doctors reported this as a good source of calories), "Majority Good" (the majority of

[5] Income, age and education are given in categories. For the purposes of summary statistics, I recode at the median of the categories. I will use the categories directly in any demographic analyses later.

[6] More information is available at http://www.gladson.com/.

[7] I mark products whose nutrition per size is more than 3 standard deviations away from the mean as outliers. I calculate averages ignoring these outliers. In addition, I can impute values for an unmatched product using matched products with identical product description or, more broadly, identical product module. I choose the criterion with the lower variance in nutrient values within matched products.

doctors report this as a good source of calories), "Majority Bad" (majority of doctors report this as a bad source of calories; this category includes foods with an equal number of good and bad rankings) and "All Bad" (all 17 of the doctors report this food as a bad source of calories). Appendix B lists the full set of items and their rankings.

3.2 Data Limitations

This data has some significant advantages in addressing the questions here. The monitoring is passive, so we worry less about Hawthorne effects. I observe food choices before and after diagnosis for the same individual, which has not been possible in large-scale data before. Finally, the data is available at a very detailed food level. However, there are a number of limitations in the data which deserve discussion.

A first central issue is that I observe only a subset of what households buy and consume. This is true for two reasons. First, Nielsen panelists do not scan food purchased away from home. Second, even within the subset of food at home, it is very likely that individuals do not record all purchases. Einav et al (2010) validate the HomeScan data using a match with records from a retailer and suggest slightly less than half of trips are not recorded at all; among trips which are recorded, they find a high level of accuracy.

To get a sense of the magnitude of this issue, I compare with food diary data from the National Health and Nutrition Examination Survey (NHANES). Although the food diaries recorded in the NHANES are likely also be subject to under-reporting, the issue is likely to be less significant. Using the 2007-2008 NHANES (the date is chosen as the midpoint of the Nielsen sample) I find adults report approximately 1862 daily calories in total. The calorie levels in HomeScan therefore represent approximately 80% of calories (taking the NHANES as a baseline). An alternative baseline is to evaluate this relative to the calorie level which an average diabetic would require to maintain weight. I do a calculation in this spirit later and conclude this figure is approximately 2194. With this baseline, HomeScan records about 68% of calories.

A second issue is that for sample size reasons it is infeasible to limit to single-person households and I will use two-person households. It seems likely that in nearly all cases it is only one household member who is diagnosed, but what I observe is the overall household change. When I come to magnitudes I will again suggest bounding arguments based on assuming that the diabetic individual is responsible for as little as half of the change or as much as all of it. I will also show robustness analysis with single-person households.

Finally, as discussed, non-UPC coded items are recorded only by a subset of households. I will show results for these households separately.

For all of these reasons, the level of calories, quantities and expenditures is somewhat difficult to interpret. I will also report the changes in percentages, which may have an easier interpretation. When discussing magnitudes I will make a set of assumptions which allow me to scale the changes to the overall diet, and compare the predictions to evidence from external data on weight loss after diabetes.

3.3 Empirical Strategy

The key empirical challenge here is identifying the timing of diabetes diagnosis. I do this using information on purchases of glucose testing products. Individuals with diagnosed diabetes need to monitor their blood sugar; doing so requires a glucose monitor and accompanying test strips. Individuals put a drop of blood on the test strip and it is read by the monitor, which reports blood sugar levels. This information is required for individuals to know if they are managing their disease effectively. Test strips are discarded after a single use; the monitor is a durable good.

The identifying assumption is that observing the purchase of any glucose testing product after a period of at least nine months of observing no such purchase is a marker of a new diagnosis. This assumption is consistent with medical guidance. I validate it using a small online survey of diabetics. Among a sample of 43 individuals with Type 2 diabetes who engage in glucose monitoring, 90% reported acquiring either a glucose monitor or test strips within the first month of diagnosis.

It is worth noting that I do not directly observe health information and it is possible that the purchasing behavior observed represents news about diabetes rather than a new diagnosis. In the most general sense, we can think of this as marking some diabetes-related event. Given the exclusion period, however, this event seems most likely to be a diagnosis. I will refer to this event as "diagnosis" for linguistic simplicity, but with this caveat in mind.

Having identified the timing of diagnosis using this procedure, the empirical strategy is fairly straightforward. I use an "event study" method within the household to estimate the response to diagnosis timing. It is possibly important to adjust for other non-diagnosis time effects - in particular, time in the Nielsen sample (which could increase or decrease recorded purchases) and month-year effects. Doing this within the household fixed effects framework generates within household co-linearity and makes the results difficult to interpret. It also constrains our estimation of these time effects to the small set of timing around diagnosis events.

Given this, I first use the entire sample - including individuals who are not diagnosed ever during this period - to residualize the outcomes with respect to month-year fixed effects and a linear control for time in the sample.[8] I use these residualized variables in the estimation.

Defining Y_{it} as the residualized outcome for household i in year t, I run regressions of the form:

$$Y_{it} = \beta \mathbf{D_{it}} + \gamma_i \tag{1}$$

where $\mathbf{D_{it}}$ is a vector of indicators for diabetes status for household i in month t and γ_i is a household fixed effect. In the primary analyses, $\mathbf{D_{it}}$ includes indicators for 1 to 5 months before diagnosis (as measured by test

[8] Controlling more flexibly (i.e. quadratic, cubic) for this makes no difference.

strip purchases), first month after diagnosis, 2 to 4 months after diagnosis and 5 to 7 months after diagnosis. Standard errors are clustered at the household level.

Note that I include the month before purchase of monitoring products in the "pre-period" even though individuals are likely to have been diagnosed sometime during this month. In a robustness check I will exclude this month from the analysis. In other robustness checks I will show results in which I vary the way I control for calendar time (excluding time controls or including more detailed time controls), results where I divide the pre-period or lengthen the post-period and results in which I adjust for household-specific pre-trends.

Figure 1 shows the change in spending on testing supplies based on the definition of diagnosis timing used. By construction, the period before the first month of purchases is at zero. The very large spike in the first month is reflective of the fact that by definition individuals purchase some product in this month. In the following months, we see continued purchase of testing products at a lower level.

Table 2 shows a regression of the form described in Equation (1) with testing product spending as the Y_{it} variable. The regression results are consistent with the evidence in Figure 1.

One concern here is that we may not identify all diagnosed individuals. In fact, this is likely given that a share of individuals (about 40% in the online survey) get their monitoring or testing equipment through their doctor or insurer. This will mean we estimate our results from a sub-sample of diabetics, although it does not invalidate the interpretation of the results within this sample. A second concern is that this purchase behavior occurs for reasons other than diabetes diagnosis; this seems unlikely given that there is no other use for these products. A final issue is that this identifies a diagnosis *in the household*, but does not pinpoint an individual. I limit to two person households, but in the end can say concretely only what happens to household behavior after one member is diagnosed. This relates to the data limitations above.

The evidence in the paper primarily makes use of the sample of individuals who fit our diagnosis criteria. However, in a number of places it will be useful to have a sample of individuals who do not fit this criteria for comparison. I generate a sample of this type using the HomeScan data and limiting to individuals who are never observed purchasing diabetes testing products. I randomly choose a "diagnosis" month for these households and then create pre- and post-periods as for the diabetic sample.

4 Evidence on Behavioral Response

This section presents the new facts in the paper. I begin in the first subsection by showing aggregate calorie changes and discussing the magnitudes. The second subsection uses the detailed food-level data to disaggregated the calorie changes. The third subsection discusses heterogeneity in calorie changes across households.

4.1 Aggregate Changes in Calories

Figure 2a shows the change in total calories per month around the inferred diabetes diagnosis; this figure replicates the form of Figure 1. The numbers reported are coefficients in a regression of calendar time-adjusted calories purchased on month-from-diagnosis dummies and household fixed effects.

In the very first month after diagnosis, calories purchased are roughly stable; if anything, increasing a bit.[9] In the months following diagnosis they decline by 1500 to 1800 calories per household per month; this represents a decline of about 2% from the pre-period mean. This decline is fairly stable over the period considered. There is no visual evidence of a pre-trend in the series prior to the inferred diagnosis.

I also consider food quantities and expenditures. The direction of the impacts on these variables is unclear. The recommendation is to decrease calories. This could be accompanied by either an increase or decrease in quantities and expenditures, depending on how the mix of foods changes. Figures 2b and 2c show these results. Quantities and, especially, expenditures increase significantly in the first month. Quantities decline in later months and expenditures return to baseline.

In Table 3, I show the results of estimating Equation (1). Column 1 shows the impact on calories. The evidence in this column echoes Figure 2a: an increase in calories in the first month, and a persistent decrease of about 2% after. Columns 2-3 show impacts on quantities and expenditures for the whole sample; Column 4 shows the expenditure effects for the Magnet households, who also report non-UPC coded items. Again, the evidence in these columns is consistent with Figure 2: slight increases in the first month, followed by decreases (quantities) and no change (expenditures) in the following period. The changes for Magnet households, in Column 4, are larger in magnitude due to the overall higher expenditures in this group (as would be expected since they scan a larger share of purchases). However, the percent changes are extremely close to the overall sample.

I explore a number of robustness checks. I focus on the primary results on calories in Column 1 of Table 3. The regressions appear in Table 4.

The first three columns estimate impacts with varying approaches to time. Recall the primary results residualize everything with respect to month-year fixed effects and a control for time in HomeScan. Column (1) estimate the impacts with no time controls at all. Column (2) estimates the impact with the same controls but dropping the month of diagnosis. Column (3) uses the pre-period to estimate a household-specific pre-trend and adjust for that in the analysis. The results are extremely similar to the baseline in all cases.

Columns (4) and (5) vary the household set. Column (4) looks at single person households. The sample size is smaller and the data is noisier, but the basic patterns remain. The changes are similar when we consider them as shares. In Column (5) I drop the bottom 25% of households based on pre-period

[9] Going forward, I will refer to the first test strip month as the time of diagnosis, with the understanding that this is only an inferred timing based on the strategy described in Section 3.3.

expenditures.[10] Einav et al (2010) suggest a bimodal distribution of reporting quality across households, so dropping the bottom households in terms of expenditures may eliminate some households with poor reporting behavior. The results are similar.

Columns (6) and (7) include either divide the pre-period into two (Column 6) or another post-period (Column 7). The pre-periods are relatively flat in Column (6) and there is no evidence of a drop off in the effect in the longer post-period in Column (7).

Finally, Column (8) attempts to address the concern raised in the data discussion that we do not observe food away from home. I use the NHANES dietary data to identify a subset of foods for which at least 85% of consumption reports indicate are purchased at a store - that is, 85% of the time when I observe the food in the NHANES, it is reported as purchased at a store. The foods included in this sample are not surprising - milk, cereal, frozen dinners, etc. I then limit the analysis to only these foods, to see if behavior change differs. As Column (8) demonstrates, the changes in shares are almost exactly the same as the changes in the full sample of products.

In general, the results in Table 4 suggest that the changes in calories observed in Table 3 are robust across a variety of specifications.

4.1.1 Magnitudes of Weight Loss

The evidence above suggests a 2% reduction in calories in response to diagnosis, but is not sufficient to comment on the magnitude of these changes for overall weight loss. Although the conversion between calories and weight loss is fairly straightforward, it is complicated here because we observe only household-level changes and do not observe all foods individuals purchase. In this section I describe and implement a scaling procedure to comment on magnitudes.

The first issue in scaling is the use of household-level data. It seems reasonable to assume that at least half of the changes in food intake should be assigned to the diagnosed individual. For scaling, I adopt bounds and assume the affected individual accounts for between half and all of the calorie reduction. This means that when we observe a 2% reduction in the overall calories purchased by the household (i.e. as in Table 3 Column 1, averaging the post-periods) the bounds on change for the diagnosed individual are 2% to 4%.

The second issue in scaling is that we do not observe all foods people consume. Even if individuals accurately scan all foods that they purchase at the grocery store, we do not see foods consumed outside the home. Further, if households fail to scan some of their purchased foods, those will not be observed. On average, individuals record 1491 calories purchased per household member per day. I will adopt the simple scaling assumption that the percent change on the items we observe is the same as on the items we do not observe.

[10] I use the 12 month pre-period to get a fuller picture of purchases.

There is some empirical support for this assumption at least as it applies to total grocery purchases. Magnet households, which are asked to record a larger share of purchases, have share changes similar to the overall sample. In Table 4, when I drop households with very limited reporting, we again see very similar changes in shares. Further, when I limit to foods which are consumed largely at home, the share changes remain the same. All of these facts suggest that the share assumption may reasonably describe overall changes in grocery purchases.

These assumptions together imply a range of percent change in calories. I apply these to an estimate of the total caloric intake of the average person in this sample. I generate this based on medical estimates of caloric intake required to maintain weight[11], and use weight estimates for diabetics in a matched age range from the NHANES. This procedure suggests a baseline of 2194 calories on average (2513 for men, 1875 for women).

Using the results in Column 1 of Table 3 and applying the scaling described above, I estimate the overall caloric reduction in the range of 2% to 4%, or between 42 and 84 calories per day. This would translate to between 0.4 and 0.7 pounds per month, or 4.3 to 8.3 pounds per year assuming these changes occur in all months of the year.

It is useful to compare this figure to data on measured weight loss among diabetics after diagnosis. In general, individuals diagnosed with diabetes do seem to lose some weight after diagnosis. I consider two points of comparison. First, Feldstein et al (2008) use electronic medical records to analyze weight change among individuals newly diagnosed with diabetes. Second, I analyze data on weight from the Health and Retirement Survey (HRS) for individuals who change reported diabetes status between survey waves.

The data from Feldstein et al (2008) suggests a weight loss of 5.1 pounds at 8 months; the predicted range from Nielsen is 2.9 to 6.3 pounds. The HRS shows a change of 7.8 pounds after the first wave, comparable to a predicted change of 4.3 to 8.3 pounds in the first year in Nielsen. The match suggests these changes are roughly the right order of magnitude.

It is worth noting that these changes (in these data and in the comparable data) are much smaller than what would be medically recommended for most diabetes patients. The American Diabetes Association (Franz et al, 2002) recommends a caloric deficit of at least 500 calories per day. This reduction would lead to a weight loss of approximately 50 pounds per year. This is of course far above what most individuals achieve.

4.2 Disaggregated Changes by Food Group

The previous section suggests overall calorie changes are small. It is possible, however, that this masks a larger change in diet quality. If there are large increases in consumption of good foods - fruit, vegetables - and decreases in consumption of less healthy foods - candy, cookies, etc - we could see relatively small calorie

[11] Source: HTTP://www.bcm.edu/research/centers/childrens-nutrition-research-center/caloriesneed.cfm

changes but larger health impacts. There is good evidence that even conditional on caloric intake some dietary patterns are better than others (see, for example, Estruch et al (2013) on the Mediterranean diet) so such changes could matter for health.

A key feature of the Nielsen data is that I observe detailed information about which foods are purchased and, by extension, where calorie reductions stem from. In this section I use this detailed food-level data to estimate whether there is heterogeneity across food groups in the diabetes response.

As a first step, I simply estimate changes by product group. There are sixty-four product groups in the data (examples: cookies, dairy desserts, carbonated soft drinks). Of the 64 groups, 15 show significant decreases in calories in the two-to-seven month period. A slightly larger number show significant changes (including some increases) in the first month after diagnosis. Figure 3 shows the magnitude of the impact for these groups. In the short run there are decreases in some unhealthy foods (soda in particular) and increases in some healthy foods (fresh produce, vegetables, yogurt). In the longer run the increases disappear but there are significant decreases in a larger number of product groups. These are virtually all "bad" foods - soda, candy, shortening and oil, etc. On inspection, it appears that the declines in calories stem mostly from decreases in unhealthy foods, and there are limited increases in healthy foods in the long run.

It is possible to apply some more structure to this analysis using the data from the doctor survey. I use information from the doctor survey to define a group of "All Good" foods which all doctors in the survey report as a good source of calories and a group of "All Bad" foods which all doctors report as a bad source of calories. Figure 4 shows the evolution of calories from the two groups over time. There are virtually no changes in good food calories at any time; bad food calories are stable in the first month, and then show a large decrease.

In Table 5, I show regression evidence on these changes for good foods (Panel A) and bad foods (Panel B). I look at calories, quantities and expenditures. The data is consistent with the evidence in the figure. Overall, the ratio of good to bad foods increases in the early months, although by the end of the period considered it is back to baseline.

This analysis uses only a subset of foods. I can also look in more detail across all food groups ranked by surveyed doctors. As specified in the data section, I define four groups: "All Good", "Majority Good", "Majority Bad" and "All Bad" based on the doctor rankings. For each food group I estimate changes in calories and calculate the changes as a share of the baseline by group. The results are shown in Figures 5. In both the short run (the first month) and the longer run (two to seven months) there is a gradient in doctor advice. This gradient is especially strong in the later months where there are reductions in all food groups, but much larger ones in the foods that doctors perceive as worse.

The overall picture is consistent with what we see in Table 5 and Figures 3 and 4. In the short-run, individuals change their behavior in ways very consistent with what would be recommended by a doctor. In

the longer run they sustain the reductions in unhealthy foods, but the increases in good foods do not persist.

In addition to differing in their health status, food groups differ in their importance in the pre-period diet. Some product groups and modules account for a much larger share of the diet than others. I explore whether there is heterogeneity across groups by the initial importance of the category. It is mechanical that that larger categories will see larger absolute changes, but it is not obvious they should have larger changes as a percent of the initial level.

Panel A of Table 6 uses the results by product group to describe how the percent change varies by initial level. There is significant variability. The largest percentage changes appear in the most commonly purchased "bad food" categories. These changes are larger in terms of percent than the changes in common good food categories or in less common bad food categories.

Panel B of Table 6 shows the same analysis but focusing on product module rather than product group; product modules are smaller than product groups. Within the "Candy" product group, for example, example modules are chocolate candy, non-chocolate candy, etc. Again, the pare similar. The largest percent changes are in the large "bad" food categories. In this case, the change in the largest bad food module (which is salad and cooking oil) is a full six percent reduction in calories for the average household.

The overall picture from this analysis points to the importance of changes in unhealthy foods - especially those which are heavily consumed prior to diagnosis - in driving behavior change. What these results do not suggest is any evidence that increases in good foods are masking larger decreases in unhealthy foods. The total calorie response is very moderate, at least for the average household. The next section turns to heterogeneity across households, rather than across foods groups.

4.3 Heterogeneity Across Households in Calorie Reductions

There is significant heterogeneity in the changes in calories after diagnosis in this sample. In this section I ask whether this heterogeneity is predictable. I explore two possible predictors: standard demographic variables and characteristics of the pre-diagnosis diet. These relationships have the potential, first, to provide guidance as to which individuals are likely to be most successfully targeted by diet advice. In addition, to the extent we can identify households who are engaging in more successful behavior change, this may suggest a theory for why behavior change is on average limited.

For the purposes of this analysis I slightly alter the empirical strategy. I define a new variable at the household level which is the percent change in calories from the pre-period to the 2-to-7 month post period. I regress this outcome on either demographics or on the pre-period diet characteristics. An alternative, which yields similar results, would be intact the characteristics with the timing measures in a household fixed effects framework. The approach used here provides a more straightforward way to present the results.

4.3.1 Demographics

I consider a set of standard demographics: education, income and age. In addition, I use individual zip code to match each individual to whether or not they live in a "food desert" as defined by the USDA. I regress the percent change in calories, as defined above, on each of these variables separately and then include them together. The results are shown in Table 7. Demographic characteristics are mostly unrelated to behavior change. Income, education, age and race are not predictive. There is a marginally significant relationship between behavior change and living in a zip code that is classified as a food desert, with those in food deserts showing smaller reductions. This effect is not especially large, however, and is significant only at the 10% level.

It is worth noting that although these demographics do not predict behavior *change*, they do predict diagnosis in this sample, as everywhere else (see Appendix Table C.1 for details). Individuals with less income and education, and those who are older, are more likely to be diagnosed, they are just not more likely to respond with positive changes. One possible explanation is that the selection into the sample in the first place differs. If those individuals with high education are generally healthier, then those who develop diabetes despite this may be worse in some unobservable way.

4.3.2 Baseline Diet

Individuals differ considerably in their pre-diagnosis diet. The goal in this section is to ask whether this characteristic is predictive of subsequent behavior change.

Perhaps the most straightforward aspect of the pre-period diet is quality: some individuals consume a larger share of their calories in healthy foods than others. Table 8 shows evidence on the relationship between diet quality and calorie reduction. This table reports a regression of percent change in calories on the share of calories from healthy and unhealthy foods. The omitted category are foods that all doctors surveyed say are high quality. There is perhaps some limited evidence that having a worse diet prior to diagnosis correlates with larger behavior change. In particular, having a larger share of food in any group other than the "all good" group correlates with larger subsequent changes. This is perhaps somewhat encouraging as it suggests those who are worse off *ex ante* have greater success later.

The second characteristic I explore is diet concentration. This measure captures the share of calories which are accounted for by the most commonly purchased pre-period groups. I control in this analysis for both total number of pre-period calories (flexibly, in 100 groups) and for the total number of product groups that household purchases positive amounts of. The thought experiment this captures is as follows. Consider two households A and B, who each purchase a total of 100,000 calories in the average month in the pre-period from 100 total product groups. Household B purchases 1000 calories from each product group. Household A purchases 10,000 calories from one group, and then 909 from each of the additional groups. In the language I

use here, household A has a more concentrated diet.

I define concentration based on both product groups (larger categories) and product modules (smaller categories) and using both the share of the diet in the most purchased group and the share of the diet in the top 5 most purchased groups. The initial analysis simply regresses the percent change in calories on a measure of concentration.

The initial result can be seen in Figure 6a, which shows coefficients on dummies for quintile of diet concentration, where concentration is defined as the share in the most purchased group, and the two lines represent either product group or product module level definitions. Diet concentration strongly relates to calorie reductions. Relative to the group with the least concentrated diet, the 20% of individuals with the most concentrated diets decrease their calories by an additional 15%. This group is predicted to have about an 8% reduction in calories; using the aggregation described in Section 4.1.2 this suggests a decline of between 175 and 350 calories per day for the affected individual. Although this still falls short of the doctor-recommended level it is much closer.

It is important to note that this result is not mechanical - that is, in the non-diabetic population we do not see this relationship between concentration and subsequent reductions. This is demonstrated in Figure 6b which replicates Figure 6a using individuals *without* a diagnosis.[12] There is virtually no relationship in this control population between the pre-period concentration and subsequent changes in calories.

Table 9 shows statistical evidence related to this point. I estimate the effect of concentration defined either based on product groups or product modules, and based on concentration in the top category or the top 5 categories. In all regressions we see a strong relationship between initial concentration and later reductions in calories.

The table demonstrates that concentration correlates with calorie reductions. Figure 7 provides some evidence for the pathway. This figure shows the impact of concentration quintile (defined based on product group and the single most purchased category) on subsequent reductions in the most purchased group and in all the other groups combined. The effect of a concentrated diet is seen only in the top group, not in the remainder.

Taken together, the results here suggest an important element of predictability in calorie reductions. Households whose diets are heavily concentrated in one or a few items are significantly more successful at subsequent reductions. Importantly, as evidenced in Figure 7, they seem to achieve these reductions largely through these most heavily purchased items.

Without further interpretation, these results suggest a method for targeting particular individuals with dietary advice. Perhaps counter-intuitively, households with quite bad seeming diets *ex ante* - those who

[12]Recall I define a "diagnosis" date for these individuals at a random time in their participation and structure the data identically based on this date.

consume a large share of their calories in items like candy or soda - tend to be the more successful reducers later, even conditional on total calories. To the extent that doctors and policy-makers may view these households as intractable and not attempt dietary advice, this analysis suggests that may be a poor strategy.

5 Interpretation

The section above outlines a number of facts in the data. Overall behavior change is limited in response to diagnosis; this is consistent with expectations from other literature. The changes by food type show that reductions in "bad" food groups are much more sustained than increases in good food groups, although the latter do appear in the first month after diagnosis. The sharp timing and initial patterns of change by food group - for example, the fact illustrated in Figure 5 that the changes by group line up closely with doctor advice, even when comparing across categories where doctors disagree - suggest that individuals are absorbing medical advice about dietary changes, at least qualitatively.

Perhaps most striking is the heterogeneity in food group response by pre-period diet importance and the heterogeneity across households by initial concentration. The caloric reductions appear disproportionately in the largest unhealthy food groups, and individuals whose diets are heavily concentrated in those groups actually have *more* dietary success than those with more balanced purchasing behaviors. Put differently, the households who achieve the biggest reductions in calories purchased are those with concentrated diets, and the reductions are achieved in large part through decreases in these initially heavily purchased food groups. These facts alone gives a sense of which individuals we would expect to be most successful at losing weight; in this section I reference the psychology literature to provide an interpretation of the facts in terms of a theory of behavior change. This may be a first step toward using these results to inform diet strategies.

There is a body of literature in psychology which invokes a "two-systems model" to explain limited behavior change in many dimensions (Fishbein and Ajzen, 1975). The general theory is that while the more rational "system 1" would like to undertake positive behavior change, it is difficult to constrain the impulsive "system 2" to comply. Overeating is among the central examples of sub-optimal behavior change with this model (Koritzky et al, 2014). Some authors (Hatch and Hatch, 2010) have argued that a key aspect of achieving behavior change in this framework is to have a specific, black-and-white, action path.

In the context of diet in particular, there is evidence from both the psychological and medical literature on the success of interventions which target specific changes. Booth-Butterfield and Reger (2004) report on a nutritional campaign which targeted a single behavior: switching from whole or 2% milk to 1% milk. In contrast to the limited success of many broad nutrition education campaigns (Snyder et al, 2000), this intervention produced large and sustained changes in purchasing behavior (a 17 percentage point increase in low-fat milk purchasing in the intervention communities). Similarly, there is evidence for successful weight loss

in interventions which target only reductions in soda (Tate et al, 2012).

A natural interpretation of the results here is as a reflection of this theory and evidence. Consider the following stylized account of the evidence described above. There are two individuals, A and B, with identical calorie consumption levels prior to diagnosis, both 10% above the recommended level. Individual A eats a well balanced and calorically appropriate diet, with the exception of a single category (say, soda) which accounts for the excess 10% of calories consumed. Individual B consumes 10% more calories than recommended on all food groups. Note that assuming there are a large number of foods groups consumed, individual A in this example will have a more concentrated diet prior to diagnosis by my definition.

The literature cited above suggests that it may be easier to achieve dietary success if you can do so through a focused and specific rule. Individual A in this example could reduce the entire excess 10% of their calories by adopting a strict rule of "No Soda". In contrast, focusing on any specific food group for individual B would lead to only a small decrease in calories. In other words, the focused diet option will not lead to a large change in calories.

This basic description lines up well with the facts presented above. Holding constant calories, households with concentrated diets are more successful at reducing overall. Moreover, the reductions they achieve are very heavily focused in the problematic groups. Individuals with less concentrated diets are less successful at reducing overall. It is notable that they are also less successful at reducing even on the largest food category, as shown in Figure 7. Those individuals with a concentrated diet are successful both because there is a lot of room to move on the largest category *and* because their reductions on that category are large. If those with less concentrated diets were able to achieve the same percentage reductions on their largest category this would increase their calorie reductions considerably, even if those reductions would remain smaller than for the concentrated group.

In terms of guidance for increasing behavior change, this may provide support for interventions which provide focused dietary advice. Helping individuals identify their problematic food groups and targeting those may be of particular value.

6 Discussion and Conclusion

The results in this paper show that households respond to a negative obesity-related health shock by changing their dietary choices. These changes are statistically significant, but they are quite small relative to what a doctor would recommend. The pattern of changes in the period immediately following diagnosis suggest good information about what foods are recommended, and which are not. In the longer run, households do not seem to persist with increases in healthy foods, although decreases in unhealthy foods persist. There is significant heterogeneity across the sample in dietary success, with dietary concentration playing an important

predictive role.

Section 5 suggests one interpretation for the patterns in the data - namely, that dietary strategies targeting a single food group are more effective than a general attempt to change overall diet. This may suggest more effective information strategies for encouraging weight loss in this and similar populations.

As noted in the introduction, a commonly discussed alternative to information campaigns designed to improve diet is a policy of taxes and subsidies (for a discussion of this in policy circles, see Leonhardt (2010)). Such a policy could come in a general form (e.g. a broad "soda tax") or in a more targeted way (e.g. subsidized fruits and vegetables for WIC or SNAP recipients). Given the general observation there that a treatment which provides very salient information and advice nevertheless achieves fairly small responses, it seems useful to ask how it compares to a policy of taxes or subsidies.

Evaluating the tax or subsidy equivalent of the diagnosis-produced changes in demand requires estimates of the price elasticity by food group. I use estimates from a review article (Andreyeva, Long and Brownell, 2010). These authors aggregate evidence from 160 studies on price elasticity to produce mean elasticity estimates for 16 groups, including soda, sugar and sweets, vegetables, eggs, etc. A full list and the elasticity estimates are reported in Appendix Table C.2. I match these groups to product modules in Nielsen, using the same product groups I estimate effects for in Section 4.2. Not all products can be matched to an elasticity estimate; for example, there is no elasticity estimate reported for nuts, reflecting the fact that no studies have estimated price elasticity for nuts. In these cases, I exclude the module. The second column of Appendix Table C.2 lists the product groups which are matched to each elasticity category.

Given these estimates, it is straightforward to generate a tax or subsidy equivalent. Price elasticity is known and, from the data here, I have an estimate of the percentage change in quantity. I use these together to calculate the percentage change in price which would produce the equivalent change, which is the tax (or subsidy) equivalent. I focus on product groups with significant changes in either the short run (first month) or long run (two-to-seven months) changes. The results are shown in Figure 8. The significant changes in the short run are somewhat mixed. The evidence suggests that a subsidy of about 10% would produce similarly sized changes in produce purchases. The long run results are perhaps more relevant. Since in the long run the only significant changes are decreases, the groups illustrated all have tax rather than subsidy equivalents. The changes observed would be equivalent to what would be produced by a tax in the range of 5 to 15% depending on the product.

The primary policy target for taxes on unhealthy foods is soda. The results here suggest a soda tax of about 12% would be produce an overall change similar to what is seen in response to diagnosis. Similarly, fruits and vegetables are the most common subsidy targets. Given the changes in these categories, virtually *any* subsidy would preform better at increasing purchases.

The conclusions here suggest that moderate taxes would be required to produce behavioral response

similar to what we observe from this "intervention." This is certainly in the range of what policy has discussed and implemented (Mytton, Clarke and Rayner, 2012). Whether this suggests taxes are better than intensive educational campaigns depends on how distortionary we think taxation is, as well as how close a broad education campaign could get to the treatment effects observed here. On the flip side, the evidence suggests that increasing consumption of healthy food may be better accomplished with a subsidy-type approach.

References

American Diabetes Association et al., "Economic costs of diabetes in the US in 2012," *Diabetes Care*, 2013, *36* (4), 1033–1046.

Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell, "The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food," *American journal of public health*, 2010, *100* (2), 216.

Booth-Butterfield, Steve and Bill Reger, "The message changes belief and the rest is theory: the 1 percent or less milk campaign and reasoned action," *Preventive Medicine*, 2004, *39* (3), 581–588.

Breyer, B. N., S. Phelan, P. E. Hogan, R. C. Rosen, A. E. Kitabchi, R. R. Wing, and J. S. Brown, "Intensive Lifestyle Intervention Reduces Urinary Incontinence in Overweight/Obese Men with Type 2 Diabetes: Results from the Look AHEAD Trial," *J. Urol.*, Feb 2014.

Caldwell, John, Pat Caldwell, John Anarfi, Kofi Awusabo-Asare, James Ntozi, I.O. Orubuloye, Jeff Marck, Wendy Cosford, Rachel Colombo, and Elaine Hollings, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

Centers for Disease Control and Prevention, "National Diabetes Statistics Report: Estimtes of Diabetes and Its Burden in the United States, 2014," Technical Report, US Department of Health and Human Services 2014.

Cummings, Linda and Gregory Cooper, "Colorectal Cancer Screening: Update for 2011," *Seminars in Oncology*, 2011, *38*, 483–489.

DeSantis, Carol, Rebecca Siegel, Priti Bandi, and Ahmedin Jemal, "Breast Cancer Statistics, 2011," *CA Cancer J Clin*, 2011, *61*, 409–418.

Diabetes Prevention Program Research Group et al., "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin," *The New England Journal of Medicine*, 2002, *346* (6), 393.

Dubois, Pierre, Rachel Griffith, and Aviv Nevo, "Do Prices and Attributes Explain International Differences in Food Purchases?," *American Economic Review*, March 2014, *104* (3), 832–67.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo, "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?," *QME*, 2010, *8* (2), 207–239.

Espeland, Mark, "Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the look AHEAD trial," *Diabetes care*, 2007.

Estruch, R., E. Ros, J. Salas-Salvado et al., "Primary prevention of cardiovascular disease with a Mediterranean diet," *N. Engl. J. Med.*, Apr 2013, *368* (14), 1279–1290.

Feldstein, A. C., G. A. Nichols, D. H. Smith, V. J. Stevens, K. Bachman, A. G. Rosales, and N. Perrin, "Weight change in diabetes and glycemic and blood pressure control," *Diabetes Care*, Oct 2008, *31* (10), 1960–1965.

Fishbein, Martin and Icek Ajzen, *Belief, attitude, intention and behavior: An introduction to theory and research* 1975.

Foster, G. D., K. E. Borradaile, M. H. Sanders et al., "A randomized study on the effect of weight loss on obstructive sleep apnea among obese patients with type 2 diabetes: the Sleep AHEAD study," *Arch. Intern. Med.*, Sep 2009, *169* (17), 1619–1626.

Franz, Marion J, John P Bantle, Christine A Beebe, John D Brunzell, Jean-Louis Chiasson, Abhimanyu Garg, Lea Ann Holzmeister, Byron Hoogwerf, Elizabeth Mayer-Davis, Arshag D Mooradian et al., "Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications," *Diabetes care*, 2002, *25* (1), 148–198.

Gregg, E. W., H. Chen, L. E. Wagenknecht et al., "Association of an intensive lifestyle intervention with remission of type 2 diabetes," *JAMA*, Dec 2012, *308* (23), 2489–2496.

Hatch, Chip and Dan Hatch, *Switch: How to Change Things When Change is Hard*, Crown Business, 2010.

Koritzky, Gilly, Camille Dieterle, Chantelle Rice, Katie Jordan, and Antoine Bechara, "Decision-making, sensitivity to reward and attrition in weight management," *Obesity*, 2014, *22* (8), 1904–1909.

Leonhardt, David, "The battle over taxing soda," *The New York Times*, 2010, *18*.

Lindström, Jaana, Pirjo Ilanne-Parikka, Markku Peltonen, Sirkka Aunola, Johan G Eriksson, Katri Hemiö, Helena Hämäläinen, Pirjo Härkönen, Sirkka Keinänen-Kiukaanniemi, Mauri Laakso et al., "Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study," *The Lancet*, 2006, *368* (9548), 1673–1679.

Mytton, Oliver T, Dushy Clarke, and Mike Rayner, "Taxing unhealthy food and drinks to improve health," *BMJ*, 2012, *344*, e2931.

Ogden, C. L., M. D. Carroll, M. A. McDowell, and K. M. Flegal, "Obesity among adults in the United States–no statistically significant chance since 2003-2004," *NCHS Data Brief*, Nov 2007, (1), 1–8.

Oster, Emily, "HIV and sexual behavior change: Why not Africa?," *Journal of Health Economics*, 2012, *31* (1), 35–49.

Phelan, S., A. M. Kanaya, L. L. Subak, P. E. Hogan, M. A. Espeland, R. R. Wing, K. L. Burgio, V. DiLillo, A. A. Gorin, D. S. West, and J. S. Brown, "Weight loss prevents urinary incontinence in women with type 2 diabetes: results from the Look AHEAD trial," *J. Urol.*, Mar 2012, *187* (3), 939–944.

Snyder, LB, MA Hamilton, EW Mitchell, J Kiwanuka-Tondo, F Fleming-Milici, and D Proctor, "The Effectiveness of Mediated Health Communication Campaigns: Meta-Analysis of Differences in Commencement, Prevention, and Cessation Behavior Campaigns," *Meta-analysis of media effects*, 2000.

Tate, Deborah F, Gabrielle Turner-McGrievy, Elizabeth Lyons, June Stevens, Karen Erickson, Kristen Polzien, Molly Diamond, Xiaoshan Wang, and Barry Popkin, "Replacing caloric beverages with water or diet beverages for weight loss in adults: main results of the Choose Healthy Options Consciously Everyday (CHOICE) randomized clinical trial," *The American journal of clinical nutrition*, 2012, *95* (3), 555–563.

Taylor, Kathryn S., Carl J. Heneghan, Andrew J. Farmer, Alice M. Fuller, Amanda I. Adler, Jeffrey K. Aronson, and Richard J. Stevens, "All-Cause and Cardiovascular Mortality in Middle-Aged People With Type 2 Diabetes Compared With People Without Diabetes in a Large U.K. Primary Care Database," *Diabetes Care*, 2013, *36* (8), 2366–2371.

Wang, Y. C., K. McPherson, T. Marsh, S. L. Gortmaker, and M. Brown, "Health and economic burden of the projected obesity trends in the USA and the UK," *Lancet*, Aug 2011, *378* (9793), 815–825.

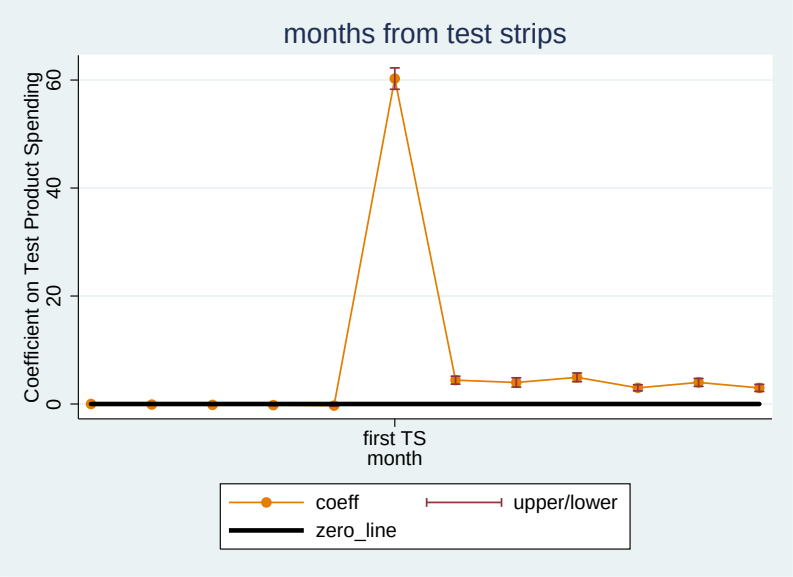
Wilding, JPH, "The importance of weight management in type 2 diabetes mellitus," *International Journal of Clinical Practice*, 2014, *68* (6), 682–691.

Williamson, D. F., T. J. Thompson, M. Thun, D. Flanders, E. Pamuk, and T. Byers, "Intentional weight loss and mortality among overweight individuals with diabetes," *Diabetes Care*, Oct 2000, *23* (10), 1499–1504.

Wing, R. R., R. C. Rosen, J. L. Fava, J. Bahnson, F. Brancati, I. N. Gendrano Iii, A. Kitabchi, S. H. Schneider, and T. A. Wadden, "Effects of weight loss intervention on erectile function in older men with type 2 diabetes in the Look AHEAD trial," *J Sex Med*, Jan 2010, *7* (1 Pt 1), 156–165.

Wing, RR, Paula Bolin, Frederick L Brancati et al., "Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes," *The New England Journal of Medicine*, 2013, *369* (2), 145.

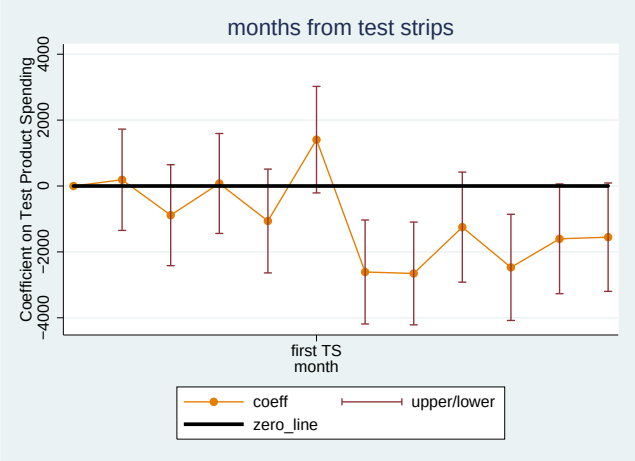
Figure 1: **Testing Supply Purchases**



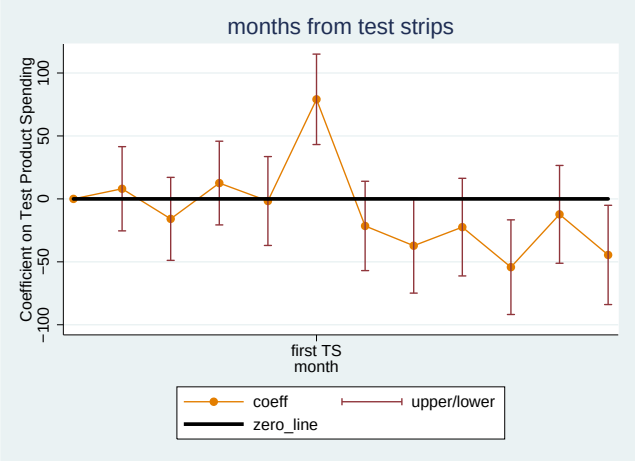
Notes: This figure shows data on purchasing any test strip products around the inferred diagnosis timing. Coefficients are from a regression which uses time-adjusted data and controls for household fixed effects.

Figure 2: **Behavior Change: Calories, Quantities and Spending**

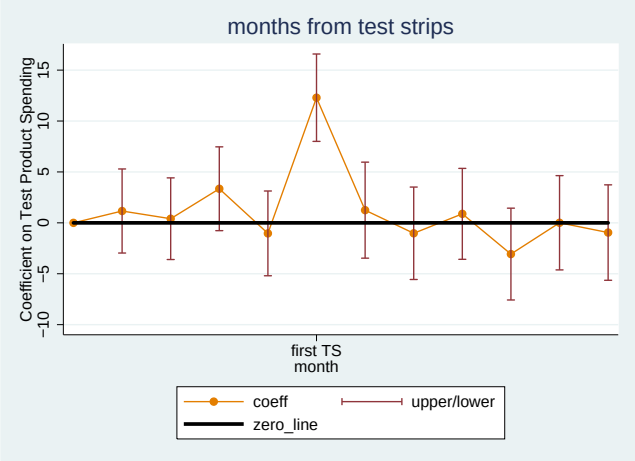
(a) Calories



(b) Quantity in Ounces



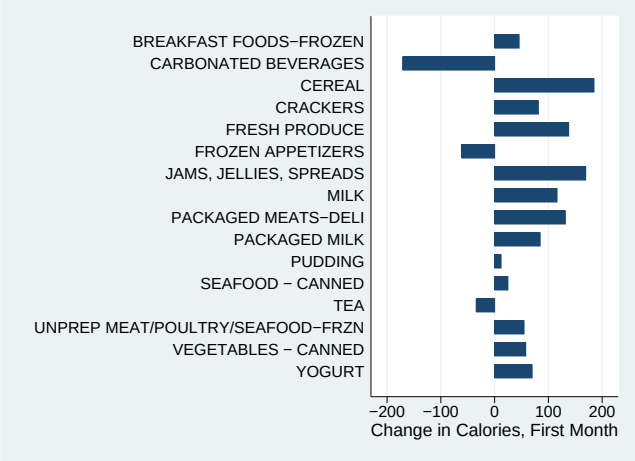
(c) Expenditures



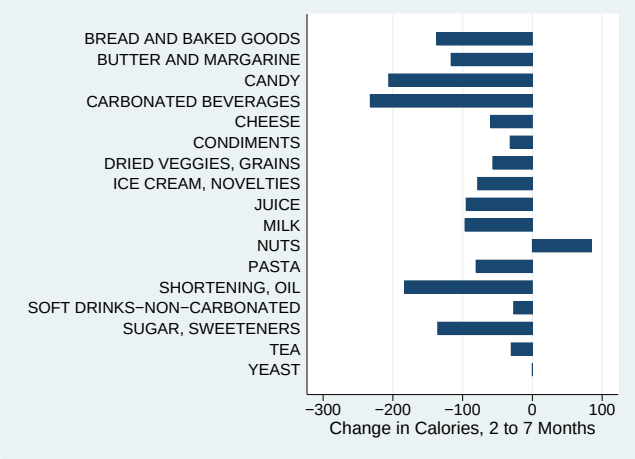
Notes: These figures show coefficients from regressions of the various outcome variables on months from inferred diagnosis. All outcomes are residualized with respect to month-year fixed effects and a linear control for time in sample and all regressions include household fixed effects. Error bars show 90% confidence intervals.

Figure 3: **Significant Group Changes**

(a) First Month After



(b) Two to Seven Month After

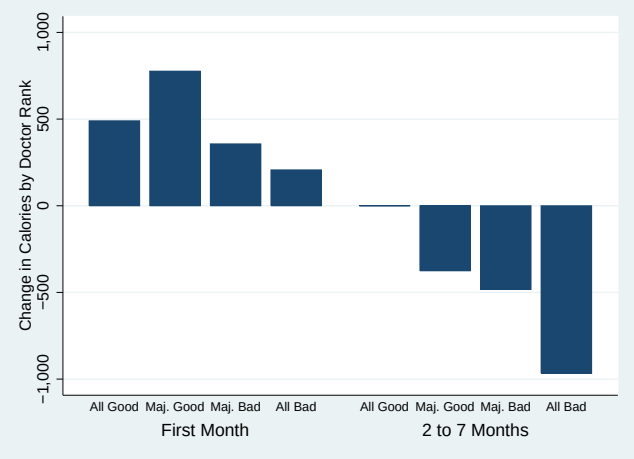


Notes: This graph shows changes in calories for the product groups which show significant changes in the first month (Sub-Figure a) and in the two-to-seven months later period (Sub-Figure b). The changes come from regressions of changes in calories by food group on timing dummies.

Figure 4: **Changes in "Good" and "Bad" Foods**

Notes: This figure shows coefficients from regressions of good and bad food calories and quantities on time from inferred diagnosis. Outcome measures are residualized with respect to month-year fixed effects and a linear control for time in Nielsen sample. Good foods are defined as those which all doctors surveyed say are a good source of calories; bad foods are defined as those which all doctors surveyed say are a

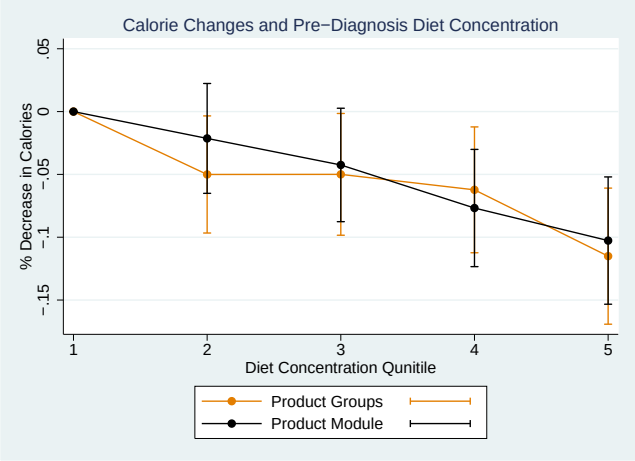
Figure 5: **Behavior Change by Doctor Rankings**



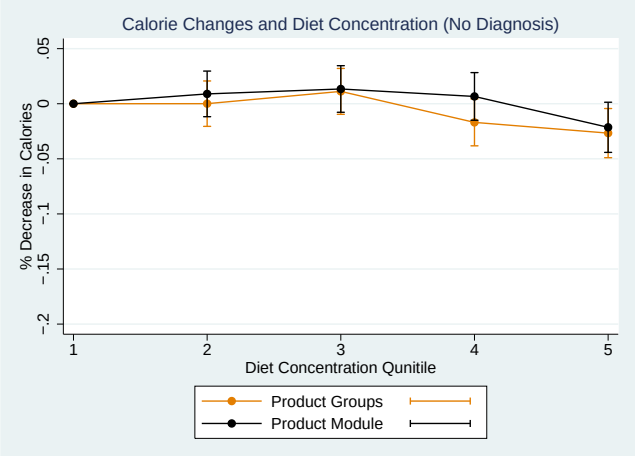
Notes: These graphs show changes in calories and quantities on foods with varying doctor rankings. Changes are reported as a share of the mean. Measures are all residualized with respect to month-year fixed effects and a linear control for time in Nielsen sample. "All Good" are foods which all doctors in the sample reported as good sources of calories; "Maj. Good" are those items which more doctors report as a good source of calories than a bad source. The corresponding "Bad" labels are defined in the same way. The data is constructed by regressing each item on diagnosis timing measures separately and then summing the coefficients and mean expenditures by group.

Figure 6: **Dietary Concentration and Behavior Change**

(a) Diabetics

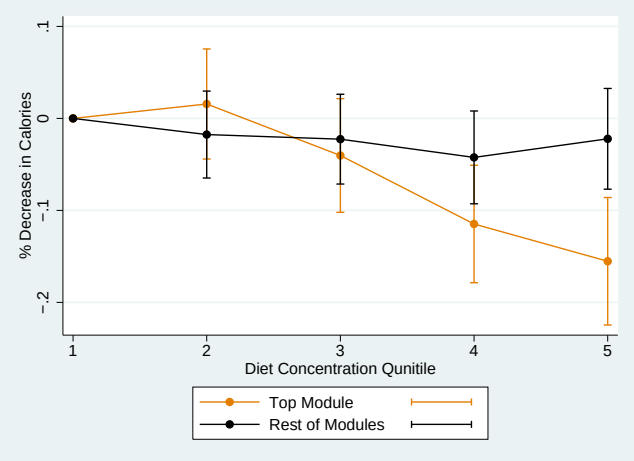


(b) Non-Diabetics



Notes: This shows the relationship between pre-period diet concentration and subsequent behavior chagne. Figure (a) is for the diabetics, the pouplation of interest. Figure (b) is for non-diabetics and is used as a falsification test.

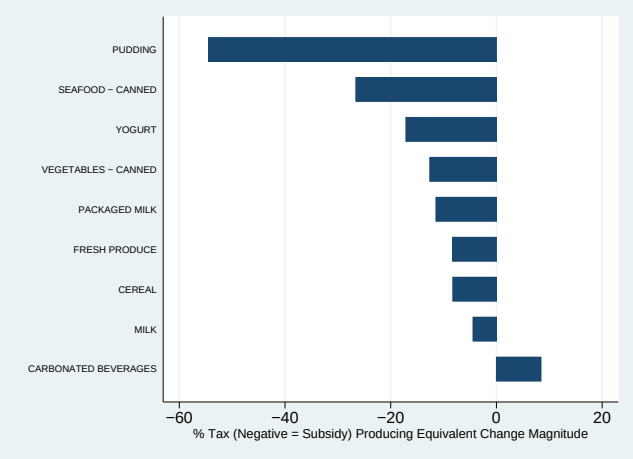
Figure 7: **Dietary Concentration and Behavior Change by Module**



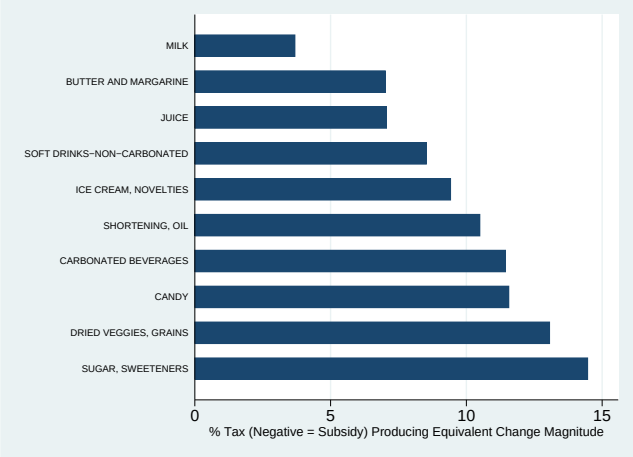
Notes: This shows the relationship between pre-period diet concentration and subsequent behavior change on the top ranked module and the rest.

Figure 8: **Tax Equivalents to Behavior Changes**

(a) First Month After



(b) Two to Seven Month After



Notes: This graph shows tax rates which would produce the same magnitude change as produced by the diagnosis event. The change I consider is the change between the pre-period and the late post-period (2 to 7 months after diagnosis). Elasticity estimates come from Andreyeva, Long and Brownell (2010).

Table 1: **Summary Statistics**

Panel A: Panelist Demographics			
	Mean	*Standard Deviation*	*Sample Size*
HH Head Age	61.8	11.8	3990
HH Head Years of Education	13.9	2.32	3601
HH Income	$65,814	$52,552	3938
White (0/1)	0.85	0.36	4007
In Food Desert (0/1)	0.36	0.48	3982
Panel B: Panelist Shopping Behavior			
Avg. Number of Trips/Month	11.2	7.2	43,026
Shopping Behavior (Per Household/Month):			
Quantity in Ounces	2078.8	1341.4	43,026
Expenditures	$262.08	$185.54	43,026
Calories (Gladson Data)	87,555	54,248	43,026
Share Carbohydrates (Gladson Data)	0.53	0.11	42,894
Share Protein (Gladson Data)	0.11	0.03	42,894
Share Saturated Fat (Gladson Data)	0.13	0.06	42,894

Notes: This table reports summary statistics on demographics (Panel A) and panelist shopping behavior (Panel B). Household age, income and education are computing at the median of reported categories. Quantity and expenditure data come from Nielsen data directly. Quantities are in ounces and items which are not reported in ounces are converted to ounces. Calories and nutrients are generated by merging the Nielsen panel with Gladson data. The details of this merge are in Section 3.1.2.

Table 2: **Test Supply Purchases By Inferred Diagnosis Time**

Outcome:	*Testing Supply Spending*
First Month After	60.43***
	(1.21)
Two-Four Months After	4.60***
	(0.35)
Five-Seven Months After	3.48***
	(0.27)
Household Fixed Effects	YES
R-squared	0.38
Number of Obs.	43,142

Notes: This table reports evidence from regression of testing supply purchase on timing from diagnosis. Diagnosis is defined as the first month in which any testing supplies are purchased. Purchases measure is residualized with respect to for month-year fixed effects and a linear trend for time in sample. The omitted category is 1 to 5 months before diagnosis.

Table 3: **Behavior Change After Inferred Diabetes Diagnosis**

Outcome:	*Calories*	*Quantity in Oz.*	*Spending ($)* *All*	*Spending ($)* *Magnet HH*
First Month After	1747.09**	78.5***	11.5***	13.2***
	[0.020]	[0.038]	[0.044]	[0.042]
	(780.5)	(17.1)	(2.10)	(3.37)
Two-Four Months After	-1845.64***	-27.7**	-0.39	0.10
	[-0.021]	[-0.013]	[-0.002]	[0.000]
	(532.1)	(13.1)	(1.56)	(2.53)
Five-Seven Months After	-1555.19***	-38.0***	-2.12	-2.71
	[-0.018]	[-0.018]	[-0.008]	[-0.009]
	(561.9)	(14.1)	(1.63)	(2.63)
Household Fixed Effects	YES	YES	YES	YES
R-squared	0.51	0.59	0.66	0.65
Number of Obs.	43,026	43,026	43,026	22,804

Notes: This table shows the primary results on aggregate changes. The omitted category is 1 to 5 months before diagnosis. All coefficients are reported in levels. Figures in square brackets represent the change as a share of baseline average. Standard errors are in parentheses. All outcome measures are residualized with respect to month-year fixed effects and a trend in time since enrollment. Magnet households are those who also scan and report prices for non-UPC coded goods. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table 4: **Robustness Checks: Calories**

	Time Robustness			**Household Type Robustness**		**Alternate Timing**		**Home Food**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No Time Controls	*Exclude Diag. Month*	*Add HH Pre-Trends*	*Single Person Households*	*Exclude Low Spenders*	*Divide Pre-Period*	*Add Months After*	
1-3 Months Before						-242.5		
						(607.8)		
First Month After	1476.2*	1597.2*	1613.0*	1196.5 [0.023]	-111.11 [-0.001]	1648.0**	1748.5**	426.0 [0.014]
	(788.2)	(799.3)	(844.9)	(851.8)	(970.2)	(822.9)	(768.5)	(332.6)
2-4 Months After	-3104.7***	-2030.0***	-2148.5***	-1299.7**[-0.025]	-3867.1***[-0.037]	-1944.9***	-188.5***	-730.9***[-0.025]
	(541.4)	(564.9)	(764.4)	(598.3)	(658.3)	(600.2)	(523.6)	(253.9)
5-7 Months After	-3637/6***	-1777.1***	-1718.3*	-606.8 [-0.012]	-3215.6***[-0.031]	-1654.6***	-1732.2***	-751.0***[-0.025]
	(578.8)	(593.2)	(980.9)	(661.0)	(692.6)	(619.4)	(550.3)	(280.1)
8-12 Months After							-2369***	
							(515.1)	
Household FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.53	0.52	0.32	0.52	0.43	0.51	0.49	0.49
Number of Obs.	43,026	39,302	40,437	17,109	32,050	43,026	61,620	31,797

Notes: This table replicates the results in Columns (1) Table 3, under varying robustness checks. Household pre-trends (Column (3)) are estimated from pre-diagnosis data. Low spenders (Column (5)) are those in the bottom 25% of the spending distribution. Omitted category in Column (6) is four to five months before. Column (8) limits to food categories where at least 85% of the NHANES consumption of the category is purchased at a store. Square brackets (Columns (4), (5) and (8)) show percentage changes from baseline; these are provided because the baseline levels differ. The omitted category is 1 to 5 months before diagnosis except in Column (6). All coefficients are reported in levels. Calorie outcome measure is residualized with respect to month-year fixed effects and a trend in time since enrollment. Standard errors are in parentheses. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table 5: **Effects for "Good" and "Bad" Foods**

Panel A: Good Foods			
Outcome:	*Calories*	*Quantity in Oz.*	*Spending ($)*
First Month After	490.1***	20.7***	2.90***
	(130.3)	(4.39)	(0.40)
Two-Four Months After	26.1	-3.09	0.80***
	(110.4)	(3.63)	(0.31)
Five-Seven Months After	-29.3	-4.21	0.11
Household FE	YES	YES	YES
R-squared	0.51	0.66	0.60
Number of Obs.	43,026	43,026	43,026
Panel B: Bad Foods			
Outcome:	*Calories*	*Quantity in Oz.*	*Spending ($)*
First Month After	207.0	19.4***	3.65***
	(471.4)	(6.64)	(0.64)
Two-Four Months After	-1094.9***	-9.98*	-0.35
	(310.0)	(5.12)	(0.45)
Five-Seven Months After	-824.7**	-7.57	-0.25
	(317.7)	(6.36)	(0.49)
Household FE	YES	YES	YES
R-squared	0.44	0.60	0.52
Number of Obs.	43,026	43,026	43,026

Notes: This table reports the impact of diagnosis timing on purchases of good and bad foods. The omitted category is 1 to 5 months before diagnosis. Good foods are defined as those which all doctors surveyed say are a good source of calories; bad foods are defined as those which all doctors surveyed say are a bad source of calories. Outcomes are residualized with respect to month-year fixed effects and a linear control for time in Nielsen sample. Standard errors are in parentheses. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table 6: **Changes and Initial Diet Importance**

Panel A: Group Level Analysis		
Group:	**Bad Foods**	**Non-Bad Foods**
Most Purchased Group	-3.9%	-2.1%
Top 5 Groups	-4.9%	-1.9%
Rank > 5 Groups	-0.05%	-1.3%
Panel B: Module Level Analysis		
Group:	**Bad Foods**	**Non-Bad Foods**
Most Purchased Group	-6.0%	-2.0%
Top 5 Groups	-5.2%	-1.3%
Rank > 5 Groups	-1.5%	-1.6%

Notes: This table reports the percent change in purchases by the popularity in the pre-period diet. "Bad" food groups are groups where all doctors surveyed say the foods are a bad source of calories. "Non-Bad" foods are all others.

Table 7: **Demographic Heterogeneity**

Outcome	*Percent Change in Calories*					
Education	-0.001					-0.0004
	(.007)					(0.008)
Income		-0.0002				-0.0003
		(.001)				(.0015)
Age			0.002			0.003
			(.006)			(.006)
White (0/1)				-0.026		-0.029
				(.022)		(.022)
Food Desert (0/1)					0.026*	0.028*
					(0.016)	(0.015)
R-squared	0.00	0.00	0.00	0.00	0.00	0.00
Number of Observations	3297	3297	3297	3297	3281	3281

Notes: This table reports interactions between behavior change and demographics. The dependent variable is the percent change in total calories from baseline. Education and age are based on the household head. Income is for the entire household. Food desert status is defined based on the USDA data on food access. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table 8: **Baseline Diet Quality and Behavior Change**

Outcome:	*Percent Change in Calories*
Share of Calories in :	
Majority Good Foods	-0.339**
	(0.140)
Majority Bad Foods	-0.100
	(0.133)
All Bad Foods	-0.326***
	(0.121)
Controls for # of Calories	YES
R-squared	0.11
Number of Observations	3297

Notes: This table reports the relationship between baseline diet quality, as measured by the share of calories in each quality group, and subsequent behavior change. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table 9: **Baseline Diet Concentration and Calorie Changes**

Concentration Level	*Product Group Level*		*Product Module Level*	
Concentration Group Count	*Top Group*	*Top 5 Groups*	*Top Group*	*Top 5 Groups*
Concentration Quintile = 2	-0.050***	-0.066***	-0.021	-0.064***
	(0.023)	(0.024)	(0.022)	(0.023)
Concentration Quintile = 3	-0.050***	-0.094***	-0.042*	-0.098***
	(0.025)	(0.027)	(0.023)	(0.024)
Concentration Quintile = 4	-0.062***	-0.091***	-0.076***	-0.105***
	(0.025)	(0.029)	(0.024)	(0.025)
Concentration Quintile = 5	-0.115***	-0.143***	-0.103***	-0.150***
	(0.027)	(0.033)	(0.026)	(0.028)
Controls for # of Categories	YES	YES	YES	YES
Controls for # of Calories	YES	YES	YES	YES
R-squared	0.19	0.19	0.28	0.29
Number of Observations	3297	3297	3297	3297

Notes: This table shows the relationship between pre-period diet characteristics and behavior change. The dependent variable is the percent change in total calories from baseline. The first two columns define concentration based on the product group level, the second two define it based on the product module level. Standard errors are in parentheses. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Appendix A: Health and Weight Calculations

This appendix describes evidence on the link between weight loss and health for diabetics. To make the comparisons concrete, I estimate the impact of losing approximately 10 pounds per year; this amounts to a reduction of approximately 100 calories per day.

Mortality The mortality data come from Williamson et al (2000). This is a twelve year studying following overweight individuals with diabetes. They estimate a linear impact of weight loss on mortality rate from 0 to 30 pounds. The estimate is a -0.33 reduction in death rate over this range. Due to linearity, this translates to a -0.11 reduction in death rate by reducing a further 10 pounds from baseline. To estimate this in life years I use someone age 50 as an example and use life table data from the CDC to estimate the impact of this reduction in death rates in each year on total survival. The range of values provided in the paper represent either the assumption that the benefit only accrues for 12 years (the length of the study) or the assumption that it accrues for the rest of life.

Remission Remission data come from Gregg et al (2012). The authors report the impact of tercile of weight loss in the first year on diabetes remission. I use evidence from Espeland (2007) on weight loss in the first year to calculate the midpoints within each tercile. I then estimate the impact of a 1% weight loss on the 1 year remission chance (it is 0.8 percentage points). I translate this to the impact of increasing weight loss by 10 pounds from baseline pounds using an estimate of initial weight.

Sleep.Apnea Sleep apnea data come from Foster et al (2009). The authors report a reduction in sleep apnea events of 0.6 events per hour per kilogram lost; the range of weight loss contains the range from 6 to 20 pounds pound range. I multiply by 10 pounds (4.53 kg).

Erectile Function Erectile function data comes from Wing et al (2010). The authors report a -0.148 change in erectile function measure for each 1% weight loss. I use information on the baseline weight of the sample to estimate the impact of increasing weight loss by 10 pounds (4.53kg).

Urinary Incontinence (Men) Estimates come from Breyer et al (2014). The authors report 1kg of weight loss reduces the odds of having weekly incontinence by 3%; the base rate is 10.2% . This translates to a reduction of 0.30 percentage points per kilogram, which I scale up to 10 pounds (4.53kg).

Urinary Incontinence (Women) Estimates come from Phelan et al (2012). The authors report 1kg of weight loss reduces the odds of developing weekly incontinence by 3%; the base rate is 12.2% . This translates to a reduction of 0.36 percentage point per kilogram, which I scale up to 10 pounds (4.53kg).

Appendix B: Doctor Survey Results

The table below lists each food group which was covered in the doctor survey, the number of doctors who voted it a "Good Source of Calories" and those who voted a "Bad Source of Calories". Although all 17 doctors were asked about each group not all rows add to 17 because doctors could also indicate the product was neutral.

Product	*Number Report "Good Source"*	*Number Report "Bad Source"*	*Product*	*Number Report "Good Source"*	*Number Report "Bad Source"*
Frozen Pizza	0	17	Lite Dressing	4	3
Cookies	0	17	Cold Cereal	7	4
Chocolate chips	0	17	Olives	7	4
Cookie mix	0	17	Canned Vegetables	8	3
Soda	0	17	Ground Beef	9	6
Flavored Syrup	0	17	Canned Beans	9	5
Frozen Biscuits	0	17	Soup	9	5
Ice Cream	0	17	Frozen Fruit	9	5
Cake mix	0	17	Natural Cheese	10	5
Slice-n-Bake Cookies	0	17	Breakfast Bars	10	4
Sugar	0	17	Salsa	10	3
Mayonnaise	0	16	Olive Oil	11	0
Spam	0	14	Peanut Butter	12	3
Butter	0	14	Dried Fruit	12	3
Creamer	0	12	Tuna	12	1
Potato Chips	1	16	Cottage cheese	13	2
Jam	1	16	Eggs	13	1
Salad Dressing	1	15	Frozen Vegetables	14	1
Pasta Dinner	1	15	Yogurt	14	0
Snack Crackers	1	14	Shrimp	15	1
Bread	1	12	Hot Cereal	15	0
Margerine	1	12	Fresh Fruit	15	0
Juice	2	13	Chicken	16	0
Flour	2	8	Fish	17	0
Regular Milk	3	11	Low Fat Milk	17	0
Potatoes	3	9	Vegetables	17	0
Applesauce	3	8	Nuts	17	0
Pretzels	4	9	Dried Beans	17	0
Pasta	4	9			
Rice	4	9			
Pickles	4	3			

Appendix C: Figures and Tables

Table C.1: **Demographic Predictors of Diagnosis**

	In Diabetic Sample
Highest Household Education	-0.005**
	(0.002)
Not White	0.043***
	(0.007)
Household Income Category	-0.0019***
	(0.0004)
Household Age Category	0.008***
	(0.001)
R-squared	0.004
Number of Observatoins	24,069

Notes: This table reports the demographic predictors of appearing in the diabetic sample. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Table C.2: **Elasticity Matches**

Food Group	*Matched Elasticity Category*	*Price Elasticity Estimate*
alcohol	alcohol	-0.60
cereal	cereal	-0.60
cheese	cheese	-0.44
dessert	sugar/sweets	-0.34
diet dessert	sugar/sweets	-0.34
diet shortening	fats	-0.48
diet soda	soft drinks	-0.79
eggs	eggs	-0.27
fish	fish	-0.50
fruit	fruits	-0.70
fruit juice	juice	-0.76
meat	beef/poultry/pork	-0.72
milk	milk	-0.59
potatoes	vegetables	-0.58
prep. food	food away from home	-0.81
rice/beans	vegetables	-0.58
shortening/oil	fats	-0.48
soda	soft drinks	-0.79
soft drinks	soft drinks	-0.79
sugar	sugar/sweets	-0.34
sweet baked goods	sugar/sweets	-0.34
vegetables	vegetables	-0.58
yogurt	dairy	-0.65

Notes: This table reports the food groups and the elasticity groups they are matched to, along with the price elasticity. Elasticity groups and estimates come from Andreyava et al, 2010.