I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13851

Price and Saliency in Health Care: When Can Targeted Nudges Change Behaviors?

Brigham Walker

NOVEMBER 2020

DISCUSSION PAPER SERIES

IZA DP No. 13851

Price and Saliency in Health Care: When Can Targeted Nudges Change Behaviors?

Brigham Walker
Tulane University, McKesson and IZA

NOVEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Price and Saliency in Health Care: When Can Targeted Nudges Change Behaviors?[1]

This paper takes advantage of a natural experiment to examine the relationship between the price and saliency of health services. A large employer e-mailed individually-targeted health education encouraging high-value care to high-risk employees. Weeks before the program launched, a company reorganization affecting about a quarter of employees resulted in that group not receiving the intervention. Using event study, difference-in-differences, and triple differences methods, I find that costlier services are associated with relatively less utilization and that prior use was associated with relatively more utilization following the campaigns. These results may inform employer, governmental, and health insurer choices concerning low-cost interventions seeking to shift health behaviors, and may also be relevant in other settings in which targeted informational nudges are deployed.

JEL Classification: I1, D8

Keywords: nudges, health care, price, saliency, information, prior beliefs

Corresponding author:
Brigham Walker
Department of Health Policy and Management
Tulane University
1440 Canal Street
New Orleans, LA 70112
USA
E-mail: bwalker6@tulane.edu

1 I thank Mary K. Olson, Jim Alm, Kevin Callison, Patrick Button, participants at the American Society of Health Economists (ASHEcon) annual conference, at the Association for Public Policy Analysis and Management (APPAM) annual fall research meeting, and at the Murphy Institute Health Policy Seminar at Tulane University for helpful comments.

I. Introduction

Can sending a targeted e-mail change health care behavior? Information is an essential input to rational decision-making, and if information alone can alter health care choices, its deployment may represent an effective tool to influence health behaviors. Despite growing support for behavioral interventions in health settings (Reisch et al, 2017), it is unclear whether these programs are effective (Marteau et al, 2011). Many results are mixed with some studies finding intended effects (Stone et al., 2002; McCaul et al., 2002) and others finding mixed or no effects (Chen et al., 2019; Bourgeois et al., 2008) despite similar interventions in similar settings. Many studies also rely on recruited volunteers, who may be more motivated to act on health information than the average patient, raising questions about validity and representativeness (Roni and Fry et al., 2009). Recent randomized controlled studies further suggest that information alone may be insufficient to change behavior (Leight and Safran, 2019; Leight and Wilson, 2019). Given these challenges, more evidence on the effectiveness of health nudges is needed.

This paper examines the effects of a nudge campaign whose deployment was plausibly exogenous and well-suited for this topic. Here, a large national self-insured employer targeted high-risk employees enrolled in the health plan to receive an e-mail pertaining to their health care options. Individuals were targeted according to their specific claims history and health search queries on a wellness platform. There were several campaigns spanning preventative and screening care, emergency care, pain and chronic disease management, and cancer care. These campaigns were generally seeking to increase preventative care (for example, office visit consultation, physical therapy, or chiropractic services for lower back pain) and to reduce high-cost acute care (for example, steroid injections or surgery for lower back pain). Weeks before deployment of the program, an unanticipated company re-organization was announced that resulted in about a quarter of the company being excluded from the program. Since each member had already been individually targeted for each campaign, there is a list of individuals who were planned to receive the e-mail but did not. This business unit level variation, plausibly exogenous to the individual, is what I exploit to assess the effect of targeted health information against a control group who was identified according to the same inclusion criteria and who otherwise would have received the same information but did not.

Conceptually, this paper evaluates these programs through the lens of a model developed by Bordalo, Gennaioli, and Shleifer (2013). Intuitively, this model describes utility as a linearly separable function of quality and price for a given consumable against all other consumables. As the quality or price for that consumable is made more salient – that is, disproportionately weighted in the choice environment (Taylor and Thompson, 1982), consistent with an informational nudge – the relative utility weight for the salient attribute of price or quality for that consumable adjusts. This leads to a change in consumption relative to a rational choice environment free from salience considerations. This could explain a relative change in consumption in either direction following a nudge.

For many health interventions assessed in the literature, the nudges were deployed in low-cost (for example, up to a few hundred dollars) and low-intensity (for example, non-acute and ambulatory) settings including vaccinations, medicine adherence, or screening. In these settings, information may be a margin on which patient care choices can change. Little is known about nudges in higher cost health settings and related studies on the effects of information nudges in non-health settings concerning more involved choices like college selection or tax filing have often found that information alone does not influence behaviors (Bhargava and

Manoli, 2015; Manoli and Turner, 2014; Chetty and Saez, 2013; Bettinger et al., 2012; Booij et al., 2012, Darolia and Harper, 2018; Hoxby and Turner, 2013; Bergman et al., 2019).

A potential link between these literatures is the intensity of the choice contemplated. Built into the Bordalo et al. model is Weber's law, which describes changes in stimuli as being perceived in relation to the magnitude of the overall stimuli. So, a fixed difference in price is less salient when it is contemplated at higher price levels than at lower price levels. Here, I evaluate a range of nudges with a range of costs for the services targeted to shed some light on this predicted property. While I cannot directly measure quality or price salience, I can evaluate the effects of these nudges considering the variation in reimbursed costs for the services contemplated. Strong prior beliefs regarding cost and quality also inform responses. While I do not have access to these beliefs, I do have access to prior utilization as a proxy for those who previously decided that the targeted care was worth pursuing, so I additionally evaluate how those with prior utilization are differently affected by the nudge campaign.

Overall, I find that many of these programs did lead to relative increases in utilization in many preventative services, but not in acute care services or some chronic care. These preventative services include increased chiropractic visits and physical therapy visits for back pain in response to a back pain targeted intervention. Similarly, increased primary care visits and imaging events were observed in response to a hip and knee pain campaign. The weight management campaign led to increases in weight-based appraisals and dietician visits. The outcomes with estimates that were not statistically significant generally included higher intensity services (for example, injections and surgeries for pain or emergency room visits) or select chronic conditions (for example, all outcomes evaluated for cardiac and diabetes care). Pooling all observations from across all campaigns and using triple differences models that include high-cost versus low-cost service variation, I find that high-cost services are less likely to be used compared with low-cost services following these campaigns. Using separate triple differences specifications that include an indicator variable for prior utilization, I find that prior use is associated with higher relative utilization following the informational nudge. This suggests that prior experience does inform one's response to a nudge in ways that appear to activate prior beliefs consistent with increased saliency.

This paper contributes to the health economics literature that examines both informational nudges and saliency. Most papers in this field focus on low-cost preventative care like immunizations (Stone et al., 2002; McCaul et al., 2002; Chen et al., 2019; Bourgeois et al., 2008; Leight and Safran, 2019) with few, if any, considering the cost variation of the targeted health service. While prior work has focused on the programmatic effects of informational nudges on health behaviors, few link the programs to economic theory. To date, while there are nearly five hundred references to the salience model outlined by Bordalo, Gennaioli, and Shleifer (2013), only one appears to directly reference health care interventions (Baicker et al., 2015), which references the model in terms of behavioral hazards (for example, misutilization mistakes) in an effort to incorporate behavioral considerations in developing optimal co-pay structures. Here, providing some evidence of the model's predictions can help policymakers, employers, and health insurers better decide on or design similarly motivated health programs.

The remaining sections of this paper will outline in more detail the relevant literature, elaborate on the conceptual model that motivates this empirical work, describe the program, detail the data and identification strategy, summarize the results, and briefly discuss conclusions.

II. Background

A. Motivation

While there is growing support for behavioral interventions like informational nudges in health settings, there remains little consensus about how well these programs work. In a survey of civilians, a majority supported the use of educational pushes around healthy behaviors (Reisch et al, 2017). The potential for encouraging patients to pursue lower cost options appears to have some support for common medical services like clinical office visits (where it had modest effects) and advanced imaging services (where it had larger effects) (Whaley et al., 2014), but there generally remains little evidence on the effectiveness for these interventions in health care settings (Marteau et al, 2011).

Many of the results are mixed across studies, even for similar interventions in similar settings. Patient reminders appear to increase cancer screening by an adjusted odds ratio of between 1.74 and 2.75 (Stone et al., 2002) and increase flu vaccinations from 20% to 28% (McCaul et al., 2002) in some settings, while another study of the effects of a postcard vaccine reminder campaign for the elderly found modest affects that varied by the season (Chen et al., 2019) and another found no effects (Bourgeois et al., 2008). One meta-analysis found across eleven studies that 67% of the treatment group adhered to their medications compared with 55% of the control group (Fenerty et al., 2012). Report cards in health care have been found to increase vaccination rates in hospital settings when coupled with other programs (Findley, 2003), but have not been found to work in schools when randomizing just the report card (Leight and Safran, 2019). Further, a recent randomized controlled field study including over 11,000 federal employees found that e-mail-based nudges did not increase the utilization of flexible spending accounts (FSAs) (Leight and Wilson, 2019).

The issues concerning bundled interventions in evaluating informational pushes are common with programs outside of health economics. Informational pushes bundled with simplified worksheets in the tax setting appear to affect tax credit utilization (Bhargava and Manoli, 2015; Manoli and Turner, 2014), but sharing just information about the Earned Income Tax Credit (EITC) does not systematically affect earnings on average (Chetty and Saez, 2013). Similarly, providing information and helping fill out Free Application for Federal Student Aid (FAFSA) forms in the education setting increased enrollment, while only telling students about their eligibility (Bettinger et al., 2012) or only sharing information about loan amounts (Booij et al., 2012, Darolia and Harper, 2018) did not affect participation levels. Information pushes affected college outcomes but so did the concurrent application fee waivers (Hoxby and Turner, 2013) and varying only the informational intervention appears to not affect enrollment for high schoolers, currently enrolled students, or previous applicants (Bergman et al., 2019). In many of these cases, information does not appear to be the binding constraint.

B. The Nudge

The campaigns sought to connect employees with resources relevant to their health conditions with the goal of reducing costs, improving health outcomes, and increasing employee satisfaction with their overall health benefits package. The administrators sought to accomplish this by sending targeted e-mails to employees to provide relevant information to support informed care decisions, especially to help employees make cost-mitigating choices. Employees were targeted based upon their care history in the claims record and health search query history on a wellness platform. Those meeting the criteria for high-risk were targeted for a diagnosis-specific campaign. My analysis considers eight different informational programs in the following areas: emergency room (ER) prevention, back pain, cancer screening, cardiac care, diabetes care, preventative care, hip and knee pain, and weight management. A list of programs and targeted utilization and cost outcomes can be found in Table 1.

Table 1: Campaigns and Outcomes Evaluated

Campaign	Outcomes
ER Prevention	ER Visits
Back	Back Visits, Chiropractor Visits, Back Physical Therapy (PT), Back Image, Back Injections, Back Surgery, Back Prescriptions (Rx), ER Visits
Cancer	Cancer E&M Visits, Total Costs, Medical Costs, Drug Costs, Colonoscopy
Cardiac	Overall Visits, Primary Care Physician (PCP) Visits, Lab Testing, Blood Pressure (BP) Readings, Electrocardiogram (EKG) Reading, Cardiac Surgery, Heart Attacks
Colon	Colonoscopy Ordered
Diabetes	Visits, ER Visits, Surgery, Non-Surgery, Acute Visits, Non-Acute Visits All Scripts, Total Rx Costs, Diuretics Rx, Non-Insulin Rx, Insulin Rx
Prevention	E&M and Visits
Hip and Knee	Injections, Surgeries, Image Studies, PCP Visits
Weight	Weight Appraisal, BMI, Dietician Visits

Coming from an employer with visibility into individual-level health histories, this nudge has two distinctive features: (1) the sender is a known informed authority on the recipient's health expenditures, and (2) the message potentially prompts focus or provides information on the recipient's expected health state. Either feature may inform how these programs operate and I cannot formally decompose these effects. This is in part because I was only able to recover three of the e-mail interventions. For the diabetes campaign, the headline approximately reads "How much do you spend on diabetes care?", which leads into a brief description of the cost implications of poorly managed disease. For the low back pain campaign, the lead-in is approximately "Low back pain doesn't need to lead to the emergency room", which is followed by an appeal to plan for unexpected flare ups. Finally, the hip and knee pain campaign appeals to informed exercise leading with something akin to “Ways to be active while managing hip or knee pain”.

While none of these e-mails direct the recipient to specific treatments, each e-mail does indirectly suggest certain types of care. For example, back pain planning plausibly implies considering physical therapy, chiropractic care, or office visits for back care. Some of this information may be explicitly described for those who click-through on the e-mails. While I do not have the content behind the click-through, between 15% and 18% of patients did click-through to learn more for each campaign (see Table 2).

Table 2: Pre-Period Differences in Risk Scores and Use

Campaign	% Opened (Treated)	Risk Score			Variable	Utilization		
		Treated	Control	P-Value		Treated	Control	P-Value
Hip/Knee	17.8%	2.28	2.69	0.05	Hip/Knee PCP	0.49	0.65	0.06
ER	15.9%	0.98	1.11	<0.01	ER Visits	0.23	0.27	<0.01
Back	16.2%	1.09	1.26	<0.01	Back Visits	0.33	0.41	<0.01
Cancer	18.4%	3.76	4.22	0.22	Cancer E&M	0.76	0.70	0.68
Cardiac	16.9%	3.01	3.38	0.18	Cardiac Visits	0.74	0.75	0.91
Diabetes	16.0%	1.20	1.21	0.95	Diabetes Visits	0.41	0.59	<0.01
Weight	16.7%	0.93	0.97	0.03	Weight Appraisal	0.96	1.00	<0.01
Colon	21.3%	0.93	0.89	0.21	Colonoscopy	0.01	0.00	0.01
Preventative	(no data)	0.44	0.35	<0.01	Physical Exams	1.79	1.49	<0.01

Notes: Risk Scores are generated as a combination of demographic variables (for example, age, race, sex, etc.) and past utilization values meant to be predictive of total costs.

III. Conceptual Framework

A. Salience and the Re-Weighting of Preferences

The health nudges described above increase the salience of the targeted health condition. People are heavily influenced by what is novel and relevant and this consequently focuses their attention and decision-making framework (King et al, 2013). In this regard, salience represents a disproportionate weighting in the choice environment (Taylor and Thompson, 1982). The conceptual framework for my analysis relies upon a model that incorporates salience by

Bordalo, Gennaioli, and Shleifer (2013). In this model, individuals make consumption choices by assessing the relative quality/price ratios of consumables in their choice set. Here, people seek “bargains” where, against a reference good, the relative quality is perceptibly increased or the relative price is perceptibly decreased, leading to “context effects”. Utility is a function of quality and price and salience on quality or price changes the relative utility weight for the salient attribute of price or quality for a given good or service.

Consider the back-pain campaign as an example: An employee has a history of back pain and is e-mailed by their employer about managing back pain given their known history with the condition. That employee may, because of the e-mail, newly focus on care options such as chiropractic care, physical therapy, pain injections, or surgery measured against a potential flare up leading to an emergency room visit. All options can mitigate pain and avoid a costly emergency room visit, but they each have different quality and price features. In the case of chiropractic care or physical therapy, these are plausibly high-quality and low-price options compared with injections or surgery. If this new sense of quality dominates their prior sense of quality and their new sense of cost, then they will increase their use of chiropractic care or physical therapy relative to having not been e-mailed. Similarly, if the patient is reminded about the costliness of injections or surgeries, and that sense of costliness is greater than their previous sense of cost or their new sense of the quality of these services as alternatives to pain, they may decrease their use of injections or surgery. The Bordalo, Gennaioli, and Shleifer (2013) model helps to formalize this dynamic.

In the model, adapted to this situation, patients are sensitized to the quality attributes through the health informational outreach, but they are also sensitized to the perceived cost of the related health interventions, and these two sensitivities inform the new quality/price ratio. If the ratio is large enough relative to all other things available to the patient, then those related health services will become relative *bargains*. Since this considers a quality/price ratio, this suggests that new sensitivities to quality are measured against new sensitivities to price. Conceptually, this provides space for heterogeneous effects, with two services of equal quality sensitivity potentially yielding different effects if their cost sensitivities varied.

The Bordalo, Gennaioli, Shleifer (2013) model can be summarized as follows: the individual considers multiple consumables in the choice set $\mathbf{C} \equiv \{(q_k, p_k)\}_{k=1,\ldots,N^*}$, where each consumable k has a nonnegative price (p_k) and quality (q_k), and the price and quality are known to the consumer and in dollar terms. In this case, absent a "salience distortion", the consumer values a targeted health service (k) with the following linear utility function:

$$u_k = q_k - p_k \qquad [A]$$

The generic reference consumable is $(\overline{p},\overline{q})$, which has the average attributes in quality and price from among **C**. For k, the price or quality attributes are denoted generally as a_k with $\overline{a}$ being that attribute's average in **C** (which simplifies the following general statements). The salience of quality for k is given by $\sigma(q_k, \overline{q})$ and by $\sigma(p_k, \overline{p})$ for price, which together comprise the salience function $\sigma(\cdot,\cdot)$. Then:

$$\sigma(a_k + \mu\varepsilon, \overline{a} - \mu\varepsilon') > \sigma(a_k, \overline{a}) \qquad [B]$$

$$\sigma(a_k + \varepsilon, \overline{a} + \varepsilon) < \sigma(a_k, \overline{a}) \qquad [C]$$

where $\mu = \text{sgn}(a_k - \overline{a})$ for any $\varepsilon, \varepsilon' \geq 0$ with $\varepsilon + \varepsilon' > 0$ for [B], and for any $a_k, \overline{a} > 0$ and all $\varepsilon > 0$ for [C]. Accordingly, quality is more salient for k when $\sigma(q_k, \overline{q}) > \sigma(p_k, \overline{p})$, price is more salient when $\sigma(q_k, \overline{q}) < \sigma(p_k, \overline{p})$, and both are equally salient when $\sigma(q_k, \overline{q}) = \sigma(p_k, \overline{p})$. Equation [B] affords people the ability to detect changes in prices or quality for k that are different from their respective averages in the choice set, with bigger differences leading to higher saliency. Equation [C] describes Weber's law of diminishing sensitivity, which generally notes that as everything gets more expensive or of higher quality, a fixed difference between consumables in price or quality will become relatively smaller.

With all this in place, salience operates by changing the relative utility weight for the salient attribute of price or quality for k. This distorts the valuation of k, formally noted as:

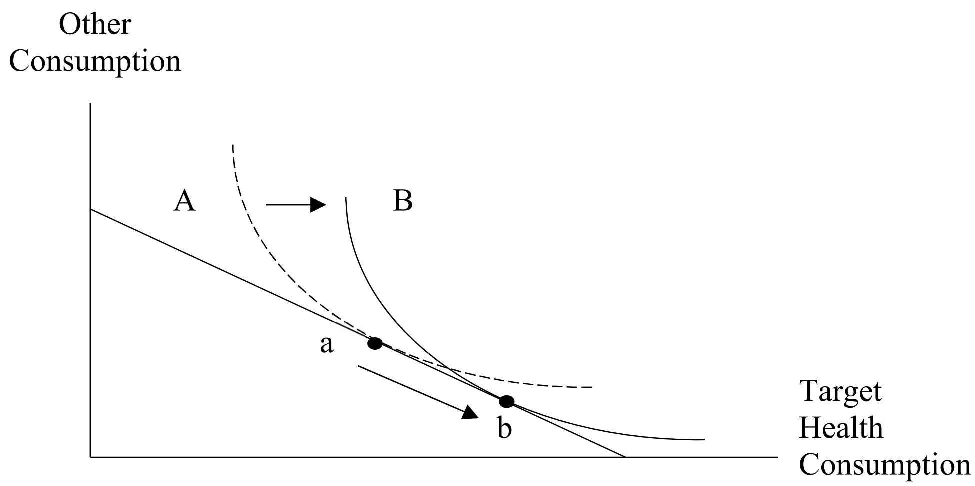
$$u_k^s = \begin{cases} \frac{2}{1+\delta} q_k - \frac{2\delta}{1+\delta} p_k & \text{if } \sigma(q_k, \overline{q}) > \sigma(p_k, \overline{p}) \\ \frac{2\delta}{1+\delta} q_k - \frac{2}{1+\delta} p_k & \text{if } \sigma(q_k, \overline{q}) < \sigma(p_k, \overline{p}) \\ q_k - p_k & \text{if } \sigma(q_k, \overline{q}) = \sigma(p_k, \overline{p}) \end{cases} \qquad \text{[D]}$$

where $\delta \in (0,1]$ becomes smaller with salience, converging with rational thinking as $\delta \rightarrow 1$. So, salience pushes preferences towards the highest quality/price ratios. If an individual is choosing between multiple consumables along a rational indifference curve, each consumable has the same quality/price ratio.

This leaves us with the governing intuition for this study: with a targeted health nudge, the quality salience ($\sigma(q_k, \overline{q})$) of the targeted health services (k) increases and re-weights preferences (δ) towards those services relative to all other services. However, price saliency ($\sigma(p_k, \overline{p})$) is also inherent as part of the campaign. As an example, the back-pain campaign promotes planning for flare ups which entails considering lower-intensity visits and more costly options like injections and discectomies. Here, the quality of mitigating emergency room visits and the prices of these options are both plausibly salient. This does not require the patient to be newly convinced of this, but rather that this sense of quality is newly top-of-mind (that is, salient). If this quality saliency is larger than the associated price saliency (for example, $\sigma(q_k, \overline{q}) > \sigma(p_k, \overline{p})$), then the informational push would re-weight the value of the targeted health service to become relatively more desirable and more likely to be used (for example, $\frac{2}{1+\delta}qk - \frac{2\delta}{1+\delta}pk$, per Equation [D] above). If price saliency is larger than quality saliency, then the reverse effect would instead take place also according to per Equation [D]. Given Weber's law of diminishing sensitivity that is built into the model, at lower price levels there are steeper trade-offs between quality and price. So, lower-cost targeted services may be more sensitive to quality salience.

Intuitively, we can adapt an indifference curve framework to illustrate these effects (see Figure 1). Here, individuals have an initial indifference curve (A) subject to a budget constraint leading to consumption bundle (a). A health nudge does not affect prices, the budget constraint, or the perceived value of all goods and services other than the health services targeted by the nudge, but it does make the perceived value and price of the targeted health service more salient. In this case, the quality is more salient than price (for example, $\sigma(q_k, \overline{q}) > \sigma(p_k, \overline{p})$), so utility is positively re-weighted to relatively prefer that service (for example, $\frac{2}{1+\delta}q_k - \frac{2\delta}{1+\delta}p_k$) as opposed to without the nudge (for example, $q_k - p_k$). This leads to a new indifference curve (B), not a different level set of the previous indifference curve (A), which implies a new consumption bundle (b). This new bundle reflects a relatively increased preference for more of the targeted health services stemming from the targeted nudge which increased quality saliency and re-weighted the utility function.

Figure 1: Stylistic Representation of Change in Indifference Curve



Notes: As the exogenous shock provides information about a target health service, relative preferences change inducing a new indifference curve that prefers more (or less) of at service. The indifference curves here are not level sets of the same structure, but instead entirely new curves.

B. Testable Hypotheses Stemming from this Theory

I evaluate a range of informational interventions with visibility into between- and within-service reimbursed cost variation. While I do not have visibility into beliefs about quality, I do have visibility into prior use of health services related to the informational intervention, reflective of prior quality/price assessments that led to utilization choices. Accordingly, I test two hypotheses predicted by the above theory: (1) responses to informational nudges are on average sensitive to price regardless of unobserved variations in quality, with lower priced services leading to relatively higher utilization patterns following the nudge, and (2) informational nudges are more effective among individuals with prior utilization as their past choices reflect their prior opinions of service value brought to the forefront by the nudge.

I make several assumptions in this evaluation. The first assumption is that the choice sets among recipients are uniform within each campaign, especially given that the informational nudges were not explicit in suggesting the same choice set. The second assumption, related to the first, is that the effectiveness of these messages are, at least in part, a function of the attributes of the related services and not simply reflective of differences in the messages themselves. The third assumption is that people are aware of the prices of the services contemplated, at least ordinally. The fourth assumption is that price saliency generally increases at the extremes – that is, higher and lower – of reimbursed costs *across* all campaigns (and thus across all choice sets) and not just *within* each campaign. This allows me to pool all campaign outcomes and evaluate the full set of price variation. The fifth assumption, embedded in the hypothesis, is that at least some of the variation in price is independent of unobserved variation in quality.

IV. Data and Empirical Strategy

A. Overview

I utilize four years of health care claims data in which 31,797 individuals were specifically targeted for at least one program intervention and 9,179 individuals were specifically targeted (but ultimately excluded) for at least one program intervention (see list of programs with measured outcomes in Table 1 and details on the target populations by group in Table 2). The data used span three years pre- and one-year post-program launch. This leaves the unit of measurement as member-by-year observations.

B. Merits and Challenges to Identification

There are several attributes that make the treatment and control group plausibly comparable. All employees work for the same employer and had the same uniform set of insurance offerings in each year of the dataset. So, any annual changes in insurance offerings and their associated prices are uniform across treatment and controls for each year of the data. All employees were targeted at the individual level with full intent to treat up until the point when the campaign was canceled for the control group. The panel is longitudinal and includes only employees. I balanced the panel to ensure that any measured differences are estimated off the same population rather than off cohort differences. The decision to exclude controls was according to the organizational situation and not due to members' individual traits. Since the choice of who to receive the nudge was unrelated to those individuals save for their business unit affiliation, there is have a potentially credible control group for comparison.

Given that the control group emerged due to a re-organization, this may imply a fear of losing one's job and benefits among that group in which the control group may tend towards overutilization in comparison. In this case, the treatment group may appear to have a relative decrease in utilization across broad categories of care relative to the control group. The implication here is that positive estimates would experience downward bias and thus would be lower bound estimates. All negative estimates, in contrast, would have upward bias in magnitude and these estimates would be upper bounds and should be viewed with caution. Finally, if there are few negative estimates or a high share of estimates with no measured effects, then this suggests that bias stemming from the re-organization is minimal.

It is also possible, though unlikely, that there are spillovers between treatment and control groups. Employees in the treatment group could discuss health decisions informally with employees in the control group. If present, this would attenuate the effects observed as the control group would increase or decrease care in the same direction that the treatment group changes their consumption. In this case, the estimates would represent a lower bound since their relative effects are muted. This is unlikely, though, since these business units operate largely in different facilities and locations.

C. Formal Control Group Tests

Although the employees were targeted in the same manner across treatment and control groups, there are still several potential threats to the valid estimation of casual effects. Chief among these concerns is that it is possible that there were cohort differences through time between employees across the treatment and control groups. Compounding this concern is that I am unable to directly analyze demographic information necessary to verify potential group compositional differences due to limitations in data access. In lieu of this, however, I was provided an individual time-varying risk score variable that is generated to predict the costliness of each individual member. This variable takes into consideration the member's age, sex, race, location, diagnoses, and past health care consumption. In case there are cohort differences through time between employees across the treatment and control groups, I can include these risk scores. However, these ideally would not differ across groups.

To formally test the validity of the control groups, as an alternative to testing whether a vector of covariates between treated and control groups are jointly zero, I instead pursue two approaches to assess the suitability of the control group: (1) conduct t-tests to compare pre-period mean differences for risk scores and visits associated with each campaign, and (2) do an event study to assess pre-period parallel trends between treatment and control groups for all statistically significant estimates generated under a differences-and-differences model, and further conduct a joint F-test for the pre-period treatment group indicator variables.

I find significant differences in the mean risk scores for patients for five of the nine conditions. To do this, I assessed whether the pre-period risk-scores are statistically significantly different between treatment and control groups using a t-test with a 90% or higher confidence level cutoff (see Table 2). Doing the same for care utilization, all but two p-values are statistically significant at the 90% or higher confidence level. This suggests that the pre-period levels are not comparable. In my difference-in-differences specification, I include the time-varying patient risk scores and patient fixed effects to account for cohort differences.

While these intercept values may differ, I am primarily concerned about the parallel trends between treatment and control groups in the pre-period. To evaluate this and to see if there is enough motivation for the inclusion of group-specific time-trends in my difference-in-differences specifications described below (see Equation [2]), I conduct an event study for each estimate that is statistically significant under the difference-in-differences framework. The methodology used here is as follows:

$$Y_{ijt} = \sum_{t=-2}^{+1} \alpha_{jt} \text{ x } Policy_{jt} \text{ x } 1\ [(t - T_i) = t] + \tau_t + \sigma_i + \lambda\ Policy_j + \delta\ State_i + \varphi\ Covariates_{it} + \varepsilon_{ijt} \quad [1]$$

where Y_{ijt} is the outcome of interest (that is, annual utilization, which is imputed to a full twelve month average in the event that enrollment was less than twelve months), α_{jt} is a series of relative time fixed effects (for example, the effect of being in the target group for each year), T_i is when the shock occurred for individual (i), τ_t captures year fixed effects (which includes the post period), σ_i captures individual fixed effects, λ is the average effect of being part of the targeted group for the program, δ is the average effect for being in a given state, φ covers the average patient risk score (comprised of disease, gender, and age considerations), number of overlapping programs, and indicators for the specific programs, and ε_{ijt} is the error term.

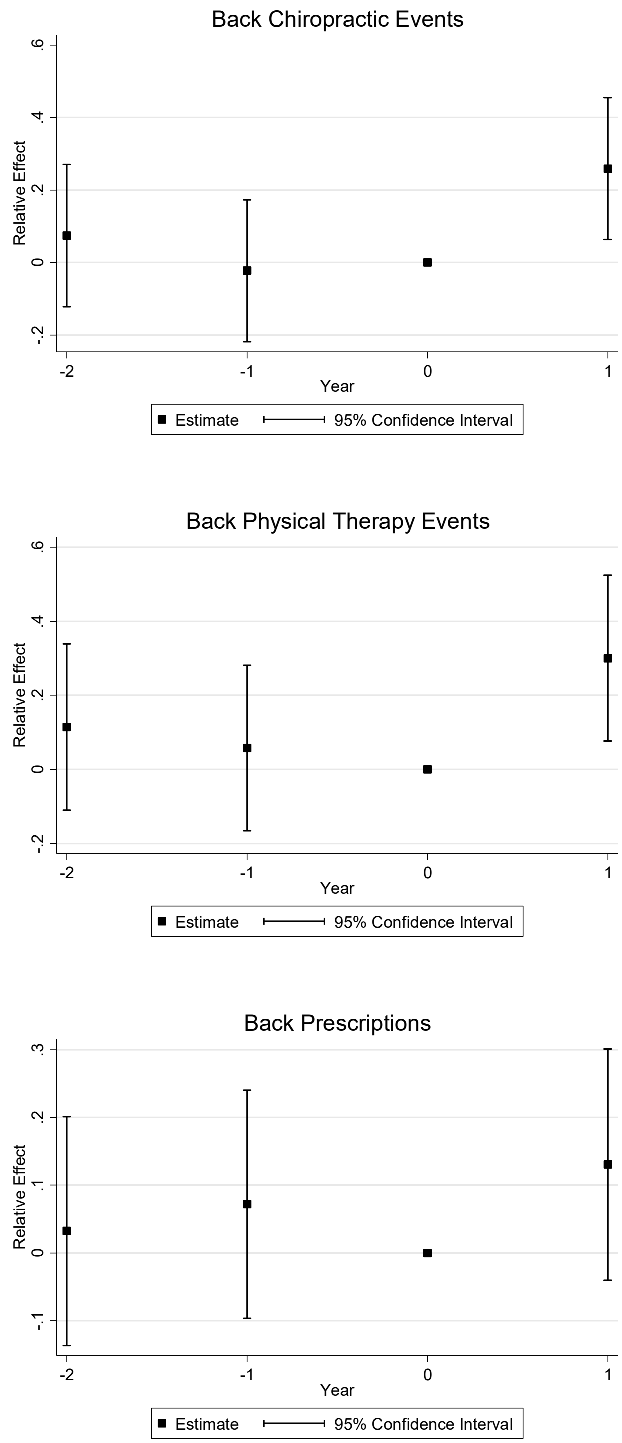
By and large, the pre-period trends appear flat for most outcomes in which the difference-in-differences specifications yielded statistically significant estimates (see Figures 2 through 4). I also tested whether the pre-period treatment indicator estimates were jointly zero using an F-test approach (see Table 3). In short, the pre-period estimates are not jointly zero at the 90% or greater confidence level for preventative physicals and the two hip and knee programs, but are jointly zero for the remaining scenarios. The preventative physical pre-period values, as seen in the event study, are not parallel. This is due to a pre-period drop in the control group and is accordingly not a valid comparison. The hip and knee campaigns, however, show slightly decreasing pre-trends followed by discontinuous increases following the campaign launch and can thus still provide insights as controls. This pattern suggests that the estimates summarized in a difference-in-differences framework with this data are biased towards zero without the inclusion of trends based only upon pre-period values (Goodman-Bacon, 2018). Accordingly, any estimates generated from this comparison are likely underestimates and can still be meaningfully considered. All other pre-periods appear suitable under both the joint F-test and event study frameworks.

Table 3: Pre-Period F-Tests for Statistically Significant Event Study Results

	Back			Hip and Knee		Preventative	Weight		
	(1) Chiro	(2) PT	(3) Rx	(4) PCP	(5) Image	(6) Visits	(7) Appraisals	(8) BMI	(9) Dietician
$p > F$	0.60	0.61	0.70	0.05	0.05	<0.01	0.82	0.63	0.44

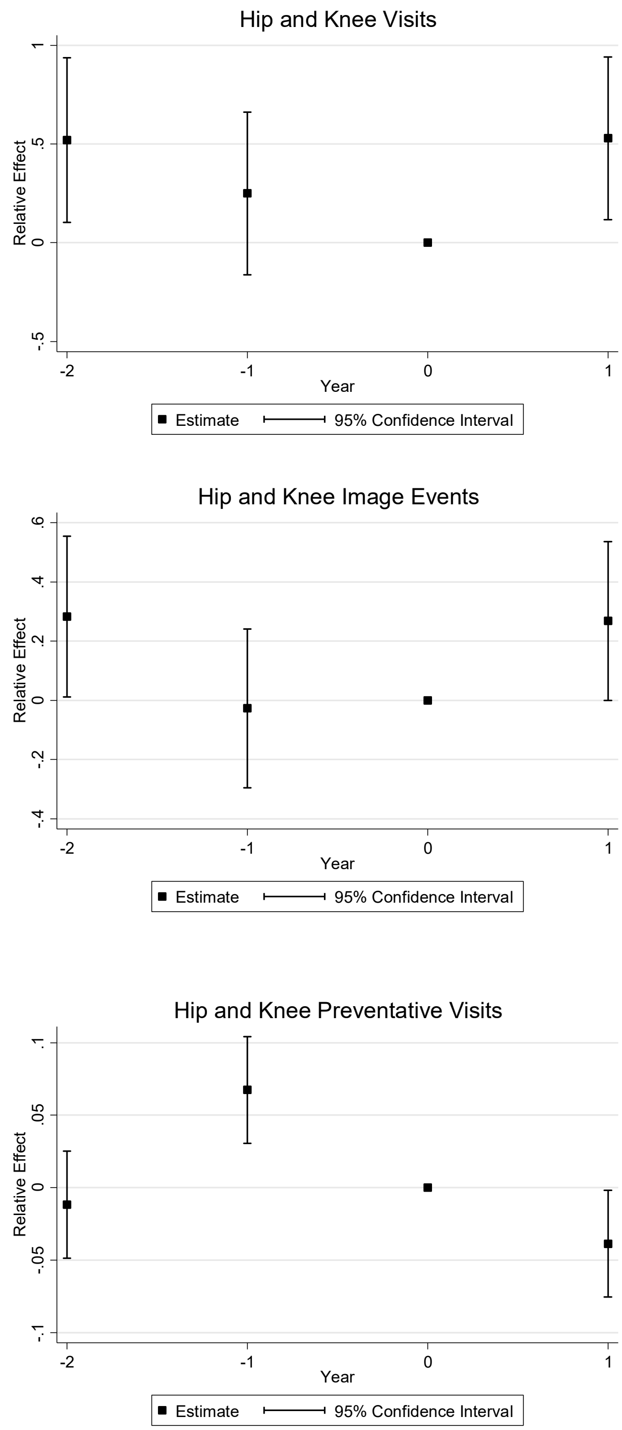
Notes: F-test tests the joint probability of the pre-period event study estimates as being zero (the third pre-period estimate for the period immediately prior to the program launch is the omitted variable in the regression).

Figure 2: Event Study Charts for the Back Campaign



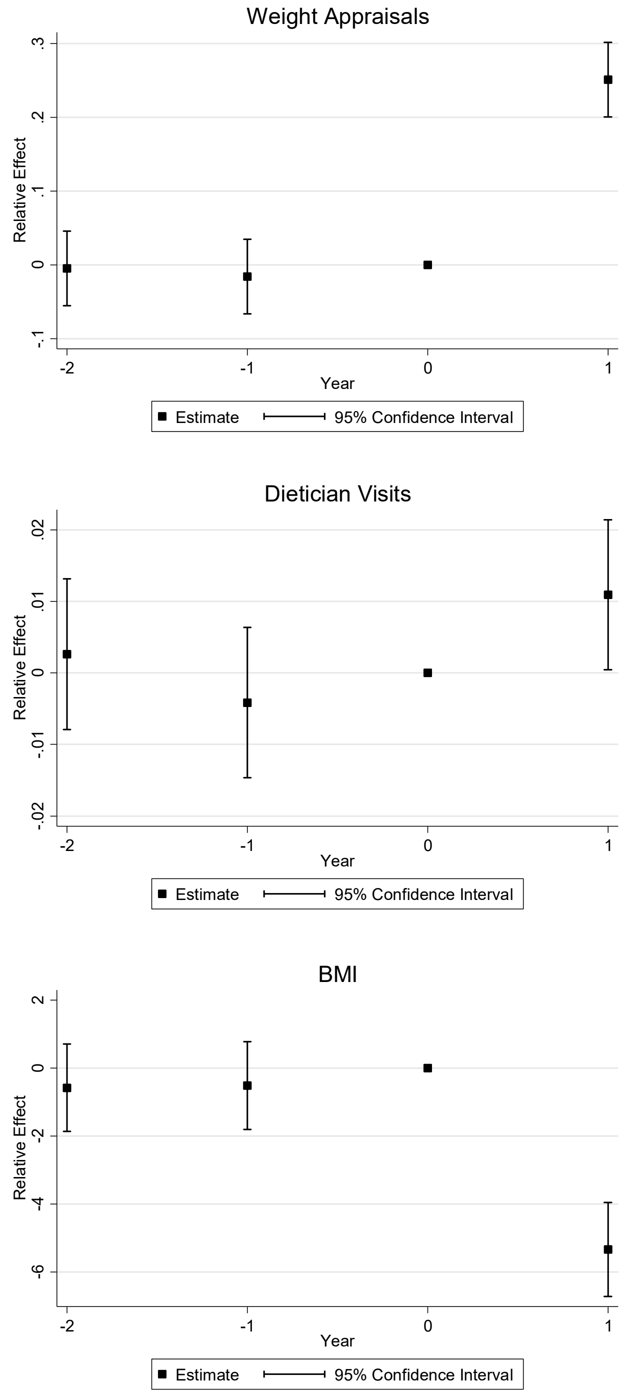
Notes: Y-axis values are differences in per-employee annual mean utilization measures. Estimates come from Equation [1]. Outcomes evaluated under the back campaign appear flat in the pre-period and show discontinuous increases upon campaign launch.

Figure 3: Event Study Charts for the Hip and Knee and Preventative Campaigns



Notes: Y-axis values are differences in per-employee annual mean utilization measures. Estimates come from Equation [1]. Outcomes evaluated under the hip and knee campaigns show slightly decreasing pre-trends followed by discontinuous increases after campaign launch. The preventative physical pre-period values are not parallel due to a pre-period drop in the control group.

Figure 4: Event Study Charts for the Weight Campaign



Notes: Y-axis values are differences in per-employee annual mean utilization measures (and, for BMI, differences in that measure). Estimates come from Equation [1]. Outcomes evaluated under the weight campaign show that weight appraisals and dietician visits have relatively flat pre-period trends with increases after the campaign launched. BMI shows flat pre-trends followed by a discontinuous decrease following the campaign launch.

D. Difference-in-Differences

This regression specification seeks to measure the average effect of being in the treatment group in the post-period for a given campaign and outcome while controlling for remaining relevant cohort differences. This uses employee-by-year observations and is specified as follows:

$$Y_{ijt} = \alpha + \beta \text{ Policy}_j \text{ x Post}_t + \tau_t + \sigma_i + \lambda \text{ Policy}_j + \delta \text{ State}_{it} + \varphi \text{ Covariates}_{it} + \gamma \text{ Trends}_{jt} + \varepsilon_{ijt} \quad [2]$$

where Y_{ijt} is the outcome of interest (that is, annual utilization) varying by individual (i) in business unit (j) across time (t) and β is the treatment effect. The Post_t variable is an indicator variable that is zero for observations prior to the launch of the program and one for observations after the launch of the program. The Policy_j variable varies at the business unit level and is one for the treated group and zero for the control group. The interaction of the two comprise the treatment variable from which β is estimated. As for the rest of the equation, α is the intercept, τ_t captures year fixed effects, σ_i captures individual fixed effects, λ is the average effect of being part of the targeted group for the program, δ is the average effect for being in a given state, φ covers the average patient risk score (comprised of disease, gender, and age considerations), number of overlapping programs, and indicators for the specific programs, γ covers treatment group-specific linear trends, and ε_{ijt} is the error term. Regression estimates are generated for each set of relevant utilization outcomes for each campaign as outlined in Table 1.

E. Price and Utilization

To compare effects across campaigns and outcomes, I conduct a triple differences regression where I interact treatment, post, and an indicator variable for high cost services. I do this for three different levels of high cost: $\geq 25^{th}$ Percentile, $\geq 50^{th}$ Percentile, and $\geq 75^{th}$ Percentile. These percentiles are based upon the annual mean costs for each outcome evaluated. I evaluate two outcomes: an indicator variable for the use of the targeted service and total units of that service utilized. Since there are several campaigns and several outcomes, the sample is constructed as person-by-campaign-by-outcome-by-year rather than person-by-year (and evaluated for each outcome). I then control for campaign and person using fixed effects for each. Finally, people face prices that they often do not act upon and thus do not appear in claims data. To address this, I imputed the average reimbursed cost-by-state-by-year-by-service outcome. For observations where the service was rendered and reimbursed, that cost value was retained. All other observations used the imputed costs. The regression was specified as follows:

$$\begin{aligned} Y_{ijkt} = {} & \alpha + \beta \text{ Policy}_{jk} \text{ x Post}_t \text{ x High Cost}_i + \rho \text{ Policy}_{jk} \text{ x Post}_t + \phi \text{ Policy}_{jk} \text{ x High Cost}_i \\ & + \mu \text{ Post}_t \text{ x High Cost}_i + \lambda \text{ Policy}_{jk} + \theta \text{ High Cost}_i \\ & + \tau_t + \sigma_i + \delta \text{ State}_{it} + \gamma \text{ Campaigns}_{ik} + \varphi \text{ Covariates}_{it} + \varepsilon_{ijkt} \end{aligned} \tag{3}$$

Where Y_{ijt} is the outcome of interest (that is, any utilization or total utilization) varying by individual (i) in business unit (j) for campaign (k) across time (t), α is the intercept, β is the estimated treatment effect driven by cost, ρ is the estimate on being in the treatment group in the post period, ϕ is the estimate on being in the treatment group and high cost, μ is the estimate on being high cost in the post period, λ is the average effect of being part of the targeted group for the program, θ is the effect of being high cost, and γ is the campaign fixed effect. As above, τ_t captures year fixed effects, σ_i captures individual fixed effects, δ is the average effect for being in a given state, φ covers the average patient risk score, number of overlapping programs, and indicators for the specific programs, and ε_{ijt} is the error term.

F. Priors Use and Utilization

To assess the role of prior beliefs, I conduct a separate triple differences regression where I interact treatment, post, and an indicator variable for any utilization in the pre-period. As with the previous triple differences regression, the sample is person-by-campaign-by-outcome-by-year. The regression was specified as follows:

$$\begin{aligned} Y_{ijkt} = \; & \alpha + \beta \text{ Policy}_{jk} \text{ x Post}_t \text{ x Prior Use}_i + \rho \text{ Policy}_{jk} \text{ x Post}_t + \phi \text{ Policy}_{jk} \text{ x Prior Use}_i \\ & + \mu \text{ Post}_t \text{ x Prior Use}_i + \lambda \text{ Policy}_{jk} + \theta \text{ High Cost}_i \\ & + \tau_t + \sigma_i + \delta \text{ State}_{it} + \gamma \text{ Campaigns}_{ik} + \varphi \text{ Covariates}_{it} + \varepsilon_{ijkt} \end{aligned} \tag{4}$$

Where Y_{ijt} is the outcome of interest (any utilization or total utilization) varying by individual (i) in business unit (j) for campaign (k) across time (t), α is the intercept, β is the estimated treatment effect driven by prior utilization, ρ is the estimate on being in the treatment group in the post period, ϕ is the estimate on being in the treatment group and with prior use, μ is the estimate on those with prior use in the post period, λ is the average effect of being part of the targeted group for the program, θ is the effect of having prior use, and γ is the campaign fixed effect. The rest are identical as in equation [3].

V. Results

In this section, I summarize in the difference-in-differences estimates from equation [2] and the triple differences estimates from equation [3] and equation [4]. In all, seven of the 34 utilization outcomes evaluated using difference-in-differences had statistically significant effects and all were increases. When evaluating the overall effect of price in a triple differences framework, on average, the higher the price the lower the effects of the program. When considering prior utilization, on average, those with prior use are more likely to utilize care following the program.

Many of the programs with statistically significant estimates in the difference-in-differences framework are associated with increases in utilization in lower-cost preventative services (for example, average cost of $237 among statistically significant estimates vs. full sample average cost of $5,344 among all services targeted). However, many other lower-cost preventative services were not associated with statistically significant estimates. A summary of the statistically significant results can be found in Table 4, and all estimates can be found organized by campaign in Section VIII Tables A1-A7. Briefly, the following services saw relative increases among the treatment group: chiropractic services and back physical therapy events for low-back pain, primary care visits for hip and knee pain and related imaging events, weight appraisals[2], and dietician visits.

[2] This result led to an increase of in body mass index (BMI). This is not to say that the intervention increased BMI. Instead, this suggests that those who might otherwise not have sought an appraisal had relatively higher BMIs than those who did.

Table 4: Summary of Statistically Significant Results

	Back			Hip and Knee		Preventative	Weight		
	(1) Chiro.	(2) PT	(3) Rx	(4) Image	(5) PCP	(6) Visits	(7) Weight Appraisals	(8) BMI	(9) Dietician Visits
Policy * Post	0.316** (0.129)	0.358** (0.147)	0.190* (0.111)	0.464*** (0.177)	0.793*** (0.272)	-0.069*** (0.024)	0.253*** (0.033)	2.24*** (0.779)	0.014** (0.007)
Mean	0.883	0.904	0.896	0.316	0.530	0.150	0.956	23.4	0.00749
Mean Cost	$332	$246	$256	$180	$191	-	-	-	-
Policy / Mean	36%	40%	21%	147%	150%	-46%	26%	10%	187%
Valid Pre-Trends?	✓	✓	✓	✓	✓		✓	✓	✓
N	32,396	32,396	32,396	1,416	1,416	22,176	48,208	48,208	48,208

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses. From Equation [2].

The estimates that are not statistically significant generally included higher-intensity services (for example, injections or surgeries for pain or emergency room visits) or non-pain related chronic conditions (for example, all outcomes evaluated for cardiac and diabetes care). Despite increases in lower-intensity care for back and hip and knee pain campaigns, higher-intensity treatments that the program sought to mitigate like injections and surgeries were not statistically significantly affected. The cardiac campaign was associated with no statistically significant effects for visits, lab tests, blood pressure readings, electro-cardiogram readings, surgeries, or heart attacks. The diabetes campaign also logged no statistically significant estimates for medical services including office visits, emergency room visits, surgeries, and other acute and non-acute visits. Pharmacy use and spend also did not change for insulin, non-insulin, or overall scripts or costs in response the diabetes campaign. Preventative services did not have valid parallel trends and so the effects are less clear, while emergency room visits were not statistically significantly affected. Cancer visits and costs were unaffected.

When using the triple differences specifications for high cost, price does appear to influence how the program functioned with higher cost services seeing reductions in the post period for the treatment group relative to lower cost services for both any utilization and total utilization outcome variables. For the outcome variable any utilization, point estimates range from a relative reduction for higher cost services of between 16.2% and 16.9% relative to the mean depending on whether high cost was defined as 25^{th}, 50^{th}, and 75^{th} percentile cut-offs. For total utilization, point estimates range from a relative reduction for higher cost services of between 18.5% and 28.9% relative to the mean depending on whether high cost was defined as 25^{th}, 50^{th}, and 75^{th} percentile cut-offs. All estimates are statistically significant at the 95% or 99% confidence level (see Table 5).

Table 5: Triple Differences for Costs and Utilization

	Any Utilization			Total Utilization		
	(1)	(2)	(3)	(4)	(5)	(6)
Policy * Post * High Cost	-0.026*** (0.004)	-0.028*** (0.005)	-0.025*** (0.006)	-0.142*** (0.029)	-0.105*** (0.031)	-0.091** (0.038)
Mean	0.154	0.154	0.154	0.491	0.491	0.491
DDD / Mean	-16.9%	-18.2%	-16.2%	-28.9%	-21.4%	-18.5%
High Cost:						
≥ 25^{th} Percentile = $548	✓			✓		
≥ 50^{th} Percentile = $4,926		✓			✓	
≥ 75^{th} Percentile = $17,591			✓			✓

Notes: N = 616,104. * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses. From Equation [3].

When using the triple differences specifications for prior utilization, prior use also appears to influence how individuals responded to the program with those with prior use beings between 39.0% and 41.5% more likely to respond relative to the mean. This varied, if only slightly, according to whether different levels of outliers were removed from the sample. Similarly, those with prior utilization increased total utilization levels between 19.1% and 26.7% relative to mean total utilization, again varying according to the degree to which outliers are excluded from the sample (see Table 6).

Table 6: Triple Differences for Prior Utilization

	Any Utilization				Total Utilization			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy * Post * Prior Use	0.060*** (0.004)	0.061*** (0.004)	0.061*** (0.004)	0.061*** (0.004)	0.094*** (0.031)	0.110*** (0.024)	0.084*** (0.014)	0.111*** (0.028)
Mean	0.154	0.153	0.147	0.151	0.491	0.440	0.315	0.453
DDD / Mean	39.0%	39.9%	41.5%	40.4%	19.1%	25.0%	26.7%	24.5%
Full Sample	✓				✓			
≤ 99^{th} Percentile Among Non-Zero Any Use = 26		✓				✓		
≤ 99^{th} Percentile Any Use = 11			✓				✓	
≤ 99^{th} Percentile Any, Balanced				✓				✓
Observations	616,104	615,188	610,105	610,032	616,104	615,188	610,105	610,032

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses. From Equation [3]. Balanced (Panel) is constructed by removing the full set of observations contributed by individuals with greater than 99^{th} percentile use across the average of all their observations.

VI. Discussion

Despite the increasing interest in understanding behavioral interventions in health care, the effects of these programs have remained unclear, and recent randomized studies in other settings find information to be insufficient by itself. I take advantage of a natural experiment with a plausibly exogenous assignment between treatment and control group to assess when targeted health information can change behavior by including price variation of the health services that are related to the nudge. I find that this informational nudge alone appears to have increased the utilization of many but not all lower-intensity preventative services, and that it did not affect the utilization of costlier services. These heterogeneous effects hold in a triple differences framework across multiple high-cost cut-off thresholds where higher cost services in the post-period of the treatment group are associated with lower relative levels of utilization.

These results appear to be non-random. In all, seven of the thirty-five difference-in-differences utilization outcomes evaluated had statistically significant effects with valid pre-trends and all were increases. This is about twice the number of statistically significant estimates expected at random with a 90% confidence level, and most of these estimates are statistically significant at the 95% or 99% confidence levels. The seven statistically significant utilization estimates with valid pre-trends are also amassed among only three campaigns (back, hip and knee, and weight) in which only thirteen utilization outcomes are evaluated. This implies that these three campaigns were particularly effective given that one would expect only one or two outcomes to be statistically significant at the 90% confidence level by random chance. This also suggests that the other campaigns were generally ineffective: the ER campaign had one utilization outcome evaluated while the preventative campaign had two, the diabetes campaign had ten, and the cardiac campaign had six and no estimates were statistically significant for any of these campaigns (save for the ER campaign whose pre-period trends invalided the control group and estimation). Given the concern noted above about potential bias stemming from control group relative overutilization, the fact that all statistically significant estimates with valid pre-trends are positive suggests that these estimates may be lower bounds.

My hypotheses stemming from the theory were that relatively lower priced services and those with prior use would be associated with relatively higher utilization patterns following the informational nudge. I find some evidence that both predictions bear out. First, all statistically significant difference-in-differences utilization estimates are positive and associated with lower-cost preventative services whose average is only a few hundred dollars compared with the roughly five-thousand-dollar average cost from the full sample. Second, higher cost services do predict lower relative utilization compared with lower cost services. Finally, prior utilization is associated with higher utilization levels following the nudge.

A. Limitations

I am limited in the amount of follow-up data with just one year. Where there are increases in ambulatory services, there is little follow-up time to adequately assess how these increases may have mitigated related acute services (for example, increased back physician therapy and chiropractic services may reduce injections and surgeries in subsequent years). This suggests that the program effects may occur outside the window of data I am afforded, which is possible. The fact that these are targeted campaigns introduce a selection concern pertaining to external validity. These are not randomly assigned campaigns within a general population. Instead, participants are selected into the intervention as being high-risk for the campaign of interest as employees on private health insurance within this large firm. So, the estimates generated are not to be interpreted as the overall likely effect for a general population, but rather a potential effect for high-risk privately insured and employed individuals. As noted above, I make several assumptions regarding price, quality, prior beliefs, and choice sets to interpret my estimations. Finally, the lack of visibility into all e-mail templates used as well as the additional information accessed after each click-through limits my ability to comment on the quality of the nudge and how it varies by campaign.

B. Policy Implications

This paper provides evidence that informational nudges targeting employees according to their past health records can increase their subsequent consumption choices for some lower-cost preventative services, but it is unlikely to have short-term effects in high-cost care. Consequently, it is also unlikely to decrease costs overall in the short term as all estimates implied increases in utilization and not decreases. Considering prior use will likely help to improve predictability of the expected effects. Finally, this information is most likely useful for self-insured employers, governments, or health insurers considering low-cost interventions to shift health behaviors, but may be relevant in other settings in which targeted informational nudges are deployed.

VII. References

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. "Behavioral hazard in health insurance." *The Quarterly Journal of Economics* 130, no. 4 (2015): 1623-1667.

Bergman, Peter, Jeffrey T. Denning, and Dayanand Manoli. "Is information enough? The effect of information about education tax benefits on student outcomes." *Journal of Policy Analysis and Management* 38, no. 3 (2019): 706-731.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment." *The Quarterly Journal of Economics* 127, no. 3 (2012): 1205-1242.

Bhargava, Saurabh, and Dayanand Manoli. "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment." *American Economic Review* 105, no. 11 (2015): 3489-3529.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek. "The role of information in the take-up of student loans." *Economics of Education Review* 31, no. 1 (2012): 33-44.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. "Salience and consumer choice." *Journal of Political Economy* 121, no. 5 (2013): 803-843.

Bourgeois, Florence, William Simons, Karen Olson, John Brownstein, and Kenneth Mandl. "Evaluation of influenza prevention in the workplace using a personally controlled health record: randomized controlled trial." *Journal of medical Internet research* 10, no. 1 (2008): e5.

Chen, Nuole, Kris-Stella Trump, Stacy Hall, and Quan Le. "The Effect of Postcard Reminders on Vaccinations Among the Elderly: A Block-Randomized Experiment."

Chetty, Raj, and Emmanuel Saez. "Teaching the tax code: Earnings responses to an experiment with EITC recipients." *American Economic Journal: Applied Economics* 5, no. 1 (2013): 1-31.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103, no. 7 (2013): 2683-2721.

Darolia, Rajeev, and Casandra Harper. "Information use and attention deferment in college student loan decisions: Evidence from a debt letter experiment." *Educational Evaluation and Policy Analysis* 40, no. 1 (2018): 129-150.

Fenerty, Sarah D., Cameron West, Scott A. Davis, Sebastian G. Kaplan, and Steven R. Feldman. "The effect of reminder systems on patients' adherence to treatment." *Patient preference and adherence* 6 (2012): 127.

Findley, Sally E., Matilde Irigoyen, Donna See, Martha Sanchez, Shaofu Chen, Pamela Sternfels, and Arturo Caesar. "Community-provider partnerships to reduce immunization disparities: field report from Northern Manhattan." *American Journal of Public Health* 93, no. 7 (2003): 1041-1044.

Goodman-Bacon, Andrew. “Difference-in-differences with variation in treatment timing”. No. w25018. National Bureau of Economic Research (2018).

Hoxby, Caroline, and Sarah Turner. “Expanding college opportunities for high-achieving, low income students.” *Stanford Institute for Economic Policy Research Discussion Paper* 12-014 (2013).

King, Dominic, Felix Greaves, Ivo Vlaev, and Ara Darzi. “Approaches based on behavioral economics could help nudge patients and providers toward lower health spending growth.” *Health Affairs* 32, no. 4 (2013): 661-668.

Leight, Jessica, and Elana Safran. “Increasing immunization compliance among schools and day care centers: Evidence from a randomized controlled trial.” *Journal of Behavioral Public Administration* 2, no. 2 (2019).

Leight, Jessica, and Nicholas Wilson. “Framing Flexible Spending Accounts: A Large-Scale Field Experiment on Communicating the Return on Medical Savings Accounts.”

Manoli, Dayanand S., and Nicholas Turner. *Nudges and learning: evidence from informational interventions for low-income taxpayers*. No. w20718. National Bureau of Economic Research, 2014.

Marteau, Theresa M., et al. "Judging nudging: can nudging improve population health?." Bmj 342 (2011): d228.

McCaul, Kevin D., Rebecca J. Johnson, and Alexander J. Rothman. “The effects of framing and action instructions on whether older adults obtain flu shots.” *Health Psychology* 21, no. 6 (2002): 624.

Reisch, Lucia A., Cass R. Sunstein, and Wencke Gwozdz. "Beyond carrots and sticks: Europeans support health nudges." Food Policy 69 (2017): 1-10.

Neff, Roni, and Jillian Fry. "Periodic prompts and reminders in health promotion and health behavior interventions: systematic review." Journal of medical Internet research 11.2 (2009): e16.

Stone, Erin G., Sally C. Morton, Marlies E. Hulscher, Margaret A. Maglione, Elizabeth A. Roth, Jeremy M. Grimshaw, Brian S. Mittman, Lisa V. Rubenstein, Laurence Z. Rubenstein, and Paul G. Shekelle. “Interventions that increase use of adult immunization and cancer screening services: a meta-analysis.” *Annals of internal medicine* 136, no. 9 (2002): 641-651.

Taylor, Shelley E., and Suzanne C. Thompson. “Stalking the elusive” vividness” effect.” *Psychological review* 89, no. 2 (1982): 155.

VIII. Appendix

Table A1: Responses to the Back-Related Campaign

	(1) Visit	(2) Chiro.	(3) PT	(4) Image	(5) Injections	(6) Surgery	(7) Rx	(8) ER
Policy * Post	0.066 (0.044)	0.316** (0.129)	0.358** (0.147)	-0.006 (0.008)	-0.006 (0.011)	-0.012 (0.011)	0.190* (0.111)	-0.022 (0.034)
Mean	0.344	0.883	0.904	0.030	0.024	0.024	0.896	0.0295
Mean Cost	$225	$332	$246	$981	$2,030	$17,302	$256	$633
Policy / Mean	19%	36%	40%	-20%	-25%	-50%	21%	-74.6%
N	32,396	32,396	32,396	32,396	32,396	32,396	32,396	32,396

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A2: Responses to the Hip and Knee Campaign

	(1) Injections	(2) Surgeries	(3) Image	(4) PCP
Policy * Post	-0.004 (0.018)	0.084 (0.086)	0.464*** (0.177)	0.793*** (0.272)
Mean	0.00430	0.0847	0.316	0.530
Mean Cost	$256	$28,563	$180	$191
Policy / Mean	-93%	99%	147%	150%
N	2,068	1,416	1,416	1,416

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A3: Responses to the Preventative, Weight, and ER Campaigns

	Preventative		Weight			ER
	(1) E&M	(2) Visits	(3) Weight Appraisals	(4) BMI	(5) Dietician Visits	(6) Visits
Policy * Post	-0.087 (0.123)	-0.069*** (0.024)	0.253*** (0.033)	2.24*** (0.779)	0.014** (0.007)	0,004 (0,023)
Mean	1.80	0.150	0.956	23.4	0.00749	0.241
Mean Cost	-	-	-	-	-	$3,585
Policy / Mean	-5%	-46%	26%	10%	187%	1667%
N	22,176	22,176	48,208	48,208	48,208	54,404

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A4: Medical Service Responses to the Diabetes Campaign

	(1) Visits	(2) ER Visits	(3) Surgery	(4) Non-Surgery	(5) Acute Visits	(6) Non-Acute Visits
Policy * Post	0.050 (0.059)	-0.003 (0.006)	0.013 (0.018)	-0.003 (0.026)	0.010 (0.032)	-0.006 (0.052)
Mean	0.456	0.005	0.00794	0.0189	0.0305	0.0588
Mean Cost	-	$1,936	$11,162	$2,677	$4,773	$1,927
Policy / Mean	11%	-60%	164%	-16%	33%	-10%
N	15,248	15,248	15,248	15,248	15,248	15,248

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A5: Drug Treatment Responses to the Diabetes Campaign

	(1) All Scripts	(2) Total Rx Costs	(3) Diuretics Rx	(4) Non-Insulin Rx	(5) Insulin Rx
Policy * Post	-0.044 (0.189)	-$186 (390)	0.143 (0.065)	-0.138 (0.155)	-0.063 (0.061)
Mean	2.35	$2,612	0.286	1.73	0.333
Mean Cost	$2,612	-	$76	$1,772	$4,819
Policy / Mean	-2%	-7%	50%	-8%	-19%
N	15,248	3,652	15,248	15,248	15,248

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A6: Cost Effects of the Cancer Campaign and Colon Campaign

	Cancer				Colon
	(1) E&M Visits	(2) Total Costs	(3) Medical Costs	(4) Drug Costs	(5) Colonoscopy
Policy * Post	-0.273 (0.394)	-$5,273 (8,758)	-$7,011 (8,599)	$9,497 (6,552)	-0.001 (0.009)
Mean	0.840	$26,547	$23,667	$10,071	0.019
Mean Cost	$26,547	-	-	-	-
Policy / Mean	-33%	-20%	-30%	94%	-5.3%
N	1,632	1,056	1,056	302	15,428

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.

Table A7: Responses to the Cardiac Campaign

	(1) Overall Visits	(2) PCP	(3) Lab Tests	(4) BP Events	(5) EKG Events	(6) Surgery	(7) Heart Attacks
Policy x Post	-0.291 (0.306)	0.062 (0.265)	0.047 (0.355)	-2.72 (2.53)	-0.161 (0.149)	-0.010 (0.148)	0.005 (0.056)
Mean	0.827	0.861	0.771	4.90	0.288	0.182	0.0568
Mean Cost	$261	$238	$150	$834	$854	$33,599	$31,720
Policy / Mean	-35%	7%	6%	-56%	-56%	-5%	9%
N	1,248	1,248	1,248	1,248	1,248	1,248	1,248

Notes: * p<0.1, ** p<0.05, and *** p<0.01. Standard errors in parentheses.