Received: 18 April 2019 | Revised: 30 August 2019 | Accepted: 4 September 2019

DOI: 10.1002/hec.3965

RESEARCH ARTICLE

WILEY Health Economics

Framing Flexible Spending Accounts: A Large-Scale Field Experiment on Communicating the Return on Medical Savings Accounts

Jessica Leight[1] | **Nicholas Wilson**[1,2]

[1] International Food Policy Research Institute, Washington, DC, USA

[2] Economics, Reed College, Oregon, USA

Correspondence
Economics, Reed College. USA.
Email: nwilson@reed.edu

Abstract

Tax-preferred health savings devices such as Flexible Spending Accounts (FSAs) and Health Savings Accounts (HSAs) offer employees potentially valuable financial instruments for directing pre-tax earnings to eligible medical expenses. Despite their increasing popularity as an employee benefit, however, there is little causal evidence around individual demand for these accounts. This paper seeks to address this gap in the literature, reporting on a randomized controlled field experiment conducted with over 11,000 U. S federal employees in 2017 in order to evaluate the effectiveness of targeted messages designed to increase FSA contributions. Our results suggest that the provision of basic information about FSAs delivered via an emailed employee newsletter did not affect the likelihood of contribution or the contribution level. The addition of statements about the absolute returns or relative returns offered by the accounts similarly had no significant effects, and these null effects are observed despite relatively high email open rates. We discuss explanations for the null results and the policy implications of findings from what appears to be the first health economics experiment analyzing tax incentives around health care savings.

KEYWORDS
health, savings, taxation

JEL CLASSIFICATION
D14; D91; H24; I11; I18

1 | INTRODUCTION

Tax-preferred health savings devices such as Flexible Spending Accounts (FSAs) and Health Savings Accounts (HSAs) offer employees financial instruments for spending pre-tax earnings on eligible medical expenses. These devices reduce the effective price of eligible medical expenses, and thus may increase the quantity of health services demanded for households who are price-sensitive or otherwise constrained in investing in such services; from a policymaking perspective, promoting the use of such accounts may be a viable strategy to increase household investment in health inputs. FSAs are a common employee benefit in the U.S., with three-quarters of large firms offering the accounts (Kaiser Family Foundation, 2016). Yet fewer than half of eligible employees participate (Mercer, 2008), and there is little large-scale evidence on

methods designed to increase participation in these savings devices. From a welfare perspective, the tax subsidies of FSAs may promote overuse (Pauly et al., 1995) or conversely, FSAs may help counteract the fact that behavioral hazard may lead individuals with insufficient precautionary savings to underuse health care (Baicker, Mullainathan, & Schwartzstein, 2015). Addressing part of this gap in the literature, this paper reports on a large-scale randomized controlled field experiment conducted with over 11,192 federal employees in 2017 in order to evaluate the effectiveness of three different targeted messages delivered via an employee email newsletter and designed to increase FSA contribution levels.

FSAs allow employees to contribute pre-tax earnings from their paycheck to an account from which they subsequently request reimbursement for eligible expenses. The United States created FSAs in the Revenue Act of 1978 with the goal of increasing household purchasing power for medical goods.[1] In 2010, the passage of the Patient Protection and Affordable Care Act (ACA) created a contribution limit, raised periodically and currently equal to $2,650 for an individual or married couples, and allowed for up to $500 to be rolled over into the subsequent year.[2] The Kaiser Family Foundation Employer Health Benefits 2016 Annual Survey of U.S. firms indicates that 13% of all firms and 75% of firms with 200 employees or more offer FSAs (Kaiser Family Foundation, 2016). Participation rates among eligible employees appear to be below 50%, and are closer to 25% at large firms (Mercer, 2008).[3]

This evaluation sought to analyze the effectiveness of three different messages delivered to employees and designed to increase FSA contributions. The first message, denoted the basic treatment, stated that an employee could "save money" by using a FSA and provided information about the ease of enrollment in the benefit as well as its utilization. The second message, denoted the absolute return treatment, provided the same information about the benefit, but also highlighted that an average employee would "save $949", given a specified estimated level of out-of-pocket expenses and an estimated marginal tax rate. The third message, denoted the relative return treatment, again provided basic benefit information, but framed the savings as "savings of 36.5%" for the same hypothetical employee. These messages were disseminated through a tri-weekly electronic mail newsletter on the same day at the start of the annual open enrollment period and again four weeks later; each employee assigned to a treatment arm received the same information twice. The fourth study arm received none of these communications and constitutes our control arm.

Our experiment reveals three key facts about FSAs and framing the return on savings via this tax device in this study population. First, participation rates are relatively low; approximately 25% of employees contributed to a FSA, and the mean contribution in the full sample was approximately $450; the mean contribution conditional on any positive contribution is $1775. Second, communicating FSA eligibility and the return on savings in the basic treatment did not increase use or amount contributed. Third, highlighting the absolute or relative savings similarly had no additional effect on increasing contributions. The estimated treatment effects for all three intervention arms are precisely estimated zeros.

Beyond the analysis specified in our pre-analysis plan, we conducted an exploratory analysis of email open and click-through rates.[4] Although approximately 50% of recipients opened the electronic mail newsletter, only around 1% clicked through on the FSA links embedded in the emails. Accordingly, the null result could reflect the fact that recipients did not find the framing statements particularly salient, or were already fully informed. Alternatively, the experimental evidence is consistent with the hypothesis that the perceived utility of the FSA among those not already using a FSA is limited, especially for employees whose out-of-pocket medical expenses are low or variable; the results are also consistent with the hypothesis that low participation rates reflect behavioral barriers linked to default risk and uncertainty around transaction costs that the experiment did not fully target. We explore these channels further in the Discussion section.

Our study provides some of the first causal evidence on demand for medical savings accounts of any form.[5] We build on a small set of studies of health savings accounts, including FSAs, HSAs, and other medical savings accounts. Existing

[1]The United States tax code also allows employers to offer Dependent Care FSAs and Limited Expense FSAs. Health Care FSAs appear to be more common than these other two types of FSAs. In our analysis, when we refer to FSA we are restricting our focus to Health Care FSAs.

[2]A maximum of $500 may be rolled over into the subsequent year. The ability to roll-over some funds may increase the probability of non-zero contributions. Yet the limit on the rollover amount may promote wasteful spending near the end of the fiscal year or inhibit any FSA contributions at all.

[3]While this information is relatively out of date, there is no large-scale published data that is more recent.

[4]At the time at which we had to file a pre-analysis plan we had not been able to confirm that we could access email open and click-through data. We discuss our pre-analysis plan in more detail in the Experimental Design section.

[5]There is a closely related body of literature on demand for retirement savings accounts (e.g., Individual Retirement Accounts (IRAs) and 401ks). For example, Duflo and Saez (2003), Thaler and Benartzi (2004), Chan and Stevens (2008), Choi, Laibson, and Madrian (2011), Collins and Urban (2016), Beshears, Choi, Laibson, and Madrian (2017), Dolls, Doerrenberg, Peichl, and Stichnoth (2018), and Messacar (2018) all provide evidence on demand for retirement savings accounts. Dolls et al. (2018) appears to be the only study demonstrating that a light-touch intervention like ours can increase participation in savings programs. Dolls et al. (2018) uses a regression discontinuity design to show that the annual German pension plan letter increased tax-deductible private savings.

economic analyses of FSAs study optimal participation levels (Bhattacharya, Schoenbaum, & Sood, 2002; Cardon & Showalter, 2007), actual participation and its correlates (Levy 1998, Hamilton & Marton, 2008, Cardon, Moore, & Showalter, 2012), and the effects of FSAs on coinsurance (Jack, Levinson, & Rahardja, 2006). The initial HSA literature examines the effects of HSAs on savings and insurance (Peter, Soika, & Steinorth, 2016; Steinorth, 2011; Ye, 2015). More recently, Leive (2018) uses a regression discontinuity design exploiting an employer's matching rate to measure the marginal propensity to consume from HSAs. Literature on other medical savings accounts examines the welfare effects of medical savings accounts (Hurley, Emmanuel Guindon, Rynard, & Morgan, 2008; Pauly & Herring, 2000; Zabinski, Selden, Moeller, & Banthin, 1999) and whether medical savings accounts in Singapore are sufficient to meet retirement needs (Chia & Tsui, 2005).

In addition, we expand the economic literature on framing the return on savings in health accounts. Schmitz and Ziebarth (2017) analyze a 2009 German regulatory reform in which insurers were required to change from expressing premium differences in terms of percentage point payroll tax differences to stating premium differences in absolute terms (i.e., euros). It found that the switch to absolute terms (euros) greatly increased willingness to switch plans by increasing the salience of the differences. We find that framing savings in absolute terms is no more and no less effective than framing savings in percentage terms and that neither communication was more effective than no communication. The magnitudes of the potential gains – on the order of 10–20 euros per month in Schmitz and Ziebarth (2017) and as much as $80 per month in our setting – do not appear to explain the differences in our findings.[6] One possible explanation is that the decision to choose a health insurance plan is much more salient than the decision to choose a medical savings account such as a FSA. Although Schmitz and Ziebarth (2017) find that the absolute framing increased the salience of the premium differences, the baseline attention to the decision to participate in health insurance at all is likely quite high. In contrast, our findings are consistent with high baseline inattention to FSAs and continued inattention despite direct communication to eligible employees, possibly because among the pool of individuals who do not participate in a FSA, the potential benefits of FSAs are not large.

Finally, we add to the broader economic literature on nudges designed to increase take-up of public benefits, including tax-based benefits (e.g., Bhargava & Manoli, 2015). These studies largely test the effects of paper-based nudges. We demonstrate that email-based nudges in and around the brief window of open enrollment are not sufficient to increase take-up. Furthermore, we demonstrate that re-framing the return on savings was similarly insufficient to stimulate take-up.

Our study appears to be the first health economics experiment using the income tax system, complementing other experiments and nudges in tax systems (Finkelstein et al. 2009, Chetty & Saez, 2013). It also appears to be one of the first large-scale health consumer nudge field experiment in economics, complementing existing large-scale experimental evidence on nudging health outcomes using implementation prompts (Milkman, Beshears, Choi, Laibson, & Madrian, 2011) and on nudging health providers using social comparisons (Sacarny, Yokum, Finkelstein, & Agrawal, 2016). Similarly, our analysis of demand for a government health program complements large-scale experimental evidence on other government health programs such as Medicaid (Finkelstein et al., 2012).

The rest of the analysis is organized as follows. Section 2 describes the experimental design. Section 3 discusses the data and statistical methods. Section 4 presents the results. Section 5 discusses the main findings and evaluates several hypotheses about the mechanism underlying the null results. Section 5 concludes.

2 | EXPERIMENTAL DESIGN

The experiment analyzed in this section is a randomized controlled trial conducted by the researchers as fellows with the Office of Evaluation Sciences (OES), working in collaboration with a large federal agency. OES is housed at the U.S. General Services Administration (GSA), and works to translate and tests evidence-based insights into concrete recommendations for how to improve government. In this project, OES partnered with the GSA Human Resources (HR) department in order to test strategies to increase utilization of and contribution to FSAs.

The sample for the evaluation included the universe of GSA employees, and each employee was randomly assigned to one of four experimental arms. Randomization was conducted by the research team in Stata using the list of recipients of the agency's tri-weekly information newsletter, GSA Today. Individuals assigned to the three treatment arms received an additional message included in the GSA Today newsletter focusing on FSAs, while individuals assigned to the control arm received no additional information about FSAs. The messages were sent twice in the first and last week of open enrollment, approximately one month apart.

[6]Given that $80 a month corresponds to maximum savings, however, the average employee can probably expect a more modest benefit.

In the first, basic treatment arm – an arm in which employees received information about the FSA benefit – the primary statement included in the body of the newsletter included the following information.

> "Do you have medical, dental or vision expenses? Save money by enrolling in a Health Care FSA today! A Health Care FSA (HCFSA) is a pre-tax benefit account that's used to pay for eligible medical, dental, and vision care expenses that are not covered by your health care plan or elsewhere. Enroll by December 11, 2017 to claim your savings!"

Employees could then click through to a more detailed one-page summary that provided some examples of expenses eligible for a FSA, noted that contributions were pre-tax, and highlighted that up to $500 in the account could be rolled over into the next year. The Appendix provides the full text of the newsletter messages, as well as the content included in the detailed one-page summary and a sample of the actual newsletter e-mail.

In the second and third treatment arms, employees received the same basic information, but the messages also sought to explicitly highlight the potential savings from utilizing a FSA. In the second absolute returns arm, the text included the same basic information about the account's functioning, but also noted that "A GSA employee with a marginal state and federal income tax rate of 36.5% spending $2600 per year for out of pocket medical expenses would save $949 over the course of a year." In the third relative returns arm, the text noted that "A GSA employee with a marginal state and federal income tax rate of 36.5% would save that rate on out of pocket medical expenses." All three messages also highlighted the deadline for enrollment. Importantly, in each of the treatment arms, the email subject heading was aligned with the distinct content of the email. Thus, the subject heading in the basic arm was "Save money", while the absolute returns arm used the heading "Save up to $949," and the relative returns arm used the heading "Save up to 36.5%". Importantly, these savings estimates were calculated using a relatively high level of projected medical expenditure and a high marginal tax ratio; for the average employee, savings may be meaningfully lower, and we certainly cannot rule out that employees were aware that their own projected savings were significantly lower than the figures cited.

These communication strategies were designed to address two key factors affecting employees' utilization of FSAs. The first is limited salience of the accounts or low levels of information; some employees may not have been aware of the benefits of enrollment. The second is procrastination, as employees may have been aware of FSAs and had the objective of enrolling, but failed to do so. In this agency, information about enrollment in FSAs is disseminated during on-boarding for new employees, and thereafter is shared periodically via employee message boards. However, it was hypothesized that these messages may not have been widely read. At the beginning of our study, around 29% of GSA employees were also enrolled using a specific benefits manager (FSAFEDs), and this vendor also directly contacted employees with information about open enrollment. While the presence of other communications is not a source of bias in our experimental design, it potentially narrows the scope for these targeted communications to have a significant effect.

3 | DATA AND STATISTICAL METHODS

3.1 | Data

Our study utilizes administrative data available to the human resources department of the participating agency, and anonymized for the purposes of the evaluation. While this data has the advantage of not relying on self-reports by employees, a relatively limited set of variables are available as compared to standard individual and household surveys.

More specifically, data available includes whether the employee contributed to a health spending FSA for the benefit year of 2018 – the cycle that is the target of this intervention – as well as the amount of the contribution. We also have an indicator variable for whether the employee contributed in the previous year, though the amount of the contribution is unreported. The only additional covariates available are basic characteristics of the employee's position at the agency (duration of employment, office, civil service GS rank, etc.) Demographic characteristics such as income, familial status, etc. are not available. As is standard OES practice, a time-stamped pre-analysis plan specifying the outcomes of interest for this evaluation was pre-filed on our website.[7]

Table 1 reports summary statistics for FY18 FSA contributions. Approximately 25% of employees contributed to a FSA in FY18 and the mean unconditional contribution was approximately $450; conditional on contributing, however,

[7]Our pre-analysis plan (Project 1733) may be found here: https://github.com/gsa-oes/office-of-evaluation-sciences/tree/master/assets/analysis

TABLE 1 Summary Statistics

	Mean	Standard deviation
	(1)	(2)
Panel A: Full Sample		
Contribution amount ($)	445.87	876.64
Contribution conditional on > $0	1775.23	836.53
Contribution > $0	0.251	0.443
Contribution > $500	0.224	0.417
Observations	11,192	
Panel B: FY17 Non-Contributors		
Contribution amount ($)	83.70	402.84
Contribution conditional on > $0	1536.33	865.04
Contribution > $0	0.054	0.227
Contribution > $500	0.044	0.206
Observations	8,425	
Panel C: FY17 Contributors		
Contribution amount ($)	1548.61	999.50
Contribution conditional on > $0	1821.86	822.99
Contribution > $0	0.850	0.357
Contribution > $500	0.772	0.420
Observations	2,767	

Notes: "Contribution > $0" and "Contribution > $500" are indicator variables equal to 1 if true and 0 if false.

the mean contribution is approximately $1775.[8] One potential explanation for the seemingly low overall participation rate is that employees who are interested in utilizing health savings accounts participate in a HSA, and thus are not eligible for contributions to a FSA; in the absence of data on employee HSA participation, however, we cannot directly examine this hypothesis. Contribution patterns are very different for previous contributors and previous non-contributors. Only 5% of the previous year's non-contributors contribute a non-zero amount, while 85% of those who contributed last year continue to contribute a positive amount. Accordingly, the correlation of the probability of contributing across years is around 0.8.

Figure 1 displays the distribution of contribution amounts conditional on contributing more than $0; the contribution amounts are collapsed to $100 increments. A notable fraction of the sample contributed the maximum amount for one person for a Health Care FSA (i.e. $2,650). There is also a smaller mass at $500, the maximum amount that can be rolled over into the following year. Many employees contributed multiples of $500, consistent with employees using this as a heuristic for choosing the contribution amount.[9]

We do not report additional demographics in Table 1, but it may be useful to briefly note that the sample is characterized on average by a relatively high GS rank, and thus a relatively high estimated household income. Only 15% of the sample is at or below GS grade 11; approximately 60% are at grades 12 or 13; and 25% are at grades 14 or higher. Each GS grade corresponds to a salary range that varies depending on location, step ranking and other characteristics, but in 2017 an employee at grade 11 earned at least $52,000 and potentially up to $68,000; an employee at grade 14 earned between $88,000 and $114,000. While a variety of other information would be required to impute marginal household tax rates, and we do not have access to any such information, it is reasonable to expect that a substantial fraction of the targeted employees do in fact face relatively high marginal tax rates.

3.2 | Statistical Methods

The primary regression specification can be written as follows:

$$contribution_i = \gamma + \beta\, treatment_i + \varepsilon_i \quad (1)$$

[8]The median contribution conditional on contributing is $2000; the 25th percentile is $1000, and the 75th percentils is $2650, the statutory maximum.

[9]Our data are administrative data on exact contribution amount, not self-reported contribution amount. Thus, "heaping" at multiples of $500 cannot reflect employee-driven measurement error in contribution amount, as is often the case in self-reported expenditures.

[10]Approximately one-quarter of employees in our study contributed to a GSA FSA in FY17.

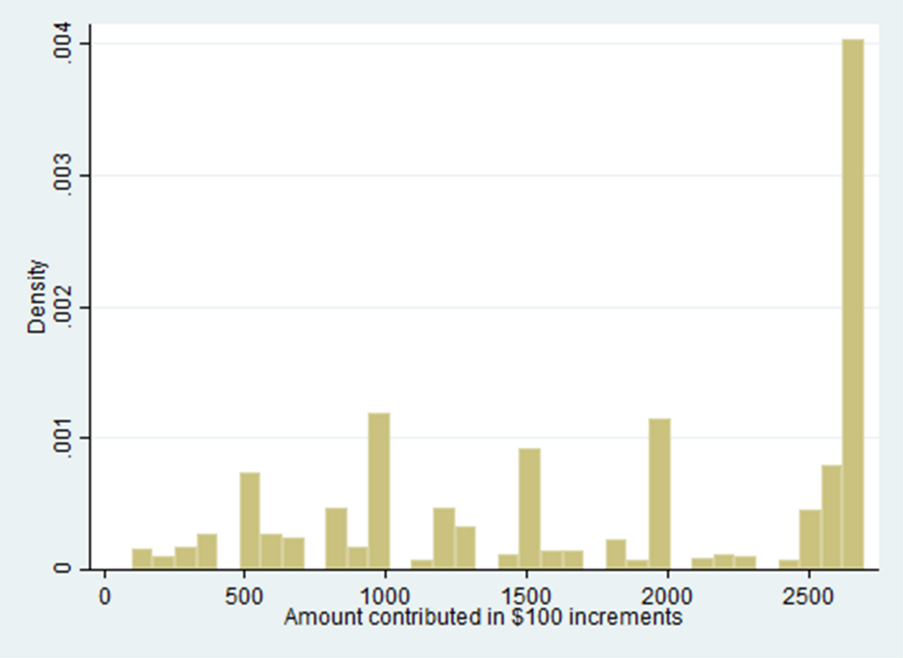

FIGURE 1 Distribution of FY18 Contributions ($) Conditional on Contribution>$0 [Colour figure can be viewed at wileyonlinelibrary.com]

In this specification, $contribution_i$ is the amount the individual contributed to a FSA (defined equal to zero if s/he did not contribute), $treatment_i$ is an indicator variable equal to one if individual i was assigned to the treatment group, and ε_i is an idiosyncratic error term. We use ordinary least squares (OLS) regression to estimate the parameters of Equation (1) and employ heteroskedasticity-robust standard errors. We also estimate the same equation using two additional dependent variables: an indicator variable equal to one if the employee reports any contributions to a FSA, and an indicator variable equal to one if the employee reports contributing more than $500. In addition, we estimate a parallel specification including indicator variables for all three treatment arms, and we estimate separate effects for employees who did and did not report contributions to a FSA in the previous year.

4 | REGRESSION RESULTS

4.1 | Main results

Figure 2 conveys the primary results in graphical format, depicting the mean level of contributions by experimental arm (in Figure 2a) and the probability of contributing more than zero or more than $500 by arm (in Figures 2b and 2c). It is evident that there are no significant differences in the level or the probability of contributions across arms.

Expanding on these results and presenting the results of formal hypothesis tests, Table 2 reports the results of estimating the simple specification using a single indicator variable for assignment to treatment. In Panel A, the specification of interest is estimated using the full sample; in Panel B, a dummy control variable for whether or not the individual contributed in the previous year is included; Panel C reports the results for the subsample of employees who had not contributed to a GSA FSA in the previous fiscal year, and Panel D reports results for the subsample of employees who had previously contributed.[10] It is evident that the coefficients of interest are consistently small in magnitude and statistically insignificant. The coefficient in Column (1) of Panel A suggests employees assigned to receive targeted information about FSAs contributed approximately $4.05 more on average relative to a mean contribution level of nearly $450; this is a proportional effect of less than 1%. Similarly, the coefficient estimated in Column (2) suggests a decline in the probability of any contribution of 0.6 percentage points; relative to a mean contribution probability of 25%, a proportional effect of around 2%. There is no discernible difference in this pattern evident in Column (3), capturing the probability of contributing at higher levels, or for the results estimated using subsamples.

[10]Approximately one-quarter of employees in our study contributed to a GSA FSA in FY17.

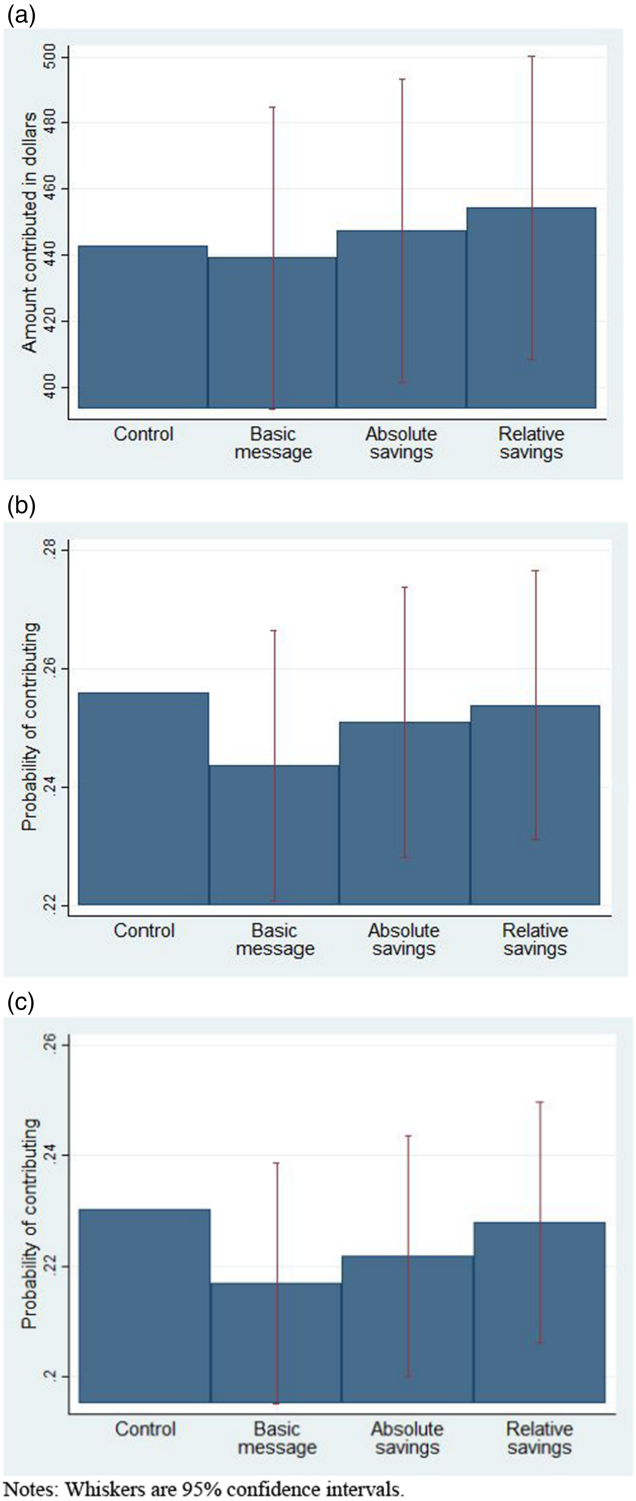

FIGURE 2 a: Contribution Amount by Study Arm. b: Probability of Contribution>$0. c: Probability of Contribution>$500Notes: Whiskers are 95% confidence intervals [Colour figure can be viewed at wileyonlinelibrary.com]

Importantly, the estimated coefficients are also relatively precise, allowing us to rule out even quite small positive effects of the messaging intervention on contributions. Examining the 95% confidence intervals, we can rule out an increase in contributions larger than around $42 for the pooled treatment, or an increase in the probability of making any contribution larger than around 1.4 percentage points.

Table 3 reports parallel results employing indicator variables for each treatment arm. The results reveal a consistent pattern: the point estimates are small in absolute terms, they are small relative to sample mean outcomes, and they are

TABLE 2 Pooled Effect of Messaging on FY18 FSA Participation

Dependent variable:	Contribution ($)	Contribution>$0	Contribution>$500
	(1)	(2)	(3)
Panel A: Full Sample			
Any messaging	4.046 (18.952)	−0.006 (0.010)	−0.008 (0.009)
Observations	11,192	11,192	11,192
Panel B: Full Sample (including controls)			
Any messaging	4.401 (13.208)	−0.006 (0.006)	−0.008 (0.006)
Observations	11,192	11,192	11,192
Panel C: FY17 Non-Contributors			
Any messaging	−6.172 (10.183)	−0.006 (0.006)	−0.007 (0.005)
Observations	8,425	8,425	8,425
Panel D: FY17 Contributors			
Any messaging	36.572 (43.484)	−0.008 (0.016)	−0.01 (0.018)
Observations	2,767	2,767	2,767

Notes: "Any messaging" is an indicator variable equal to 1 if the employee was assigned to receive any study messaging. "Contribution > $0" and "Contribution > $500" are indicator variables equal to 1 if true and 0 if false. FY17 Non-Contributors are individuals who did not contribute to a GSA FSA in the previous year.

*** Significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level.

TABLE 3 Effects of Specific Messages on FY18 FSA Participation

	Contribution ($)	Contribution>$0	Contribution>$500
Dependent variable:	**(1)**	**(2)**	**(3)**
Panel A: Full Sample			
Basic message	−3.776 (23.301)	−0.012 (0.012)	−0.013 (0.011)
Absolute savings	4.451 (23.333)	−0.005 (0.012)	−0.008 (0.011)
Relative savings	11.46 (23.347)	−0.002 (0.012)	−0.002 (0.011)
F test	.927	.727	.227
Observations	11,192	11,192	11,192
Panel B: Full Sample			
Basic message	4.88 (16.177)	−0.008 (0.007)	−0.009 (0.007)
Absolute savings	0.659 (16.533)	−0.007 (0.007)	−0.010 (0.008)
Relative savings	7.668 (15.919)	−0.004 (9.007)	−0.004 (0.007)
F test	.959	.704	.485
Observations	11,192	11,192	11,192
Panel C: FY17 Non-Contributors			
Basic message	−8.250 (12.394)	−0.007 (0.007)	−0.010 (0.006)
Absolute savings	4.329 (12.886)	0.000 (0.007)	−0.003 (0.007)
Relative savings	−14.587 (12.035)	−0.01 (0.007)	−0.009 (0.006)
F test	.420	.366	.327
Observations	8,425	8,425	8,425
Panel D: FY17 Contributors			
Basic message	45.336 (53.987)	−0.010 (0.019)	−0.007 (0.023)
Absolute savings	−10.019 (53.878)	−0.027 (0.020)	−0.034 (0.023)
Relative savings	74.767 (52.66)	0.014 (0.018)	0.010 (0.022)
F test	.345	.179	.254
Observations	2,767	2,767	2,767

Notes: "Basic message" is an indicator variable equal to 1 if the employee was assigned to receive the basic information about FSAs. "Absolute savings" is an indicator variable equal to 1 if the employee was assigned to receive messaging highlighting the absolute savings on FSA contributions. "Relative savings" is an indicator variable equal to 1 if the employee was assigned to receive messaging highlighting the relative return on FSA contributions. "Contribution > $0" and "Contribution > $500" are indicator variables equal to 1 if true and 0 if false. FY17 Non-Contributors are individuals who did not contribute to a GSA FSA in the previous year.

*** Significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level.

not statistically different from zero. In addition, the hypothesis that the treatment effects are identical comparing across different treatment arms cannot be rejected.

Finally, Table 4 examines heterogeneous effects with respect to employee characteristics: in particular, whether the employee is based in the capital (Washington, D.C. region); the employee's length of tenure; and an indicator variable

TABLE 4 Other Heterogeneous Effects of Pooled Messaging on FY18 FSA Participation

Dependent variable:	Contribution ($)	Contribution>$0	Contribution>$500
	(1)	(2)	(3)
Panel A: Geographic heterogeneity			
Any messaging	16.64 (23.250)	0.002 (0.012)	0.001 (0.011)
Capital region	14.755 (34.667)	0.009 (0.017)	0.011 (0.017)
Any messaging * Capital region	−37.116 (40.139)	−0.025 (0.020)	−0.027 (0.019)
Observations	11,192	11,192	11,192
Panel B: Experience heterogeneity			
Any messaging	−8.654 (27.754)	−0.014 (0.014)	−0.015 (0.014)
Tenure	2.832* (1.569)	0.000 (0.001)	0.001 (0.001)
Any messaging * Tenure	0.93 (1.811)	0.001 (0.001)	0.001 (0.001)
Observations	11,192	11,192	11,192
Panel C: Experience heterogeneity (alternative specification)			
Any messaging	−3.101 (24.931)	−0.011 (0.013)	−0.011 (0.012)
Tenure>10 years	99.958*** (32.683)	0.026 (0.016)	0.038** (0.016)
Any messaging * (Tenure>10 years)	12.981 (37.905)	0.009 (0.019)	0.006 (0.018)
Observations	11,192	11,192	11,192

Notes: "Any messaging" is an indicator variable equal to 1 if the employee was assigned to receive any study messaging. "Capital region" and "Tenure>10 years" are indicator variables equal to 1 if true and 0 if false. "Contribution > $0" and "Contribution > $500" are indicator variables equal to 1 if true and 0 if false. FY17 Non-Contributors are individuals who did not contribute to a GSA FSA in the previous year.

*** Significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level.

equal to one for tenure of ten years or more. The objective of these tests is to explore whether employees with certain characteristics are more likely to be responsive to messaging around FSAs. For example, a location in Washington, D.C. may be associated with a more senior position and thus a higher marginal tax rate, rendering use of a FSA a more attractive benefit. Tenure is positively correlated with age, and this may be a (weak) proxy for a higher level of medical expenses. However, estimated coefficients on the treatment variable as well as the interaction term are consistently insignificant, suggesting that there is no evidence that particular subgroups of employees are differentially responsive to the messages provided around the benefits of FSAs.

4.2 | Exploratory analysis

In addition to the data evaluated above following our pre-analysis plan, we have access to individual-level data capturing whether employees clicked on the embedded links in the informational treatments to access more information. Using the "click" variable, we estimate the following specification to evaluate whether there is any evidence of differential click-through rates by treatment arm:

$$click_i = \gamma + \beta_1 treatment_{1i} + \beta_2\ treatment_{2i} + \beta_3 treatment_{3i+}\ \varepsilon_i \quad (2)$$

Table 5 reports the results of these regressions. The first column employs a pooled variable for assignment to treatment, and indicates that the average click-through rate across all treatment arms is 0.7%.[11] The second column estimates the specification including three indicator variables for treatment assignment; the results indicate that the increase in click-through rates is observed only among individuals who received the basic message around the benefits of FSAs. Of the 59 unique individuals who clicked on a message, 58 of them received the basic message, and 1 received the message highlighting the relative savings from a FSA; there were no click-throughs among those who received the message highlighting absolute savings.

In additional exploratory analysis, we evaluate the effects of the intervention on two subgroups of employees who may have been particularly likely to show a significant response. The first is employees who are relatively new; they may have been less likely to be aware of the FSA, and would thus be more responsive to the information. The second is employees who are characterized by a relatively higher marginal tax rate. As previously noted, we do not have access to specific information about employees' income or household income, but we can analyze heterogeneous effects for the

[11]In contrast, email open rates are quite high, around 50%. However, individual-level data on open rates are not available.

TABLE 5 Effects of Messaging on Click-Throughs

Dependent variable:	Click-through	
	(1)	(2)
Any messaging	0.007*** (0.001)	
Basic message		0.021*** (0.003)
Absolute savings		0.000 (0.000)
Relative savings		0.000 (0.000)
Observations	11,192	11,192

Notes: "Click-through" is an indicator variable equal to 1 if the individual clicked-through the link embedded in the email messaging and 0 if they did not. "Any messaging" is an indicator variable equal to 1 if the employee was assigned to receive any study messaging. "Basic message" is an indicator variable equal to 1 if the employee was assigned to receive the basic information about FSAs. "Absolute savings" is an indicator variable equal to 1 if the employee was assigned to receive messaging highlighting the absolute savings on FSA contributions. "Relative savings" is an indicator variable equal to 1 if the employee was assigned to receive messaging highlighting the relative return on FSA contributions.

***Significant at the 1% level,

** Significant at the 5% level, * Significant at the 10% level.

highest-grade and thus highest-salaried employees; approximately 25% of sampled individuals are at GS rank 14 and above.

These results are reported in Table 6. In Panel A, we can observe that there is no evidence of any significant effects of the treatment for new employees, suggesting that there is no evidence the information was differentially useful for employees who may have been less exposed to information in the past. In Panel B, the results suggest that if anything, the treatment may have had a negative effect on the probability of contributing for relatively high-salaried individuals. When exposed to the treatment, these individuals are marginally less likely to contribute any positive amount (a decrease of 3.6 percentage points, significant at the 10% level), and significantly less likely to contribute more than $500 (a decrease of 4.2 percentage points, significant at the 5% level). Channels for this observed effect will be explored further in the next section.

5 | DISCUSSION

Our pre-specified analysis yielded three main facts. First, approximately 25% of employees contributed to a FSA and the mean contribution in the full sample, conditional on contributing, was approximately $1775. Second, communicating FSA eligibility and basic information about the accounts did not increase utilization of FSAs or the amount contributed. Third, highlighting the absolute savings or relative savings similarly had no additional effect on increasing contributions.

Data tracked by the communications team that manages this newsletter reports email open rates of approximately 50% for the first email campaign and approximately 50% for the second email campaign. Each recipient received our messaging twice, during the most read day of the tri-weekly newsletter, and our messaging was one of only four items highlighted in the main section of each newsletter. All of this suggests that recipients may have been aware that they

TABLE 6 Pooled Effect of Messaging on FY18 FSA Participation for New Employees and High-Income Employees

Dependent variable:	Contribution ($) (1)	Contribution>$0 (2)	Contribution>$500 (3)
Panel A: New Employees (Joined > = 2015)			
Any messaging	−10.897 (35.876)	−0.011 (0.019)	−0.013 (0.018)
Observations	2,368	2,368	2,368
Panel B: High-salary employees (GS grade > =14)			
Any messaging	−42.256 (44.185)	−0.036* (0.021)	−0.042** (0.021)
Observations	2,829	2,829	2,829

Notes: "Any messaging" is an indicator variable equal to 1 if the employee was assigned to receive any study messaging. "Capital region" and "Tenure>10 years" are indicator variables equal to 1 if true and 0 if false. "Contribution > $0" and "Contribution > $500" are indicator variables equal to 1 if true and 0 if false. FY17 Non-Contributors are individuals who did not contribute to a GSA FSA in the previous year.

*** Significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level.

had received information about potential savings associated with using FSAs, and that the null effect does not simply reflect low levels of information penetration.

Contextual details and the existing literature suggest several possible hypotheses about why our messages did not change behavior. We will broadly explore three hypotheses: first, individuals may have rationally chosen not to contribute to FSAs, and thus were indifferent to further information; second, low contribution rates reflect behavioral biases that were not successfully targeted by this intervention; and third, individuals simply did not notice or read the additional information provided, and thus did not show any response to the intervention.

The first hypothesis is that although recipients may have researched FSAs because of our emails, they may have concluded that low and/or uncertain medical expenses meant that the net benefit of participation was low, as suggested in Bhattacharya et al. (2002). Individuals who were not already contributing to a FSA may have found that it was not beneficial to participate. However, in general if individuals were responding rationally to the costs and benefits of FSA contribution, a differentially positive response would be expected from higher salaried individuals, who ceteris paribus have a higher marginal tax rate. In fact, exactly the opposite response is observed, as is evident in Table 6; there is some evidence of a negative treatment effect for higher income individuals. This suggests that a fully rational model in which individuals are optimally choosing not to use FSAs may not be consistent with the observed pattern. However, in the absence of more detailed information about savings forgone by employees at different levels of income, this evidence should be interpreted tentatively.

The second hypothesis is that individuals' enrollment choices around FSAs reflect a range of behavioral biases that were not targeted, or imperfectly targeted, by the intervention. For example, individuals may be deterred by the risk of defaulting or losing their upfront contribution if they fail to incur eligible expenses, even though this risk is likely low (and is reduced to zero if the contribution is $500 or less). Alternatively, individuals may overestimate or fixate on the upfront transaction costs of setting up the FSA, or the learning costs associated with using the FSA and redeeming eligible expenses (Bhattacharya et al., 2002, Handel 2013). More generally, individuals may have procrastinated on the enrollment decision or failed to devote sufficient attention to the participation decision (Handel 2013).

The third hypothesis is that individuals may have received information about FSAs from other sources, either prior to or during our experiment, and thus information provision was already at a saturation level. Employees already enrolled in a FSA may receive additional information from the FSA vendor (as distinct from the human resources department). There are also other sources of information about employee benefits at the agency. Evidence from other settings suggests that more frequent reminders may not be more effective than less frequent reminders (e.g., Damgaard & Gravert, 2018; Pop-Eleches et al., 2011); accordingly, in the context of widespread communication from GSA HR and the third party contractor, including in the brief window leading up to and during Federal open enrollment period, our additional communication may not have increased information or salience. While we do not see any differential effect of our intervention on employees who joined relatively recently (since 2015) and who may have been less exposed to previous sources of information, it may be that even new employees are relatively well-informed.

At the same time, employees may have received information from other sources that led them to conclude it is advantageous to contribute to a HSA instead of a FSA. HSAs offer two advantages over FSAs: individuals can roll over their entire HAS indefinitely and HSAs earn interest. On the other hand, many, if not all HSA third party administrators charge an annual fixed fee, meaning small balance HSAs may yield a negative rate of return (ignoring the tax savings).[12]

Alternatively, our results may reflect in part contamination bias, if employees in the treatment and control arms directly shared information. We considered a cluster randomized design, yet the only feasible level at which to cluster (i.e. employee federal region) included only 11 clusters and still risked contamination given that employees often collaborate across regions.[13] In general, we interpret the absence of any response to the interventions conducted as evidence that important behavioral barriers were not targeted by the communication, information provision was already at a high level, or both.

Power calculations conducted as part of our pre-analysis plan indicated minimum detectable effects (MDEs) for contribution likelihood of approximately 2.5 to 3 percentage points.[14] MDEs for contribution amount were approximately

[12]Estimates indicate that only 3% of HSAs have invested assets other than cash (Fronstin, 2016).

[13]The emails may have stimulated attention to FSAs and recipients procrastinated on plans to "do it tomorrow". Evidence on procrastination in health (e.g., DellaVigna and Malmendier (2006) highlights this possibility, yet the low click through rate suggests any immediate attention to new FSA information was limited.

[14]Our pre-analysis power calculations assumed a baseline contribution likelihood of 23% (or contribution amount of $250 and standard deviation of $500), power of 80%, significance of 0.05, and two-sided hypothesis tests.

$30. Virtually all GSA employees were included in the study, with the goal of maximizing statistical power. The size of this evaluation was large enough to generate relatively precisely estimated zero coefficients for the coefficients of interest. Yet given that the nudge intervention was virtually free of cost, even very small effects could be of interest from a policy perspective and our evaluation was not powered to detect these very small effects.

Our exploratory analysis (i.e. analysis not included in our pre-analysis plan) revealed one striking fact. Among the 59 individuals who clicked through on the embedded link, 58 were in the basic information study arm. Thus, framing savings as "save money" generated greater interest than framing savings in relative (i.e. percent terms) or in absolute (i.e. dollar terms). One explanation for this finding is that message recipients were already aware of an average level of savings, yet the "save money" framing suggested that it may be more than this amount because it did not state the maximum savings. An alternative explanation – one we find more persuasive yet cannot adjudicate empirically – is that the coarse framing of "save money" was clearer and more direct than the detailed framing statements, and thus more effectively drew respondents' attention; this may be particularly plausible if recipients do not find the hypothesized higher level of savings to be a credible claim. More broadly, the overall low click through rates suggest that individuals did not value the embedded information (e.g., because they may have already been informed) or any operative behavioral biases were so strong that they overwhelmed the potential benefit of clicking through. The low click through rates also highlight the possibility that study participants may not have read the newsletter closely and did not internalize the FSA information.

6 | CONCLUSION

Many policymakers view health savings devices as an important policy tool to be deployed as part of a broader agenda of health care reform (e.g., see Lieber, 2017). Yet take-up among eligible employees in the United States appears to be low, with estimates of participation in Flexible Spending Accounts (FSAs), the most common health-savings device ranging from 10% to as high as 40% in some sub-groups (Mercer, 2008). Our results indicate that email-based communication at the start of open enrollment period framing the return on savings in FSAs did not increase the likelihood of participation or contribution levels. We explore several explanations for this null result. Our inference is that individuals may already be informed and do not appear to exhibit behavioral biases in health savings calculations.

Future research examining the effects of changes in incentives to participate in medical savings accounts could reveal whether the magnitude of the financial benefits affects demand for these accounts. Another policy proposal that may merit exploration would entail defaulting employees into contributing the minimum amount that can be rolled over ($500), since this contribution is essentially riskless. Framing the return on savings in medical saving accounts may not be sufficient to change health savings behavior.

ORCID


Nicholas Wilson https://orcid.org/0000-0002-2632-704X

REFERENCES

Baicker, K., Mullainathan, S., & Schwartzstein, J. (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, *130*(4), 1623–1667.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2017). Does front-loading taxation increase savings? Evidence from Roth 401 (k) introductions. *Journal of Public Economics*, *151*, 84–95.

Bhargava, S., & Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, *105*(11), 3489–3529.

Bhattacharya, J., Schoenbaum, M., & Sood, N. (2002). Optimal contributions to flexible spending accounts. *Economics Letters*, *76*(1), 129–135.

Cardon, J. H., Moore, J., & Showalter, M. H. (2012). The distributional effects of health reform limits on flexible spending accounts. *Forum for Health Economics & Policy*, *15*(2).

Cardon, J. H., & Showalter, M. H. (2007). Insurance choice and tax-preferred health savings accounts. *Journal of Health Economics*, *26*(2), 373–399. https://doi.org/10.1016/j.jhealeco.2006.10.010

Chan, S., & Stevens, A. H. (2008). What you don't know can't help you: Pension knowledge and retirement decision-making. *The Review of Economics and Statistics*, *90*(2), 253–266.

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.

Chia, N.-C., & Tsui, A. K. C. (2005). Medical savings accounts in Singapore: how much is adequate? *Journal of Health Economics*, *24*(5), 855–875. https://doi.org/10.1016/j.jhealeco.2005.01.005

Choi, J. J., Laibson, D. I., & Madrian, B. (2011). $100 bills on the sidewalk: Violations of no-arbitrage in 401 (k) accounts. *Review of Economics and Statistics*, *93*(3), 748–763.

Collins, J. M., & Urban, C. (2016). The role of information on retirement planning: Evidence from a field study. *Economic Inquiry*, *54*(4), 1860–1872.

Damgaard, M. T., & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, *157*, 15–26.

DellaVigna, S., & Malmendier, U. (2006). Paying not to go to the gym. *The American Economic Review*, 694–719.

Dolls, M., Doerrenberg, P., Peichl, A., & Stichnoth, H. (2018). Do retirement savings increase in response to information about retirement and expected pensions? *Journal of Public Economics*, *158*, 168–179.

Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, *118*(3), 815–842.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., ... Oregon Health Study Group (2012). The Oregon health insurance experiment: evidence from the first year. *The Quarterly Journal of Economics*, *127*(3), 1057–1106.

Fronstin, Paul. 2016. "Health savings account balances, contributions, distributions, and other vital statistics, 2015: Estimates from the EBRI HSA Database." EBRI Issue Brief, November 29, 2016.

Hamilton, B. H., & Marton, J. (2008). Employee choice of flexible spending account participation and health plan. *Health Economics*, *17*(7), 793–813. https://doi.org/10.1002/hec.1296

Hurley, J., Emmanuel Guindon, G., Rynard, V., & Morgan, S. (2008). Publicly funded medical savings accounts: expenditure and distributional impacts in Ontario, Canada. *Health Economics*, *17*(10), 1129–1151.

Jack, W., Levinson, A., & Rahardja, S. (2006). Employee cost-sharing and the welfare effects of flexible spending accounts. *Journal of Public Economics*, *90*(12), 2285–2301.

Kaiser Family Foundation. 2016. Employer Health Benefits 2016 Annual Survey.

Leive, A. (2018). *Health insurance design meets tax incentives for saving: Consumer responses to complex contracts*. University of Virginia Working Paper.

Lieber, Ron. 2017. "Republicans Want You in a Health Savings Account. So Now What?" The New York Times, February 24, 2017.

Mercer (2008). *National Survey of Employer-Sponsored Health Plans 2008*. New York, New York: Mercer Incorporated. March 2009

Messacar, D. (2018). Crowd-out, education, and employer contributions to workplace pensions: Evidence from canadian tax records. *Review of Economics and Statistics*, *100*(4), 648–663.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415–10420.

Pauly, M. V., & Herring, B. J. (2000). An efficient employer strategy for dealing with adverse selection in multiple-plan offerings: an MSA example. *Journal of Health Economics*, *19*(4), 513–528.

Pauly, Mark V., John C. Goodman, Judith Feder, Levitt Larry, Butler Stuart M., Cutler David M., and Wilensky Gail R.. 1995. "Incremental steps toward health care reform." Health Affairs (Spring): 125–139.

Peter, R., Soika, S., & Steinorth, P. (2016). Health insurance, health savings accounts and healthcare utilization. *Health Economics*, *25*(3), 357–371.

Pop-Eleches, C., Thirumurthy, H., Habyarmina, J., Graff Zivin, J., Goldstein, M., DeWalque, D., ... Bangsberg, D. (2011). Mobile phone technologies improve adherence to antiretroviral treatment in resource-limited settings: a randomized controlled trial of text message reminders. *Aids*, *25*(6), 825–834.

Sacarny, A., Yokum, D., Finkelstein, A., & Agrawal, S. (2016). Medicare letters to curb overprescribing of controlled substances had no detectable effect on providers. *Health Affairs*, *35*(3), 471–479. https://doi.org/10.1377/hlthaff.2015.1025

Schmitz, H., & Ziebarth, N. R. (2017). Does price framing affect the consumer price sensitivity of health plan choice? *Journal of Human Resources*, *52*(1), 88–127.

Steinorth, P. (2011). Impact of health savings accounts on precautionary savings, demand for health insurance and prevention effort. *Journal of Health Economics*, *30*(2), 458–465. https://doi.org/10.1016/j.jhealeco.2010.12.007

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(S1), S164–S187.

Ye, J. (2015). The effect of Health Savings Accounts on group health insurance coverage. *Journal of Health Economics*, *44*, 238–254. https://doi.org/10.1016/j.jhealeco.2015.09.009

Zabinski, D., Selden, T. M., Moeller, J. F., & Banthin, J. S. (1999). Medical savings accounts: microsimulation results from a model with adverse selection. *Journal of Health Economics*, *18*(2), 195–218.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Leight J, Wilson N. Framing Flexible Spending Accounts: A Large-Scale Field Experiment on Communicating the Return on Medical Savings Accounts. *Health Economics*. 2019;1–14. https://doi.org/10.1002/hec.3965