Health Recommendations and Selection in Health Behaviors*

Emily Oster, Brown University and NBER

August, 2019

Abstract

Consider a case in which a new research finding links a health behavior with good health outcomes. A possible consequence is take-up of this behavior among individuals who engage in other positive health behaviors. If this occurs, later analyses of observational data may be biased by the change in selection. This paper evaluates these dynamic biases in empirical settings. Using data from vitamin supplementation and diet I show that selection responds endogenously to health recommendations. Further, the relationship between these behaviors and health outcomes changes over time. I find adjustment for selection on observables is insufficient to address the bias.

1 Introduction

The starting point for this paper is two facts about health behavior.

First, adherence to health recommendations varies systematically across people. Positive health behaviors tend to cluster – people who exercise are also less likely to smoke, for example – and they also correlate with education and income (e.g. Berrigan et al, 2003; Friel, Newell and Kelleher, 2005; Finke and Huston, 2003; Kirkpatrick et al, 2012; Cutler and Lleras-Muney, 2010; Cutler, Lleras-Muney and Vogl, 2008; Goldman and Smith, 2002). These adherence differences may generate bias in estimated links between behaviors and health outcomes in observational data (e.g. Greenland et al, 1999; Vandenbroucke et al, 2007).

*I am grateful for helpful comments from Isaiah Andrews, Amy Finkelstein, Matthew Gentkzow, Florian Gunsilius, Emir Kamenica, Ilyana Kuziemko, Matthew Notowidigdo, Jesse Shapiro, Andrei Shleifer, Heidi Williams and participants seminars at the US Census, Yale University, the University of Michigan and the NBER. I am grateful to Valeria Zurla, Marco Petterson, Claire Hug, Julian DeGeorgia, Sofia LaPorta, Cathy Yue Bai, James Okun and Geoffrey Kocks for outstanding research assistance. This paper benefits from the Nurses' Health Study data, funded with NIH grant UM1 CA186107.

Second, health recommendations – about the best diet, the optimal amount of exercise, vitamin supplements, and other things – often change over time in response to new information or changes in expert opinion.

This paper connects these two facts and argues that in the presence of differential *response* to new health advice, bias in estimates of the impact of health behaviors on health outcomes may be dynamic and endogenously respond to changes in recommendations.

To be concrete, consider a hypothetical case in which researchers are evaluating the relationship between pineapple and cardiovascular health. Imagine that although the true effect is zero, sampling variability leads to a study showing that pineapples significantly reduce heart attacks. One result may be positive pineapple-related news coverage or even a change in official guidelines about pineapple consumption. In response to this, some people will increase their consumption of pineapple. It is plausible that these would be the people who are most concerned about their health. But this group is also likely to be engaged in other heart-healthy behaviors like exercise, not smoking and eating other fruits and vegetables. As a result of this differential adoption of the recommendation, later studies of the pineapple-heart health relationship may see a more substantial link between pineapple and health, even though the initial effect was a statistical accident, since a bias has now been created by changes in selection.[1]

That these dynamics *could* arise follows directly from the two motivating facts. It is less obvious that these dynamics will be quantitatively important. It is possible that the behavioral response to changes in recommendations may be limited or that standard approaches to addressing bias in these settings – for example, regression adjustment for selection on observables – fully eliminate these dynamics.

The goal of this paper is to explore these dynamics empirically, to ask whether they occur in data and whether they are quantitatively important.

I present data from a number of settings. I first describe detailed evidence on a single case – vitamin E – where there have been sharp changes in recommendations over time and I observe detailed evidence on behavior, selection patterns and health outcomes. I then show corroborating evidence from other examples, including Vitamin D and several dietary patterns.

I begin in Section 2 by describing an empirical model of the research process which highlights the circumstances under which these dynamics will occur and may be important. Of note in this

[1]This discussion, and indeed this paper overall, presumes that the actual size of the causal effects is the same in each time period. This seems reasonable since these are intended as biological relationships, unlikely to change substantially within a population on a year to year time frame.

discussion is that a key assumption underlying my interpretation of the results is that the treatment effects considered are homogeneous. In the absence of this assumption, changes in the apparent effects of these treatments over time could reflect heterogeneous treatment effects, not changes in selection. The empirical results may still be of interest in this case, but they would not have the interpretation I give them.

Section 3 describes the data used in the paper. I make use of two survey datasets, the National Health and Nutrition Examination Survey ("NHANES") and the Nurses' Health Study ("NHS"). In addition, I use data from the Nielsen HomeScan scanner panel.

Section 4 then turns to developing the evidence in the case of vitamin E. This case is notable because of significant changes in the recommendation over time. First, in the early 1990s several studies suggested vitamin E supplements could prevent heart disease and cancer; this led to an uptick in health advice around vitamin E supplementation. However, in 2004 new evidence - this time from a meta-analysis of randomized trials - suggested vitamin E supplements not only didn't help but might actually *increase* mortality. Overall consumption of vitamin E responds strongly to these changes in information, increasing after the early 1990s and decreasing after 2004.

The selection patterns in vitamin E consumption also change over time. In the period when vitamin E is more recommended, there is a much stronger positive relationship between vitamin E consumption and education, income, exercise and diet quality; the relationship with smoking becomes more negative. To give one example, in the period before the two positive vitamin E studies, those who take vitamin E are 0.7 percentage points less likely to smoke. During the 1993 to 2004 period, those who take vitamin E are 4 percentage points less likely to smoke. After 2004, when the supplement became less recommended, this falls again to 1.6 percentage points.

These changes in selection are reflected in changes in the relationship between supplementation and health outcomes. In the NHANES, I estimate a relationship between vitamin E and heart health, and show some (large but statistically imprecise) evidence that heart health is more positively related to vitamin E consumption in the 1993 to 2004 period than in the period before or after. In the Nurse Health Study, I estimate the relationship between vitamin E consumption and short-term mortality. In the period before 1993 taking vitamin E is associated with a 10 percent reduction the risk of death over the next two years. After the positive health recommendation, this jumps to a significant 25 percent reduction in the mortality risk. Surveys conducted after the negative vitamin E news in 2004 shows that the effect of supplementation on mortality is again around 10 percent.

In both cases - heart health in the NHANES and mortality in the Nurse Health Study - I show the results in raw correlations (adjusting only for age and gender) and with comprehensive controls. These include all of the selection variables discussed, plus additional ones (race, marital status in the case of NHANES and a full set of disease controls for the Nurse Health Study). Including these controls does lower the variance across time, but it leaves the overall message virtually unchanged.

In Section 5 I turn to analysis of several other settings: Vitamin D, sugar consumption, fat consumption and the Mediterranean diet. In these cases the movements in recommendations are more gradual, and the data is less complete. Nevertheless, I am able to look at many of the same dynamics.

Vitamin D become more favored in the mid-2000s with release of evidence suggesting a wide range of benefits, but in the early 2010s some additional findings suggested this might have been overblown. Consumption of the supplement in the data increases and then decreases corresponding to this timing. Selection in the behavior also follows this pattern, with initial increases in the link between Vitamin D and other positive health behaviors, income and education and then later decreases.

Diet recommendations have changed over time: sugar and fat have become less recommended, and the Mediterranean diet more so. There are corresponding changes in the selection of behavior. Those who decrease their sugar consumption over time are more likely to exercise, less likely to smoke, have higher income and more education. Similar patterns show up for fat and the Mediterranean diet (opposite in the latter case as it becomes more recommended over time).

When diet behaviors are more recommended, they are more strongly associated with lower BMI and better heart health. These effects are large. In the case of sugar, in the earliest period of the data, higher sugar consumption has no effect on BMI; the point estimate is even negative. By the latest period of the data it is strongly associated with a higher BMI. This suggests that the changes in selection are large enough to change not just the magnitude but the *sign* of the relationship.

These changes do not seem to be significantly affected by the inclusion of standard observable controls for demographics and other health behaviors. Even if we consider regressions with comprehensive controls, large, visible, changes in the links between behavior and outcomes persist.

Overall, the evidence in these empirical sections suggests that these dynamics are qualitatively important for treatment effect estimates. Perhaps most importantly, they suggest that controlling for observed covariates – at least in these settings – is insufficient to eliminate these dynamics. This can be interpreted to say that the changes in selection on *unobservables* is sufficient to generate large

changes in treatment effects. This is important precisely because in many observational studies we are relying on adjustments for selection on observables to isolate causal effects.

The findings bode especially poorly for our ability to learn about null effects. A true effect of zero will be unstable, as statistical false positives and false negatives will be self-reinforcing. Once a significant result has been obtained in some setting, biases may be enduring until better data (for example, from a randomized trial) become available.

A natural question, which I take up briefly in the conclusion, is whether there are solutions to these issues. Randomized controlled trials are the gold standard for causal inference and, where feasible, would dramatically improve the evidence in these settings. But they are expensive, difficult to run and in some of these settings may be impossible to implement. Better research designs using non-randomized data, which have not received the attention in the public health literature that they have in economics (Angrist and Pischke, 2010), could significantly improve the conclusions here. This could include using food-specific taxes, or discontinuities in vitamin recommendations across age groups. Finally, when these are infeasible it may be possible to use changes in selection patterns directly to evaluate robustness (i.e. Heckman, 1978; Altonji et al, 2005; Oster, 2019).

A primary contribution of this paper is to extend work on the limits of observational data. Many authors have noted that observational evidence in health settings often appears biased and may be contradicted by randomized trials (Autier et al, 2014; Maki et al, 2014; Brownlee et al, 2010; Jones et al, 2019). The results here suggest that the observational findings themselves may contribute to the creation of bias.

The paper also contributes to a large literature in economics on the relationship between socioeconomic status and adherence to health recommendations (e.g. Berrigan et al, 2003; Friel, Newell and Kelleher, 2005; Finke and Huston, 2003; Kirkpatrick et al, 2012; Cutler and Lleras-Muney, 2010; Cutler, Lleras-Muney and Vogl, 2008; Goldman and Smith, 2002; Kowalski, 2018) and on consumer response to health information (e.g. Cutler, 2004; Chern et al, 1995; Brown and Schrader, 1990; Chang and Just, 2007; Roosen et al, 2009; Kinnucan et al, 1997; Ippolito and Mathios, 1995; Einav et al, 2019).

2 Empirical Framework

In this section I briefly formalize the statistical model for the dynamics described in the introductory example. Underlying this model is, of course, a model in which individuals with varying

characteristics react differently to new health information. I focus here on a statistical model of these dynamics. This approach is similar to that taken by Einav et al (2019) in their analysis of mammograms. In Appendix B, I outline one (by by no means the only) utility model which would deliver these implications.

Consider the empirical problem of evaluating the effect of some health behavior Λ_j on outcome Y. There is a vector of other behaviors, denoted $\mathring{\Lambda}$, indexed by $k = 1...n$.

I assume that Y is fully determined by behavior Λ_j and the vector of other behaviors $\mathring{\Lambda}$. Given the full vectors of behaviors I assume we could estimate the causal effect of Λ_j on Y through Equation (1).

$$Y_i = \eta + \beta\Lambda_{j,i} + \mathring{\vartheta}\mathring{\Lambda}_i + \epsilon_i \tag{1}$$

Researchers may not observe all of the elements of $\mathring{\Lambda}$. If the unobserved elements are correlated with Λ_j then the feasible estimates of β will be biased. Imagine researchers draw a sample of individuals and collect data on behavior Λ_j, outcome Y and a set of other variables Θ. This vector Θ may include some elements of $\mathring{\Lambda}$. It may also include some other variables (demographics, for example) which correlate with the behaviors in $\mathring{\Lambda}$. Following the data collection, the researchers estimate feasible equation (2) below.

$$Y_i = \alpha_t + \hat{\beta}\Lambda_{j,i} + \varsigma\Theta_i + \epsilon_i \tag{2}$$

It is well understood that if Θ does not fully absorb $\mathring{\Lambda}$, then $\beta \neq \hat{\beta}$.

The bias in the estimate of $\hat{\beta}$ will be controlled by the degree of correlation between Λ_j and the unobserved elements of $\mathring{\Lambda}$. In this paper, I consider a situation in which between periods t and $t+1$, the perceived health benefits of behavior Λ_j go up. If there is selection in the new adoption of behavior Λ_j, this may increase the covariance between Λ_j and the unobserved set. In turn, this increase in covariance will increase the bias in the estimated impact of Λ_j on outcome Y. The key argument in the paper is that observing the change in $\hat{\beta}$ over time around changes in recommendations is a way to understand how large the selection problem is.

The empirical work below focuses on analyzing what happens when there is a change in recommendation about some behavior Λ_j. I look for several signatures in the data. First, a change in the incidence of behavior Λ_j after the recommendation changes. Second, a change in the correlation between Λ_j and other health behavior or demographic correlates of health behaviors. Finally, a

change in the estimated relationship between behavior Λ_j and outcome Y, both with and without adjustments for selection on observables.

Stability Before moving on to the empirical results, it is worth noting that under this model of the research process a true treatment effect of zero will be fragile and difficult to sustain relative to a true positive or negative treatment effect.

To see why, consider the case where $\beta = 0$, and imagine that initially selection is minimal so in the population the estimated treatment effect would also be zero. Due to sampling variability, with repeated sampling, 5% of the time the data will yield a significant positive or negative effect. When this happens, and if people do not recognize the changes in selection, the endogenous behavioral reaction will reinforce that finding. In later periods the estimated treatment effect in the population will be biased.

In contrast, a true treatment effect that is positive (or, conversely, negative) will be self-reinforcing. We expect a move away from a positive effect (for example) *only if* the sampling variability in the data generates a significant negative effect despite the true significant positive effect. This will happen strictly less than 2.5% of the time. As a result, the null effects will be less stable.

Treatment Effect Heterogeneity In this discussion I am assuming there is a single true β. In the presence of treatment effect heterogeneity (i.e. as in Einav et al, 2019) and if the heterogeneity is related to the selection variables, then it is possible that changes in coefficients over time could reflect not changes in selection but this heterogeneity. Given the settings here, I argue that the movements in selection are a more likely interpretation, but I cannot rule this out.

3 Data

3.1 NHANES

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey which has been run, in some form, since the 1960s. In this project, I use data from the NHANES III (1988 through 1994) and from the continuous NHANES (beginning in 1999/2000 through 2014/2015).

Target Behavior Data Information on vitamin supplementation is obtained from the vitamin supplement modules. I focus on individual vitamin supplements - that is, if someone is taking a single-ingredient Vitamin D or E supplement. Information on diet is generated from the daily dietary recalls in the study. I generate a Mediterranean diet score as described in Trichopoulou et al (2003). For both sugar and fat I focus on the shares of carbohydrates and total fat here rather than the share of calories since the recommendations for total fat and carbohydrates move over time, and this share focus better isolates "sugar" and "saturated fat" as the key target components.

Selection Variables: Other Health Behaviors and Demographics The NHANES measures exercise and smoking behavior, which are used as measures of health behaviors. In addition, I use a simple measure of diet quality (vegetable consumption relative to the median) when analyzing vitamins. Finally, I extract data on education, income and other demographics from the demographic survey portion of the NHANES.

Health Outcomes All health outcomes in the NHANES are collected objectively. For vitamin E, I look at an index of heart health based on blood pressure and cholesterol. In the case of diet, I look at BMI and heart health.

3.2 Nurses' Health Study

The Nurses' Health Study (NHS) is a panel dataset of female nurses. The NHS recruited a cohort of approximately 120,000 female nurses in 1976 and conducted mail surveys of the cohort every two years. The study is ongoing and response rates are very high. The cohort is described in more detail in Colditz et al (1997). I use data from 1984 through 2010.

Target Behavior Data The NHS asks participants about vitamin E supplementation.

Selection Variables: Other Health Behaviors The primary selection behavior is smoking, which is collected consistently over time in these data; note that because the sample is all nurses, there is limited variation in education.

Health Outcome The health outcome is mortality, measured as the chance of death before the next survey (in two years).

3.3 HomeScan

The Nielsen HomeScan panel tracks consumer purchases using at-home scanner technology. Households that are part of the panel are asked to scan purchases, and the Nielsen data records the UPC of items purchased. I use Nielsen data for 2004 through 2016, available through the Kilts Center at the University of Chicago Booth School of Business. There is no information on health outcomes in these data.

Target Behavior Data The HomeScan data will be used to augment the analysis of Vitamin D. The target behavior is household purchases of Vitamin D supplements by year.

Selection Variables: Other Health Behaviors and Demographics I extract information on cigarette purchases and construct a diet quality score (as in Hut and Oster, 2019) as measures of health behaviors. I use household head education and income from the panelist surveys.

4 Results: Evidence from Vitamin E

I begin by analyzing the case of vitamin E.

Recommendations about vitamin E consumption have changed over time. In the early 1990s, a number of studies pointed to significant health benefits to vitamin E. Most notable was a pair of studies in NEJM in 1993 which cited large benefits of vitamin E in preventing heart disease in men and women (Rimm et al, 1993; Stampfer et al, 1993). There was significant media attention (i.e. Brody, 1993; Miller, 1993).

In 2004, however, new evidence suggested that excess vitamin E supplementation could actually *increase* mortality (Miller et al, 2005).[2] This evidence consisted of a meta-analysis of small randomized controlled trials, and was similarly widely covered by media organizations (Kolata, 2004).

These changes are reflected in trends in vitamin E consumption over time, which are illustrated in Figure 1. There is clear evidence that purchase and consumption of these products vary around the events identified above: vitamin E consumption increases after 1993, and then decreases, with a significant drop after 2004.

[2]This study was released in 2004 but the publication was in print in January, 2005.

4.1 Selection in Vitamin E Consumption

The first suggestion in the statistical model above is that the group who consumes vitamins *after* they are more recommended will be differently selected than those who consume before. I estimate the variation in the relationship between vitamin E consumption and various selection variables over time. These correlations are adjusted for age and gender controls but nothing else.

The results are shown in Figure 2; I divide the data in to three time periods around the changes in recommendations. Panel A uses the NHANES data, where I observe both socioeconomic status measures and other health behaviors. Panel B uses the Nurse Health Study data to look at smoking (yearly estimates for the NHS appear in the Appendix). In the case of the NHS, the coefficients are reported relative to the average level of smoking in these data so are interpreted as percent changes.

There is clear positive selection in vitamin E consumption increases in the middle period relative to the other two. These differences are statistically significant and the effects are large. The relationship between education and vitamin E consumption, for example, is twice as large after the positive recommendation as before. In the case of the NHS, the data is a panel, so I can directly interpret this to to say (for example) that the new adopters of vitamin E in the period after 1993 are less likely to be smokers than those who do not adopt.

The patterns in this table are consistent with the selection posited in the statistical model above. In the periods in which vitamin E is more recommended the consumers are less likely to smoke, more likely to exercise, eat a better quality diet and are richer and better educated.

4.2 Vitamin E and Health Outcomes

I turn now to changes in the relationship between vitamin E and health *outcomes.* There are three key questions. First, is there a change in the raw correlation between vitamin consumption and outcomes? Second, is the degree of change lessened or eliminated by inclusion of controls? And, third, is the change important in magnitude? The findings above suggest the first is very likely to be true. The latter two are less clear.

I begin with the relationship between vitamin E and a measure of heart health in the NHANES. Panel A of Figure 3 shows the estimated impacts of vitamin E supplementation on heart health in the same time periods used in Figure 2. The first set of bars look at the correlations with no controls other than age and gender. The second set of bars includes a full set of demographic and

other health behavior controls in the regressions.

Qualitatively, the patterns echo the patterns in selection, with vitamin E supplementation being more strongly associated with better heart health in the middle period (when it is more recommended) than in the surrounding periods. The differences are statistically noisy, although in the fully controlled regression we can reject equality between the latter two bars at the 10% level.

Despite the lack of statistical precision, two thing are worth noting. First, the inclusion of the full set of controls does not seem to bring the estimates closer together over time. If anything, the effects are more different in the controlled regression than in the uncontrolled case. Second, the size of the differences are large. For example, between the middle and the latter period in the fully controlled regression, the estimated effect falls from about 0.095 to 0.015, a decline of 0.080. This is equivalent to the effect of moving a full education category in the data and just a bit less than the difference between white and black participants.

Using the Nurse Health Study I can analyze the relationship between vitamin E and mortality, in this case with a larger sample and more precision. The results are shown in Panel B of Figure 3.[3] The outcome in this regression is a dummy for dying within two years of the survey (that is, before the next survey round) and the coefficient of interest is on vitamin E supplementation. I divide this coefficients by the average death rate over that period; magnitudes are interpreted as the percent decrease in mortality rate. I again look at these regressions with limited controls (only age) and with a full set of controls which would be commonly used in analysis of these data, including smoking behavior and measures of other health conditions.[4]

The evidence shows the effects on mortality moving significantly with the recommendations. In the early period, taking vitamin E is associated with approximately a 10 percent reduction in death risk. In the period from 1993 through 2003 this jumps to a 30 percent reduction in mortality risk. In the time period after the 2004 release of evidence undermining the value of vitamin E, the effect is again smaller at around 10 percent. The data reject equality of coefficients across periods with a high level of significance.

The inclusion of controls here – even the comprehensive set which capture a significant amount of the variation in mortality – makes relatively little difference to these conclusions. The estimated effect of vitamin E in the period from 1993 through 2004 is about a third of the size of the impact of smoking on mortality. Given the fairly simple nature of the health intervention "take vitamin E

[3]Results by year are shown in the Appendix.

[4]The full set of controls is listed in the table notes. This set is similar to what is used in analyses of these data (i.e. Stampfer et al, 1993).

supplements", if the effect were actually a third of the size of quitting smoking, this would be quite notable.

Together, these findings suggest that the changes in selection in this behavior *are* reflected in a changing relationship in the data between behavior and outcomes. These changes seem to be large in magnitude, and they largely persist with inclusion of a standard control set. Put simply, depending on which time period you draw data from, the conclusions about the impacts of vitamin E on health in observational data could be very different.

5 Results: Auxiliary Evidence from Vitamin D, Diet

Vitamin E is an effective case study for several reasons, including the availability of rich data and the fact that it becomes both more and less recommended over the period of the data . However, it is only a single example. I turn now to explore whether corroborating evidence appears in another vitamin (Vitamin D) and in three diet measures (sugar, fat and the Mediterranean diet).

5.1 Vitamin D

Background and Setup In 2007 there were a number of pieces of positive news about vitamin D supplementation. These included a widely cited New England Journal of Medicine (NEJM) article on the benefits of vitamin D (Holick, 2007), coverage in the NY Times and other outlets (e.g. Nagourney, 2007) and a corresponding spike in Google searches. On the other side, around 2012 there was a push back on much of this. An Institute of Medicine report (Rosen et al, 2012) suggested the purported benefits of vitamin D were overblown, and media coverage in the same period reinforced this (e.g. Bakalar, 2012a; Kolata, 2012; Bakalar, 2012b).

Using the NHANES and HomeScan data I will be able to explore trends in purchases over time and changes in selection of these behaviors. There is no linked outcome data to use here so I will focus only on the changes in selection.

Results

Panel A of Figure 4 shows the time trends in consumption (NHANES) or purchases (HomeScan) of Vitamin D around these events. There is evidence of increasing purchases of Vitamin D in the period where the news is more positive. The HomeScan evidence clearly points to a decrease in purchases after 2012.

Panels B and C of Figure 4 show changes in selection in the NHANES (Panel B) and the HomeScan (Panel B). In both datasets, across all the metrics, the evidence points to changes in selection consistent with the theory and the effects in vitamin E. When Vitamin D is more recommended it is more strongly associated with high education and income, a good diet, less smoking and more exercise. These associations are significantly different over time.

The difference across education groups in Vitamin D consumption doubles from the late 1990s to the late 2000s; by the late 2000s, increasing one education category increases the chance of consuming vitamin D supplements by about 25%.

5.2 Dietary Patterns

Background and Setup I analyze three dietary patterns: sugar intake, saturated fat intake and adoption of a Mediterranean diet.

Recommended levels of sugar and fat intake have fallen over time. In the case of sugar, the revision of the US Dietary guidelines in 2000 marks the first mention of avoiding sugar for health reasons (Krauss et al, 2000). Following this, in 2011/2012 there was a surge of popular media interest in the dangers of sugar (e.g. Taubes, 2011; Chang, 2012; 60 Minutes "Sugar").

Saturated fat was first restricted in the US dietary guidelines in 1990, with a suggested limit of 10% of calories (Peterkin, 1990). In 2005 this limit was lowered to 7% (Thompson and Veneman, 2005).

The Mediterranean diet has been, conversely, increasingly recommended over time, largely due to positive findings from major research studies. The first of these was in 2004, when two articles in the Journal of the American Medical Association (JAMA) showed positive health benefits from the diet (Knoops et al, 2004; Esposito et al, 2004). In the 2009 period additional studies argued for cognitive benefits (Feart et al, 2009). Largely after these data, a large randomized trial released in 2013 showed significant cardiovascular benefits (Estruch et al, 2013).

In analyzing diet I focus on the NHANES data. I am able to use these data to look at trends over time, changes in selection and changes in links with health outcomes, specifically BMI and heart health. A downside of the diet analysis relative to vitamins is that in all three cases the trends are secular. As a result, it is not possible to look for the reversals of selection that are a feature of the vitamin analysis.

Results Table 1 shows the full set of results on diet, including evidence on changes in selection and changes in the relationship between dietary choices and outcomes. For each dietary pattern I divide the data into the three periods. Panel A shows the mean levels of behavior (detailed yearly evidence on behavior is in Appendix Figure **??**). Panel B shows correlations with smoking, exercise, education and income, and Panel C shows the relationship with BMI and heart health, both with and without controls.

Panel A documents that these dietary patterns do respond to the changes in recommendations, with sugar and fat consumption declining over time and adherence to a Mediterranean diet increasing.

Panel B documents corresponding changes in selection. In the case of sugar, for example, consuming more sugar (measured as a share of carbohydrates) becomes, over time, more positively associated with smoking and more negatively associated with exercise, education and income. It is notable that in the earliest periods of the data those with higher income and education actually consume on average *more* sugar. In the latest periods of the data, they consume much less.

There are similar trends for fat and the Mediterranean diet. Over time, consuming more saturated fat becomes more negatively selected, while consuming a Mediterranean diet becomes more positively selected. The evidence on fat is noisier than the others, perhaps reflecting the generally smaller movements in this behavior, but the patterns are the same.

Panel C of Table 1 shows links with weight and heart health. These patterns echo the selection results. Over time sugar becomes *much* more positively associated with having a high BMI. Indeed, this link is roughly zero in the earliest periods of the data, and only becomes large and positive over time. There are similar patterns for fat. The Mediterranean diet is always associated with a lower weight and better heart health, but it becomes more so over time as the diet becomes more positively selected.

The second row in each section of Panel C shows these regressions including a large number of controls - income, education, exercise, smoking, marital status and race. These include all the selection variables analyzed in Panel B, plus additional controls. The inclusion of these controls does impact the coefficient dynamics, but only minimally. The very large changes in the relationship over time persist.

6 Discussion and Conclusion

In this paper I analyze the role of behavior change in driving biases in estimates of the impact of health behaviors on outcomes. I outline a simple data generating process in which changes in health recommendations differentially change health behaviors for different groups and show that these changes may influence estimated relationships between behavior and health over time. Using data on vitamin supplementation and diet I demonstrate that these dynamics occur in data. The degree of selection in behaviors varies over time, and the relationship between behavior and health also varies with these changes in selection. These dynamics are quantitatively important.

A primary contribution is to add to existing caution in interpreting observational results in settings like these. The problem of omitted variable bias is well known, but these results suggest such bias may be dynamic and, indeed, may *respond* to research findings. This suggests that awareness of the changes in recommendations over time should inform discussions about the plausible degree of bias in estimates. Further, the data clearly points to the conclusion that that current approaches in this literature – typically, using standard selection on observables adjustments – are very unlikely to yield causal results.

A natural question is how we can improve our understanding of causality given these challenges. One clear answer is more use of randomized controlled trials (RCTs). Indeed, in several of the settings here, our best evidence does come from RCTs.

However, expanding the scope of RCTs is not a panacea. These studies are expensive, it is often difficult to impact subjects' behavior around diet and it can be especially challenging to evaluate long-term outcomes like mortality. A growing literature (see e.g. Angrist et al, 2017) seeks to combine RCTs with observational estimates of the kind developed here. These approaches would be challenged, however, by the endogenously changing bias demonstrated above. This leaves us seeking solutions beyond expansion of randomized trials.

Angrist and Pischke (2010) argue that among the central advances in empirical economics in the past decades has been an improvement in non-experimental research design for causal infernence. These advances – more use of difference in difference, sharp and fuzzy RD, event studies – have been slower to spread into the type of public health problems addressed here. In light of the findings above it seems even more crucial to consider the scope for improved research designs in these settings.

A third option, perhaps feasible even without new research designs, would be to *use* the dy-

namic selection to adjust the estimates. One approach in this spirit would rely on a Heckman selection framework (e.g. Heckman, 1978) again using multiple time periods to pin down causal parameters. A second would rely on an assumption about selection on observed and unobserved factors, as in Altonji et al (2005) and Oster (2018). In either case, it would be feasible to combine these assumptions with multiple periods of data to ask what causal effect is consistent with the combination of observed treatment effects and changes in selection.

As a final note, although this paper focuses in particular on individual health behaviors, the dynamics here may be present in other settings (parental behaviors, for example) where individual choices vary over time.

References

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 2005, *113* (1), 151–184.

Angrist, Joshua D and Jörn-Steffen Pischke, "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of economic perspectives*, 2010, *24* (2), 3–30.

___ **, Peter D Hull, Parag A Pathak, and Christopher R Walters**, "Leveraging lotteries for school value-added: Testing and estimation," *The Quarterly Journal of Economics*, 2017, *132* (2), 871–919.

Autier, Philippe, Mathieu Boniol, Cécile Pizot, and Patrick Mullie, "Vitamin D status and ill health: a systematic review," *The lancet Diabetes & endocrinology*, 2014, *2* (1), 76–89.

Bakalar, Nicholas, "Risks: Too Much Vitamin D May Backfire," *New York Times*, Jan 17, 2012a.

___ , "Regimens: Study Questions Benefit of Extra Vitamin D," *New York Times*, October 23, 2012b.

Berrigan, David, Kevin Dodd, Richard P Troiano, Susan M Krebs-Smith, and Rachel Ballard Barbash, "Patterns of health behavior in US adults," *Preventive medicine*, 2003, *36* (5), 615–623.

Brody, Jane, "Vitamin E Greatly Reduces Risk Of Heart Disease, Studies Suggest," *New York Times*, May 1993.

Brown, Deborah J and Lee F Schrader, "Cholesterol information and shell egg consumption," *American Journal of Agricultural Economics*, 1990, *72* (3), 548–555.

Brownlee, Iain A, Carmel Moore, Mark Chatfield, David P Richardson, Peter Ashby, Sharron A Kuznesof, Susan A Jebb, and Chris J Seal, "Markers of cardiovascular risk are not changed by increased whole-grain intake: the WHOLEheart study, a randomised, controlled dietary intervention," *British Journal of Nutrition*, 2010, *104* (1), 125–134.

Chang, Hung-Hao and David R Just, "Health Information Availability and the Consumption of Eggs: Are Consumers Bayesians?," *Journal of Agricultural and Resource Economics*, 2007, pp. 77–92.

Chang, Kenneth, "What we Eat: Doubts by the Teaspoonful," *New York Times*, June 12, 2012.

Chern, Wen S, Edna T Loehman, and Steven T Yen, "Information, health risk beliefs, and the demand for fats and oils," *The Review of Economics and Statistics*, 1995, pp. 555–564.

Colditz, Graham A, Joann Manson, and Susan Hankinson, "The Nurses' Health Study: 20-year contribution to the understanding of health among women," *Journal of Women's Health*, 1997, *6* (1), 49–62.

Cutler, David M, "Behavioral health interventions: what works and why," *Critical perspectives on racial and ethnic differences in health in late life*, 2004, *643*, 674.

___ **, Adriana Lleras-Muney, and Tom Vogl**, "Socioeconomic status and health: dimensions and mechanisms," Technical Report, National Bureau of Economic Research 2008.

___ **and** ___ , "Understanding differences in health behaviors by education," *Journal of health economics*, 2010, *29* (1), 1–28.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams, "Screening and Selection: The Case of Mammograms," Technical Report, MIT Working Paper 2019.

Esposito, Katherine, Raffaele Marfella, Miryam Ciotola, Carmen Di Palo, Francesco Giugliano, Giovanni Giugliano, Massimo D'armiento, Francesco D'andrea, and Dario Giugliano, "Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial," *JAMA*, 2004, *292* (12), 1440–1446.

Estruch, Ramón, Emilio Ros, Jordi Salas-Salvadó, Maria-Isabel Covas, Dolores Corella, Fernando Arós, Enrique Gómez-Gracia, Valentina Ruiz-Gutiérrez, Miquel Fiol, José Lapetra et al., "Primary prevention of cardiovascular disease with a Mediterranean diet," *New England Journal of Medicine*, 2013, *368* (14), 1279–1290.

Feart, Catherine, Cécilia Samieri, Virginie Rondeau, Hélène Amieva, Florence Portet, Jean-François Dartigues, Nikolaos Scarmeas, and Pascale Barberger-Gateau, "Adherence to a Mediterranean diet, cognitive decline, and risk of dementia," *JAMA*, 2009, *302* (6), 638–648.

Finke, Michael S and Sandra J Huston, "Factors affecting the probability of choosing a risky diet," *Journal of Family and Economic Issues*, 2003, *24* (3), 291–303.

Friel, Sharon, John Newell, and Cecily Kelleher, "Who eats four or more servings of fruit and vegetables per day? Multivariate classification tree analysis of data from the 1998 Survey of Lifestyle, Attitudes and Nutrition in the Republic of Ireland," *Public health nutrition*, 2005, *8* (2), 159–169.

Goldman, Dana P and James P Smith, "Can patient self-management help explain the SES health gradient?," *Proceedings of the National Academy of Sciences*, 2002, *99* (16), 10929–10934.

Greenland, Sander, James M Robins, and Judea Pearl, "Confounding and collapsibility in causal inference," *Statistical science*, 1999, pp. 29–46.

Heckman, James, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–959.

Holick, Michael F, "Vitamin D deficiency," *New England Journal of Medicine*, 2007, *357* (3), 266–281.

Ippolito, Pauline M and Alan D Mathios, "Information and advertising: The case of fat consumption in the United States," *The American Economic Review*, 1995, *85* (2), 91–95.

Jones, Damon, David Molitor, and Julian Reif, "What do workplace wellness programs do? Evidence from the Illinois workplace wellness study," Technical Report, National Bureau of Economic Research 2018.

Kinnucan, Henry W, Hui Xiao, Chung-Jen Hsia, and John D Jackson, "Effects of health information and generic advertising on US meat demand," *American Journal of Agricultural Economics*, 1997, *79* (1), 13–23.

Kirkpatrick, Sharon I, Kevin W Dodd, Jill Reedy, and Susan M Krebs-Smith, "Income and race/ethnicity are associated with adherence to food-based dietary guidance among US adults and children," *Journal of the Academy of Nutrition and Dietetics*, 2012, *112* (5), 624–635.

Knoops, Kim TB, Lisette CPGM de Groot, Daan Kromhout, Anne-Elisabeth Perrin, Olga Moreiras-Varela, Alessandro Menotti, and Wija A Van Staveren, "Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project," *JAMA*, 2004, *292* (12), 1433–1439.

Kolata, Gina, "Large Doses of Vitamin E May Be Harmful," *New York Times*, Nov 2004.

___ , "Healthy Women Advised Not to Take Calcium and Vitamin D to Prevent Fractures," *New York Times*, Jun 13, 2012.

Kowalski, Amanda E, "Behavior within a Clinical Trial and Implications for Mammography Guidelines," Technical Report, National Bureau of Economic Research 2018.

Krauss, Ronald M, Robert H Eckel, Barbara Howard, Lawrence J Appel, Stephen R Daniels, Richard J Deckelbaum, John W Erdman Jr, Penny Kris-Etherton, Ira J Goldberg, Theodore A Kotchen et al., "AHA Dietary Guidelines: revision 2000: A statement for healthcare professionals from the Nutrition Committee of the American Heart Association," *Circulation*, 2000, *102* (18), 2284–2299.

Maki, Kevin C, Joanne L Slavin, Tia M Rains, and Penny M Kris-Etherton, "Limitations of observational evidence: implications for evidence-based dietary recommendations," *Advances in nutrition*, 2014, *5* (1), 7–15.

Miller, Edgar R, Roberto Pastor-Barriuso, Darshan Dalal, Rudolph A Riemersma, Lawrence J Appel, and Eliseo Guallar, "Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality," *Annals of internal medicine*, 2005, *142* (1), 37–46.

Miller, Leslie, "Lowly vitamin supplements may pack a big health punch," *USA Today*, Jun 1993.

Nagourney, Eric, "Shortage of Vitamin D May Weaken the Elderly," *New York Times*, May 1, 2007.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2018, pp. 1–18.

Peterkin, BB, "Dietary guidelines for Americans," *Journal of the American Dietetic Association*, 1990, *90* (12), 1725–1727.

Rimm, Eric B, Meir J Stampfer, Alberto Ascherio, Edward Giovannucci, Graham A Colditz, and Walter C Willett, "Vitamin E consumption and the risk of coronary heart disease in men," *New England Journal of Medicine*, 1993, *328* (20), 1450–1456.

Roosen, Jutta, Stéphan Marette, Sandrine Blanchemanche, and Philippe Verger, "Does health information matter for modifying consumption? A field experiment measuring the impact of risk information on fish consumption," *Review of Agricultural Economics*, 2009, *31* (1), 2–20.

Rosen, Clifford J, Steven A Abrams, John F Aloia, Patsy M Brannon, Steven K Clinton, Ramon A Durazo-Arvizu, J Christopher Gallagher, Richard L Gallo, Glenville Jones, Christopher S Kovacs et al., "IOM committee members respond to Endocrine Society vitamin D guideline," *The Journal of Clinical Endocrinology & Metabolism*, 2012, *97* (4), 1146–1152.

Stampfer, Meir J, Charles H Hennekens, JoAnn E Manson, Graham A Colditz, Bernard Rosner, and Walter C Willett, "Vitamin E consumption and the risk of coronary disease in women," *New England Journal of Medicine*, 1993, *328* (20), 1444–1449.

Sugar, *60 Minutes*, *2012.*

Taubes, Gary, "Is Sugar Toxic?," *New York Times*, April 17, 2011.

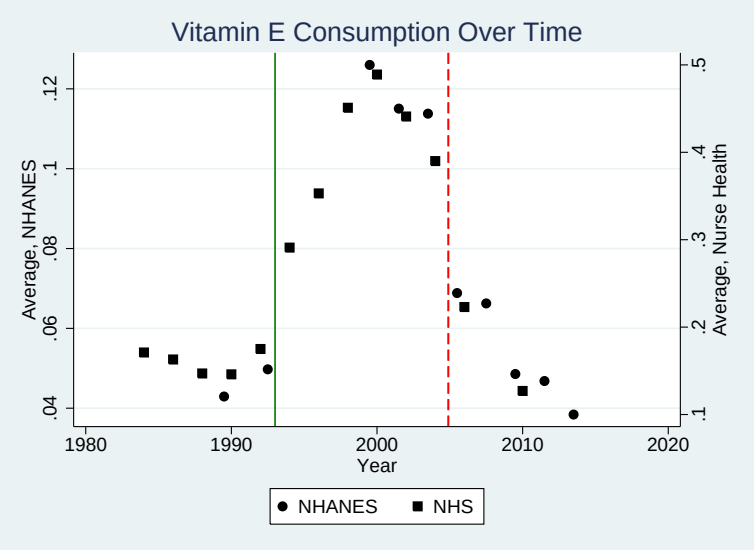
Thompson, Tommy G and Ann M Veneman, "Dietary guidelines for Americans 2005," *United States Department of Health and Human Services and United States Department of Agriculture*, 2005.

Thorisson, Hermann, "Coupling methods in probability theory," *Scandinavian journal of statistics*, 1995, pp. 159–182.

Trichopoulou, Antonia, Tina Costacou, Christina Bamia, and Dimitrios Trichopoulos, "Adherence to a Mediterranean diet and survival in a Greek population," *New England Journal of Medicine*, 2003, *348* (26), 2599–2608.

Vandenbroucke, Jan P, Erik Von Elm, Douglas G Altman, Peter C Gøtzsche, Cynthia D Mulrow, Stuart J Pocock, Charles Poole, James J Schlesselman, Matthias Egger, Strobe Initiative et al., "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration," *PLoS medicine*, 2007, *4* (10), e297.

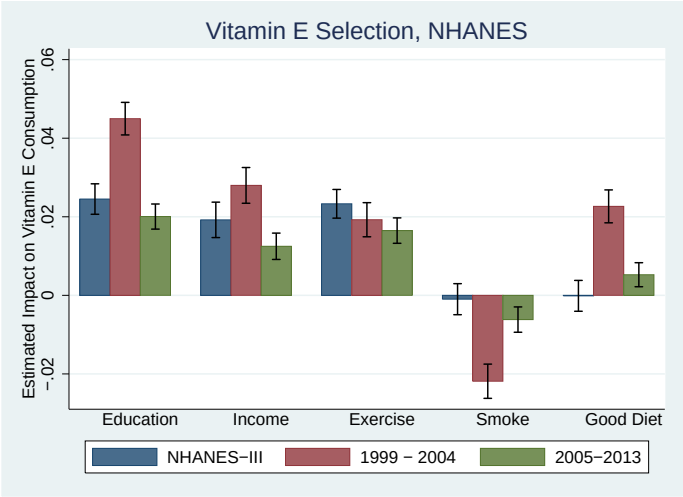
Figure 1: **Vitamin E Consumption Levels over Time**



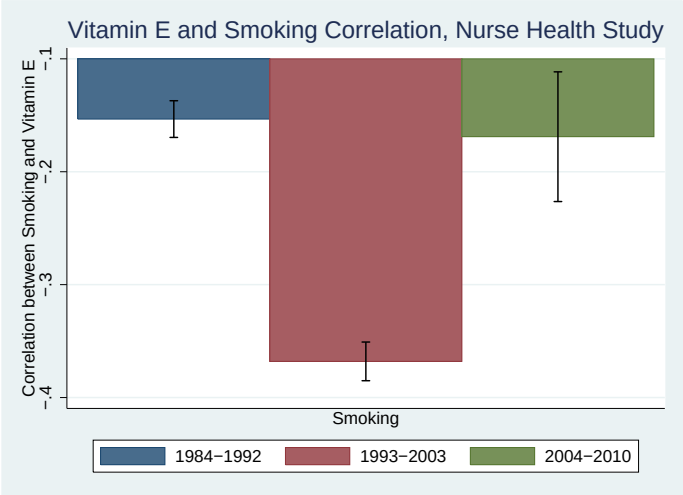
Notes: These graphs show evolution of reported vitamin E consumption in the NHANES and Nurses' Health Study. Events are marked with vertical lines; details of the events appear in Appendix Table **??**. The solid green line indicates the release of positive news about vitamin E; dotted red lines indicate the release of negative news.

Figure 2: **Vitamin Consumption, Socioeconomic Status and Health Behaviors**

Panel A



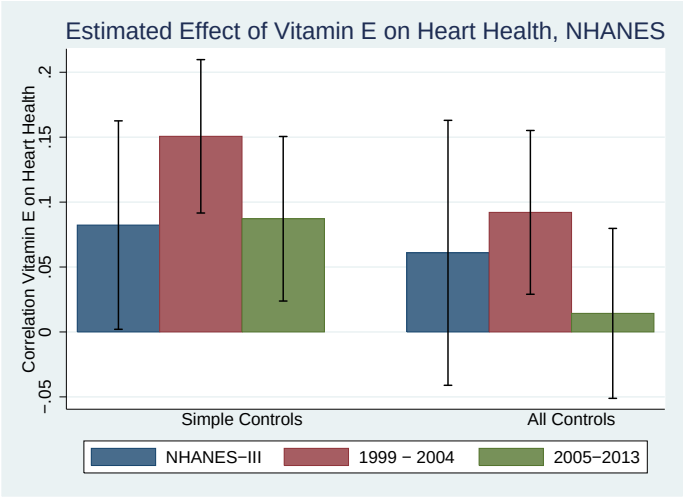
Panel B



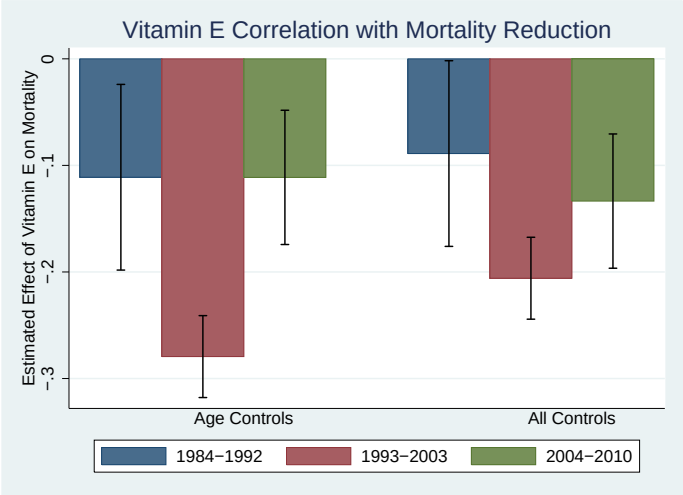
Notes: These figures show changes in the relationship between Vitamin E consumption and health behavior or socioeconomic status over time. Panel A uses data from the NHANES. In Panel A all effects can be interpreted as the impact of a one standard deviation change in the selection variable on the probability of reporting Vitamin E supplement consumption. Diet quality is based on vegetable consumption relative to the mean. The NHANES-III data covers the period between 1988 and 1994. Panel B uses data from the Nurse Health Study. Effects are reported relative to the mean; these can be interpreted as the percent decrease in smoking among those who take Vitamin E supplements. Events are marked with vertical lines; details of the events appear in Appendix Table **??**. Error bars in both cases show 95% confidence intervals.

Figure 3: **Vitamin E and Health Outcomes**

Panel A

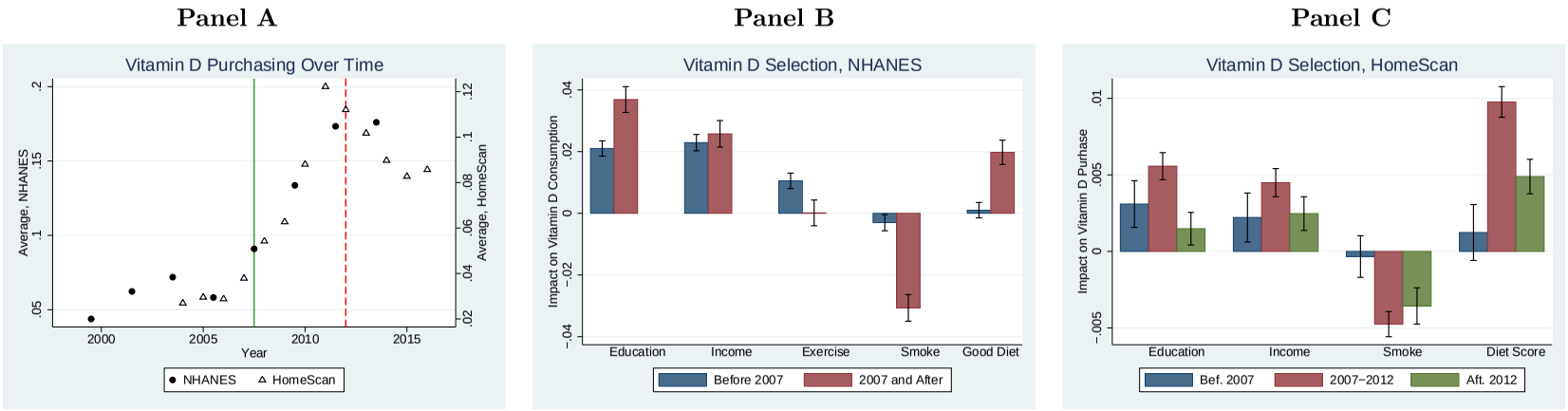


Panel B



Notes: This figures shows changes in the Vitamin E-health outcome relationship over time. Panel A shows changes in the relationship between Vitamin E consumption and heart health in the NHANES data over time. Heart health is the first principal component of a high blood pressure indicator and LDL and total cholesterol measures. Regressions with "Simple Controls" adjust for age and gender. "All Controls" include education, income, marital status, race, smoking behavior, exercise and the diet quality measure. Panel B shows the estimated impact of vitamin E consumption on two-year mortality in the Nurses' Health Study data over time. All results are based on regressions of a two-year-death indicator on a dummy for reported Vitamin E consumption. All regressions in Panel B control for age dummies. Controlled regressions include BMI and indicators for hypertension, high cholesterol, diabetes. cancer diagnosis post-menopausal hormone use, angina, myocardial infarction, and menopausal status. Error bars show 95% confidence intervals.

Figure 4: **Evidence on Vitamin D**



Notes: This figure shows evidence on Vitamin D. Panel A shows data on vitamin D consumption in NHANES or purchasing (in HomeScan data). Events are marked with vertical lines; details of the events appear in Appendix Table **??**. Panels B and C show the relationship between Vitamin E consumption and selection variables in both the NHANES and HomeScan data. In both cases the coefficients can be interpreted as the effect of a one standard deviation change in the selection variable on vitamin D consumption or purchase. In the NHANES data diet quality is based on vegetable consumption. In HomeScan the diet score is constructed as in Hut and Oster (2019). The smoking measure in these data is an indicator for purchasing cigarettes. Error bars show 95% confidence intervals.

Table 1: **Correlation between Dietary Choices, Health Behaviors/Socioeconomic Status and Outcomes**

	Panel A: Levels of Behavior								
	Sugar Share of Carbohydrates			**Saturated Fat (Share of Fat)**			**Med. Diet Score**		
	Before 2000	2000-2012	After 2012	Before 1990	1990-2005	After 2005	Before 2005	2005-2010	After 2010
Average	$0.456^{\ddagger,\ddagger\ddagger}$	$0.444^{\ddagger}$	0.417	$0.344^{\ddagger,\ddagger\ddagger}$	$0.328^{\ddagger}$	0.322	$3.67^{\ddagger\ddagger}$	$3.68^{\ddagger}$	3.78
	Panel B: Correlations between Dietary Choices & Health Behaviors/Socioeconomic Status								
	Sugar Share of Carbohydrates			**Saturated Fat (Share of Fat)**			**Med. Diet Score**		
	Before 2000	2000-2012	After 2012	Before 1990	1990-2005	After 2005	Before 2005	2005-2010	After 2010
Smoking	$0.005^{\ddagger,\ddagger\ddagger}$	$0.0110^{\ddagger}$	0.0165	$0.0026^{\ddagger\ddagger}$	0.0043	0.005048	-0.187	-0.173	-0.1820
Exercise	$0.0031^{\ddagger,\ddagger\ddagger}$	-0.0052	-0.0068	$-0.0021^{\ddagger\ddagger}$	$-0.0006^{\ddagger}$	-0.0042	$0.087^{\ddagger\ddagger}$	$0.106^{\ddagger}$	0.138
Education	$0.011^{\ddagger,\ddagger\ddagger}$	$-0.0008^{\ddagger}$	-0.005	-0.0012	$-0.0008^{\ddagger}$	-0.0023	$0.088^{\ddagger,\ddagger\ddagger}$	0.126	0.1527
Income	$0.0044^{\ddagger,\ddagger\ddagger}$	$-0.0064^{\ddagger}$	-0.0115	-0.0018	-0.0023	-0.0036	$0.120^{\ddagger\ddagger}$	$0.120^{\ddagger}$	0.177
	Panel C: Correlations between Dietary Choices and Health Outcomes								
	Sugar Share of Carbohydrates			**Saturated Fat (Share of Fat)**			**Med. Diet Score**		
	Before 2000	2000-2012	After 2012	Before 1990	1990-2005	After 2005	Before 2005	2005-2010	After 2010
BMI									
Raw	$-0.160^{\ddagger\ddagger}$	$-0.081^{\ddagger}$	2.62	$-1.399^{\ddagger\ddagger}$	$-0.594^{\ddagger}$	2.398	$-0.205^{\ddagger\ddagger}$	$-0.236^{\ddagger}$	-0.524
Adjusted	$0.835^{\ddagger\ddagger}$	$0.040^{\ddagger}$	2.94	$-0.761^{\ddagger\ddagger}$	$0.083^{\ddagger}$	2.76	$-0.256^{\ddagger\ddagger}$	$-0.263^{\ddagger}$	-0.530
Heart Health									
Raw	$-0.107^{\ddagger,\ddagger\ddagger}$	-0.575	-0.412	$-0.350^{\ddagger\ddagger}$	$-0.301^{\ddagger}$	-1.17	$0.024^{\ddagger,\ddagger\ddagger}$	0.056	0.045
Adjusted	$-0.265^{\ddagger}$	-0.531	-0.333	$-0.469^{\ddagger\ddagger}$	$-0.234^{\ddagger}$	-0.918	$0.012^{\ddagger,\ddagger\ddagger}$	0.041	0.031

Notes: This table shows changes in average levels of diet behaviors (Panel A), selection in behaviors (Panel B) and the relationship between behaviors and outcomes (Panel C). All data come from the NHANES for 1988 through 2015. The periods are divided based on the events detailed in Appendix Table **??**. To generate the results in Panel B, in the NHANES I use the full dataset to regress diet on the behavior or socioeconomic status variable, interacted with period. I include controls for the period, as well as for age, age square and gender (all interacted with the period to allow for flexibility). Panel C is constructed by running similar regression of the outcome variable (BMI or heart health) on diet metrics interacted with period. The raw correlations include only age and gender controls. The adjusted regressions also control for all of the selection variables, plus marital status and race. ‡ significantly different from next period at 5% level; ‡‡significantly different from two periods later at 5% level.