

Health Recommendations and Selection in Health Behaviors[†]

By EMILY OSTER*

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. These results highlight how spurious findings on health behaviors can be self-reinforcing. (JEL I12)

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, Robins, and Pearl 1999; Vandenbroucke et al. 2007).

Second, health recommendations—about the best diet, the optimal amount of exercise, vitamin supplements—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.

*Brown University and NBER (email: emily_oster@brown.edu). Amy Finkelstein was coeditor for this article. I am grateful for helpful comments from Isaiah Andrews, Matthew Gentzkow, Florian Gunsilius, Emir Kamenica, Ilyana Kuziemko, Matthew Notowidigdo, Jesse Shapiro, Andrei Shleifer, Heidi Williams, and seminar participants 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 UMI CA186107. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Conclusions calculated based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business.

[†]Go to <https://doi.org/10.1257/aeri.20190355> to visit the article page for additional materials and author disclosure statement(s).

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, some people will increase their consumption of pineapple. These may be the people who are most concerned about their health. But this group is also likely to be engaged in other heart-healthy behaviors (exercise, not smoking, etc.). As a result of this differential adoption of the recommendation, later observational studies of the pineapple–heart-health relationship may see a more substantial link between pineapple and health, since a bias has now been created by changes in selection.

If dynamics like this are present and quantitatively important, it bodes especially poorly for our ability to learn about null effects. False-positive or negative results are an inevitable part of statistical analysis, even with randomized trials. If such results are self-reinforcing, we cannot rely on later observational evidence to correct these statistical accidents.

That these dynamics *could* arise follows directly from the two motivating facts. It is less obvious that they do arise or will be quantitatively important. The goal of this paper, therefore, is to explore these dynamics empirically.

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

Section I describes an empirical model of the research process that highlights the circumstances under which these dynamics may be important. Of note in this 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 relationship between the behavior and health outcomes over time could reflect heterogeneous treatment effects, not changes in selection. I discuss this issue in more detail below.

Section II describes the data used in the paper. I make use of three survey datasets. First, the National Health and Nutrition Examination Survey (NHANES), which is a repeated cross-sectional survey of a nationally representative sample of individuals. It includes data on health behaviors, objectively measured health outcomes, and demographics. Second, the Nurses' Health Study (NHS), which is a panel survey of female nurses that began in the 1970s. Data are collected by mail survey every two years, and information is available on health behaviors and some health outcomes, notably mortality. Third, the Nielsen HomeScan scanner panel, which is a panel dataset of grocery, drugstore, and other purchases among a (nonrepresentative) set of households.

Section III analyzes vitamin E. There are significant changes in recommendations over time. 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. 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.

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, not smoking, and diet quality. For example, in the period before 1993, 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, 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 NHS, 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 in 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. After 2004, the coefficient on supplementation in the mortality regression is again around 10 percent.

In both cases—heart health in the NHANES and mortality in the NHS—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 the NHANES, and a full set of disease controls for the NHS). Including these controls does lower the variance across time, but it leaves the overall message virtually unchanged.

In Section IV I analyze 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 are less complete. Nevertheless, I am able to look at many of the same dynamics.

Vitamin D becomes more favored in the mid-2000s with the 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, and 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 body mass index (BMI) and better heart health. In the case of sugar, in the earliest period of the data, my sugar measure and BMI are uncorrelated. By the latest period of the data, it is strongly associated with a higher BMI.

Overall, the evidence in these empirical sections suggests that these dynamics are quantitatively important for estimated correlations. As noted above, this may be especially problematic for null results, given that false-positive and negative results are statistically inevitable, and this suggests that they may be self-reinforcing.

The paper 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, Loehman, and Yen 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).

I. Empirical Framework

In this section I formalize the statistical model for the dynamics described in the introductory example. I focus here on the statistical framework. In online Appendix B, I outline one (by no means the only) utility model that 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 and characteristics (i.e., demographics), denoted $\mathring{\Lambda}$, indexed by $k = 1, \dots, n$.

I assume that Y is fully determined by behavior Λ_j and $\mathring{\Lambda}$. If both were fully observed, I assume we could estimate the causal effect of Λ_j on Y through equation (1):

$$(1) \quad Y_i = \eta + \beta\Lambda_{j,i} + \vartheta\mathring{\Lambda}_i + \epsilon_i.$$

I assume a homogeneous treatment effect. Later, I discuss the possible role of treatment effect heterogeneity.

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 that researchers draw a sample of individuals and collect data on behavior Λ_j , outcome Y , and a set of other variables Θ , which are a subset of $\mathring{\Lambda}$. The researchers estimate feasible equation (2):

$$(2) \quad Y_i = \alpha_i + \hat{\beta}\Lambda_{j,i} + \varsigma\Theta_i + \epsilon_i.$$

It is well understood that if some elements of $\mathring{\Lambda}$ are omitted from Θ and correlated with Λ_j , 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 increase. 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 behaviors

or demographic correlates of health behaviors; finally, a change in the estimated relationship between behavior Λ_j and outcome Y .

Instability of Null Effects.—Under this model of the research process, a true treatment effect of zero will be fragile and difficult to sustain.

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 percent of the time the data will yield a significant positive or negative effect. When this happens, the endogenous behavioral reaction will reinforce that finding. In later periods the estimated treatment effect in the population will be biased by selection.

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 percent of the time. As a result, the null effects will be less stable than non-null.

Treatment Effect Heterogeneity.—The empirical framework focuses on the case with homogeneous treatment effects and in which selection drives changes in estimated coefficients over time. The empirical component of the paper will focus on estimating variation in selection and outcome–treatment relationships over time and interpreting them as connected. However, observing changes in outcome correlations over time could alternatively reflect treatment effect heterogeneity, if that heterogeneity is correlated with selection variables.

Imagine that the true treatment effect varies across groups and is larger for individuals who select into the behavior after the recommendation.¹ In this case, a larger estimated treatment effect in the “more recommended” period could reflect a larger causal effect for this group. Put differently, in this case we could observe movement in coefficients over time even if the regressions all yielded unbiased causal effects for the relevant treated population.

It is impossible to fully rule out this type of heterogeneity as an explanation for the patterns I observe. However, the data may be informative about the degree to which this is more or less plausible. More specifically, I can ask whether estimated heterogeneity on *observable* dimensions can explain some of the results. To do so, consider the underlying assumption that the treatment effects estimated in each period represent the causal effect for the treated population. Under this assumption, it is possible to estimate variation in treatment effects along *observable* dimensions in the baseline period. I can then reweight along observable dimensions to match the distribution of observables in later periods. If treatment effect heterogeneity is important in explaining these movements over time, I would expect the reweighting to explain some of the coefficient movements between periods. If this is not the case, it would be less consistent with this explanation for the results.

¹In other selection-in-health-behaviors settings, this may be fundamental. For example, Einav et al. (2019) shows mammography recommendations are disproportionately adopted by otherwise healthy individuals, and the resulting mammography results are more likely to be false positives or noninvasive tumors.

II. Data

A. NHANES

The NHANES is a nationally representative survey that has been run, in some form, since the 1960s. In this project, I use data from the NHANES III (1988 through 1994) and the continuous NHANES (beginning in 1999/2000 through 2014/2015). Summary statistics for all NHANES variables appear in panel A of online Appendix Table A2.

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 sugar I use the share of carbohydrates that are sugar; for fat, the share of total fat that is saturated fat. I do this rather than total calorie shares since the recommendations for total fat and carbohydrates are also moving over time, and this approach 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. In the case of exercise, I focus on measures of vigorous exercise, which is reported in episodes per month or week. Due to some variations in question structure, I standardize relative to the mean for each year. Smoking is a measure of current smoking behavior. I create a simple measure of diet quality: specifically, whether the individual consumes above or below the median number of vegetable calories.

Finally, I extract data on education (in five categories), income (in eight bins), age, and gender from the demographic survey portion of the NHANES.

Health Outcomes.—I create an index of heart health based on the first principal component of blood pressure, total cholesterol, and good cholesterol. BMI is defined (as is usual) as weight in kilograms divided by height in meters-squared. All variables are measured by the surveyor, not self-reported.

B. NHS

The 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; as of 2012 the average per-wave response rate was 86.2 percent (Bao et al. 2016). The cohort is described in more detail in Colditz, Manson, and Hankinson (1997). I use data from 1984 through 2010. Summary statistics for the variables used in the NHS appear in online Appendix Table A2, panel B.

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

Selection Variables: Other Health Behaviors.—The NHS asks nurses about their smoking behavior.

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

C. 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 universal product code 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. Summary statistics for HomeScan appear in online Appendix Table A2, panel C.

Target Behavior Data.—The target behavior is household purchases of vitamin D supplements by year. I use product descriptions to identify the set of products that are vitamin D supplements.

Selection Variables: Other Health Behaviors and Demographics.—Household smoking is identified based on household purchases of cigarettes. I construct a diet quality score based on Hut and Oster (2018); this approach uses information from a survey of doctors to code food quality and aggregate to a score.

Household head education and income (both in bins) are drawn from yearly panelist surveys.

III. Results: Evidence from 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 the *New England Journal of Medicine* in 1993 that 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.

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

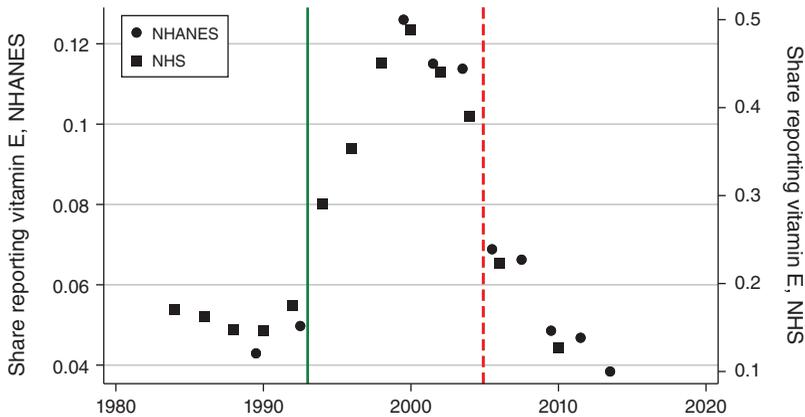


FIGURE 1. VITAMIN E CONSUMPTION LEVELS OVER TIME

Notes: This graph shows the evolution of reported vitamin E consumption in the NHANES and NHS. In both cases, the variable measures the share of survey respondents who report consuming any vitamin E supplements. In the NHS respondents are asked about vitamin E directly. For the NHANES I code this based on the reported list of vitamin supplements, including only individual vitamin E supplements, not multivitamins. Events are marked with vertical lines; details of the events appear in online Appendix Table A1. The solid green line indicates the release of positive news about vitamin E; dotted red lines indicate the release of negative news. Number of observations per year: NHANES (4,252 to 7,921); NHS (76,343 to 99,382).

A. Selection in Vitamin E Consumption

The first suggestion in the statistical model above is that the group that consumes vitamins *after* they are more strongly recommended will be differently selected than those who consume before. To explore this, I define three time periods around the changes in recommendations: before 1994, between 1994 and 2004, and after 2004.³ For each selection variable X_i I estimate equation (3) in each time period:

$$(3) \quad \text{Vit}E_i = \alpha + \beta^T X_i + \Pi A_i + \varepsilon_i.$$

The dependent variable, $\text{Vit}E_i$, is an indicator for whether the respondent reports taking a vitamin E supplement. Each time period produces a coefficient β^T ; these are the results of interest. The term A_i is a vector including only age, age-squared, and gender (in the NHANES; the NHS has only women). I estimate this using a standard linear probability model with homoskedastic standard errors.

The results are shown in Figure 2. Panel A uses the NHANES data, where I observe both socioeconomic status measures and other health behaviors. Panel B uses the NHS data to look at smoking (yearly estimates appear in the online Appendix). For the NHS, the coefficients are reported relative to the average level of smoking in these data, so they are interpreted as percent changes.

There is clear positive selection in vitamin E as consumption increases in the middle period relative to the other two. These differences are statistically

³Due to the timing of the NHANES-III, which covers 1988 through 1994, the first time period in the NHANES analysis will include some observations from 1994. It is not possible to pull out the 1994 observations.

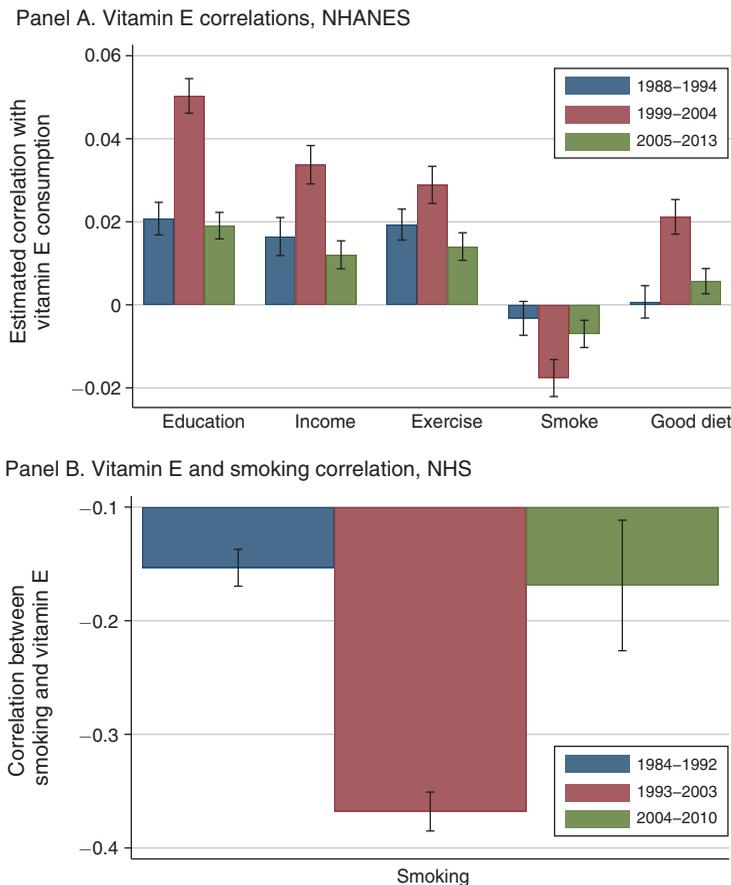


FIGURE 2. VITAMIN CONSUMPTION, SOCIOECONOMIC STATUS, AND HEALTH BEHAVIORS

Notes: This figure shows changes in the relationship between vitamin E consumption and health behavior or socioeconomic status over time; it is the result from estimating equation (3) in the text. The regression outcome is an indicator for vitamin E consumption, and I create these figures by extracting coefficients on each selection variable in each time period. Panel A uses data from the NHANES. In panel A all effects can be interpreted as the estimated relationship between a one standard deviation change in the selection variable on the probability of reporting vitamin E supplement consumption. A good diet is defined as having above-median calories from vegetables. Panel B uses data from the NHS. In this case I divide the coefficients from the regression by the mean level of smoking, which addresses the overall reduction in smoking rates over time. The bars can be interpreted as the percent decrease in smoking among those who take vitamin E supplements. Error bars show 95 percent confidence intervals. Number of observations by time period: NHANES ($N_1 = 15,748$; $N_2 = 13,421$; $N_3 = 24,524$); NHS, in person-years ($N_1 = 446,700$; $N_2 = 513,518$; $N_3 = 140,614$).

significant, and the coefficients 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 are a panel, so I can directly interpret this 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 figure 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, are more likely to exercise and eat a better quality diet, and are richer and better educated.

B. Vitamin E and Health Outcomes

I now look at the relationship between vitamin E consumption and health outcomes. I estimate equation (4) for each time period:

$$(4) \quad Y_i = \gamma + \Psi^T(\text{VitE}_i) + \Pi\mathbf{A}_i + \varepsilon_i.$$

As above, the coefficients of interest are the Ψ^T —the by-period correlations. I estimate this regression first including only age, age-squared, and gender (where appropriate) in the \mathbf{A}_i vector. I then estimate the regression including some additional demographic controls—the selection variables used above and some additional health variables for the NHS. Full lists of controls are included in the figure notes. The outcome Y_i is the index of heart health in the NHANES and two-year mortality in the NHS. In the latter case, the coefficients are scaled by the mortality rate and so are interpretable as percentage changes.

Figure 3, panel A shows the NHANES results. Qualitatively, both with and without controls, 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 more precise controlled regression we can reject equality between the latter two bars at the 10 percent level.

Despite the lack of statistical precision, it is worth noting that the size of the differences is large. For example, between the middle and latter period in the controlled regression, the estimated coefficient falls by 0.080. Because this is an index, this has no direct interpretation, but I can compare it to other variables. This decline is equivalent to moving a full education category in the data and just a bit less than the difference between white and black participants.

Figure 3, panel B shows the NHS results.⁴ The evidence shows the size of the correlation between supplementation and 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 25 percent reduction in mortality risk (the mean is approximately 2 percentage points over this period). In the time period after the 2004 release of evidence undermining the value of vitamin E, the relationship is again smaller at around 10 percent. The data reject equality of coefficients across periods with a high level of significance. The dynamics are similar with and without the controls included.

The variation across time periods is large. The estimated coefficient on 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 supplements,” if the mortality effect were actually a third of the size of quitting smoking, this would be quite notable. Put simply, depending on which time period you draw observational data from, the conclusions about the impacts of vitamin E on health could be very different.

⁴Results by year are shown in the online Appendix.

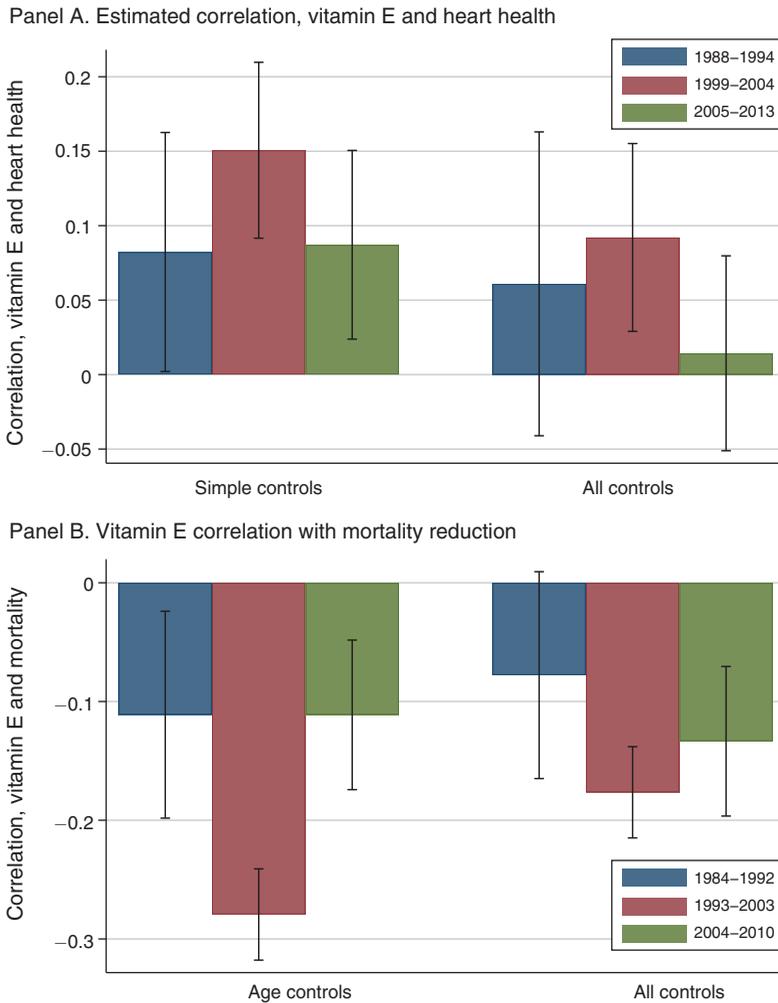


FIGURE 3. VITAMIN E AND HEALTH OUTCOMES

Notes: This figure shows changes in the vitamin E–health outcome relationship over time; it results from estimation of equation (4) in the text. In panel A (NHANES) the outcome is an index of heart health, and the key independent variable is reporting consumption of a vitamin E supplement. The first set of bars controls for only age, age-squared, and gender; the second set includes controls for education, income, marital status, race, smoking behavior, exercise, and vegetable consumption relative to the mean. In panel B (NHS) the outcome is two-year mortality, and the key independent variable is reported consumption of vitamin E. The coefficients are scaled by the death rate, so they can be interpreted as a percent change. The first set of bars controls only for age dummies. The second set includes controls for smoking, BMI, and indicators for hypertension, high cholesterol, diabetes, cancer diagnosis post-menopausal hormone use, angina, myocardial infarction, and menopausal status (this is a set similar to what is used in Stampfer et al. 1993). Error bars show 95 percent confidence intervals. Observations: NHANES, simple controls ($N_1 = 14,770$; $N_2 = 11,986$; $N_3 = 20,152$); NHANES, all controls ($N_1 = 10,199$; $N_2 = 10,666$; $N_3 = 18,554$); NHS, person-years, age controls ($N_1 = 470,264$; $N_2 = 521,313$; $N_3 = 144,715$); NHS, person-years, all controls: ($N_1 = 390,355$; $N_2 = 470,587$; $N_3 = 123,802$).

The variation in coefficients over time seems consistent with variation in selection, but as I note in Section II these change could also be driven by treatment effect heterogeneity. As I suggest there, it may be possible to get a sense of the scope of this possibility by using the baseline period to estimate treatment effect heterogeneity

in the observed data and then calculate the implied treatment effects in later periods by reweighting to match the observables. I do that in these examples; in general, the reweighting does not suggest a treatment effect heterogeneity explanation.

In the case of the NHANES, if we take the cross-sectional results as causal, the baseline effect of vitamin E on heart health is 0.060. Using the estimated heterogeneity in this effect in the baseline period, applied to the observable distribution in the intermediate period, I would predict a second-period coefficient of 0.064. In fact, the coefficient is 0.094. A similar result is obtained in the NHS. The predicted intermediate period effect would actually be smaller in magnitude than the baseline effect in this analysis, whereas the actual coefficient is much larger in magnitude. While this does not rule out treatment effect heterogeneity driving the effects, it does not support that theory either.

IV. 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 therefore present suggestive, corroborating evidence in another vitamin (vitamin D) and three diet measures (sugar, fat, and the Mediterranean diet).

A. 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* article on the benefits of vitamin D (Holick 2007), coverage in *The New York Times* and other outlets (e.g., Nagourney 2007), and a corresponding spike in Google searches. On the other side, around 2012 there was pushback. 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, b; Kolata 2012).

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. I will estimate equations of the form in equation (3) but with vitamin D rather than vitamin E as the outcome.

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 in which 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 HomeScan (panel C). In both datasets, across all the metrics, the evidence points to changes in selection consistent with the theory and patterns 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.

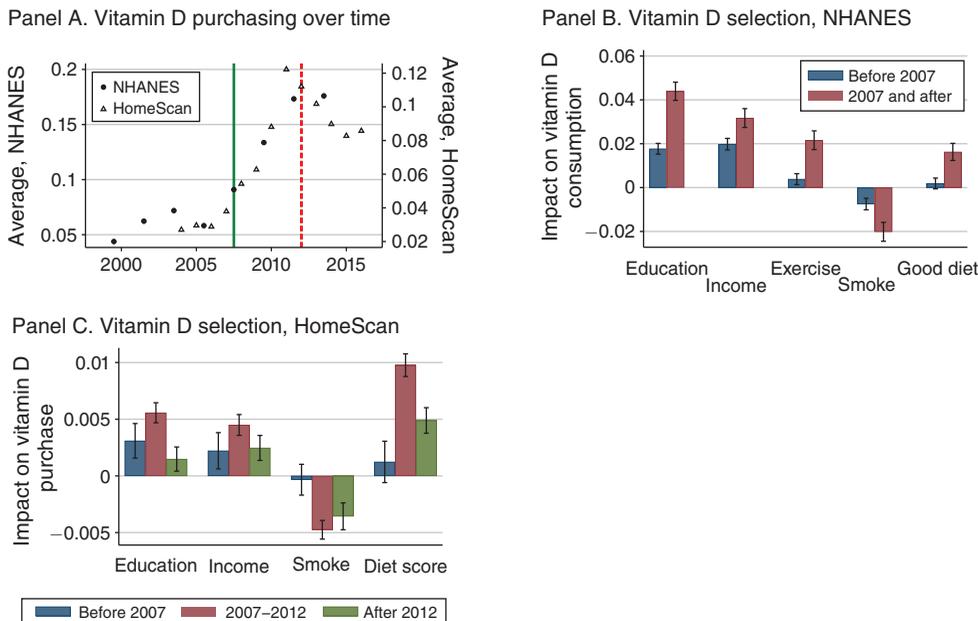


FIGURE 4. EVIDENCE ON VITAMIN D

Notes: This figure shows evidence on vitamin D. Panel A shows data on an indicator for vitamin D consumption in NHANES or purchasing in HomeScan. Events are marked with vertical lines; details of the events appear in online Appendix Table A1. The NHANES data range from 1998 through 2013; the 2014–2015 NHANES does not have dietary supplement data available. HomeScan data range from 2004 through 2016. Vitamin D consumption in the NHANES is based on reported supplements taken; in the HomeScan this is a measure of whether the household records a purchase of any single vitamin D supplement during the calendar year. Panels B and C show the relationship between vitamin D consumption and selection variables in both the NHANES and HomeScan data. These coefficients come from regressions of the form of equation (3), with vitamin D rather than vitamin E as the outcome. The NHANES regressions control for age, age-squared, and gender. HomeScan regressions include controls for household composition and household size. 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 a good diet is defined as having above-median calories from vegetables; in the HomeScan diet score is constructed based on Hut and Oster (2018). Smoking in the HomeScan data is defined based on whether the household purchases cigarettes. Error bars show 95 percent confidence intervals. Number of observations in panel A: NHANES (4,252 to 7,827), HomeScan (38,863 to 69,247). Panel B: $N_1 = 38,706$, $N_2 = 14,841$. Panel C: $N_1 = 131,536$, $N_2 = 399,313$, $N_3 = 264,228$.

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 percent.

B. Dietary Patterns

Background and Setup.—I analyze three dietary patterns: sugar intake, saturated fat intake, and adoption of a Mediterranean diet. I focus on presenting the results on the case of sugar, for illustrative purposes, but include all three in the online Appendix.

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 percent of calories (Peterkin 1990). In 2005 this limit was lowered to 7 percent (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 (Féart et al. 2009). A large randomized trial released in 2013 showed significant cardiovascular benefits (Estruch et al. 2013), although this largely postdates my data.

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.

For each dietary measure, I will estimate equation (3) and equation (4), replacing the vitamin E measures in each case with diet measures.

Results.—I present the results for sugar in the main text, focusing on the BMI outcome. The full set of diet results, including both health outcomes and all three dietary measures, are shown in online Appendix Table A3. The overall patterns are similar across dietary choices and outcomes.

The sugar results are shown in Figure 5. Panel A shows changes in sugar intake over time; it declines with the changes in recommendations.

Panel B documents corresponding changes in selection. 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.

Figure 5, panel C shows links with BMI. 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. The second set of columns in panel C shows the regressions including controls and demonstrates that the patterns persist.

V. 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 describe a simple data-generating process in which new health recommendations are more likely to be adopted by individuals who invest more in their health in other ways. Under this process, the bias in observational data-based estimates of the relationship between health behaviors and health outcomes will be endogenous to these recommendations.

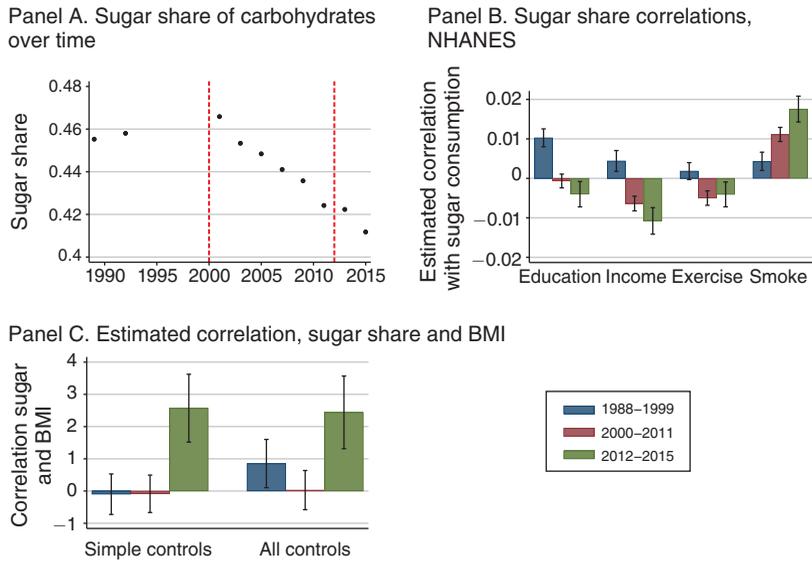


FIGURE 5. EVIDENCE ON SUGAR

Notes: This figure shows evidence on sugar. Panel A shows data on the sugar share of carbohydrates over time in the NHANES. Events are marked with vertical lines; details of the events appear in online Appendix Table A1. The NHANES data range from 1998 through 2015. Sugar share is measured based on dietary recall data. Panel B shows the relationship between the sugar share variable and selection variables in the NHANES. These coefficients come from regressions of the form of equation (3), with sugar share rather than vitamin E as the outcome. The NHANES regressions control for age, age-squared, and gender. The coefficients can be interpreted as the effect of a one standard deviation change in the selection variable on sugar share. Panel C shows changes in the sugar–BMI relationship over time; these result from estimation of equation (4) in the text. The outcome is BMI (weight in kilograms divided by height in meters-squared), and the key independent variable is sugar share of carbohydrates. The first set of bars controls for only age, age-squared, and gender; the second set includes controls for education, income, marital status, race, smoking behavior, and exercise. Error bars show 95 percent confidence intervals. Number of observations in panel A: (4,064 to 7,777). Panel B: $N_1 = 12,004$ – $15,682$, $N_2 = 24,807$ – $26,954$, $N_3 = 7,872$ – $8,683$. Panel C: Simple controls ($N_1 = 15,651$; $N_2 = 26,423$; $N_3 = 8,606$); all controls ($N_1 = 10,800$; $N_2 = 24,308$; $N_3 = 7,800$).

That this is a possibility follows directly from the theory; using data on vitamin supplements and diet, I demonstrate that this is not just a possibility but is reflected by dynamics in the data. The degree of selection in behaviors varies over time, and the relationship between behavior and health also varies with these changes in selection. The changes over time—both in the patterns of selection and the behavior–outcome relationships—are large.

These results have several implications that may be worth highlighting.

First, I will reiterate the possible implications for null effects. The nature of statistical precision is that—with a 5 percent significance cutoff—5 percent of the time treatment effect analysis will yield a significant result even if the true effect is zero. This is true even when the analysis is done through a randomized controlled trial. The dynamics here can reinforce such false-positive results. This issue is exacerbated by publication bias (i.e., Ioannidis 2005). In the presence of a publication process that favors significant outcomes, we may be even more likely to observe false-positive results published and, subsequently, reinforced by these patterns.

Second, I note that the results on selection in uptake of recommended behaviors may be independently interesting. In particular, to the extent that the recommended behaviors *are* beneficial, this provides insight into how improvements in knowledge about health behaviors may exacerbate health inequality.

Finally, although this paper focuses on diet and vitamin supplementation, it seems possible that similar dynamics could show up in other behaviors where health recommendations vary over time. Linking back to the first point above, this seems most likely in cases where the true treatment effects are likely to be (at best) small.

REFERENCES

- Bakalar, Nicholas.** 2012a. “Risks: When Too Much Vitamin D Is Too Much.” *New York Times*, January 12. <https://www.nytimes.com/2012/01/17/health/research/risks-when-too-much-vitamin-d-is-too-much.html>.
- Bakalar, Nicholas.** 2012b. “Study Questions Benefit of Extra Vitamin D.” *New York Times*, October 23. <https://well.blogs.nytimes.com/2012/10/22/study-questions-benefit-of-extra-vitamin-d/>.
- Bao, Ying, Monica L. Bertoa, Elizabeth B. Lenart, Meir J. Stampfer, Walter C. Willett, Frank E. Speizer, and Jorge E. Chavarro.** 2016. “Origin, Methods, and Evolution of the Three Nurses’ Health Studies.” *American Journal of Public Health* 106 (9): 1573–81.
- Berrigan, David, Kevin Dodd, Richard P. Troiano, Susan M. Krebs-Smith, and Rachel Ballard Barbash.** 2003. “Patterns of Health Behavior in US Adults.” *Preventive Medicine* 36 (5): 615–23.
- Brody, Jane E.** 1993. “Vitamin E Greatly Reduces Risk of Heart Disease, Studies Suggest.” *New York Times*, May 20. <https://www.nytimes.com/1993/05/20/us/vitamin-e-greatly-reduces-risk-of-heart-disease-studies-suggest.html>.
- Brown, Deborah J., and Lee F. Schrader.** 1990. “Cholesterol Information and Shell Egg Consumption.” *American Journal of Agricultural Economics* 72 (3): 548–55.
- Centers for Disease Control and Prevention.** 2014. “National Health and Nutrition Examination Survey (NHANES) Data.” CDC, National Center for Health Statistics (NCHS), <https://wwwn.cdc.gov/nchs/nhanes/Default.aspx> (accessed February 17, 2020).
- Chang, Hung-Hao, and David R. Just.** 2007. “Health Information Availability and the Consumption of Eggs: Are Consumers Bayesians?” *Journal of Agricultural and Resource Economics* 32 (1): 77–92.
- Chang, Kenneth.** 2012. “What We Eat: Doubts by the Teaspoonful.” *New York Times*, June 12. <https://well.blogs.nytimes.com/2012/06/11/which-sweetener-should-you-choose/>.
- Chern, Wen S., Edna T. Loehman, and Steven T. Yen.** 1995. “Information, Health Risk Beliefs, and the Demand for Fats and Oils.” *Review of Economics and Statistics* 77 (3): 555–64.
- Colditz, Graham A., Joann Manson, and Susan Hankinson.** 1997. “The Nurses’ Health Study: 20-Year Contribution to the Understanding of Health among Women.” *Journal of Women’s Health* 6 (1): 49–62.
- Cutler, David M.** 2004. “Behavioral Health Interventions: What Works and Why?” In *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*, edited by Norman B. Anderson, Rodolfo A. Bulatao, and Barney Cohen, chapter 17. Washington, DC: National Academies Press.
- Cutler, David M., and Adriana Lleras-Muney.** 2010. “Understanding Differences in Health Behaviors by Education.” *Journal of Health Economics* 29 (1): 1–28.
- Cutler, David M., Adriana Lleras-Muney, and Tom Vogl.** 2008. “Socioeconomic Status and Health: Dimensions and Mechanisms.” NBER Working Paper 14333.
- Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi L. Williams.** 2019. “Screening and Selection: The Case of Mammograms.” NBER Working Paper 26162.
- Esposito, Katherine, Raffaele Marfella, Miryam Ciotola, Carmen Di Palo, Francesco Giugliano, Giovanni Giugliano, Massimo D’Armiendo, Francesco D’Andrea, and Dario Giugliano.** 2004. “Effect of a Mediterranean-Style Diet on Endothelial Dysfunction and Markers of Vascular Inflammation in the Metabolic Syndrome: A Randomized Trial.” *JAMA* 292 (12): 1440–46.
- Estruch, Ramón, Emilio Ros, Jordi Salas-Salvadó, Maria-Isabel Covas, Dolores Corella, Fernando Arós, Enrique Gómez-Gracia, et al.** 2013. “Primary Prevention of Cardiovascular Disease with a Mediterranean Diet.” *New England Journal of Medicine* 368 (14): 1279–90.
- Féart, Catherine, Cécilia Samieri, Virginie Rondeau, Hélène Amieva, Florence Portet, Jean-François Dartigues, Nikolaos Scarmeas, and Pascale Barberger-Gateau.** 2009. “Adherence to a Mediterranean Diet, Cognitive Decline, and Risk of Dementia.” *JAMA* 302 (6): 638–48.

- Finke, Michael S., and Sandra J. Huston.** 2003. "Factors Affecting the Probability of Choosing a Risky Diet." *Journal of Family and Economic Issues* 24 (3): 291–303.
- Friel, Sharon, John Newell, and Cecily Kelleher.** 2005. "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* 8 (2): 159–69.
- Goldman, Dana P., and James P. Smith.** 2002. "Can Patient Self-Management Help Explain the SES Health Gradient?" *Proceedings of the National Academy of Sciences of the United States of America* 99 (16): 10929–34.
- Greenland, Sander, James M. Robins, and Judea Pearl.** 1999. "Confounding and Collapsibility in Causal Inference." *Statistical Science* 14 (1): 29–46.
- Holick, Michael F.** 2007. "Vitamin D Deficiency." *New England Journal of Medicine* 357 (3): 266–81.
- Hut, Stefan, and Emily Oster.** 2018. "Changes in Household Diet: Determinants and Predictability." NBER Working Paper 24892.
- Ioannidis, John P. A.** 2005. "Why Most Published Research Findings Are False." *PLOS Medicine* 2 (8): e124.
- Ippolito, Pauline M., and Alan D. Mathios.** 1995. "Information and Advertising: The Case of Fat Consumption in the United States." *American Economic Review* 85 (2): 91–95.
- Kinnucan, Henry W., Hui Xiao, Chung-Jen Hsia, and John D. Jackson.** 1997. "Effects of Health Information and Generic Advertising on US Meat Demand." *American Journal of Agricultural Economics* 79 (1): 13–23.
- Kirkpatrick, Sharon I., Kevin W. Dodd, Jill Reedy, and Susan M. Krebs-Smith.** 2012. "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* 112 (5): 624–35.
- Knoops, Kim T. B., Lisette C. P. G. M. de Groot, Daan Kromhout, Anne-Elisabeth Perrin, Olga Moreiras-Varela, Alessandro Menotti, and Wija A. van Staveren.** 2004. "Mediterranean Diet, Lifestyle Factors, and 10-Year Mortality in Elderly European Men and Women: The HALE Project." *JAMA* 292 (12): 1433–39.
- Kolata, Gina.** 2004. "Large Doses of Vitamin E May Be Harmful." *New York Times*, November 11. <https://www.nytimes.com/2004/11/11/health/large-doses-of-vitamin-e-may-be-harmful.html>.
- Kolata, Gina.** 2012. "Healthy Women Advised Not to Take Calcium and Vitamin D to Prevent Fractures." *New York Times*, June 12. <https://www.nytimes.com/2012/06/13/health/calcium-and-vitamin-d-ineffective-for-fractures-us-preventive-services-task-force-says.html>.
- Kowalski, Amanda E.** 2018. "Behavior within a Clinical Trial and Implications for Mammography Guidelines." NBER Working Paper 25049.
- Krauss, Ronald M., Robert H. Eckel, Barbara Howard, Lawrence J. Appel, Stephen R. Daniels, Richard J. Deckelbaum, John W. Erdman, Jr., et al.** 2000. "AHA Dietary Guidelines: Revision 2000: A Statement for Healthcare Professionals from the Nutrition Committee of the American Heart Association." *Circulation* 102 (18): 2284–99.
- Miller, Edgar R., Roberto Pastor-Barriuso, Darshan Dalal, Rudolph A. Riemersma, Lawrence J. Appel, and Eliseo Guallar.** 2005. "Meta-Analysis: High-Dosage Vitamin E Supplementation May Increase All-Cause Mortality." *Annals of Internal Medicine* 142 (1): 37–46.
- Miller, Leslie.** 1993. "Lowly Vitamin Supplements May Pack a Big Health Punch." *USA Today*, June.
- Nagourney, Eric.** 2007. "Shortage of Vitamin D May Weaken the Elderly." *New York Times*, May 1. <https://www.nytimes.com/2007/05/01/health/01nutr.html>.
- Nielsen Company.** 2018. "HomeScan Data." Nielsen Company, <https://www.chicagobooth.edu/research/kilts/datasets/Nielsen>.
- Nurses' Health Study Research Team.** 2019. "Nurses' Health Study." Centers for Disease Control and Prevention, National Center for Health Statistics, <https://www.nurseshealthstudy.org/researchers>.
- Oster, Emily.** 2020. "Health Recommendations and Selection in Health Behaviors: Dataset." *American Economic Review: Insights*. <https://doi.org/10.1257/aeri.20190355>.
- Peterkin, B. B.** 1990. "Dietary Guidelines for Americans." *Journal of the American Dietetic Association* 90 (12): 1725–27.
- Rimm, Eric B., Meir J. Stampfer, Alberto Ascherio, Edward Giovannucci, Graham A. Colditz, and Walter C. Willett.** 1993. "Vitamin E Consumption and the Risk of Coronary Heart Disease in Men." *New England Journal of Medicine* 328 (20): 1450–56.
- Roosen, Jutta, Stéphan Marette, Sandrine Blanchemanche, and Philippe Verger.** 2009. "Does Health Information Matter for Modifying Consumption? A Field Experiment Measuring the Impact of Risk Information on Fish Consumption." *Applied Economic Perspectives and Policy* 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, et al.** 2012. "IOM Committee Members Respond to

- Endocrine Society Vitamin D Guideline.” *Journal of Clinical Endocrinology and Metabolism* 97 (4): 1146–52.
- Stampfer, Meir J., Charles H. Hennekens, JoAnn E. Manson, Graham A. Colditz, Bernard Rosner, and Walter C. Willett.** 1993. “Vitamin E Consumption and the Risk of Coronary Disease in Women.” *New England Journal of Medicine* 328 (20): 1444–49.
- Taubes, Gary.** 2011. “Is Sugar Toxic?” *New York Times*, April 13. <https://www.nytimes.com/2011/04/17/magazine/mag-17Sugar-t.html>.
- Thompson, Tommy G., and Ann M. Veneman.** 2005. *Dietary Guidelines for Americans*. Washington, DC: United States Department of Health and Human Services and United States Department of Agriculture.
- Trichopoulou, Antonia, Tina Costacou, Christina Bamia, and Dimitrios Trichopoulos.** 2003. “Adherence to a Mediterranean Diet and Survival in a Greek Population.” *New England Journal of Medicine* 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, and Matthias Egger.** 2007. “Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration.” *Epidemiology* 18 (6): 805–35.