

NBER WORKING PAPER SERIES

CHAIN RESTAURANT CALORIE POSTING LAWS, OBESITY, AND CONSUMER
WELFARE

Charles J. Courtemanche
David E. Frisvold
David Jimenez-Gomez
Mariétou Ouayogodé
Michael Price

Working Paper 26869
<http://www.nber.org/papers/w26869>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2020

We thank Chelsea Crain and Abigail Kappelman for research assistance. We thank Melissa Haberlin for her legal expertise and assistance interpreting local and state calorie mandate legislation. We thank participants from numerous seminars and conferences for helpful comments. We acknowledge funding from the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number R01DK107686. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Charles J. Courtemanche, David E. Frisvold, David Jimenez-Gomez, Mariétou Ouayogodé, and Michael Price. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Chain Restaurant Calorie Posting Laws, Obesity, and Consumer Welfare
Charles J. Courtemanche, David E. Frisvold, David Jimenez-Gomez, Mariétou Ouayogodé,
and Michael Price
NBER Working Paper No. 26869
March 2020
JEL No. D91,I12,I18,I31

ABSTRACT

The 2010 Patient Protection and Affordable Care Act (ACA) introduced a mandate requiring chain restaurants to post calorie counts on menus and menu boards. This paper investigates whether and why calorie posting laws work. To do so, we develop a model of calories consumed that highlights two potential channels through which mandates influence choice and outlines an empirical strategy to disentangle these alternatives. We test the predictions of our model using data from the Behavioral Risk Factor Surveillance System to compare changes in body mass index (BMI), obesity, and consumer well-being in locations that implemented calorie-posting laws between 2008 and 2011 to those in neighboring locations without such laws. We find that calorie mandates lead to a small but statistically significant reduction in average BMI of 0.2 kg/m² (1.5 pounds) and reductions in self-reported measures of life satisfaction. Quantile regressions provide evidence that reductions in BMI and life satisfaction are concentrated among those with healthy weight. Viewed in its totality, the pattern of results is consistent with an economic model in which calorie labels influence consumers both by providing information and by imposing a welfare-reducing moral cost on unhealthy eating.

Charles J. Courtemanche
Department of Economics
Gatton College of Business and Economics
University of Kentucky
Lexington, KY 40506-0034
and NBER
courtemanche@uky.edu

Mariétou Ouayogodé
Department of Population Health Sciences
School of Medicine and Public Health
University of Wisconsin-Madison
610 Walnut St, 785 WARF
Madison, WI 53726
marietou.ouayogode@wisc.edu

David E. Frisvold
Department of Economics
University of Iowa
21 East Market Street
Iowa City, IA 52242
and NBER
david-frisvold@uiowa.edu

Michael Price
Department of Economics, Finance,
and Legal Studies
The University of Alabama
250 Alston Hall
Box 870224
Tuscaloosa, AL 35487
and NBER
mkprice2@cba.ua.edu

David Jimenez-Gomez
Department of Economics
University of Alicante
Spain
davidjimenezgomez@ua.es

I. Introduction

The concurrent increases in restaurant meals, caloric intake, and obesity have spurred a literature exploring the causal link between the consumption of meals at restaurants and increases in obesity (e.g., Currie et al., 2010; Dunn, 2010; Anderson and Matsa, 2011; Courtemanche et al., 2016).² From a policy perspective, this connection raises the question of whether market failures are at work – a concern that is enhanced by evidence that consumers underestimate the calories contained in restaurant meals (e.g., Kozup et al., 2003; Burton and Creyer, 2004; CCPHA, 2007; Wisdom et al., 2010; Block et al., 2013). Attempting to address such externalities and reverse trends in obesity, policymakers have introduced mandates that require chain restaurants to post calorie counts on menus and menu boards. Such mandates are designed to “nudge” healthier food choice by making calories more salient at the point of purchase.

Despite the increased prevalence of such mandates, the existing literature has focused on estimating their impacts on either calories consumed (e.g., Elbel et al., 2009 and 2011; Bollinger et al., 2011; Finkelstein et al., 2011; Auchincloss et al., 2013; Krieger et al., 2013) or BMI (e.g., Yelowitz, 2016; Deb and Vargas, 2016; Restrepo, 2017). Given this focus, the existing literature provides little to no guidance as to the channels through which “nudges”, such as calorie mandates, influence choice or the associated welfare effects. It is this gap in the literature that our study aims to fill. In doing so, we seek to move the literature from studies designed to evaluate whether a given “nudge” works to studies designed to understand why the “nudge” works.

We begin by developing a model of calories consumed and how this choice is impacted by calorie mandates. Our model assumes that calorie choice is influenced by three considerations; (i)

² For example, Currie et al. (2010) find that having a fast food restaurant within 0.1 miles of a school results in an approximate 5.2 percent increase in obesity rates amongst ninth grade students. However, Anderson and Matsa (2011) find that such effects are less pronounced in rural areas.

consumption (or experienced) utility, which depends on both calories consumed and an individual's weight, (ii) an “internality” or bias that influences choice but not subsequent consumption utility, and (iii) psychic (or moral) costs that influences both decision and consumption utility. In our model, internalities could reflect imperfect information about calorie counts or other sources of “bias” such as inattention, “cognitive bandwidth”, or time inconsistency that lead to the overconsumption of calories (e.g., Mullainathan and Shafir, 2013; Courtemanche et al., 2015a; Allcott et al., 2019; Sadoff et al., forthcoming). Psychological (or moral) costs depend on the environment, i.e., whether one is aware of the relative calories for different menu items, and reflects the notion that individuals may feel guilty when selecting high calorie items.

Our model shows that calorie mandates can influence choice through two distinct channels. First, by providing information on caloric content on menus and increasing the salience of healthy food options at the time of purchase, calorie mandates might reduce internalities and lead to improved choice for “biased” agents. Second, such labels may impose a “shadow tax” on higher calorie items and hence distort the relative “price” of such items vis-à-vis lower calorie alternatives.³ Our model thus predicts that calorie mandates should cause a decrease in calories consumed but that the subsequent effects on the decision-maker's welfare depend on the relative size of the internality foregone and the psychic costs (guilt) triggered by the presence of calorie labels. Further, our model provides a way to distinguish between these channels by examining heterogeneity in the effects on BMI and changes in stated wellbeing.

We use data from the 1994-2012 waves of the Behavioral Risk Factor Surveillance System (BRFSS) to test the predictions of our model and uncover the channels through which calorie

³ Intuitively, individuals can avoid the “shadow tax” by selecting lower calorie items. However, if such items are less preferred than higher calorie alternative, the presence of calorie labels can lead to lower levels of consumption utility.

mandates impact choice. Rather than relying on a single test to identify the channels, we analyze the effects of calorie mandates on multiple outcomes using the BRFSS and other secondary datasets, examine different subsets of the population based on our theoretical predictions, and weigh the totality of the evidence.⁴ To do so, we investigate the impacts of local and state chain restaurant calorie labeling mandates on BMI and overall consumer well-being, as measured by life satisfaction. As the BRFSS data do not include information on calories consumed, we also investigate the impact of the calorie mandates on other factors that could influence BMI—smoking, alcohol consumption, and exercise. Such analysis provides indirect evidence that any observed changes in BMI reflect changes in calories consumed and allows us to better link our theory and empirical findings.

Identifying variation comes from differences in calorie-labeling mandate implementation and enforcement across geographic areas and over time. We provide two pieces of evidence highlighting the validity of our identification strategy: differences in the prevalence and awareness of calorie labels across treated areas with different levels of enforcement and control areas. First, we summarize data from phone calls to chain restaurants showing that restaurants in treated areas were more likely to report listing calories on menus and menu boards than counterparts in control areas. Second, we summarize survey data showing a substantial difference in the likelihood individuals in treated areas notice calorie labels at chain restaurants than counterparts in control areas.

Turning to our econometric results, we find that calorie mandates lead to small but significant reductions in BMI of 0.19 kg/m² (corresponding to about 1.5 pounds and 0.7% of the sample mean). Interestingly, we find that the effects on BMI are driven entirely by those locations

⁴ In this regard, our strategy is similar to Porter and Zona (1993), who rely upon multiple tests rather than a single test to detect bid rigging in procurement auctions.

who actively enforced the local mandates throughout our sample period. In fact, we find no effects in locations where enforcement was “weak”. Such heterogeneity is consistent with differences in the likelihood that restaurants across these different locations report posting calories on menus and consumers living in these locations report having noticed calories on menus when eating out.

We next explore the impact of the calorie mandates on compensating behaviors that could influence BMI independent of any change in calories consumed. Specifically, we explore if the enactment of the local mandates had any impact on three compensating behaviors: (i) the likelihood an individual smokes, (ii) the number of alcoholic drinks consumed per month, and (iii) the likelihood an individual exercises along with the time spent exercising each month. We find that the enactment of the local mandates had no effect on any of these metrics. Viewed in conjunction with the effects on BMI, such patterns suggest that the observed changes in BMI are driven by the presence of calorie counts and subsequent changes in calories consumed as opposed to other factors.

Our theoretical model suggests that calorie labels could reduce BMI because of the provision of information and the presence of moral costs. Evidence on the importance of moral costs comes from our analysis of life satisfaction. Specifically, we find that the presence of calorie labeling mandates leads to a *reduction* in self-reported life satisfaction that is concentrated among people in the healthy weight category, as opposed to those who are overweight or obese. These findings are consistent with our model and suggest that calorie labels impact consumers both by providing information and by inflicting a welfare-reducing moral cost on unhealthy eating for consumers with relatively accurate information on caloric content prior to the introduction of the labels.

To further understand the importance of moral costs as a channel through which calorie labels influence choice, we estimate quantile regressions that allow heterogeneous response to local mandates across the BMI distribution. Results from this exercise show that the calorie labeling laws have statistically significant effects in quantiles that correspond to BMIs in the healthy weight category. That we find effects for healthy weight individuals who are unlikely affected by internalities provides evidence consistent with the notion of moral costs. Even in the absence of new information, consumers would reduce calories consumed when faced with calorie labels as a way to avoid the psychic costs (or guilt) triggered by the purchase of high calorie items. This is a theoretical channel introduced by Jimenez-Gomez's (2019) model of nudges that we are the first to explore empirically in a relevant context.

Our paper contributes to several literatures. First, our paper contributes to a vast body of work exploring how simple messages that provide consumer information about misperceived or shrouded attributes of a product or service impact subsequent choices. For example, there is a growing body of work exploring how simple messages that educate individuals about non-linear tax incentives such as the Earned Income Tax Credit (Chetty and Saez, 2013) or Social Security benefits (Liebman and Luttmer, 2011) impact subsequent labor supply.⁵ A related body of work examines the impact of providing information on misperceived or shrouded attributes such as fuel efficiency (Allcott and Knittel, 2019), energy savings (Allcott and Taubinsky, 2015; Davis and Metcalf, 2016; Allcott and Sweeney, 2017), the benefits of school choice or post-secondary education (Hastings and Weinstein, 2008; Jensen, 2010; Barr and Turner, 2018), the returns to job search or retirement savings plans (Duflo et al., 2016; Saez, 2009; Altmann et al., 2019), restaurant

⁵ There is similar work in the context of environmental economics showing that educating consumers about non-linear tariffs for electricity and the prevailing marginal prices impact subsequent patterns of consumption (Kahn and Wolak, 2013).

hygiene grades (Jin and Leslie, 2003), and truth-in-lending styles disclosures (Seira et al., 2017) on subsequent choice. Amongst this body of work, our study is most closely related to explorations of the impact of calorie labels on subsequent calories purchased and BMI (e.g., Bollinger et al., 2011; Deb and Vargas, 2016; Restrepo, 2017). However, our paper extends this earlier work on calorie mandates by developing a theoretical model and using it to identify the channels through which calorie labels impact choice behavior.

Second, our paper contributes to a growing body of literature using measures of subjective well-being to measure the effects of government policies or other economic shocks. For example, recent work has used measures of subjective well-being to place monetary value on public goods (or public bads) such as airport noise (Van Praag and Baarsma, 2005), flood disasters (Luechinger and Raschky, 2009), terrorism (Frey et al., 2009; Metcalfe et al., 2011), traffic congestion (Anderson et al., 2016), the Chernobyl disaster (Danzer and Danzer, 2016), and air pollution (Luechinger, 2009; Levinson, 2012). A related body of work has used subjective well-being to evaluate the impacts of policies or economic shocks such as inflation targeting (Di Tella et al., 2001), German re-unification (Frijters et al., 2004), creative destruction (Aghion et al., 2016), cigarette taxes (Gruber and Mullainathan, 2005), and the provision of improved housing for the poor (Galiani et al., 2018). Our paper extends this line of inquiry by using measures of subjective well-being to test theory and identify the channels through which a policy impacts choice.

The literature on behavioral welfare economics has emphasized the importance of heterogeneity in the population (Allcott and Taubinsky, 2015; Taubinsky and Rees-Jones, 2018; Allcott and Kessler, 2019) but has largely ignored the potential psychological costs (and heterogeneity therein) of the nudge. One exception is Thunström (2019), who measured hypothetical willingness-to-pay (WTP) and emotional reactions to a calorie information nudge in

an online experiment, and found a heterogeneous effect: those with higher self-control reacted more to the nudge and felt positive emotions, while those with lower self-control reacted less and had negative emotions. We also find that those we identify with higher self-control (i.e., those with BMI in the healthy weight range) react to the nudge; however those individuals exhibit lower life satisfaction, implying that they suffer psychological costs (rather than having positive emotions from the nudge). Our paper thus contributes by measuring the heterogeneity of the effect of the nudge and its psychological costs in the field.⁶

Finally, our research contributes to a growing literature exploring how best to use insights from behavioral economics to achieve policy goals. To date, such work has focused on areas such as educational attainment (Levitt et al., 2016), female labor force participation rates (Bursztyn et al., 2018), electricity and water conservation (Allcott, 2011; Costa and Kahn, 2013; Ferraro and Price, 2013; Allcott and Rogers, 2014), tax compliance (Fellner et al., 2013; Dwenger et al., 2016; Hallsworth et al., 2017), traffic violations (Chen et al., 2017), or the use of credit cards (Seira et al., 2017).⁷ Unlike this prior work, however, our paper goes beyond simply evaluating whether or not the given policy works. We set forth to understand why the policy works and determine if the reason why the policy works differs across segments of the population. In this regard, our paper shares similarity with work on the design of “sin” taxes (Gruber and Koszegi, 2004; Taubinsky and Rees-Jones, 2018; Allcott et al., 2019) and efforts to “nudge” the repayment of credit card debt (Bursztyn et al., 2019).

⁶ In particular, we show that once calorie-information interventions are implemented in the field, psychological costs might exist even for those who had a positive WTP for the nudge in the lab, suggesting the importance of evaluating policies after implementation in the field.

⁷ For an overview of work in this area, we refer the interested reader to survey articles such as Amir (2005); Loewenstein et al., (2007); Downs et al. (2009); Volpp et al., (2011); Matjasko et al., (2015); or Chetty (2015).

II. Background on Calorie Labeling Laws

New York City became the first location in the United States to pass a chain restaurant calorie labeling law in 2006, which was implemented in April 2008. Other areas across the country (King County, WA; Philadelphia County, PA; Albany and Schenectady County, NY; Montgomery County, MD; and the state of Vermont) implemented similar laws between 2009 and 2011, as documented in Table 1.⁸ The enforcement of laws passed in other locations, such as California, was tabled due to the passage of the ACA, as local jurisdictions anticipated the implementation and enforcement of the federal mandate. Further, enforcement varied across locations with pre-ACA mandates. For example, while New York City, Philadelphia County, Schenectady County, and Montgomery County actively enforced the local mandate throughout our sample period, enforcement in King County, Vermont, and Albany County was either weak or varied throughout the sample period.⁹

The first wave of studies exploring the impacts of these laws utilized receipt data and difference-in-difference designs to estimate the effect of calorie labels on calories ordered.¹⁰ Treatment groups in these early studies were comprised of customers at restaurants exposed to a law and control groups constructed using customers eating at the same restaurant chains in neighboring areas. Results from this body of work are mixed. For example, Elbel et al. (2009, 2011) find no evidence that the New York City mandate influenced calories ordered among adults,

⁸ The 2010 Patient Protection and Affordable Care Act (ACA) introduced a national mandate requiring chain restaurants and retail food establishments with 20 or more locations to post calories on menus and menu boards. The start date of this mandate was originally set for December 2015 but was delayed until May 2018.

⁹ Based on correspondence with local officials, King County actively enforced the mandate between January 2009 and May 2010, Vermont did not actively enforce the mandate through inspection but instead reacted to customer complaints about a lack of calories on the menu, and Albany County stopped enforcement six months after the law was implemented.

¹⁰ A related literature, using laboratory and field experiments, suggests that providing nutritional information can affect meal decisions (e.g., Burton et al., 2006; Roberto et al., 2010; Wisdom et al., 2010; Ellison et al., 2013; Bedard and Kuhn, 2015; VanEpps et al., 2016; Cawley et al., 2018).

adolescents, or children in low-income neighborhoods. However, Bollinger et al. (2011) found that the New York City law reduced calories ordered at Starbucks restaurants by 6%. The evidence on the effects of the mandates using data from restaurants in and around King County, WA is also mixed – Finkelstein et al. (2011) report no effect on calories ordered at Taco Time Northwest restaurants and Krieger et al. (2013) estimate a reduction in calories ordered at taco and coffee restaurants but not burger and sandwich restaurants.¹¹

The mixed results from this first wave of studies points to limitations inherent to focusing on a small number of restaurants in a particular geographic area. Consumer preferences and knowledge vary across locations, and the extent of misinformation about calories could differ across different types of restaurants. The ambiguity could also reflect sorting, as calorie labeling could influence the types of people who eat at particular types of restaurants. Additionally, inconsistent enforcement could lead to unobserved heterogeneity in the likelihood any given establishment posted calorie counts. Finally, such analyses are susceptible to the concerns about common trends and understated standard errors in comparative case studies with a small number of treated units (Abadie et al., 2010).

A more recent set of studies has used data from the BRFSS to shift the focus of this work from estimating the impacts of calorie mandates on calories ordered to BMI. For example, Restrepo (2017) compared changes in BMI in the counties in New York State that implemented calorie labeling laws to changes in other counties in the state and found that the local laws reduced both average BMI by 0.4 kg/m² (or approximately 1.5 percent) and the probability of being obese by 3 percentage points. Yelowitz (2016) uses a sample consisting of respondents from the thirty

¹¹ Auchincloss et al. (2013) is the only paper to estimate the effect of the mandate in Philadelphia and finds that the customers eating in treated restaurants consume 151 fewer calories than do counterparts eating at the same chain in surrounding areas.

largest U.S. cities, which includes several that had passed a calorie mandate, and finds that such mandates caused small and statistically insignificant reductions in both average BMI and the probability of being obese.¹² Deb and Vargas (2016) were the first to utilize data from every location that passed a calorie mandate to estimate the effect of such policies.¹³ To do so, they implement a data-driven entropy balancing procedure to reweight the control group in a way that best matches the treatment group's pre-treatment BMI trajectory and find small reductions in average BMI for both men and women.

Our approach differs from this prior work along multiple important dimensions. First, we worked with a health policy legal expert to develop a more rigorous classification of treatment and control groups. Prior work identified treatment groups utilizing information on the passage and intended effective dates of local calorie labeling laws. Such a strategy does not account for the fact that many of these local laws were either repealed or never actually enforced due to expected preemption from the ACA's national mandate. Our strategy for identifying treated areas was to separately classify each law using a combination of legal statutes, media articles, and direct correspondence with state and local health departments. In doing so, we conclude that many of the "treated" areas of Yelowitz (2016) and Deb and Vargas (2016) were either never actually treated (e.g., Oregon) or only temporarily treated (e.g., California, whose law was repealed). We use as treated areas only those for which we were able to document that the law was actually implemented as intended and was never repealed. As shown in Table 1, these include the five counties of New York City; King County, WA; Philadelphia County, PA; Albany, Schenectady,

¹² Cities that passed a local mandate and were thus considered treated include New York, Philadelphia, Seattle, Portland, Los Angeles, San Francisco, San Jose, San Diego, and Boston.

¹³ Their treated areas include New York City; the counties of King (WA), Multnomah (OR), Albany (NY), Schenectady (NY), Suffolk (NY), Philadelphia (PA); and the states of California, Maine, Massachusetts, and Oregon. The control areas included Connecticut, Delaware, New Hampshire, Maryland, New Jersey, Rhode Island, Vermont, Arizona, Idaho, and Nevada.

Suffolk, and Westchester County, NY; Montgomery County, MD; and the state of Vermont. We further validate these choices by using phone calls to restaurants and consumer surveys to document clear differences in the prevalence of calorie labeling between these treatment sites and neighboring areas, as described below.

A second difference between our approach and the prior literature is that we evaluate calorie-labeling laws through the lens of a theoretical framework, which highlights that changes in consumer behavior need not lead to increased consumer welfare. This differs from the standard economic model whereby calorie provision can only influence choices by rectifying imperfect information. Since calorie-labeling laws do not (directly) impose a monetary cost on the consumer, prior work has assumed that any resulting changes in food selections are unambiguously welfare improving.¹⁴ In contrast, our framework allows for calorie labels to trigger moral costs that are akin to a tax on high calorie meals. The net effect of labeling laws on consumer welfare is thus theoretically ambiguous *but* can be inferred using available data. To evaluate this possibility, we go beyond prior work and look not only at changes in average BMI and across the BMI distribution but also heterogeneous changes in subjective well-being across treatment and control areas as a way to understand why calorie labels impact consumer choice.

III. Theoretical Model

It is well known that the environment (or *choice architecture*) in which people make decisions affects their choices (Thaler and Sunstein, 2008). We assume that individuals choose an element $x \in X$ (with the interpretation that x represents calories) in a given environment $e \in E$ (the environment could be multidimensional).

¹⁴ Such sentiment is highlighted in Bollinger et al. (2011) who write that, "...as far as regulatory policies go, the costs of calorie posting are very low – so even these small benefits could outweigh the costs."

We adapt the framework in Jimenez-Gomez (2019) to model the calorie nudge. First, we define “experienced utility” $u_i^{exp}(x|e)$ as the “hedonic experience associated with an outcome” (Kahneman and Thaler, 2006), that for our purposes also includes the health effects associated with consumption.¹⁵ People, however, do not always maximize their experienced utility (Kahneman et al., 1997; Berridge, 2003). Formally, we define “decision utility” $u_i^{dec}(x|e)$ as the utility that individual i maximizes, and the “internality” $\Lambda_i(x|e)$ as the difference between individual i ’s decision and experienced utility:

$$u_i^{dec}(x|e) = u_i^{exp}(x|e) + \Lambda_i(x|e).$$

The internality $\Lambda_i(x|e)$ could be due to lack of information about the calorie content of foods, but it also encompasses other sources of non-utility-maximizing behavior such as lack of attention, “cognitive bandwidth”, or time inconsistency (Mullainathan and Shafir, 2013; Courtemanche et al., 2015a).¹⁶ Calorie labels might reduce this internality by providing relevant information to individuals and/or increasing the saliency of choosing healthy food.

Critics of nudges have argued that nudges might impose a psychological cost (Glaeser, 2006). As described by Levitt and List (2007), utility depends not only on consumption but also on a “moral cost” for violating social norms. Nudges can increase this moral cost. For instance, in the charitable giving literature, Andreoni and Rao (2011) and DellaVigna et al. (2012) have argued that being asked to give imposes a “shadow tax” of guilt that can only be removed by giving. It becomes optimal to give conditional on being asked, but people may be left worse off than if they had never been asked. A parallel process could apply to calorie labels. Seeing a high

¹⁵ We define $u_i^{exp}(x|e)$ formally below, as a function of calories consumed and the weight associated with calorie consumption.

¹⁶ In particular, the bias could arise from individuals having incorrect expectations about how much they will enjoy consumption, for example, due to projection bias (Loewenstein et al., 2003).

calorie count for a preferred item may impose a “shadow tax” of guilt that induces a lower-calorie choice, but the consumer’s resulting utility may be lower than if she had never seen the information and had chosen the unhealthy item. In other words, the motivation for the healthier choice may be guilt rather than health.

We incorporate the idea of a moral cost by assuming that experienced utility u_i^{exp} , which is the relevant concept for welfare analysis, is

$$u_i^{exp}(x|e) = v_i(x) - c_i(x|e) - p(x),$$

where $v_i(x)$ is the utility from choice x , which does not depend on the environment, $p(x)$ is the price of choice x , and $c_i(x|e)$, is the psychological/moral cost from consuming highly caloric foods, which depends on the environment, i.e., whether one is aware of the amount of calories. Following the literature on the economics of obesity (Philipson and Posner 1999), we assume $v_i(x) = V_i(x, w_i(x))$, where $w_i(x)$ is the individual’s weight, which is increasing in x . Utility V_i depends on the consumption of x and also the individual’s weight. We assume that v_i is increasing and concave in x .¹⁷ Therefore, the individual’s decision utility, which is the relevant one for understanding her choices, is

$$u_i^{dec}(x|e) = u_i^{exp}(x|e) + \Lambda_i(x|e) = v_i(x) - c_i(x|e) - p(x) + \Lambda_i(x|e).^{18}$$

Let e_0 be the status quo (i.e., no nudge), and e_1 be an environment where there are calorie labels.

We make the following assumptions:

¹⁷ We assume that V_i is increasing and concave in x , concave in w . Moreover, we assume for simplicity that weight $w_i(x)$ increases linearly with calorie consumption. This means that

$$v_i''(x) = \frac{\partial^2 V_i}{\partial x^2} + \left(2 \frac{\partial^2 V_i}{\partial x \partial w} + \frac{\partial^2 V_i}{\partial w^2} \right) w_i'(x),$$

and we assume that the former is negative, so $v_i(x)$ is concave. This would be true if, for example, the marginal increase in weight with respect to calories, $w_i(x)$, is sufficiently small in comparison to the concavity of V_i , measured by $\partial^2 V_i / \partial x^2$.

¹⁸ We assume that the individual’s utility is quasilinear in money. Therefore, as long as the price $p(x)$ is small compared with the individual’s wealth, we can ignore her budget constraint.

1. Choice x represents a number of calories, so individuals are indifferent between foods with the same number of calories.

2. Individuals have a larger marginal internalty without the nudge, $\frac{\partial \Lambda_i(x|e_0)}{\partial x} > \frac{\partial \Lambda_i(x|e_1)}{\partial x}$ for all $x \in X$, and we assume $\Lambda_i(0|e) = 0$ for $e \in \{e_0, e_1\}$.

3. The nudge has a psychological cost: $c_i(x|e_0) = 0$ for all x , $c_i(x|e_1) \geq 0$, $\frac{\partial c_i(x|e_1)}{\partial x} > 0$ for all $x \in X$.

Let x_i^j be defined as the solution of the problem when the individual uses decision utility in environment e_j ,

$$x_i^j = \arg \max_{x \in X} v_i(x) - c_i(x|e_j) - p(x) + \Lambda_i(x|e_j) .$$

Observation: $x_i^1 < x_i^0$, i.e., the nudge reduces calorie consumed. This is so because, for all x , we have that $\Lambda_i(x|e_1) \leq \Lambda_i(x|e_0)$ and $c_i(x|e_1) \geq c_i(x|e_0)$. Therefore, a higher choice of x yields less decision utility and has a higher psychological cost under the nudge, so the optimal choice of x must be lower in that case.

The welfare gain/loss from the nudge for individual i is given by

$$\Delta W_i = u_i^{exp}(x_i^1|e_1) - u_i^{exp}(x_i^0|e_0) = v_i(x_i^1) - v_i(x_i^0) - [c_i(x_i^1|e_1) - c_i(x_i^0|e_0)] - [p(x_i^1) - p(x_i^0)].$$

Proposition 1 (Jimenez-Gomez, 2019). *A sufficient condition for the nudge to increase welfare for individual i is*

$$\Lambda_i(x_i^0|e_1) - \Lambda_i(x_i^1|e_1) \geq c_i(x_i^0|e_1). \quad (1)$$

A sufficient condition for the nudge to decrease welfare for individual i is

$$c_i(x_i^1|e_1) \geq \Lambda_i(x_i^0|e_0) - \Lambda_i(x_i^1|e_0). \quad (2)$$

So far, we have described the model using a representative individual. However, heterogeneity in the population is crucial when considering behavioral welfare economics (Allcott and Taubinsky, 2015; Taubinsky and Rees-Jones, 2018; Allcott and Kessler, 2019). Suppose that the population can be divided in two groups, H and L (for high and low information). Suppose, moreover, that individual $h \in H$ has no internality in her decision-making, because she has access to all the information and cognitive resources she needs, so $\Lambda_h(x|e_0) = \Lambda_h(x|e_1) = 0$ for all $x \in X$. Then, from Equation (2), we find that the nudge decreases her welfare as long as there is some psychological cost, since the right-hand side is 0.

On the other hand, suppose that an individual $l \in L$ has a large internality in the absence of the nudge, but the nudge reduces the internality as well as the consumption of calories, so that $\Lambda_l(x_l^0|e_1) - \Lambda_l(x_l^1|e_1)$ is large.¹⁹ Then, from Equation (1), we have that the nudge will increase her welfare as long as the psychological cost is not too large.

Note that the comparisons between the internality and the cost are relative to x . For small x (i.e., foods which are low in calories), we expect *both* the internality and the psychological cost to be low; the opposite is true for high-calorie foods. If x_l^0 , the choice in the absence of the nudge was a high-calorie food, then from Equation (1), the nudge will increase the welfare of individual l if the psychological cost for eating high calorie foods is smaller than the reduction in the internality for those high calorie foods.

Predictions. From the previous arguments, we have the following predictions:

1. Everybody (both in H and L) will decrease calorie consumption;
2. Those in H will have a welfare loss from the interventions; and

¹⁹ Because the nudge reduces the internality and increases psychological/moral costs, we have that $x_l^0 > x_l^1$.

3. The welfare effect in L is ambiguous, and it depends on the relative size of the externality and the psychological cost associated with the nudge.

The model we have described is quite general, and our results are consistent with any situation that fits the described assumptions.²⁰ In terms of welfare, the argument for critics of nudges usually rests in Equation (2): when the psychological cost is high, and people are already informed and make unbiased decisions, the nudge can only decrease welfare. However, as importantly, Equation (1) shows that for those who do not make unbiased decisions (because they suffer from internalities), the nudge can be a source of welfare. The nudge will be more likely to increase social welfare when the internalities of the L group are larger, when the relative size of the L group with respect to group H grows, and when there were negative externalities derived from consumption. It will be more likely to decrease social welfare when the psychological costs (for either group) are larger (we derive this formally in Appendix A).

IV. Data

Identification of treatment and control areas

Our identification strategy involves comparing counties that implemented calorie labeling laws to proximate areas that did not implement such laws. Our main analysis sample therefore includes the Metropolitan Statistical Areas (MSAs) of New York City (the “treated” counties are the five counties of New York City, Suffolk County, and Westchester County; the control counties are the other counties in the area), Philadelphia (treated county is Philadelphia), Seattle (treated county is King), Washington DC (treated county is Montgomery), and Albany (treated counties

²⁰ We provide the analysis of a scenario with the possibility of learning about calories for groups with different levels of bias in Appendix A.

are Albany and Schenectady) as well as the entire census division of New England (treated state is Vermont).^{21,22} We excluded observations from localities where menu labeling laws were passed but, due to anticipated preemption by the ACA, were never implemented or enforced. We do this because it is unclear that these areas can truly be considered control units since some restaurants may have started posting calories in anticipation of the laws taking effect. Within our geographic areas of interest, the dropped areas include Maine, Massachusetts and New Jersey as well as Nassau County in New York.

While using proximate geographic areas as control groups increases the likelihood that the treatment and control groups experience similar local economic conditions, one concern with this identification strategy is that restaurant chains may voluntarily post calories on menus. As a result, restaurants in the control group areas may display calorie counts on menus even though they are not required to do so. Thus, individuals in our control group areas may be exposed to calorie labels when eating at select chains regardless of location. Additionally, customers who reside in the control group areas may dine at restaurants in treated regions and, thus, be exposed to calorie counts. Alternatively, customers in the treated and control group regions may be unaware of calorie counts on menus when selecting their meals.

To the extent that these concerns are valid, they should serve to minimize differences across treated and control areas and attenuate the estimated effects. In this regard, our estimates should

²¹ Ulster County, NY also passed, implemented, and enforced a menu labeling law. However, we exclude Ulster County from the sample because it is the only county within its MSA, meaning that there are not similar controls as the other treated counties. The results are robust to including Ulster County as a treated county, without any corresponding control counties or also including adjacent Orange County, NY as an additional control county.

²² The sample spans from 1994 through 2012 for the analysis of BMI and 2005 through 2010 for the analysis of life satisfaction. The Washington DC MSA and census division of New England are excluded from the life satisfaction analysis because there are no treated observations in those areas between 2005 and 2010 but are included in the analysis of BMI because there are treated observations in 2011 and 2012. Specifically, the treated county in the Washington DC MSA – Montgomery County, MD – implemented its law in July 2010, and its 2010 BRFSS respondents were all surveyed before then. The treated state in New England – Vermont – implemented its law in January 2011.

be considered intent-to-treat effects and, hence, a lower bound on the impacts of calorie labels on food choices. To understand the severity of these concerns, we examine differences in compliance and awareness using two supplemental data sources.

To address the first concern, we called approximately 17,000 restaurants in our sample areas to inquire whether they were currently posting calories on menus. These restaurants encompassed all locations of the top 500 chains by sales volume with at least 15 locations nationwide.²³ The calls were made in early spring 2017 prior to the implementation date (May 2018) of the federal calorie mandate. Of the 2800 restaurants in the regions that were actively enforcing local mandates (New York City, Westchester County, Suffolk County, Philadelphia County, Schenectady County, and Montgomery County), approximately 82.9% of all establishments were compliant with the local mandate and posting calories on menus and menu boards. Of the 1300 restaurants in the regions that had mandates but were not actively enforcing them (King County, Albany County, and Vermont), approximately 73.4% of all establishments were posting calories. Finally, of the 13,000 restaurants in the control regions, nearly 57.9% of all establishments were voluntarily posting calories on menus.

Viewed in its totality, the call data suggest that, while a substantial fraction of restaurants are voluntarily providing caloric information on menus, there are significant differences in the prevalence of calorie posting across regions. This suggests that restaurants are responding to the stringency of local mandates, which could introduce heterogeneity in the likelihood consumers observe calorie counts across these different regions. Moreover, these differences likely understate the differences that were present during our sample period, which ends in 2012. No chain

²³ The top 500 restaurant chains by sales volume are drawn from the 2016 Technomic Top 500 Chain Restaurant Advance Report. Given that many of the chains outside of the top 400 by sales volume have less than 15 locations, these 500 chains likely encompass nearly all restaurants required to post calories by the local mandates.

restaurants voluntarily posted calories on menus in all their locations until the very end of our sample period, whereas several did by the time of our phone calls in 2017.

To examine whether consumers observe calorie counts on menus, we surveyed approximately 4000 adults in December 2016 throughout our regions in collaboration with Qualtrics Panels to construct a representative sample on age, gender, and education. We asked consumers to describe whether they had noticed calorie counts on the menu or menu board of every meal they had eaten at a restaurant during the previous 24 hours. In regions that were actively enforcing local calorie-posting mandates, 44.0% of respondents reported that they had noticed calorie counts on the menu. In regions that were not actively enforcing local mandates, 39.2% of respondents noticed calorie counts on menus, and in regions that did not have a local mandate, 35.6% of respondents reported that they had such information. It is important to note that these percentages are based on the home residence of the respondent, not the location of the restaurant where they had eaten; hence, it is possible that respondents living in the control group regions observed calorie counts while eating at a restaurant in the treated region. Despite this possibility, we still see that individuals residing in areas actively enforcing calorie-posting requirements were 24% more likely to have observed calorie counts when eating away at a restaurant than were counterparts residing in control areas.

These survey results suggest that a substantial fraction of consumers recalled observing calorie counts when dining at restaurants. While consumers who reside in the control group areas were exposed to calorie counts, consumers in the treatment regions were more likely to notice calorie counts on menus when dining at restaurants.²⁴ Given that heterogeneity in the enforcement

²⁴ These results likely understate the extent of heterogeneity in our sample period as consumers in the control group became more likely to observe calorie counts over time. The survey in December 2016 provides an estimate of consumer awareness of calorie posting during a time period that corresponds to the spring 2017 estimates of voluntary posting of calories by restaurants. In an earlier wave of the survey in December 2014, 44.1% of

of local calorie-posting requirements corresponds to differences in the likelihood that restaurants post calories on menus and heterogeneity in the likelihood that consumers residing in treatment and control groups observe calories on menus, we estimate both the average impact of calorie-posting laws and the differential impact of these laws based on the stringency of local enforcement.

BRFSS Data and Outcome Measures

Our primary data source is the Behavioral Risk Factor Surveillance System (BRFSS) for the period between 1994 and 2012. The BRFSS is a random-digit-dialing telephone survey conducted by state health departments and the Centers for Disease Control and Prevention (CDC). The BRFSS collects information about adults' health, health-related behaviors, and demographic characteristics. Importantly for our purposes, the BRFSS also includes state and county identifiers; however, county identifiers are not publicly available after 2012. In addition, as chain restaurants began voluntarily posting calorie counts on menus in all their locations in 2012, ending our sample in 2012 minimizes the likelihood of attenuation bias driven by exposure to calorie counts in restaurants voluntarily posting such information in control areas.

We construct our main analysis sample as follows. First, pooling all observations from 1994-2012 yields 5,232,786 respondents. Restricting geography to the treatment and control areas leaves 614,179 respondents. From this sample, we exclude 5,608 pregnant women and 14,207 individuals with missing information. The final sample consists of 594,364 individuals.

consumers in areas that actively enforced local mandates noticed calorie counts in restaurants, which is the same percentage as December 2016. However, fewer percentages of respondents noticed calorie counts in other regions. In December 2014, 36.2% of respondents in areas with local mandates that were not actively enforced noticed calorie counts, and 30.4% of respondents in areas without a local mandate noticed calorie counts in restaurants. Further, our BRFSS sample ends in 2012, when chain restaurants first started voluntarily posting calorie counts, so the percent of consumers in the control group who observed calorie counts in restaurants is likely much lower during the sample period.

The theoretical model from section III predicts impacts on calorie consumption and utility. While we do not observe these outcomes directly, we use BMI and life satisfaction as proxies for these outcomes in our empirical analysis.²⁵ BMI is computed based on self-reported measures of height and weight, as weight in kilograms divided by the square of height in meters.²⁶

Given that the BRFSS data do not include food consumption or calories consumed, we use information on other health-related behaviors – smoking, alcohol consumption, and exercise – that could influence BMI absent changes in calories consumed. Examining these outcomes allows us to determine whether any observed changes in BMI reflect changes in our unobserved measure of interest (calories consumed) or a broader array of behavioral change. Smoking is measured as whether the individual currently smokes cigarettes. Alcohol consumption is measured as the number of drinks consumed during the past 30 days. The physical activity outcomes are any exercise in the past thirty days besides a regular job (available in all waves), the average minutes of moderate exercise per week over the past month (collected only in odd years between 2001 and 2011), and the average minutes of vigorous exercise per week over the past month (odd years between 2001 and 2011).²⁷

²⁵ Empirically, using caloric intake as an outcome is not feasible since the only government survey dataset with a comprehensive food diary – the National Health and Nutrition Examination Surveys (NHANES) – is not nearly large enough and does not have the geographic scope to precisely identify the impact of this intervention. The predicted effect of calorie labeling on BMI is of the same sign (negative) as the predicted effect on calorie intake as long as calorie expenditure is unaffected.

²⁶ Self-reported weight and height admittedly suffer from measurement error, and Cawley (2004) and Courtemanche et al. (2015b) develop “corrections” that adjust self-reports based on the comparison of actual and self-reported values in the NHANES. We elect not to implement these corrections in this paper because the sample period does not correspond in a straightforward way to the NHANES survey periods, and assumptions made about how to reconcile the difference could potentially induce more error than the correction solves. Further, researchers studying determinants of obesity have repeatedly found that implementing such corrections does not substantively impact results (Courtemanche et al., 2015b; Courtemanche et al., 2016). Finally, in our particular context, the nature of reporting error is not likely to differ across treatment and control areas.

²⁷ Restrepo (2017) measures physical activity as any exercise in the past 30 days and the number of minutes of any form of exercise in the past 30 days; he finds no impact of the New York City law on physical activity.

The BRFSS data include a question about life satisfaction, which has been used in prior work to evaluate the effects on consumer well-being of policy changes such as German reunification (Frijters et al., 2004), cigarette taxes (Gruber and Mullainathan, 2005), and the provision of improved housing for the poor (Galiani et al., 2018). While the use of self-reported life satisfaction is not beyond reproach (Deaton and Stone, 2016), it has been proposed as a useful proxy for experienced utility (Kahneman and Sugden, 2005) and is negatively correlated with more objective measures such as demand for antidepressants (Blanchflower and Oswald, 2016).^{28,29}

The BRFSS collected information on life satisfaction in the 2005 through 2010 waves. While we are therefore only able to use those years to examine the impact on life satisfaction, this period still allows for identifying variation from all the calorie labeling laws except for Montgomery County, MD (which was implemented in July 2010) and Vermont (which was implemented in 2011). The life satisfaction question's choices are very satisfied, satisfied, dissatisfied, and very dissatisfied. We use the same ordinal definition in our analyses.

The BRFSS data also include a number of demographic and socioeconomic characteristics of the respondents, such as race/ethnicity, educational attainment, age, gender, and marital status. As these variables are correlated with our outcomes of interest, we incorporate them into our empirical model described below. In addition, as the BRFSS data include county of residence, we are able to construct county-level socioeconomic characteristics using data from the U.S. Census Bureau and Bureau of Labor Statistics that could also be correlated with our outcomes of interest.

²⁸ There is evidence that self-reported life satisfaction is correlated with experienced utility. For example, Perez-Truglia (2015) provides evidence from a validation test that expenditure shares predicted by life satisfaction questions are highly correlated with changes in actual expenditure shares on different categories of food consumption. Similarly, Benjamin et al. (2012) show that stated measures of subjective well-being on life satisfaction are a powerful predictor of stated choices, what is to be expected if life satisfaction is a good proxy for experienced utility.

²⁹ We would prefer to have a broader set of questions upon which to build our measure of well-being and recover associated measures of relative marginal utility, as in Benjamin et al. (2014). Unfortunately, our data only include a single question on life satisfaction, so we are unable to develop a more comprehensive measure.

V. Effects on Body Mass Index

Our theoretical model predicts that all consumer types will reduce calories consumed when exposed to calorie counts on menus. As noted earlier, the BRFSS data do not include information on the amount of calories consumed. To test this hypothesis, we instead rely on an alternate measure, BMI, which is directly impacted by calories consumed. Specifically, our model predicts that we should observe reductions in BMI in treated areas relative to control areas following the implementation of local mandates. We evaluate this prediction by estimating the average effect of chain restaurant calorie posting mandates on BMI using a difference-in-differences design.

In doing so, we estimate the following empirical model:

$$Y_{ict} = \beta_0 + \beta_1 LAW_{ct} + \beta_2 \mathbf{X}_{ict} + \beta_3 \mathbf{Z}_{ct} + \alpha_c + \tau_t + \gamma_{ct} + \varepsilon_{ict}, \quad (3)$$

where Y_{ict} is the outcome variable (BMI) for individual i in county c in time period t . Since survey data are collected monthly and we have the exact month and year when the laws were implemented, the time unit is “month-year” (e.g., April 2008). The treatment variable is defined as LAW_{ct} and is a dichotomous variable equal to one whenever a chain restaurant calorie labeling law is implemented in county c in month-year t . The associated coefficient β_1 measures the average treatment effect of a menu labeling law on the outcome variable and provides a test of our first hypothesis that the introduction of calorie labels leads to reductions in BMI.

We control for a number of demographic and socioeconomic characteristics of individuals, denoted by \mathbf{X}_{ict} , including race/ethnicity, educational attainment, age, gender, and marital status. The race/ethnicity variables are indicators for whether the individual is non-Hispanic white (the reference), non-Hispanic black, Hispanic, or other race/ethnicity. Educational attainment is characterized by four mutually exclusive categories for whether the individual did not obtain a

high-school degree (the reference), obtained a high-school degree, attended college but did not graduate, or obtained a college degree. We include dummy variables for each individual year of age. We also use indicator variables for females and married individuals.

County level controls (\mathbf{Z}_{ct}) include continuous variables for annual population density, monthly unemployment rate, and annual median household income. The population density measure is constructed as the number of people per square mile using U.S. Census Bureau information on population estimates for each year and land area estimates from 2000. We obtained data on median income estimates for each year from the Census Bureau and local area unemployment rates for each month from the Bureau of Labor Statistics.

Equation (3) also includes county fixed effects α_c , month-year fixed effects τ_t , and county-specific linear time trends $\gamma_c t$. The last term, ε_{ict} , is an idiosyncratic error term. Standard errors are clustered at the county level and are adjusted for heteroscedasticity. There are 110 counties in our sample so concerns about inference with a small number of clusters should not apply in our case (Cameron and Miller, 2015). We utilize sampling weights in all analyses. The BRFSS did not start to incorporate cell phones into its sampling frame until 2011, and the sampling weights are designed to adjust for the resulting changes in the composition of the sample.

In sum, our identification strategy accounts for an array of observable characteristics as well as unobservable characteristics that are constant within counties, constant within time periods, or change linearly within counties. In other words, $\hat{\beta}_1$ gives the causal effect of menu labeling laws under the relatively weak assumption that the implementation of such laws is uncorrelated with unobservable county characteristics that change non-linearly.

To examine the plausibility of this identification strategy, we examine whether the characteristics of individuals in the treated and control group regions are similar prior to the

implementation of the first calorie-posting requirement and also whether the trends in BMI are parallel for these two groups during the pre-treatment period. We additionally conduct a series of robustness checks, including an event study analysis, and falsification tests that provide further support of the plausibility of this identification strategy; such results are reported in Appendix B. Table 2 presents summary statistics for the outcome variables and covariates from the pre-treatment period (1994-2007) for the treated and control groups. The weight outcomes of the treatment and control groups were very similar prior to the implementation of calorie labeling mandates: for both groups, average BMI was around 26.1 kg/m² while 36% were overweight but not obese ($25 \leq \text{BMI} < 30$), 12%-13% were Class I obese ($30 \leq \text{BMI} < 35$), and 6% were Class II/III obese ($\text{BMI} \geq 35$). Pre-treatment life satisfaction was somewhat lower in the treatment group than the control group, though our identification strategy will be able to control for such baseline differences through the inclusion of county fixed effects and county-specific linear time trends. The treatment and control groups were broadly similar along the dimensions of education, age, and gender, but the control group was somewhat wealthier, had a smaller proportion of racial/ethnic minorities, and lived in less densely populated counties. Again, our identification strategy will be able to account for these differences through the inclusion of county fixed effects and county-specific linear time trends.

Figure 1 presents annual changes in average BMI in the treatment and control locations during our sample period. The vertical line, drawn between 2007 and 2008, reflects the timing of the first calorie labeling law, implemented by New York City in 2008. As can be observed in the figure, the treatment and control groups exhibited very similar levels and trends in average BMI in the pre-treatment period. Thus, Figure 1 provides visual evidence consistent with the identifying

assumption and suggests that the individuals in the control group provide a valid counterfactual for the outcomes of individuals in the treatment group.

Additionally, as shown in Figure 1, BMI in the treatment group drops in 2008 – right after the implementation of the New York City law – and then stabilizes throughout the rest of the post-treatment period. In contrast, BMI in the control group continues on its steady upward trajectory during the post-treatment period. This graph therefore provides evidence that menu labeling laws may decrease BMI. To more formally assess the impact of the New York City law and subsequent local mandates, we next report estimates from the difference-in-differences design of equation (3).

Estimation results for BMI are presented in Table 3. We only report the coefficient estimates for β_1 . We find that calorie posting laws lead to a decrease in BMI of 0.17 kg/m² on average, as shown in column (1). This change in BMI corresponds to a 1.5-pound reduction in weight at the mean height of the sample and a reduction of 0.7% relative to the mean of the treatment group prior to the implementation of calorie posting laws.

We show that these results are robust to utilizing alternate control variables, time periods, and control group definitions in Appendix B. Finally, in a series of falsification tests, we create a series of placebo variables as if the laws occurred in years before they were actually implemented and find that there is no influence of the placebo calorie mandates. Importantly, these falsification tests provide evidence consistent with the identifying assumption of equation (3), which is that there is a common counterfactual trend in the dependent variable between the treated and control groups.

To better relate the average treatment effects reported in column (1) to the prevalence of calorie labels on menus reported in chain restaurants, we augment equation (3) to include variables denoting that the law was passed but not yet implemented and that the law was implemented but

is not actively enforced. The first indicator allows us to examine anticipatory effects from restaurants posting calories on menus before the local mandates were implemented.³⁰ The second indicator allows us to examine whether the influence of calorie posting laws depends on the stringency with which treated regions enforce local mandates. As noted earlier in section IV, the likelihood that consumers observed calorie counts on menus was lower in areas that did not actively enforce local mandates. This later variable is equal to one in the regions of King County after May 2010, Albany County after September 2010, and Vermont; the coefficient estimate provides evidence of whether the influence of calorie posting laws depends on the intensity of enforcement.

Results from column (2) of Table 3 provide evidence of the importance of controlling for the stringency of enforcement. While our estimates show that actively enforced laws reduce BMI by 0.19 kg/m², the estimated effect for laws in regions that are enacted but not actively enforced is an increase of 0.13 kg/m², which is not statistically significant at any meaningful level. The estimated effect on the indicator for passage of the local law is near zero and positive (0.001 with a standard error of 0.099), which rules out that the average treatment effect reported in column (1) reflects anticipatory changes.

Based on the totality of the findings, the results in column (2) suggest that the average treatment effect is driven by active enforcement of local mandates and hence likely reflects exposure to calorie labels on menus in chain restaurants rather than other factors related to the passage and implementation of local mandates. To further explore this conjecture, we examine the impact of menu labeling laws on smoking, drinking, and physical activity, as these behaviors could also influence BMI in the absence of changes in calories consumed. These results are shown

³⁰ Anticipatory effects could also occur due to public health messaging and information campaigns surrounding the passage of these mandates (Rees-Jones and Rozema, 2020).

in Table 4. Across all of these outcomes, the estimates of the impact of implementation of the law are not statistically significant and small in magnitude relative to the sample mean. This reinforces our interpretation of the estimates in Table 3 that the changes in BMI reflect changes in calories consumed due to exposure to calorie labels on menus.

Before proceeding, it is interesting to note that our estimated effects for changes in BMI are significantly lower than what would be predicted based upon experimental estimates of calorie labels on calories consumed at restaurants (Wisdom et al., 2010; VanEpps et al., 2016; Cawley et al., 2018). For comparison, Wisdom et al. (2010) find that including calories and daily calorie recommendations on menus in a fast-food restaurant decreased the amount of calories ordered in a meal by 99 calories, which is predicted to decrease BMI by 0.75 kg/m^2 within one year.³¹ VanEpps et al. (2016) find that including calories on online menus for a large corporate cafeteria led to a 60 calorie reduction in the number of calories ordered, which is predicted to decrease BMI by 0.47 kg/m^2 within one year. Cawley et al. (2018) find that including calories on the menu in a full-service restaurant reduced the number of calories ordered by 45 calories, which is predicted to decrease BMI by 0.35 kg/m^2 within one year.³²

Viewed through the lens of recent work on scalability, however, such divergence is not particularly surprising (Al-Ubaydli et al., 2017a; Al-Ubaydli et al., 2017b). By design of the local mandates, individuals have the ability to avoid calorie labels by choosing to eat at restaurants that are not required to post calories on menus. To the extent that calorie labels trigger guilt/impose a psychic cost on the consumption of high calorie foods, we would expect some individuals to switch

³¹ Menu-labeling laws require both providing calorie information and the daily calorie recommendation on menus. Providing calorie information only led to a decrease of 61 calories ordered in a meal, which is predicted to decrease BMI by 0.46 kg/m^2 within one year (Wisdom et al., 2010).

³² The predicted decrease in BMI within one year is based on the estimated decrease in calories ordered in Wisdom et al. (2010), VanEpps et al. (2016), and Cawley et al. (2018); the baseline height in each sample; and the result in Hall et al. (2011) that a decrease of 10 calories per day will eventually lead to weight loss of one pound, with half of the weight change occurring within one year.

where they eat to avoid exposure to calorie labels.³³ Such sorting would attenuate the effect of the mandate by limiting the number of meals purchased at restaurants affected by the laws. Moreover, our data suggests imperfect compliance as many chain restaurants – particularly those in treated areas that did not actively enforce the local mandate – report that they do not post calories on menus. Given results in Table 3 that there was no impact on BMI in regions where the local mandate was not actively enforced, we would expect the average effect of calorie labels in our sample to be less pronounced than what would be inferred from experimental work that restricts attention to calories consumed in a restaurant that posts calories on menus.

VI. Understanding Why Calorie Labels Matter

The prior section shows evidence consistent with our theoretical prediction that calorie labels reduce BMI and calories consumed for the average consumer. To better ascertain why we observe this effect, we return to the predictions of our theoretical model, which suggest two channels through which calorie labels could impact choice: they provide information which reduces the marginal internality for low information individuals and they may impart psychological or moral costs on all decision makers.

As highlighted in our theory, the key to understanding the relative importance of these different channels is to explore the effects of calorie labels on the well-being of low and high information types. Specifically, our theory predicts that the effect for high information types is unambiguously negative provided that calorie labels trigger psychological or moral costs. In

³³ We are not the first to note the role of sorting to avoid social pressures or moral costs. Prior work in charitable giving and sharing in dictator games has shown how the ability to avoid the ask (or opportunity to share) impacts subsequent rates of giving and the interpretation of motives for doing so (Andreoni et al., 2017; DellaVigna et al., 2012; Lazear et al., 2012). Unfortunately, we do not have data on where individuals eat and if they attempt to avoid chain restaurants that post calories on menus and menu boards.

contrast, the welfare effect on low information types is ambiguous as it depends on the size of the reduction in the marginal internality and the psychological costs associated with the calorie nudge.

These predictions provide a way to establish the existence and importance of moral costs. If calorie labels work solely through the information channel, we would expect the average effect on life satisfaction to be unambiguously positive as the labels would have no impact on high information types and an unambiguously positive impact on low information types by reducing the marginal internality. In contrast, if calorie labels trigger moral or psychological costs, the effect for high information types is unambiguously negative while the effect for low information types is ambiguous. Hence, the prediction for the average effect is ambiguous and depends on the relative portion of each type and the relative strength of the competing effects for low information types.

We do not directly observe any individual's type; however, we can draw upon prior literature to infer types of consumers who are more likely to have low levels of information. In particular, the review article of Wehling and Lusher (2017) documents that individuals with BMI greater than or equal to 30 (in the obese range) systematically underreport calories consumed in meals, while there is no such pattern for individuals in healthy weight ranges.³⁴ We thus use observed BMI as a proxy for individuals' types when exploring the effects of calorie labels on life satisfaction.

To do this, we estimate ordered probit models since life satisfaction is an ordinal outcome with responses on a 4-point Likert scale ranging from very dissatisfied to very satisfied. Since the sample period for life satisfaction is much shorter than the period for BMI (2005-2010 rather than

³⁴ While it may be tempting to use additional demographic characteristics to define types, Wehling and Lusher (2017) document that there is no consistent pattern of differences in beliefs across other demographic characteristics such as race, age, and gender.

1994-2012), we omit county-specific time trends in our baseline specification and instead show the sensitivity of the results to their inclusion in a robustness check, which we show in Appendix B. The other robustness checks, also in the appendix, are similar to those used for BMI, with the only exception being that we do not vary the length of the pre-treatment periods since utilizing a lengthy pre-treatment period is not feasible for life satisfaction.

In the upper panel of Table 5, we document the estimates from ordered probit model for each life satisfaction category for the full sample. Results from this exercise provide evidence that the average consumer is worse off following the implementation of the local calorie mandates.³⁵ For the full sample, the implementation of the laws decrease the probability of being very satisfied with one's life by 2.6 percentage points, which corresponds to 6.3% of the mean. At the other end, calorie-posting requirements increase the probability of reporting being very dissatisfied by 0.2 percentage points and dissatisfied by 0.6 percentage points. For perspective, these effects correspond to an approximate 17% and 12% increase relative to the means for these categories. Further, calorie posting-requirements increase the probability of reporting being satisfied by 1.8 percentage points, corresponding to an increase of 3% relative to the baseline means.

Considering the estimates for all categories jointly, these results suggest that, on average, individuals are less satisfied with their life following the implementation of calorie mandates, as the distribution of life satisfaction shifts to the left. To better ascertain the magnitude of these changes, our estimated effects are equivalent to those estimated for small changes in air pollution (Levinson, 2012) or cigarette taxes (Gruber and Mullainathan, 2005). However, the estimates are substantially smaller than the effects of variation in daily traffic congestion in Beijing, China

³⁵ Because the sample ends in 2010 for life satisfaction, which limits the variation across regions in the stringency of enforcement, we do not examine the impact of calorie-posting requirements on life satisfaction by the stringency of the enforcement of the local laws.

(Anderson et al., 2016) or improved housing for low-income households in South America (Galiani et al., 2018). The results are again stable across the various specifications, with the estimates from all robustness checks falling within a standard error of the baseline estimate.³⁶

While the effects for the full sample are suggestive of the presence of moral costs, a sharper test of our theoretical predictions is based upon heterogeneity of the effects across high and low information types. The bottom two panels of Table 5 provide results for this exercise, where we proxy for information types using weight status. The estimates for individuals in the healthy weight category (high information types) are presented in the second panel and reinforce the notion that calorie labels impart moral costs. For example, we find a 6.4 percentage point reduction in the likelihood that an individual residing in the treated areas in the post-treatment period reports being very satisfied with their life. Further, we find a 0.4 percentage point increase and a 1.4 percentage point increase in the likelihood of reporting being very dissatisfied and dissatisfied with their life due to calorie-posting requirements. Estimates for overweight and obese individuals are not statistically significant for any life satisfaction category and are uniformly smaller in magnitude relative to the corresponding estimates for healthy weight individuals.

A potential concern is that estimates of the impact of calorie-posting laws conditional on current weight status may be biased if treatment impacts weight status. To assess the plausibility of this concern, we estimate ordered probit models analogous to equation (3) to determine whether calorie-posting requirements change the probability that an individual is classified as healthy weight, overweight, or obese. We find that the estimates for healthy weight, overweight, and

³⁶ Given the relatively short length of the pre-treatment period in the life satisfaction regressions, we conduct only one falsification test in which the sample period consists of 2005-2007 and the placebo intervention occurs midway through this period (July 2006) for all the treated areas. The estimated “effects” of the placebo calorie labeling laws on the probabilities of being in each of the life satisfaction states are small and statistically insignificant.

obesity are small in magnitude and not statistically significant, as shown in Appendix Table A6.³⁷ Hence, it does not appear as if our classification of individuals as high or low information types is itself a function of exposure to calorie-posting laws.

Another potential concern with the interpretation of these results is that there may be differences in the likelihood that overweight and obese individuals observe calorie counts on restaurant menus relative to healthy weight individuals. If so, our results could reflect differential exposure to calorie counts rather than differences in the channels through which calorie labels impact choice across these types. To assess the plausibility of such concern, we return to the survey data of restaurant consumers initially discussed in section IV. As the survey included information on height and weight for the respondent, we are able to examine if there are differences across types in the likelihood that consumers reported noticing calorie counts at restaurant meals in the previous 24 hours. The data suggest that there are no such differences; calorie counts were noticed by 38.6% of healthy weight individuals, 37.4% of overweight individuals, and 35.4% of obese individuals.

It is thus unlikely that our results reflect endogenous sample selection or differential exposure to calorie counts across consumer types. Rather, our results suggest that both information and moral costs are important channels through which calorie labels influence consumer choice and, hence, BMI. For high information types, we observe reductions in life satisfaction, which is consistent with the theoretical role of moral costs. For low information types, we observe no change in life satisfaction, suggesting reductions in the marginal internality and a compensating increase in moral costs.

³⁷ The weight loss corresponding to the average decrease in BMI shown in Table 3 is 1.5 pounds. The difference between the overweight and obesity thresholds of 5 BMI units corresponds to around 30 pounds at the sample mean height. Thus, it is not surprising that the estimates of the impact of calorie-posting requirements on overweight and obesity are small in magnitude and not statistically significant.

An alternate way to identify the importance of moral costs would be to return to the analysis of BMI to consider the effects across the distribution of consumer types using quantile regression. Recall that our theoretical model predicts a reduction in BMI across all consumer types, if calorie labels trigger moral costs. The key to testing this prediction is to examine changes in BMI for individuals in the healthy weight region of the distribution of BMI prior to calorie labels (high information types) for whom reductions in BMI should only be observed if there are moral costs. In particular, high information types (i.e., individuals with BMI < 25) are contained within the first five quantiles.

Results from the quantile regressions shown in Table 6 highlight two important patterns. First, we observe statistically significant reductions in BMI for high information consumers in quantiles 0.2 and 0.3, reinforcing the importance of moral costs identified in our estimates on life satisfaction. Second, we observe the largest point estimates in the 0.9 quantile, corresponding to low information consumers (i.e., obese individuals with BMI ≥ 30). This suggests that both moral costs and information channels are likely influential for such types.³⁸ Further, these results reinforce the interpretation of our findings for life satisfaction on the importance of moral costs as a channel through which calorie labels influence choice.

VI. Conclusion

In an attempt to provide information to consumers at the point of purchase, policymakers have introduced laws requiring chain restaurants to post calorie counts on menus. To date, research exploring such laws has focused on identifying their impact on calories consumed and/or BMI. In

³⁸ We note that the estimate in the 0.9 quantile is not statistically significantly different from zero, nor is this estimate significantly different from the estimates at any other quantile. However, it is important to note that there is large variability in BMI in this quantile relative to other quantiles; hence, given our estimated effect size, the power of tests in this quantile is low.

this paper, we examine a new question – why such laws influence consumer choice. To do so, we develop a theoretical model through which the information on restaurant menus reduces internalities and influences choice. Our model demonstrates another channel through which calorie counts influence choice – by increasing moral costs, which leads consumers to decrease calories consumed but also reduces welfare for these consumers. Importantly, we use the model to develop testable predictions that allow us to uncover the presence and importance of each channel.

We test these predictions by combining BRFSS data on BMI and life satisfaction with other secondary data sources. Specifically, we explore the impact of city, county, and state calorie posting requirements on BMI and life satisfaction by comparing changes in these outcomes over time in geographic areas that implemented such requirements to changes in the surrounding geographic areas using a difference-in-differences design. Lending credibility to this design, we show that the trends in BMI are similar in these areas prior to the enforcement of any law. Moreover, we present evidence from supplemental survey data showing that restaurants in the treated areas are more likely to report posting calories on menus and consumers in treated areas are more likely to report noticing calories on menus.

We find that calorie-posting laws lead to small reductions in average BMI and that such effects reflect both moral costs and responses to more salient information. The importance of moral costs is identified through two pieces of evidence: weight loss and reductions in life satisfaction for individuals in the healthy weight range of BMI (i.e., the high information types in our theoretical model). The importance of information comes from the results that we find reductions in BMI but no change in life satisfaction for individuals in the obese category (i.e., the low types in our model).

From a policy perspective, the estimated 0.17 kg/m^2 reduction in average BMI, which corresponds to a decrease of about 1.5 pounds at the mean height of the sample, is unlikely to impact health conditions related to BMI and obesity. Moreover, our estimated effects are smaller than what would be predicted if we were to extrapolate evidence from field experiments exploring the impact of calorie labels on the choices of consumers in restaurants. Yet, insights from the literature on scalability helps to reconcile these differences. Since calorie labels introduce moral costs, some consumers may change where they eat to avoid exposure to such labels and restaurants may choose to not comply with the law. Both effects would serve to attenuate the impact of calorie-labeling requirements on BMI relative to what would be inferred from field experiments that do not allow for sorting or imperfect compliance.

Our results are consistent with the notion that “nudges” may be welfare enhancing for some individuals but welfare reducing for others. We are not the first to demonstrate such heterogeneous welfare effects for policies motivated by behavioral insights. Similar welfare effects have been explored in the context of the introduction of active choice for health insurance plans (Handel, 2013) and the receipt of monthly home energy reports (Allcott and Kessler, 2019). Yet, one should not use the possibility that calorie labels may have adverse welfare effects for certain types of individuals to rule out the use of such nudges. Such an approach would be akin to advocating doctors forgo prescribing any medication that could have adverse side effects for some individuals. Instead, we would ideally understand the benefits and costs of nudges, and tailor them to the population that is the most likely to benefit from them.

Future work should focus on the welfare effects of calorie labels in greater detail. We envision three avenues for such work. First, we should directly estimate the welfare effects of calorie labels on individual well-being. Second, we should extend the analysis to incorporate

changes in externalities related to obesity to allow for the estimation of social welfare. Third, just as physicians have thought of ways to mitigate the side effects of certain medicines, we should explore ways to mitigate the adverse effects of “nudges” such as calorie labels without compromising their overall impact on the targeted behaviors.

An additional direction for future work is to incorporate adaptation. The current paper and sample frame do not allow us to say anything about adaptation. Our results should therefore be considered as short run estimates that are necessary for testing our theory, which is a static model. Further work should extend our theory to allow for dynamics and consider a longer-time horizon to allow for adaptation and the derivation of long-run welfare effects.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493-505.
- Aghion, P., Akcigit, U., Deaton, A., & Roulet, A. (2016). Creative Destruction and Subjective Well-Being. *American Economic Review*, 106(12), 3869-3897.
- Allcott, H. (2011). Social Norms and Energy Conservation. *Journal of Public Economics*, 95(9-10), 820-842.
- Allcott, H., & Kessler, J. (2019). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics*, 11(1), 236-276.
- Allcott, H., & Knittel, C. (2019). Are Consumers Poorly Informed about Fuel Economy? Evidence from Two Experiments. *American Economic Journal: Economic Policy*, 11(1), 1-37.
- Allcott, H., Lockwood, B.B., & Taubinsky, D. (2019). Regressive Sin Taxes, with an Application to the Optimal Soda Tax. *Quarterly Journal of Economics*, 134(3), 1557-1626.
- Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10), 3003-3037.
- Allcott, H., & Sweeney, R. (2017). The Role of Sales Agents in Information Disclosure: Evidence from a Field Experiment. *Management Science*, 63(1), 21-39.
- Allcott, H., & Taubinsky, D. (2015). Evaluating behaviorally-motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8), 2501-2538.
- Altmann, S., Falk, A., Heidhues, P., Jayaraman, R., & Teirlinck, M. (2019). Defaults and Donations: Evidence from a Field Experiment. *Review of Economics and Statistics*, 101(5), 808-826.
- Al-Ubaydli, O., List, J.A., LoRe, D., & Suskind, D.L. (2017a). Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature. *Journal of Economic Perspectives*, 31(4), 125-144.
- Al-Ubaydli, O., List, J.A., & Suskind, D.L. (2017b). What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results. *American Economic Review*, 107(5), 282-286.
- Amir, O., Ariely, D., Cooke, A., Dunning, D., Epley, N., Gneezy, U., Koszegi, B., Lichtenstein, D., Mazar, N., Mullainathan, S., Prelec, D., Shafir, E., & Silva, J. (2005). Psychology, Behavioral Economics, and Public Policy. *Marketing Letters*, 16, 443-454.

- Anderson, M.L. & Matsa, D.A. (2011). Are Restaurants Really Supersizing America? *American Economic Journal: Applied Economics*, 3(1), 152-188.
- Anderson, M.L., Lu, F., Zhang, Y., Yang, J., & Qin, P. (2016). Superstitions, Street Traffic, and Subjective Well-Being. *Journal of Public Economics*, 142, 1-10.
- Andreoni, J., & Rao, J.M. (2011). The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism. *Journal of Public Economics*, 95(7-8), 513-520.
- Andreoni, J., Rao, J.M., & Trachtman, H. (2017). Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. *Journal of Political Economy*, 125(3), 625-653.
- Auchincloss, A.H., Lally, G.G., Leonberg, B.L., Ricchezza, A., Glanz, K. & Schwarz, D.F. (2013). Customer Response to Mandatory Menu Labeling at Full-Service Restaurant. *American Journal of Preventive Medicine*, 45(6), 710-719.
- Barr, A., & Turner, S. (2018) A Letter and Encouragement: Does Information Increase Postsecondary Enrollment of UI Recipients? *American Economic Journal: Economic Policy*, 10(3), 42-68.
- Bedard, K. & Kuhn, P.J. (2015). Micro-marketing Healthier Choices: Effects of Personalized Ordering Suggestions on Restaurant Purchases. *Journal of Health Economics*, 39, 106-122.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose? *American Economic Review*, 102(5), 2083–2110.
- Benjamin, D.J., Heffetz, O., Kimball, M.S., & Szembrot, N. (2014). Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference. *American Economic Review*, 104(9), 2698-2735.
- Berridge, K. (2003). Irrational pursuits: Hyper-incentives from a visceral brain. *The psychology of economic decisions*, 1, 17–40.
- Blanchflower, D. G., & Oswald, A. J. (2016). Antidepressants and Age: A New Form of Evidence for U-shape well-being through life. *Journal of Economic Behavior & Organization*, 127, 46-58.
- Block, J.P., Condon, S.K., Kleinman, K., Mullen, J., Linakis, S., Rifas-Shiman, S. & Gillman, M.W. (2013). Consumers' estimation of calorie content at fast food restaurants: cross sectional observational study. *British Medical Journal*, 346, 1-10.
- Bollinger, B., Leslie, P. & Sorensen, A. (2011). Calorie Posting in Chain Restaurants. *American Economic Journal: Economic Policy*, 3(1), 91-128.

- Burton, S. & Creyer, E.H. (2004). What Consumers Don't Know *Can* Hurt Them: Consumer Evaluations and Disease Risk Perceptions of Restaurant Menu Items. *The Journal of Consumer Affairs*, 38(1), 121-145.
- Burton, S., Creyer, E.H., Kees, J. & Huggins, K. (2006). Attacking the Obesity Epidemic: The Potential Health Benefits of Providing Nutrition Information in Restaurants. *American Journal of Public Health*, 96, 1669-1675.
- Bursztyn, L., Gonzalez, A.L., & Yanagizawa-Drott, D. (2018). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. NBER Working Paper No. 24736.
- Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. *Journal of Political Economy*, 127(4), 1641-1683.
- California Center for Public Health Advocacy (CCPHA) (2007). Californians Overwhelmingly Support Mandatory Menu Labeling. <http://news.ufl.edu/archive/2013/10/uf-survey-adults-lack-basic-knowledge-on-caloric-intake.html>. Accessed February 4, 2016.
- Cameron, A. & Miller, D. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50, 317-372.
- Cawley, J. (2004). The Impact of Obesity on Wages. *Journal of Human Resources*, 39(2), 451-474.
- Cawley, J., Susskind, A., & Willage, B. (2018). The Impact of Information Disclosure on Consumer Behavior: Evidence from a Randomized Field Experiment of Calorie Labels on Restaurant Menus. National Bureau of Economic Research Working Paper No. 24889.
- Chen, Y., Lu, F., & Zhang, J. (2017). Social Comparisons, Status and Driving Behavior. *Journal of Public Economics*, 155, 11-20.
- Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review*, 105(5), 1-33.
- Chetty, R., & Saez, E. (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics*, 5(1), 1-31.
- Costa, D.L., & Kahn, M.E. (2013). Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association*, 11(3), 680-702.
- Courtemanche, C., Heutel, G., and McAlvanah, P. (2015a). Impatience, Incentives, and Obesity. *Economic Journal*, 125, 1-31.
- Courtemanche, C., Pinkston, J., & Stewart, J. (2015b). Adjusting Body Mass for Measurement Error with Invalid Validation Data. *Economics and Human Biology*, 19, 275-293.

- Courtemanche, C., Pinkston, J., Ruhm, C. & Wehby, G. (2016). Can Changing Economic Factors Explain the Rise in Obesity? *Southern Economic Journal*, 82(4), 1266-1310.
- Currie, J., DellaVigna, S., Moretti, E., & Pathania, V. (2010). The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy*, 2(3), 32-63.
- Danzer, A., & Danzer, N. (2016). The Long-Run Consequences of Chernobyl: Evidence on Subjective Well-being, Mental Health and Welfare. *Journal of Public Economics*, 135, 47-60.
- Davis, L.W., & Metcalf, G.E. (2016). Does Better Information Lead to Better Choices? Evidence from Energy-Efficient Labels. *Journal of the Association of Environmental and Resource Economists*, 3(3), 589-625.
- Deb, P. & Vargas, C. (2016). Who Benefits from Calorie Labeling? An Analysis of its Effects on Body Mass. National Bureau of Economic Research Working Paper No. 21992.
- Deaton, A., & Stone, A.A. (2016). Understanding Context Effects for a Measure of Life Evaluation: How Responses Matter. *Oxford Economic Papers*, 68(4), 861-870.
- DellaVigna, S., List, J.A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *Quarterly Journal of Economics*, 127(1), 1-56.
- Di Tella, R., MacCulloch, R., & Oswald, A.J. (2001). Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness. *American Economic Review*, 91(1), 335-341.
- Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for Promoting Healthier Food Choices. *American Economic Review*, 99(2), 159-64.
- Dunn, R.A. (2010). The Effect of Fast-food Availability on Obesity: An Analysis by Gender, Race, and Residential Location. *American Journal of Agricultural Economics*, 92(4), 1149-1164.
- Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2016). Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203-32.
- Duflo, E., Gale, W., Liebman, J., Orszag, P., Saez, E. (2016). Saving Incentives for Low- and Middle-Income Families: Evidence From A Field Experiment With H&R. Block. *Quarterly Journal of Economics*, 121(4), 1311-1346.
- Elbel, B., Kersh, R., Brescoll, V.L. & Dixon, L.B. (2009). Calorie Labeling And Food Choices: A First Look At The Effects On Low-Income People In New York City. *Health Affairs*, 28(6), w1110-w1121.

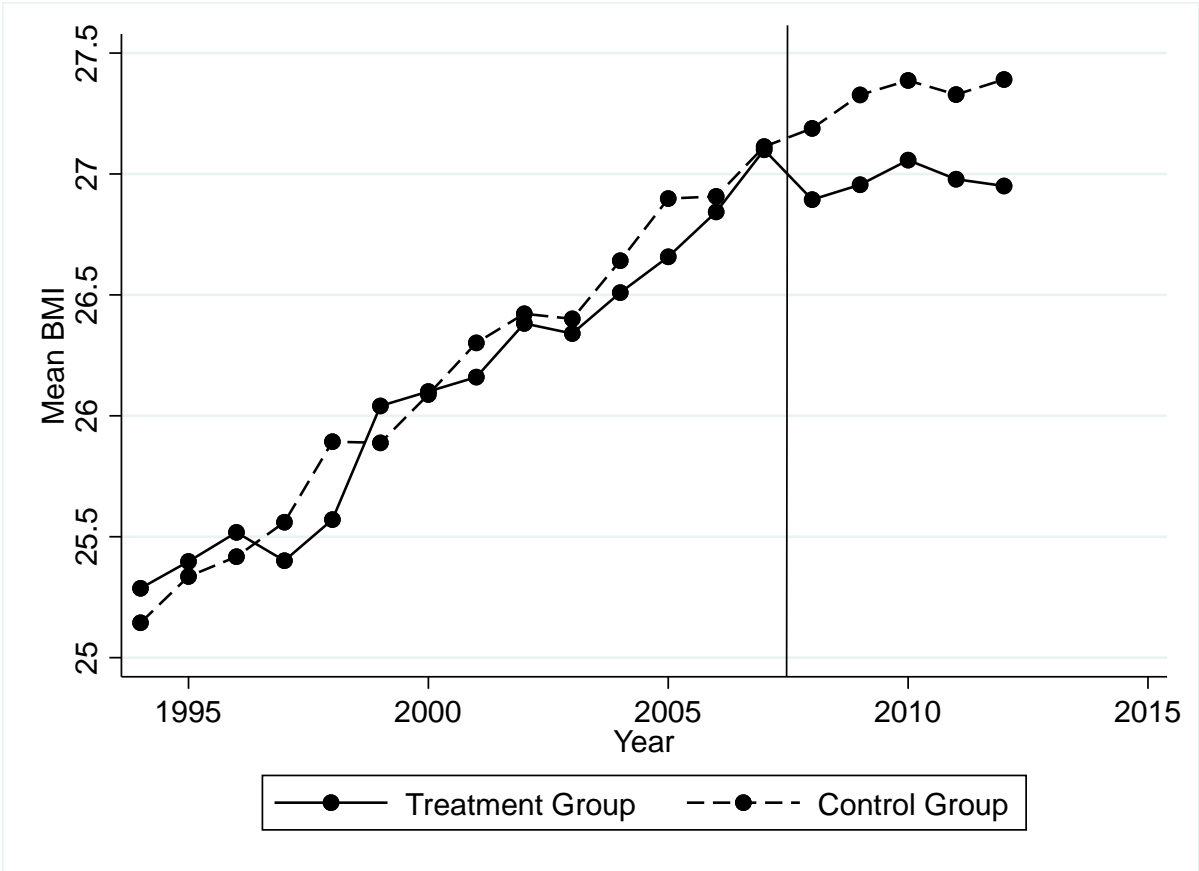
- Elbel, B., Gyamfi, J. & Kersh, R. (2011). Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment. *International Journal of Obesity*, 35, 493–500.
- Ellison, B., Lusk, J.L. & Davis, D. (2013). Looking at the label and beyond: the effects of calorie labels, health consciousness, and demographics on caloric intake in restaurants. *International Journal of Behavioral Nutrition and Physical Activity*, 10(21), 1-9.
- Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information. *Journal of the European Economic Association*, 11(3), 634-660.
- Ferraro, P.J., & Price, M.K. (2013). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *Review of Economics and Statistics*, 95(1), 64-73.
- Finkelstein, E.A., Strombotne, K.L., Chan, N.L. & Krieger, J. (2011). Mandatory Menu Labeling in One Fast-Food Chain in King County, Washington. *American Journal of Preventive Medicine*, 40(2), 122-127.
- Frey, B.S., Luechinger, S., & Stutzer, A. (2009). The Life Satisfaction Approach to Valuing Public Goods: The Case of Terrorism. *Public Choice*, 138, 317-345.
- Frijters, P., Haisken-DeNew, J.P., & Shields, M.A. (2004). Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification. *American Economic Review*, 94(3), 730-740.
- Galiani, S., Gertler, P., & Undurraga, R. (2018). The Half-Life of Happiness: Hedonic Adaptation in the Subjective Well-Being of Poor Slum Dwellers to the Satisfaction of Basic Housing Needs. *Journal of the European Economic Association*, 16(4), 1189-1233.
- Glaeser, E.L. (2006). Paternalism and psychology. *University of Chicago Law Review*, 73(1), 133-156.
- Gruber, J., & Koszegi, B. (2004). Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes. *Journal of Public Economics*, 88(9-10), 1959-1987.
- Gruber, J., & Mullainathan, S. (2005). Do Cigarette Taxes Make Smokers Happier. *Advances in Economic Analysis and Policy*, 5(1), Article 4.
- Handel, B.R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, 103(7), 2643-2682.
- Hall, K.D., Sacks, G., Chandramohan, D., Chow, C.C., Wang, Y.C., Gortmaker, S.L. & Swinburn, B.A. (2011). Quantification of The Effect of Energy Imbalance on Bodyweight. *Lancet*, 378(9793), 8363-837.

- Hallsworth, M., List, J., Metcalfe, R., & Vlaev, I. (2017). The Behavioralist As Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance. *Journal of Public Economics*, 148(C), 14-31.
- Hastings, J.S., & Weinstein, J.M. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics*, 123(4), 1373-1414.
- Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2), 515-548.
- Jimenez-Gomez, D. (2019). Nudging and Phishing: A Theory of Behavioral Welfare Economics. Working Paper.
- Jin, G.Z., & Leslie, P. (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics*, 118(2), 409-451.
- Kahn, M.E., & Wolak, F.A. (2013). Using Information to Improve the Effectiveness of Nonlinear Pricing: Evidence from a Field Experiment. Stanford University Program on Energy and Sustainable Development Working Paper.
- Kahneman, D., & Sugden, R. (2005). Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics*, 32(1), 161-181.
- Kahneman, D., & Thaler, R. H. (2006). Anomalies: Utility maximization and experienced utility. *Journal of Economic Perspectives*, 20(1), 221-234.
- Kahneman, D., Wakker, P.P. & Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *Quarterly Journal of Economics*, 112(2), 375-406.
- Kozup, K.C., Creyer, E.H. & Burton, S. (2003). Making Healthful Food Choices: The Influence of Health Claims and Nutrition Information on consumers' Evaluations of Packaged Food Products and Restaurant Menu Items. *Journal of Marketing*, 67, 19-34.
- Krieger, J.W., Chan, N.L., Saelens, B.E., Ta, M.L., Solet, D. & Fleming, D.W. (2013). Menu Labeling Regulations and Calories Purchased at Chain Restaurants. *American Journal of Preventive Medicine*, 44(6), 595-604.
- Lazear, E.P., Malmendier, U., & Weber, R.A. (2012). Sorting in Experiments with Application to Social Preferences. *American Economic Journal: Applied Economics*, 4(1), 136-163.
- Levinson, A. (2012). Valuing Public Goods Using Happiness Data: The Case of Air Quality. *Journal of Public Economics*, 96(9), 869-880.
- Levitt, S.D., & List, J.A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21(2), 153-174.

- Levitt, S.D., List, J.A., Neckermann, S., & Sadoff, S. (2016). The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy*, 8(4), 183-219.
- Liebman, J.B. & Luttmer, E.F.P. (2015). Would People Behave Differently if They Better Understood Social Security? Evidence from a Field Experiment. *American Economic Journal: Economic Policy*, 7(1), 275-299.
- Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection Bias in Predicting Future Utility. *Quarterly Journal of Economics*, 118(4), 1209–1248.
- Loewenstein, G., Brennan, T. & Volpp, K.G. (2007). Asymmetric Paternalism to Improve Health Behaviors. *JAMA*. 298(20), 2415-2417.
- Luechinger, S. (2009). Valuing Air Quality Using the Life Satisfaction Approach. *Economic Journal*. 119(536), 482-515.
- Luechinger, S., & Raschky, P. (2009). Valuing Flood Disasters using the Life Satisfaction Approach. *Journal of Public Economics*, 93(3-4), 620-633.
- Matjasko, J.L., Cawley, J.H., Baker-Goering, M.M. & Yokum, D.V. (2016). Applying Behavioral Economics to Public Health Policy: Illustrative Examples and Promising Directions. *American Journal Preventive Medicine*. 50(5 Suppl 1), S13–S19.
- Metcalfe, R., Dolan, P., Powdthavee, N. (2011). Destruction and Distress: Using a Quasi-Experiment to Show the Effects of the September 11 Attacks on Subjective Well-Being in the U.K. *Economic Journal*, 121, F81-F103.
- Mullainathan, S. & Shafir, E. (2013) Scarcity: Why Having Too Little Means So Much. New York, NY: Times Books.
- Perez-Truglia, R. (2015). A Samuelsonian validation test for happiness data. *Journal of Economic Psychology*, 49, 74–83.
- Philipson, T. J., & Posner, R. A. (1999). The long-run growth in obesity as a function of technological change. NBER Working Paper 7423.
- Porter, R. and Zona, J. (1993). Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy*. 101(3), 518-538
- Restrepo, B. (2017). Calorie Labeling in Chain Restaurants and Body Weight: Evidence from New York. *Health Economics*, 26, 1191-1209.
- Rees-Jones, A. & Rozema, K.T. (2020). Price Isn't Everything: Behavioral Response around Changes in Sin Taxes. NBER Working Paper No. 25958.

- Roberto, C.A., Larsen, P.D., Agnew, H., Baik, J. & Brownell, K.D. (2010). Evaluating the Impact of Menu Labeling on Food Choices and Intake. *American Journal of Public Health*, 100(2), 312-318.
- Sadoff, S., A. Samek, & C. Sprenger. (Forthcoming). Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts. *Review of Economic Studies*.
- Saez, E. (2009). Details Matter: The Impact of Presentation and Information on the Take-up of Financial Incentives for Retirement Saving. *American Economic Journal: Economic Policy*, 1(1), 204-228.
- Seira, E., Elizondo, A., & Laguna-Müggenburg, E. (2017). Are Information Disclosures Effective? Evidence from the Credit Card Market. *American Economic Journal: Economic Policy*. 9(1), 277-307.
- Taubinsky, D., & Rees-Jones, A. (2018). Attention variation and welfare: theory and evidence from a tax salience experiment. *Review of Economic Studies*, 85(4), 2462–2496.
- Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Thunström, L. (2019). Welfare effects of nudges: The emotional tax of calorie menu labeling. *Judgment and Decision Making*, 14(1), 11–25.
- VanEpps, E.M., Downs, J.S., & Loewenstein, G. (2016). Advance Ordering for Healthier Eating? Field Experiments on the Relationship between the Meal Order-Consumption Time Delay and Meal Content. *Journal of Marketing Research*, 53, 369-380.
- Van Praag, B.M.S., & Baarsma, B.E. (2005). Using Happiness Surveys to Value Intangibles: The Case of Airport Noise. *Economic Journal*, 115(500), 226-246.
- Volpp, K.G., Asch, D.A., Galvin, R., & Loewenstein, G. (2011). Redesigning Employee Health Incentives — Lessons from Behavioral Economics. *New England Journal of Medicine*, 365(5), 388–390.
- Wehling, H., & Lusher, J. (2017). People with a Body Mass Index ≥ 30 under-report their Dietary Intake: A Systematic Review. *Journal of Health Psychology*, 24(14), 2042-2059.
- Wisdom, J., Downs, J. S. & Loewenstein, G. (2010). Promoting Healthy Choices: Information versus Convenience. *American Economic Journal: Applied Economics*, 2(2), 164-78.
- Yelowitz, A. (2016). Menu Mandates and Obesity: A Futile Effort. Policy Analysis, Cato Institute, No. 789.

Figure 1 – Average BMI During Sample Period



Note: Observations are aggregated to the county level, with each county weighted by its population. The treatment group consists of all localities listed in Table 1, while the control group consists of other counties in MSAs with at least one “treated” county, plus New England, excluding any counties/states in these areas that passed but did not implement a calorie labeling law.

Table 1 – Menu Labeling Laws in Localities Prior to the 2014 FDA Final Rule

Localities	Date Passed	Date Implemented	Requirement on Number of Establishments
New York City (New York, Bronx, Queens, Kings, Richmond Counties, NY)	December 2006	April 2008	≥15
Seattle (King County, WA)	August 2007	January 2009	≥15
Westchester County, NY (NYC MSA)	November 2008	June 2009	≥15
Philadelphia City/County, PA	December 2008	January 2010	≥15
Albany County, NY	September 2009	April 2010	≥15
Montgomery County, MD (DC MSA)	December 2009	July 2010	≥20
Schenectady County, NY (Albany MSA)	March 2010	October 2010	≥15
Suffolk County, NY (NYC MSA)	November 2009	November 2010	≥15
State of Vermont	June 2010	January 2011	≥20

Notes: New York City, Westchester County, Suffolk County, Philadelphia County, Schenectady County, and Montgomery County actively enforced the local mandate from the date of implementation shown in the table through the implementation of the federal mandate as part of the ACA in May 2018. Based on correspondence with local officials, King County enforced the mandate between January 2009 and May 2010, Vermont did not actively enforce the mandate through inspection but instead reacted to customer complaints about a lack of calories on the menu, and Albany County stopped enforcement six months after the law was implemented.

Table 2 – Pre-Treatment Means with Standard Deviations in Parentheses

Variable	Treatment Group	Control Group
Body mass index (BMI=weight in kg/height in m ²)	26.093 (5.195)	26.177 (5.098)
<i><u>BMI Categories</u></i>		
Healthy Weight (BMI < 25)	0.466 (0.499)	0.453 (0.498)
Overweight (25 ≤ BMI < 30 kg/m ²)	0.356 (0.479)	0.362 (0.481)
Class I Obese (30 ≤ BMI < 35 kg/m ²)	0.122 (0.328)	0.128 (0.334)
Class II/III Obese (BMI ≥ 35 kg/m ²)	0.055 (0.229)	0.057 (0.231)
<i><u>Other Health Behaviors</u></i>		
Currently smokes cigarettes	0.201 (0.401)	0.208 (0.406)
Number of drinks (0 if none) in past 30 days	9.496 (18.082)	11.534 (20.485)
Any exercise (besides regular job) in past month	0.725 (0.447)	0.783 (0.412)
Minutes per week over the past month of moderate exercise (large increases in heart rate/breathing)	48.793 (66.801)	52.497 (67.213)
Minutes per week over the past month of vigorous exercise (small increases in heart rate/breathing)	30.859 (51.165)	35.113 (53.556)
<i><u>Life Satisfaction Categories</u></i>		
Very dissatisfied	0.013 (0.113)	0.010 (0.009)
Dissatisfied	0.051 (0.220)	0.036 (0.187)
Satisfied	0.537 (0.499)	0.478 (0.499)
Very satisfied	0.400 (0.384)	0.476 (0.499)
<i><u>Control Variables</u></i>		
Race/ethnicity is non-Hispanic black	0.180 (0.384)	0.101 (0.301)
Ethnicity is Hispanic	0.166 (0.372)	0.054 (0.226)
Race/ethnicity is neither white, black, nor Hispanic	0.092 (0.289)	0.049 (0.215)
High school degree but no further	0.268 (0.443)	0.282 (0.450)
Some college but no 4-year degree	0.242 (0.428)	0.257 (0.437)
College graduate or higher	0.363 (0.481)	0.381 (0.486)
Age in years	44.883 (17.357)	45.421 (17.112)
Gender is female	0.507 (0.500)	0.500 (0.500)
Married	0.497 (0.500)	0.594 (0.491)
County population density	19537 (21892)	1520 (1988)
County median household income	45602 (15335)	55868 (12749)
County unemployment rate in percent	5.809 (2.116)	4.208 (1.285)
<i>Sample size in pre-treatment period (1994-2007)</i>	120,755	246,848

Notes: Sample sizes vary for the other health behavior and life satisfaction outcomes. Observations are weighted using the BRFSS sampling weights.

Table 3 – Effects of Chain Restaurant Calorie Posting Laws on Body Mass Index

	(1)	(2)
Law Implemented	-0.174** (0.081)	-0.191* (0.099)
Law Implemented but No Longer Actively Enforced		0.126 (0.121)
Law Passed but Not Yet Implemented		0.001 (0.099)
Sample size	594,364	594,364
Mean of Outcome	26.711	26.711

Notes: Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. The control group consists of other counties in MSAs with at least one “treated” county, plus New England, excluding any counties/states in these areas that passed but did not implement a calorie labeling law. The controls include binary indicators for the individual’s sex, age, race/ethnicity, marital status, and education level, as well as county-level population density, median income, and unemployment rate. County and month*year fixed effects are also included, as well as linear county-specific time trends.

Table 4 – Effects of Chain Restaurant Calorie Posting Laws on Smoking, Drinking, and Exercising

	P(Smoker)	Drinks per Month	P(Any Exercise)	Minutes per Week Moderate Exercise	Minutes per Week Vigorous Exercise
Law Implemented	0.001 (0.007)	-0.167 (0.379)	0.001 (0.012)	-3.283 (3.419)	0.828 (1.946)
Law Implemented but No Longer Actively Enforced	0.016** (0.006)	0.064 (0.381)	0.008 (0.010)	2.083 (2.915)	1.734 (1.475)
Law Passed but Not Yet Implemented	-0.007 (0.005)	-1.037* (0.576)	0.012 (0.009)	-0.717 (2.307)	0.965 (1.749)
Sample Size	591,432	526,857	545,840	194,705	194,705
Mean of Outcome	0.192	10.979	0.762	49.600	30.093

Notes: Results shown are average marginal effects from probit models for smoker and any exercise and negative binomial models for drinks per month and minutes of exercise per week. (Coefficient estimates from linear models are nearly identical.) Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. The control group consists of other counties in MSAs with at least one “treated” county, plus New England, excluding any counties/states in these areas that passed but did not implement a calorie labeling law. The controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level, as well as county-level population density, median income, and unemployment rate. County and month*year fixed effects are also included, as well as linear county-specific time trends.

Table 5 – Effects of Chain Restaurant Calorie Posting Laws on Probabilities of Life Satisfaction States

	P(Very Dissatisfied)	P(Dissatisfied)	P(Satisfied)	P(Very Satisfied)
<i>Full Sample (n=82,203)</i>				
Law Implemented	0.0021** (0.0010)	0.0060** (0.0030)	0.0179** (0.0091)	-0.0260** (0.0131)
Law Passed but Not Yet Implemented	0.0011 (0.0012)	0.0033 (0.0035)	0.0097 (0.0104)	-0.0141 (0.0151)
Mean of Outcome	0.0121	0.0496	0.5256	0.4127
<i>Healthy Weight Subsample (n=32,082)</i>				
Law Implemented	0.0038*** (0.0013)	0.0135*** (0.0049)	0.0468*** (0.0174)	-0.0641*** (0.0235)
Law Passed but Not Yet Implemented	0.0014 (0.0010)	0.0049 (0.0040)	0.0170 (0.0139)	-0.0233 (0.0190)
Mean of Outcome	0.0085	0.0434	0.5145	0.4336
<i>Overweight or Obese Subsample (n=50,121)</i>				
Law Implemented	0.0001 (0.0015)	0.0002 (0.0039)	0.0005 (0.0104)	-0.0007 (0.0157)
Law Passed but Not Yet Implemented	0.0008 (0.0016)	0.0021 (0.0042)	0.0055 (0.0114)	-0.0084 (0.0172)
Mean of Outcome	0.0145	0.0538	0.5331	0.3986

Notes: Average marginal effects from ordered probit regressions are shown. Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. The control group consists of other counties in MSAs with at least one “treated” county, excluding any counties/states in these areas that passed but did not implement a calorie labeling law. The controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level, as well as county-level population density, median income, and unemployment rate. County and month*year fixed effects are also included.

Table 6 – Quantile Regression Results

	Quantile								
	Healthy Weight			Overweight			Obese		
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Law Implemented	-0.136 (0.089)	-0.192** (0.087)	-0.186** (0.085)	-0.133 (0.096)	-0.109 (0.108)	-0.186 (0.120)	-0.117 (0.134)	-0.255 (0.165)	-0.321 (0.223)
Law Implemented but No Longer Actively Enforced	0.095 (0.091)	0.114 (0.104)	0.118 (0.094)	0.125 (0.120)	0.117 (0.130)	0.110 (0.148)	-0.043 (0.151)	0.023 (0.209)	0.257 (0.250)
Law Passed but Not Yet Implemented	0.026 (0.096)	0.016 (0.088)	0.013 (0.096)	0.046 (0.093)	0.011 (0.118)	0.025 (0.133)	0.025 (0.135)	-0.179 (0.201)	-0.062 (0.287)
BMI at Quantile	20.80	22.35	23.54	24.61	25.80	27.02	28.40	30.36	33.55

Notes: Sample size is 593,364. Heteroskedasticity-robust standard errors are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. The control group consists of other counties in MSAs with at least one “treated” county, plus New England, excluding any counties/states in these areas that passed but did not implement a calorie labeling law. The controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level, as well as county-level population density, median income, and unemployment rate. Area (e.g. untreated portion of NYC metro area) and quarter*year fixed effects are also included, as well as linear area-specific time trends.

Online Appendix

APPENDIX A – Theoretical Appendix

In this Appendix we compute social welfare and provide a specific interpretation of the model. First, in order to compute social welfare, we need to take into account the relative weights of both groups. Let the weights be μ_L and μ_H that we assign to both groups respectively, and recall that $\Delta W_i = u_i^{exp}(x_i^1|e_1) - u_i^{exp}(x_i^0|e_0)$. Then, the change in social welfare is given by:

$$\Delta W = \mu_L * \Delta W_L + \mu_H * \Delta W_H.$$

Corollary 1A. *A sufficient condition for the nudge to increase welfare for society is*

$$\mu_L[\Lambda_L(e_1) - \Lambda_L(e_0)] \geq \mu_L c_L(e_1) + \mu_H c_H(e_1). \quad (1A)$$

A sufficient condition for the nudge to decrease welfare for society is

$$\mu_L c_L(e_1) + \mu_H c_H(e_1) \geq \mu_L[\Lambda_L(e_0) - \Lambda_L(e_1)]. \quad (2A)$$

Observe that in Equation (1A), the left-hand side represents the benefit of the nudge, which is the reduction in the internality for the L types, and the right-hand side represents the cost of the nudge, which is the sum of moral costs for both groups in society. A sufficient condition for the nudge to increase welfare is therefore that the reduction in the internality has to be sufficiently large, as to compensate for the moral costs that are being imposed to all individuals.

We turn now to one possible interpretation of the model, to make things more specific by analyzing a scenario with the possibility of learning. Suppose the internality arises because individuals have different expectations about how much they will enjoy consumption: there is a set of states of the world S , and experienced utility depends on the environment and the state $u^{exp}(x|e, s)$. An individual has beliefs $\beta_i(s)$ over the states of the world, and their decision utility is their expectation of their experienced utility:

$$u_i^{dec}(x|e) = E[u_i^{exp}(x|e, s)] = \int_S u_i^{exp}(x|e, s) d\beta_i(s).$$

Therefore, the internality in environment e and state s is

$$\Lambda_i(x|e, s) = E[u_i^{exp}(x|e, s)] - u_i^{exp}(x|e, s).$$

If individuals in H have accurate beliefs (they know s), then they are unbiased, because $E[u_h^{exp}(x|e, s)] = u_h^{exp}(x|e, s)$. Individuals in L have inaccurate beliefs: they believe that they will enjoy calories more than they actually do, i.e., $E[u_l^{exp}(x|e, s)] > u_l^{exp}(x|e, s)$, so they are biased towards overconsumption, since $\Lambda_l(x|e, s) > 0$. Suppose that individuals update their beliefs when they receive the nudge. Individuals in H already had accurate beliefs, and hence do not learn from the nudge; however, they suffer the psychological cost, and are therefore worse off because of the nudge. Individuals in L update their beliefs, which are now more accurate, and they will be better off whenever the increased utility from making a better choice is larger than the psychological cost from the nudge.

APPENDIX B – Sample Composition and Robustness Checks

Table B1 in this appendix provides detail on the composition of our sample by county. The rest of the appendix conducts robustness, placebo, and event-study analyses to evaluate the validity of our baseline estimates, shown in Tables 3 and 5, of the influence of local calorie mandates on BMI and life satisfaction.

Table B2 reports the results from robustness checks for BMI. We first include only basic demographic controls – age, gender, and race/ethnicity – along with the county and time fixed effects and county-specific time trends. The other controls are potentially endogenous and endogeneity bias could plausibly spill over to the coefficient of interest. The second robustness check drops county-specific time trends to increase the amount of identifying variation. Third and fourth, we restrict the sample period to 2001-2012 and 2005-2012 in order to test the sensitivity of the results to the length of the pre-treatment period. Fifth and sixth, we use these same shorter sample periods but drop the county trends. The seventh robustness check excludes counties adjacent to treated counties due to some of their residents likely working or visiting – and therefore eating some restaurant meals – in the treated county. If this is a major concern, then excluding those with the closest proximity to the treated county should affect the estimates. In any case, to the extent that such cross-county spillovers occur, this would likely mean our baseline results are conservative. The eighth check clusters standard errors by state rather than county, while addressing possible concerns about the small number of clusters by following Cameron and Miller’s (2015) recommendation and conducting inference using the t-distribution rather than standard normal.

The last two robustness checks in Table B2 focus on the choice of control units. First, as an alternate control group, we simply use the top 10% of the population density distribution from

the entire BRFSS sample. This strategy is based on the fact that the most glaring pre-treatment difference between the proximity-based treatment and control groups is that the individuals in the treatment group lived in much more densely populated counties. Note again that, since this control group is not based on geographic proximity to a treated county, this strategy is immune to the concern about cross-county restaurant eating.

Finally, we employ the synthetic control method of Abadie et al. (2010) to allow the data to choose a weighted combination of control counties that best matches the pre-treatment characteristics of the treatment group. The synthetic control method was originally designed for aggregate data; we follow the procedure used by Fitzpatrick (2008) and Courtemanche and Zapata (2014) to apply the method to individual data. We begin by aggregating the data to the county-by-year level, further combining all the treated counties together into a single cross-sectional unit. Our “donor pool” consists of all counties in the original proximity-based control group as well as all those in the top 10% of the population density distribution. The synthetic control procedure computes the weights for each county in the donor pool that yields the best possible match to the treatment group along dimensions chosen by the researcher. In the results shown, we match based on means of the control variables in the pre-treatment period as well as BMI in four pre-treatment years: 1994, 1998, 2003, and 2007.³⁹ We then return to the individual-level data and re-run the regressions using as weights for the control group the product of the BRFSS individual sampling weight and the county’s synthetic weight.

³⁹ The resulting synthetic control group is 19.2% Denver, CO; 0.1% Pinellas, FL; 39.4% Cook, IL; 4.8% Baltimore, MD; 0.8% Arlington, VA; 12.4% Fairfax, VA; 18.3% Norfolk, VA; 5.1% Roanoke, VA. We have also considered numerous other combinations of matching variables and the results are similar.

As shown in Table B2, the results are quite robust across these different specifications. The effect of calorie labeling laws on BMI is always statistically significant at the 10% level or less, with magnitudes ranging from -0.147 to -0.253 units.

Table B3 turns to placebo tests using just the pre-treatment portion of the sample period (1994-2007). In difference-in-difference designs, identification hinges on the assumption that there is a common counterfactual trend in the dependent variable between the treated and the control groups. We can test whether this assumption is likely to be valid by running placebo tests to evaluate whether calorie posting laws have “effects” in periods of time other than those during which they are implemented (“false” treatments). To implement these tests, we pre-date the laws to occur over each possible 4-year range during the pre-treatment period and re-estimate the baseline model for each outcome. As the results show, we find no evidence for differential counterfactual trends in the pre-treatment period. The estimates are all small and insignificant.

Table B4 considers a different method to test for problematic pre-treatment trends: event-study regression. Rather than including a single indicator for whether a calorie labeling law has been enacted, we include a series of indicators reflecting time until implementation. Specifically, in addition to the usual “law enacted” dummy variable, we also include dummy variables for whether the law has not yet been enacted in that particular county but will be in the next 25-48 months, 49-72 months, 73-96 months, 97-120 months, and more than 120 months. The reference period is therefore the omitted “next 1-24 months” category. Since Wolfers (2006) makes a compelling case that including treatment-unit-specific time trends is inappropriate in this type of dynamic specification, we exclude the county trends from the event study model. As the table shows, there is no evidence that any “effects” emerge in the years leading up to the law’s implementation. Moreover, the estimates for the “true” treatment effect remain virtually identical

to those from the analogous specification (the one without county trends) in Table B2.⁴⁰ The results from the event-study design and the similarity in the estimates between the event-study design and difference-in-differences specification suggests that the treatment effect estimates from the difference-in-differences specification with time-varying treatments are unlikely to be biased (Goodman-Bacon, 2018).

The top panel of Table B5 shows the robustness of the estimate of the influence of calorie mandates on life satisfaction. The robustness checks are similar to those used for BMI, except that, since life satisfaction is only available starting in 2005, we do not vary the length of the pre-treatment period. The results remain qualitatively similar across the various specifications, as all estimates have the same sign and are statistically significant at the 10% level or less. The magnitudes are somewhat sensitive to specification (e.g., calorie labeling laws reduce the probability of being very satisfied by between 1.7 and 3.0 percentage points), but all estimates are within each other's 95% confidence intervals.

The bottom two panels of Table B5 evaluate pre-treatment trends in life satisfaction. Given the relatively short length of the pre-treatment period in the life satisfaction regressions, we conduct only one placebo test in which the sample period consists of 2005-2007 and the placebo intervention occurs midway through this period (July 2006) for all the treated areas.⁴¹ Similarly, the event-study model only includes two treatment indicators: whether the law is law currently enacted and whether it will be enacted in more than 24 months, with future enactment of 1-23 months as the reference category. In both the placebo and event study regressions, we find no

⁴⁰ We also considered an even more flexible event study model with separate indicators for each year. While this model also revealed no evidence of problematic pre-trends, many of the estimates became quite imprecise, so the results were not particularly informative.

⁴¹ We exclude county time trends from the placebo regression since it only uses three years of data.

evidence of problematic pre-trends, as the “effects” of the pre-dated law are very small and statistically insignificant.

Appendix B References

- Abadie, A., Diamond, A. & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal for the American Statistical Association*, 105(490), 493-505.
- Cameron, A. & Miller, D. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50, 317-372.
- Courtemanche, C.J. & Zapata, D. (2014). Does Universal Coverage Improve Health? The Massachusetts Experience. *Journal of Policy Analysis and Management*, 33(1), 36-69.
- Fitzpatrick, M.D. (2008). Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement. *The B.E. Journal of Economic Analysis & Policy*, 8(1), 1-40.
- Goodman-Bacon, A. (2018). Difference-In-Differences with Variation in Treatment Timing. National Bureau of Economic Research Working Paper No. 25018.
- Wolfers, J. (2006). Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review*, 96(5), 1802-1820.

APPENDIX TABLES

Table B1 – Composition of the Sample by County

Treatment Group		Control Group	
County	Observations	County	Observations
<u>New York City MSA</u>			
Bronx, NY	3,847	Dutchess, NY	1,508
Kings, NY	7,524	Essex, NY	270
New York, NY	7,827	Orange, NY	1,579
Queens, NY	6,942	Putnam, NY	425
Richmond, NY	1,682	Rockland, NY	1,229
Suffolk, NY	6,199	Sullivan, NY	394
Westchester, NY	4,076	Pike, PA	2,096
<u>Philadelphia MSA</u>			
Philadelphia, PA	15,291	New Castle, DE	24,701
		Cecil, MD	3,305
		Bucks, PA	3,982
		Chester, PA	3,944
		Delaware, PA	3,670
		Montgomery, PA	5,396
<u>Seattle MSA</u>			
King, WA	40,565	Pierce, WA	17,487
		Snohomish, WA	17,340
<u>Washington, DC MSA</u>			
Montgomery, MD	14,370	Washington, DC	43,490
		Calvert, MD	2,907
		Charles, MD	4,116
		Frederick, MS	7,169
		Anne Arundel, MD	10,626
		Arlington, VA	2,118
		Clarke, VA	105
		Culpeper, VA	410
		Fairfax, VA	3,166
		Fauquier, VA	685
		Loudoun, VA	1,915
		Prince William, VA	2,170
		Rappahannock, VA	48
		Spotsylvania, VA	798

		Stafford, VA	757
		Warren, VA	285
		Alexandria City, VA	1,881
		Fairfax City, VA	274
		Falls Church City, VA	53
		Fredericksburg City, VA	165
		Manassas City, VA	147
		Manassas Park City, VA	23
		Jefferson, WV	1,333
		<u>Albany MSA</u>	
Albany, NY	1,855	Rensselaer, NY	950
Schenectady, NY	884	Saratoga, NY	1,287
		<u>New England</u>	
Addison, VT	5,096	Fairfield, CT	22,448
Bennington, VT	5,032	Hartford, CT	20,724
Caledonia, VT	4,273	Litchfield, CT	4,573
Chittenden, VT	17,513	Middlesex, CT	4,284
Essex, VT	493	New Haven, CT	20,410
Franklin, VT	5,777	New London, CT	6,795
Grand Isle, VT	462	Tolland, CT	4,303
Lamoille, VT	3,254	Windham, CT	4,400
Orange, VT	4,370	Belknap, NH	4,294
Orleans, VT	3,819	Carroll, NH	3,711
Rutland, VT	8,480	Cheshire, NH	5,277
Washington, VT	8,297	Coos, NH	3,609
Windham, VT	6,543	Grafton, NH	5,871
Windsor, VT	8,144	Hillsborough, NH	17,840
		Merrimack, NH	7,643
		Rockingham, NH	13,033
		Strafford, NH	6,775
		Sullivan, NH	3,508
		Bristol, RI	3,163
		Kent, RI	10,110
		Newport, RI	5,583
		Providence, RI	40,980
		Washington, RI	8,211
		<u>Total</u>	
26 counties	192,615	63 counties	401,749

Table B2 – Robustness Checks for Body Mass Index

	BMI
Baseline model (n=593,364)	-0.174** (0.081)
<i>Robustness checks</i>	
Basic demographic controls only (n=593,364)	-0.186* (0.095)
Drop county trends (n=593,364)	-0.253*** (0.078)
2001-2012 (n=475,842)	-0.147* (0.075)
2005-2012 (n=357,514)	-0.150** (0.075)
2001-2012; drop county trends (n=475,842)	-0.171** (0.080)
2005-2012; drop county trends (n=357,514)	-0.193** (0.075)
Exclude adjacent counties (n=478,867)	-0.211*** (0.076)
Cluster by state (n=593,364)	-0.174** (0.073)
Control group is pop. density > 2,000/sq. mi. (n=458,519)	-0.157* (0.088)
Synthetic control group (n=254,050)	-0.192** (0.094)

Notes: Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. Unless otherwise indicated, the control group consists of other counties in MSAs with at least one “treated” county, plus New England. Unless otherwise indicated, the controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level; county-level population density, median income, and unemployment rate; county and month*year fixed effects; and county-specific linear time trends.

Table B3 – Placebo Tests for Body Mass Index using Pre-Treatment Data (1994-2007)

	BMI
Laws start 4/2004-1/2007	0.037 (0.121)
Laws start 4/2003-1/2006	-0.063 (0.079)
Laws start 4/2002-1/2005	-0.056 (0.095)
Laws start 4/2001-1/2004	0.012 (0.056)
Laws start 4/2000-1/2003	0.076 (0.094)
Laws start 4/1999-1/2002	-0.030 (0.082)
Laws start 4/1998-1/2001	-0.045 (0.082)
Laws start 4/1997-1/2000	0.001 (0.070)
Laws start 4/1996-1/1999	0.094 (0.094)
Laws start 4/1995-1/1998	0.001 (0.079)

Notes: n=367,603. Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. Unless otherwise indicated, the control group consists of other counties in MSAs with at least one “treated” county, plus New England. Unless otherwise indicated, the controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level; county-level population density, median income, and unemployment rate; county and month*year fixed effects; and county-specific linear time trends.

Table B4 – Event-Study Results for Body Mass Index

	BMI
Law enacted	-0.253*** (0.078)
Law will be enacted in 25-48 months	-0.071 (0.092)
Law will be enacted in 49-72 months	-0.040 (0.095)
Law will be enacted in 73-96 months	0.045 (0.116)
Law will be enacted in 97-120 months	-0.031 (0.119)
Law will be enacted in more than 120 months	0.081 (0.087)

Notes: Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. Unless otherwise indicated, the control group consists of other counties in MSAs with at least one “treated” county, plus New England. Unless otherwise indicated, the controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level; county-level population density, median income, and unemployment rate; and county and month*year fixed effects.

Table B5 – Robustness, Placebo, and Event-Study Results for Life Satisfaction

	P(Very Dissatisfied)	P(Dissatisfied)	P(Satisfied)	P(Very Satisfied)
Baseline model (n=82,203)	0.002*** (0.001)	0.007*** (0.003)	0.021*** (0.008)	-0.030*** (0.011)
<i>Robustness Checks</i>				
Basic demographic controls only (n=82,203)	0.002*** (0.001)	0.007*** (0.003)	0.020*** (0.008)	-0.029*** (0.011)
Drop county trends (n=82,203)	0.001* (0.001)	0.004* (0.002)	0.012* (0.006)	-0.017* (0.009)
Exclude adjacent counties (n=54,540)	0.002** (0.001)	0.007** (0.003)	0.020** (0.010)	-0.030** (0.014)
Synthetic control group (n=99,660)	0.002** (0.001)	0.006** (0.003)	0.018** (0.008)	-0.026** (0.012)
Control group is pop. density > 2,000/sq. mi. (n=127,674)	0.002*** (0.001)	0.007*** (0.002)	0.020*** (0.007)	-0.029*** (0.010)
Cluster by state (n=82,203)	0.002*** (0.0003)	0.007*** (0.001)	0.021*** (0.004)	-0.030*** (0.005)
<i>Falsification Test using Pre-Treatment Data (2005-2007)</i>				
Laws start 7/2006	-0.0002 (0.001)	-0.001 (0.003)	-0.002 (0.009)	0.003 (0.013)
<i>Event-Study Model (Law will be Enacted in Next 24 months is Reference Group)</i>				
Law enacted	0.001* (0.001)	0.004* (0.002)	0.011* (0.007)	-0.017* (0.010)
Law will be enacted in more than 24 months	-0.0004 (0.0008)	-0.001 (0.002)	-0.003 (0.007)	0.005 (0.010)

Notes: The columns give estimated average effects from an ordered probit regression. Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. Unless otherwise indicated, the control group consists of other counties in MSAs with at least one “treated” county by 2010. Unless otherwise indicated, the controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level; county-level population density, median income, and unemployment rate; and county and month*year fixed effects. The first, third, fourth, fifth, and sixth rows under “robustness checks” also include county-specific linear time trends.

Table B6 – Estimates for Weight Categories

	P(Healthy Weight)	P(Over-weight but not Obese)	P(Class I Obese)	P(Class II/III Obese)
<i>Single Treatment Variable</i>				
Law Enacted	0.006 (0.006)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)
<i>Test for Anticipatory Effects and Weakened Effects from Lack of Active Enforcement</i>				
Law Passed	-0.005 (0.008)	0.002 (0.002)	0.002 (0.003)	0.002 (0.003)
Law Enacted	0.011 (0.008)	-0.003 (0.002)	-0.004 (0.003)	-0.003 (0.002)
Law Not Actively Enforced	-0.008 (0.011)	0.002 (0.003)	0.003 (0.004)	0.002 (0.004)
Mean of Outcome	0.442	0.360	0.134	0.064

Notes: Sample size is 594,364. The columns give estimated average marginal effects from an ordered probit regression. Standard errors, heteroskedasticity-robust and clustered by county, are in parentheses. *** indicates statistically significant at 1% level, ** 5% level, * 10% level. Observations are weighted using the BRFSS sampling weights. The control group consists of other counties in MSAs with at least one “treated” county, plus New England, excluding any counties/states in these areas that passed but did not implement a calorie labeling law. The controls include dummies for the individual’s sex, age, race/ethnicity, marital status, and education level, as well as county-level population density, median income, and unemployment rate. County and month*year fixed effects are also included, as well as linear county-specific time trends.