What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study*

Damon Jones
University of Chicago
and NBER

David Molitor
University of Illinois
and NBER

Julian Reif
University of Illinois
and NBER

April 2019

Abstract

Workplace wellness programs cover over 50 million U.S. workers and are intended to reduce medical spending, increase productivity, and improve well-being. Yet, limited evidence exists to support these claims. We designed and implemented a comprehensive workplace wellness program for a large employer and randomly assigned program eligibility and financial incentives at the individual level for nearly 5,000 employees. Over 56 percent of eligible (treatment group) employees participated in the program. We find strong patterns of selection: during the year prior to the intervention, program participants had lower medical expenditures and healthier behaviors than non-participants. The program persistently increased health screening rates, but we do not find significant causal effects of treatment on total medical expenditures, other health behaviors, employee productivity, or self-reported health status after more than two years. Our 95 percent confidence intervals rule out 83 percent of previous estimates on medical spending and absenteeism. Although our selection results indicate that wellness programs could act as a profitable screening mechanism for employers, this possibility is limited by our null retention effect.

JEL Classification: I1, M5, J3

Keywords: Health care, workplace wellness, selection, productivity, field experiments

*This research was supported by the National Institute on Aging of the National Institutes of Health under award number R01AG050701; the National Science Foundation under Grant No. 1730546; the Abdul Latif Jameel Poverty Action Lab (J-PAL) North America U.S. Health Care Delivery Initiative; Evidence for Action (E4A), a program of the Robert Wood Johnson Foundation; and the W.E. Upjohn Institute for Employment Research. This study was pre-registered with the American Economics Association RCT Registry (AEARCTR-0001368). We are grateful to Andy de Barros for thoroughly replicating our analysis and to J-PAL for coordinating this replication effort. We thank our co-investigator Laura Payne for her vital contributions to the study, Lauren Geary for outstanding project management, Michele Guerra for excellent programmatic support, and Illinois Human Resources for invaluable institutional support. We are also thankful for comments from Kate Baicker, Jay Bhattacharya, Tatyana Deryugina, Joseph Doyle, Amy Finkelstein, Eliza Forsythe, Drew Hanks, Bob Kaestner, David Meltzer, Michael Richards, Justin Sydnor, Richard Thaler, and seminar participants at AHEC, ASHEcon, Harvard, Junior Health Economics Summit, MHEC, NBER Summer Institute, Ohio State University, University of Chicago AFE Conference, University of Zurich, UPenn Behavioral Economics and Health Symposium, SEA, and SIEPR. The findings and conclusions expressed are solely those of the authors and do not represent the views of the National Institutes of Health, any of our funders, or the University of Illinois. Email: damonjones@uchicago.edu, dmolitor@illinois.edu, and jreif@illinois.edu.

1 Introduction

Sustained growth in medical spending has prompted policymakers, insurers, and employers to search for ways to reduce medical spending. One widely touted solution is to increase the use of "wellness programs," interventions designed to encourage preventive care and discourage unhealthy behaviors such as inactivity or smoking. The 2010 Affordable Care Act (ACA) encourages firms to adopt wellness programs by permitting them to offer participation incentives up to 30 percent of the total cost of health insurance coverage, and 18 states currently include some form of wellness incentives as a part of their Medicaid program (Saunders et al., 2018). Workplace wellness industry revenue has more than tripled in size to $8 billion since 2010, and wellness programs now cover over 50 million U.S. workers (Mattke, Schnyer and Van Busum, 2012; The Kaiser Family Foundation and Health Research and Educational Trust, 2016*b*). A meta-analysis by Baicker, Cutler and Song (2010) finds large medical and absenteeism cost savings, but some studies find only limited benefits (e.g., Gowrisankaran et al., 2013; Baxter et al., 2014). As these authors have noted, identification is limited in prior studies because employee participation, along with the firm's decision to adopt a wellness program, is voluntary.

Moreover, the prior literature has overlooked important questions regarding selection into wellness programs. The increasing use of large financial incentives now permitted by the ACA may redistribute resources across employees in a manner that runs counter to the intentions of policymakers.[1] For example, wellness incentives may shift costs onto unhealthy or lower-income employees if these groups are less likely to participate in wellness programs. Furthermore, wellness programs may act as a screening device by encouraging employees who benefit most from these programs to join or remain at the firm—perhaps by earning rewards for behaviors they already enjoy.

To improve our understanding of what workplace wellness programs actually achieve,

[1] Kaiser (2017) estimates that 13 percent of large firms (at least 200 employees) offer incentives that exceed $500 dollars per year, and 4 percent of large firms offer incentives that exceed $1,000 per year.

we designed and implemented the Illinois Workplace Wellness Study, a large-scale, randomized controlled trial (RCT) conducted at the University of Illinois at Urbana-Champaign (UIUC).[2] In conjunction with the director of Campus Wellbeing Services, we developed a comprehensive workplace wellness program that included an on-site biometric health screening, an online health risk assessment, and a wide variety of wellness activities (e.g., smoking cessation, stress management, and recreational classes). We invited 12,459 benefits-eligible university employees to participate in our study.[3] Study participants ($N = 4,834$) assigned to the treatment group ($N = 3,300$) were invited to take paid time off to participate in our workplace wellness program. Those who successfully completed the entire two-year program earned rewards ranging from from \$50 to \$650, with the amounts randomly assigned and communicated at the start of each program year. The remaining subjects ($N = 1,534$) were assigned to a control group, which was not permitted to participate. Our analysis combines individual-level data from online surveys, university employment records, health insurance claims, campus gym visit records, and administrative records from a popular community running event. We can therefore examine outcomes commonly studied by the prior literature (namely, medical spending and employee absenteeism) as well as a large number of novel outcomes.

In this paper, we provide the first set of findings from the Illinois Workplace Wellness Study. We address three key research questions. First, how do financial incentives affect the level of participation in wellness programs? Theory generally predicts that incentives should increase participation, but the magnitude of this increase, which matters for understanding the nature of cross-subsidization between employees, is an empirical question. If employee

[2]Supplemental materials, datasets, and additional publications from this project will be made available on the study website at http://www.nber.org/workplacewellness.

[3]UIUC administration provided access to university data and guidance to ensure our study conformed with university regulations, but did not otherwise influence the design of our intervention. Each component of the intervention, including the financial incentives paid to employees, was paid for entirely by our external funders. Participation required electronically signing an informed consent form and completing a 15-minute online survey. Because the consent form made subjects aware of the research project, our RCT can be classified as a "framed field experiment," in the parlance of Harrison and List (2004). The study was approved by the UIUC and University of Chicago Institutional Review Boards.

participation is price elastic, then increasing the size of incentives may result in more evenly distributed program benefits; if it is price inelastic, then larger incentives largely result in greater transfers from non-participants to participants. Second, what types of employees select into wellness programs? The expected direction of the effect is ambiguous. For example, while healthy employees may have low costs of participating in these programs, employees in poor health may have the most to gain from participating. Third, what are the causal effects of workplace wellness programs on medical spending, employee productivity, health behaviors, and well-being? Again, the expected signs of these effects are uncertain. For example, medical spending could decrease if wellness programs improve health, but it could increase if wellness programs and primary care are complements.

In turn, we have three main sets of results. First, 56 percent of employees in our treatment group completed the initial major component of our study, which included an on-campus health screening.[4] Completion depended on the size of the monetary incentive assigned to an employee: increasing the screening completion reward from $0 to $100 boosted the completion rate by 12 percentage points, from 47 to 59, but further increasing the reward to $200 only increased completion by 4 percentage points, to 63 percent. When combined with our accounting records, these participation rates imply that the marginal cost of using financial incentives to induce additional screening participation reaches $1,750 at the highest screening incentive level ($200). This rapidly diminishing effect of larger rewards implies that—at least in our setting—some employees face strong barriers to participating in wellness programs. As a result, large financial wellness incentives are likely to generate disparities by transferring large sums of money to workplace wellness program participants, while having little effect on who participates. We also find that incentives tied to completing downstream wellness activities are more cost-effective than up-front incentives tied to completing the initial health screening.

Second, we find evidence of significant advantageous selection into our program: at base-

[4] A 2013 survey of workplace wellness programs found average completion rates to be around 46 percent (Mattke et al., 2013).

line, average annual medical spending among participants was $1,385 less than among non-participants. A more detailed investigation reveals that this selection effect is concentrated in the middle of the spending distribution: employees in the upper *and* lower tails of the medical spending distribution were least likely to participate. Because spending is right-skewed, the net result is that average, baseline spending among participants is lower than that of non-participants. Our estimate is economically significant: considering only medical spending, if our program increased the share of participating (i.e. low-spending) workers employed at the university by 4.3 percentage points or more, then our result implies that this change in composition alone would offset the entire costs of our intervention.[5] We also find that participants were more likely to have visited campus recreational facilities prior to our study, and were more likely to have participated in prior community running events. Thus, a primary benefit of these programs to employers may be their potential to attract and retain healthy workers with low medical spending.

Third, we do not find significant effects of our intervention on 39 out of the 41 outcomes we examine in the first year following random assignment.[6] These 39 outcomes include all our measures of medical spending, productivity, health behaviors, and self-reported health. We investigate the effect on medical expenditures in detail, but fail to find significant effects on different quantiles of the spending distribution or on any major subcategory of medical expenditures (pharmaceutical drugs, office, or hospital). We also do not find any effect of our intervention on the number of visits to campus gym facilities or on the probability of participating in a popular annual community running event, two health behaviors that are relatively simple for a motivated employee to change over the course of one year.

These null estimates are meaningfully precise, particularly for two key outcomes of interest in the literature: medical spending and absenteeism. Our 95 percent confidence intervals rule out 83 percent of the effects reported in 115 prior studies, and the 99 percent confidence

[5]We estimate positive, albeit small and insignificant, effects of the intervention on retention after one year. Our study, which focuses on an employee cohort, was not designed to examine recruitment effects.

[6]Participants were assigned to treatment and control groups in August 2016. Health screenings occurred in August and September, and wellness activities ran from October 2016 to April 2017.

intervals for the return on investment (ROI) of our intervention rule out the widely cited medical spending and absenteeism ROI's reported in the meta-analysis of Baicker, Cutler and Song (2010). In addition, we show that our OLS (non-RCT) estimate for medical spending is in line with estimates from prior observational studies, but is ruled out by the 95 percent confidence interval of our IV (RCT) estimate. This demonstrates the value of employing an RCT design in this literature.

We do find two robust, positive treatment effects from the intervention, both based on responses to a one-year follow-up survey.[7] First, employees in the treatment group were more likely than employees in the control group to report that they had ever received a health screening. This indicates that the health screening component of our program did not merely crowd out health screenings that otherwise would have occurred in the absence of our intervention. Second, treatment group employees were much more likely to report that management places a high priority on worker health and safety.

After the first year of the intervention concluded, we offered the wellness program to the treatment group for a second year. We estimate longer-run effects of the two-year intervention using outcomes measured up to 30 months after the initial randomization. Our null estimates persist over these 30 months, suggesting our two-year intervention is unlikely to have long-run effects on these outcomes. Even if some effects do emerge over a longer time horizon, the benefit to the firm will be diminished because over 20 percent of the sample ceased employment with the university by the end of this time period.

Our study contributes to the economics literature on selection in labor and insurance markets. It is well known that signaling (Spence, 1973) and screening (Rothschild and Stiglitz, 1976; Wilson, 1977) can be effective responses to asymmetric information about worker productivity (e.g. Mas-Colell et al., 1995, Ch. 13; Lazear and Oyer, 2012). Because health insurance represents an increasingly large component of firm costs, prior studies have also focused on asymmetric information about worker health status (Cutler and Zeckhauser,

[7] We address the multiple inference concern that arises when testing many hypotheses by controlling for the family-wise error rate. We discuss our approach in greater detail in Section 3.4.

2000; Bhattacharya and Vogt, 2014). Workplace wellness programs have the potential to encourage workers with low medical spending to join or remain at firms, which is a novel example of a "self-selection" device (Salop and Salop, 1976). We complement prior studies that show compensation packages may be used to attract specific types of workers (Lazear, 2000; Liu et al., 2017) and provide an additional economic justification for the prevalent and growing use of non-wage employment benefits (Oyer, 2008). Moreover, because enrollment into wellness programs is often linked to discounts on insurance premiums, our work is related to a broader literature on adverse selection in insurance markets (see Chiappori and Salanié, 2013, and Geruso and Layton, 2017, for reviews). While we uncover strong selection effects of workplace wellness, our estimated retention effects are null after 30 months, implying that screening effects—if they exist in our setting—are more likely to operate through recruitment.

Our results also speak directly to the effects of workplace wellness on worker equity. When incentives are linked to pooled expenses such as health insurance premiums, wellness programs can have distributional consequences. A concern is that wellness programs may effectively increase insurance premiums for low-income workers in poor health (Volpp et al., 2011; Horwitz, Kelly and DiNardo, 2013; McIntyre et al., 2017). The results of our selection analysis provide support for these concerns: non-participating employees are more likely to be in the bottom quartile of the salary distribution, are less likely to engage in healthy behaviors, and have higher medical spending, on average.

We also contribute to the health literature evaluating the causal effects of workplace wellness programs. The large and growing wellness industry has received increasing attention from policymakers and researchers. Our randomized controlled design allows us to establish reliable causal effects by comparing outcomes across the treatment and control groups. By contrast, most existing studies rely on observational comparisons between participants and non-participants (see Pelletier, 2011, and Chapman, 2012, for reviews). Reviews of the literature have called for additional research on this topic and have also noted the potential for publication bias to skew the set of existing results (Baicker, Cutler and Song, 2010;

Abraham and White, 2017). To that end, our intervention, empirical specifications, and outcome variables were pre-specified and publicly archived.[8] In addition, the analyses in this paper were replicated by an independent graduate student, courtesy of the Jameel Poverty Action Lab (J-PAL) North America office. A number of RCTs have focused on components of workplace wellness, such as wellness activities (Volpp et al., 2008; Charness and Gneezy, 2009; Royer, Stehr and Sydnor, 2015; Handel and Kolstad, 2017), health risk assessments (Haisley et al., 2012), or particular outcomes such as obesity or health status (Meenan et al., 2010; Terry et al., 2011). To our knowledge, this study is the first large-scale RCT of a comprehensive workplace wellness program—which we define as a program including a biometric screening, health risk assessment, wellness activities, and financial incentives.

The rest of the paper proceeds as follows. Section 2 provides a background on workplace wellness, a description of our experimental design, and a summary of our datasets. Section 3 outlines our empirical methods, while Section 4 presents the results of our first-year analysis and discussion. Section 5 presents results from our longer-run analysis, and Section 6 offers concluding observations.

2 Experimental Design

2.1 Background

Workplace wellness programs are employer-provided efforts to "enhance awareness, change behavior, and create environments that support good health practices" (Aldana, 2001, p. 297). For the purposes of this study, "wellness programs" encompass three major types of interventions: (1) biometric screenings, which provide clinical measures of health; (2) health risk assessments (HRA), which identify potential health issues; and (3) wellness activities,

[8] Our pre-analysis plan is available at http://www.socialscienceregistry.org/trials/1368. We indicate in the paper the few instances in which we deviate from our pre-analysis plan. A small number of pre-specified analyses have been omitted from the main text for the sake of brevity and because their results are not informative. For completeness, we will report those omitted results in a separate appendix.

which promote a healthy lifestyle by encouraging behaviors such as smoking cessation, stress management, or fitness. Best practice guides advise employers to let employees take paid time off to participate in wellness programs, and to combine wellness program components to maximize their effectiveness (Ryde et al., 2013). In particular, it is recommended that information from a biometric screening and HRA inform the selection of wellness activities (Soler et al., 2010). Among firms with 200 or more employees, the share offering a biometric screening, HRA, or wellness activities in 2016 was 53 percent, 59 percent, and 83 percent, respectively (Kaiser, 2016a). These benefits are often coupled with financial incentives for participation, such as cash compensation or discounted health insurance premiums. A 2015 survey estimates an average cost of $693 per employee for these programs (Jaspen, 2015) and a recent industry analysis estimates annual revenues of $8 billion (Kaiser, 2016b).

A number of factors may explain the increasing popularity of workplace wellness programs. First, some employers believe that these programs reduce medical spending and increase productivity. For example, Safeway famously attributed its low medical spending to its wellness program (Burd, 2009) (although this evidence was subsequently disputed (Reynolds, 2010)), and recent work suggests wellness programs may increase productivity (Gubler, Larkin and Pierce, 2017). Second, if employees have a high private value of wellness-related benefits, then labor market competition may drive employers to offer wellness programs in order to attract and retain workers. Third, the Affordable Care Act (ACA) has relaxed constraints on the maximum size of financial incentives offered by employers. Prior to the ACA, health-contingent incentives could not exceed 20 percent of the cost of employee health coverage. The ACA increased that general limit to 30 percent, and raised it to 50 percent for tobacco cessation programs (Cawley, 2014). The average premium for a family insurance plan in 2017 was $18,764 (Kaiser, 2017), which means that many employers are permitted to offer wellness rewards or penalties in excess of $5,000.

Like other large employers, many universities also have workplace wellness programs. Of the nearly 600 universities and liberal arts colleges ranked by *U.S. News & World Report*,

over two-thirds offer an employee wellness program.[9] Prior to our intervention, UIUC's campus wellness services were run by the University of Illinois Wellness Center, which has one staff member. The Wellness Center coordinates smoking cessation resources for employees and provides a limited number of wellness activities, many of which are not free. Importantly for our study, the campus did not offer any health screenings or HRAs and did not provide monetary incentives to employees in exchange for participating in wellness activities. Therefore, our intervention effectively represents the introduction of all major components of a wellness program at this worksite.

2.2 The Illinois Workplace Wellness Study and iThrive

The Illinois Workplace Wellness Study is a large-scale randomized controlled trial designed to investigate the effects of workplace wellness programs on employee medical spending, productivity, and well-being. As part of the study, we worked with the director of Campus Wellbeing Services to design and introduce a comprehensive wellness program named "iThrive" at the University of Illinois at Urbana-Champaign. Our goal was to create a representative program that includes all the key components recommended by wellness experts: a biometric screening, a health risk assessment, a variety of wellness activities, monetary incentives, and paid time off. We summarize the program here and provide full details in Appendix D.

Figure 1 illustrates the experimental design of the first year of our study. In July 2016 we invited 12,459 benefits-eligible university employees to enroll in our study by completing a 15-minute online survey designed to measure baseline health and wellness. The invitations were sent by postcard and email. Employees were offered a $30 Amazon.com gift card to complete the survey, as well as a chance "to participate in a second part of the research study." Over the course of three weeks, 4,834 employees completed this baseline survey. Study participants, whom we define as anybody completing the 15-minute baseline survey, were then randomly

[9] Source: authors' tabulation of data collected from universities and colleges via website search and phone inquiry.

assigned to either a control group (N=1,534), or one of six treatment groups (N=3,300). Members of the control group were notified that they may be contacted for follow-up surveys in the future, and further contact with this group was thereafter minimized. Members of the treatment group were offered the opportunity to participate in iThrive.

The first step of iThrive included a biometric health screening and an online HRA. For a period of 5 weeks in August and September 2016, participants had an opportunity to schedule a screening at one of many locations on campus. They had to make an appointment in advance and fast for 12 hours prior to the screening, where a clinician measured their height, weight, waist circumference, and blood pressure. The clinician also performed a fingerstick test to measure blood cholesterol, triglycerides, and glucose levels. Finally, participants met with a health coach, who explained their health measurements to them. The entire screening process lasted about 20 minutes. A few days later, participants received an email invitation to complete an online HRA designed to assess their lifestyle habits. Upon completion of the HRA, participants were given a score card incorporating the results of their biometric screening and providing them with recommended areas of improvement. The HRA was available as early as one week after the beginning of biometric screening and remained open until two weeks after the last biometric screening. Only participants who completed both the screening and HRA were eligible to participate in wellness activities.

The second step of iThrive consisted of wellness activities. Eligible participants were offered the opportunity to participate in one of several activities in the fall and then again in the spring. Eligibility to participate in spring wellness activities was not contingent on enrollment or completion of fall activities. In the fall, activities included in-person classes on chronic disease management, weight management, tai chi, physical fitness, financial wellness, and healthy workplace habits; a tobacco cessation hotline; and an online, self-paced wellness challenge. A similar set of activities was offered in the spring. Classes ranged from 6 to 12 weeks in length, and "completion" of a class was generally defined as attending at least three-fourths of the sessions. Participants were given two weeks to enroll in wellness activities and

were encouraged to incorporate their HRA feedback when choosing a class.

Study participants were offered monetary rewards for completing each step of the iThrive program, and these rewards varied depending on the treatment group to which an individual was assigned. Individuals in treatment groups labeled A, B, and C were offered a screening incentive of $0, $100, or $200, respectively, for completing the biometric screening and the HRA. Treatment groups were further split based on an activity incentive of either $25 or $75 for each wellness activity completed (up to one per semester). Thus, there were six treatment groups in total: A25, A75, B25, B75, C25, and C75 (see Figure 1). The total reward for completing all iThrive components—the screening, the HRA, and a wellness activity during both the fall and spring—ranged from $50 to $350, depending on the treatment group. These amounts are in line with typical wellness programs (Mattke, Schnyer and Van Busum, 2012). The probability of assignment to each group was equal across participants, and randomization was stratified by employee class (faculty, staff, or civil service), sex, age, quartile of annual salary, and race (see Appendix D.1.2 for additional randomization details). We privately informed participants about their screening and wellness activity rewards at the start of the intervention (August 2016), and did not disclose information about rewards offered to others.

To help guide participants through iThrive, we developed a secure online website that granted access to information about the program. At the onset of iThrive in August, the website instructed participants to schedule a biometric screening and then to take the online HRA. Beginning in October, and then again in January, the website provided a menu of wellness activities and online registration forms for those activities. The website also provided information on a participant's current progress and rewards earned to date, answers to frequently asked questions, and contact information for participant support.

We implemented a second year of our intervention beginning in July 2017. As in the first year, treatment group participants were offered a biometric screening, a health risk assessment, and various wellness activities (see Appendix Figure A.4). Our study concluded

with a third and final health screening in July 2018. We further discuss the second year intervention and results in Section 5.

2.3 Data

Our analysis uses self-reported survey data and administrative data sources, merged at the individual level. Each data source is summarized in this section and detailed in Appendix Section D.2. Appendix Table A.9 lists and defines each variable used in the analysis.

2.3.1 University Administrative Data

We obtained university administrative data on 12,486 employees who, as of June 2016, were (1) working at the Urbana-Champaign campus of the University of Illinois and (2) eligible for part-time or full-time employee benefits from the Illinois Department of Central Management Services. We excluded 27 people who did not have a university email address or who were substantially involved with our study, yielding a final sample size of 12,459 employees.

The initial denominator file includes employee name, university identification number, contact information (email and home mailing address), date of birth, sex, race, job title, salary, and employee class (faculty, academic staff, or civil service). We used email and home mailing address to invite employees to participate in our study, and we used sex, race, date of birth, salary, and employee class to generate the strata for random sampling.

A second file includes employment history information as of July 31, 2017. This provides three employment and productivity outcomes that are measured over the first 12 months of our study: job termination (for any reason, including firings or quits), job title change (since June 2016), and salary raises. All employees were eligible for a mid-year, merit-based salary increase in February 2017. The average salary raise in our main sample was 5.9 percent after one year. For those with a job title change, the average raise was 14.5 percent. A small number (<5 percent) of employees with job title changes did not receive an accompanying salary raise. We also define an additional variable, "job promotion," which is an indicator

for receiving both a title change and a salary raise and thus omits title changes that are potentially lateral moves or demotions.[10] We obtained an updated version of this file on January 31, 2019 for the longer-run analysis presented in Section 5.

A third file provides data on sick leave. The number of sick days taken is available at the monthly level for Civil Service employees. For academic faculty and staff, the number of sick days taken is available biannually, on August 15 and May 15. We first calculate the total number of sick days taken during our pre-period (August 2015–July 2016) and post-period (August 2016–July 2017) for each employee. We then normalize by the number of days employed to make this measure comparable across employees. All specifications that include sick days taken as an outcome variable are weighted by the number of days employed. Our longer-run analysis, presented in Section 5, uses an updated version of this file that includes a post-period covering August 2016–January 2019.

A fourth file contains data on exact attendance dates for the university's gym and recreational facilities. Entering one of these facilities requires swiping an ID card, which creates a database record linked to the individual's university ID. We calculate the total number of visits per year for the pre-period (August 2015–July 2016) and the post-period (August 2016–July 2017). As with the sick leave data, our longer-run analysis uses an updated version of this file that includes a post-period covering August 2016–January 2019.

2.3.2 Online Survey Data

As described in Section 2.2, all study participants took a 15-minute online survey in July 2016 as a condition of enrollment in the study. The survey covered topics including health status, health care utilization, job satisfaction, and productivity.

Our survey software recorded that, out of the 12,459 employees invited to take the survey, 7,468 employees clicked on the link to the survey, 4,918 employees began the survey, and 4,834 employees completed the survey. Although participants were allowed to skip questions,

[10] We did not pre-specify the job promotion or job title change outcomes in our pre-analysis plan.

response rates for the survey were very high: 4,822 out of 4,834 participants (99.7 percent) answered every one of the questions used in our analysis. To measure the reliability of the survey responses, we included a question about age at the end of the survey and compared participants' self-reported ages with the ages available in the university's administrative data. Of the 4,830 participants who reported an age, only 24 (<0.5 percent) reported a value that differed from the university's administrative records by more than one year.

All study participants were also invited via postcard and email to take a one-year follow-up survey online in July 2017.[11] In addition to the questions asked on the baseline survey, the follow-up survey included additional questions on productivity, presenteeism, and job satisfaction. A total of 3,567 participants (74 percent) successfully completed the 2017 follow-up survey. The completion rates for the control and treatment groups were 75.4 and 73.1 percent, respectively. The difference in completion rates is small but marginally significant ($p = 0.079$).

Finally, we invited all study participants to take a two-year follow-up survey in July 2018. In total, 3,020 participants (62.5 percent) completed the survey. The completion rates for the control and treatment groups were 64.6 and 61.5 percent, respectively. The completion rate difference remains small but becomes more statistically significant ($p = 0.036$). Full texts of our surveys are available in our supplementary materials.[12]

2.3.3 Health Insurance Claims Data

We obtained health insurance claims data for the time period January 1, 2015, through July 31, 2017, for the 67 percent of employees who subscribe to the university's most popular insurance plan. We use the total payment due to the provider to calculate average total monthly spending. We also use the place of service code on the claim to break total spending into four major subcategories: pharmaceutical, office, hospital, and other.[13] Our spending

[11] Invitations to the follow-up survey were sent regardless of current employment status with the university.

[12] Interactive versions of the study surveys are available at http://www.nber.org/workplacewellness.

[13] Pharmaceutical and office-based spending each have their own place of service codes. Hospital spending is summed across the following four codes: "Off Campus - Outpatient Hospital," "Inpatient Hospital," "On

measures include all payments from the insurer to providers, as well as any deductibles or copays paid by individuals. We merged these data at the individual level with our other datasets for those employees who consented to participate in our study. In addition, we have access to anonymized panel data on health claims for non-participating employees who subscribe to this same plan.

Employees choose their health plan annually during the month of May, and plan changes become effective July 1. Participants were informed of their treatment assignment on August 9, 2016. We therefore define baseline medical spending to include all allowed amounts with dates of service corresponding to the 13-month time period July 1, 2015, through July 31, 2016. We define spending in the post period to correspond to the 12-month time period August 1, 2016, through July 31, 2017. For the longer-run analysis presented in Section 5, we obtained an updated version of the claims file that allowed us to define a post period corresponding to the 29-month period August 1, 2016 through December 31, 2018.

In our health claims sample, 11 percent of employees are not continuously enrolled throughout the 13-month pre-period, and 9 percent are not continuously enrolled throughout the 12-month post-period. This is primarily due to job turnover. Because average monthly spending is measured with less noise for employees with more months of claims, we weight regressions by the number of covered months whenever the outcome variable is average spending.

2.3.4 Illinois Marathon/10K/5K Data

The Illinois Marathon is a running event held annually in Champaign, Illinois. The individual races offered include a marathon, a half marathon, a 5K, and a 10K. When registering for a race, a participant must provide her name, age, sex, and hometown. That information, along with the results of the race, are published online after the races have concluded. We

Campus - Outpatient Hospital," and "Emergency Room - Hospital." All remaining codes are assigned to "other" spending, which serves as the omitted category in our analysis. We did not pre-specify subcategories of spending in our pre-analysis plan.

downloaded those data for the 2014-2018 races and matched it to individuals in our dataset using name, age, sex, and hometown.

2.4 Baseline Summary Statistics and Balance Tests

Tables 1a and 1b provide summary statistics at baseline for the employees in our sample. Columns (2)-(8) report means for those who were assigned to our control group and to each of our six treatment groups. Column (1) additionally reports summary means for employees not enrolled in our study, where available. The variables are grouped into four panels, based on the source and type of data. Panel A presents means of the university administrative data variables used in our stratified randomization, Panel B presents means of variables from our 2016 baseline survey, Panel C presents means of medical spending variables from our health insurance claims data for the July 2015–July 2016 time period, and Panel D presents baseline means of administrative data variables used to measure health behaviors and employee productivity.

Our experimental framework relies on the random assignment of study participants to the treatment and control groups. To evaluate the validity of this assumption, we first compare the means of the variables displayed in Tables 1a and 1b. For each row, we regress the study variable on seven indicators, one for the control and each of six treatment groups, and test for the joint equality of the seven coefficients. Column (9) reports the p-value from that test. We also estimate a seemingly unrelated regression model to test whether the variables listed within each panel predict enrollment into either the control or any of the six treatment groups. The bottom of Tables 1a and 1b reports the p-value from jointly testing whether all regression coefficients across all seven groups are equal to 0, within each panel.

By construction, we find no evidence of differences in means among the variables used for stratification (Panel A): all p-values in column (9) are greater than 0.97. Among all other variables listed in Panels B, C, and D, we find statistically significant differences at a 10 percent or lower level in 2 out of 34 cases, which is approximately what one would expect

from random chance. This is confirmed by our joint balance tests, which fail to reject the null hypothesis that the variables in Panel B ($p = 0.165$), Panel C ($p = 0.207$), or Panel D ($p = 0.438$) are not predictive of group assignment.

A unique feature of our study is our ability to characterize the employees who declined to participate in our experiment. We investigate the extent of this selection into our study by comparing means for study participants, reported in columns (2)-(9) of Tables 1a and 1b, to the means for non-participating employees who did not complete our baseline survey, reported in column (1). Study participants are younger, are more likely to be female, are more likely to be white, have lower incomes on average, are more likely to be administrative staff, and are less likely to be faculty. They also have lower baseline medical spending, are more likely to have participated in one of the Illinois Marathon/10K/5K running events, and have a higher rate of monthly gym visits. These selection effects mirror the ones we report below in Section 4.2, suggesting that the factors governing the decision to participate in a wellness program are similar to the ones driving the decision to participate in our study.

3 Empirical Methods

3.1 Participation

We begin by estimating the effect of our wellness program incentives on first-year participation outcomes among employees randomly assigned to a treatment group.[14] We exclude members of the control group, for whom participation is mechanically zero. First, we jointly estimate the average effects of being assigned a positive screening incentive (groups B and C) or being assigned the $75 wellness activity incentive using the following ordinary least squares (OLS) regression:

$$P_i = \alpha + \beta_{BC} T_{i,BC} + \beta_{75} T_{i,75} + \Gamma X_i + \varepsilon_i. \quad (1)$$

[14] This analysis follows our pre-analysis plan and, for the sake of brevity, only examines participation in the first year of our study. Participation rates in the second year are reported in Appendix Figure A.4.

Here, $T_{i,BC}$ is an indicator for membership in treatment groups B or C, and $T_{i,75}$ is an indicator for receiving the \$75 wellness activity incentive. The omitted category includes members of treatment group A with a \$25 wellness activity incentive.

Second, we augment equation (1) to estimate participation effects for groups B and C separately, as follows:

$$P_i = \alpha + \beta_B T_{i,B} + \beta_C T_{i,C} + \beta_{75} T_{i,75} + \Gamma X_i + \varepsilon_i. \tag{2}$$

Here, the independent variables $T_{i,B}$ and $T_{i,C}$ are indicators for membership in treatment groups B and C, respectively.

In equations (1) and (2), the outcome P_i is an indicator for one of the following three participation outcomes: completing a screening and HRA, completing a fall wellness activity, or completing a spring wellness activity. The coefficients of interest—β_{BC}, β_B, β_C, and β_{75}—represent the causal effect of increased incentives on participation. We estimate results with and without the inclusion of strata fixed effects, X_i. The identifying assumption requires that treatment be uncorrelated with unobservable determinants of participation, ε_i, which is delivered by virtue of random assignment. This assumption is supported by the balance tests across the treatment groups, reported in Section 2.4.

3.2 Selection

Next, we characterize the types of employees who are most likely to participate in or complete the various stages of our wellness program. We pool data across the six treatment groups and estimate the following OLS regression:

$$X_i = \alpha + \theta P_i + \varepsilon_i. \tag{3}$$

The left-hand side variable, X_i, is a pre-determined covariate. The regressor, P_i, is an indicator for one of the following three participation outcomes: completing a screening and

HRA, completing a fall wellness activity, or completing a spring wellness activity. The coefficient θ represents the correlation between participation and the baseline characteristic, X_i; it should not be interpreted causally.

3.3 Causal Effects

In our final analysis, we estimate the effect of our wellness intervention on a number of outcomes, including medical spending from health claims data, employment and productivity variables measured in administrative and survey data, health behaviors measured in administrative data, and self-reported health status and behaviors. We compare outcomes in the treatment group to those in the control group using the following specification:

$$Y_i = \alpha + \gamma T_i + \Gamma X_i + \varepsilon_i. \tag{4}$$

Here, T_i is an indicator variable for membership in one of our six treatment groups, and Y_i is an outcome of interest. We estimate equation (4) with and without the inclusion of controls, X_i. In one control specification, X_i includes baseline strata fixed effects. One could also include a much broader set of controls, but doing so comes at the cost of reduced degrees of freedom. Thus, our second control specification implements the Lasso double-selection method of Belloni, Chernozhukov and Hansen (2014), as outlined by Urminsky, Hansen and Chernozhukov (2016), which selects controls that predict either the dependent variable or the focal independent variable.[15] The set of potential controls includes baseline values of the outcome variable, strata variables, the baseline survey variables reported in Table 1a, and all pairwise interactions. We then estimate a regression that includes only the controls selected by double-Lasso. In our tables, we follow convention and refer to this third control strategy as "post-Lasso." As before, our main identifying assumption requires treatment

[15]No control variable will be predictive of a randomly assigned variable, in expectation. Thus, when implementing the double-selection method with randomly assigned treatment status as the focal independent variable, we only select controls that are predictive of the dependent variable. When implementing Lasso, we use the penalty parameter that minimizes 10-fold cross-validated mean squared error.

to be uncorrelated with unobserved determinants of the outcome. The key parameter of interest, γ, is the intent-to-treat effect of our intervention on the outcome Y_i.

3.4 Inference

We report conventional robust standard errors in all tables. We do not cluster standard errors because randomization was performed at the individual level (Abadie et al., 2017). Because we estimate equations (3) and (4) for many different outcome variables, the probability that we incorrectly reject at least one null hypothesis is greater than the significance level used for each individual hypothesis test. When appropriate, we address this multiple inference concern by controlling for the family-wise error rate, i.e. the probability of incorrectly rejecting one or more null hypotheses belonging to a family of hypotheses.

To control for the family-wise error rate, we first define seven mutually exclusive families of hypotheses that encompass all of our outcome variables. Each family contains all variables belonging to one of our four outcome domains (strata variables, medical spending, employment/productivity, or health) and one of our two types of data (administrative or survey).[16] When testing multiple hypotheses using equations (3) and (4), we then calculate family-wise adjusted p-values based on 10,000 bootstraps of the free step-down procedure of Westfall and Young (1993).[17]

[16] One could assign all variables to a single family of hypotheses. This is unappealing, however, because it assigns equal importance to all outcomes when in fact some outcomes (e.g., total medical spending) are of much greater interest than others. Instead, our approach groups together variables that measure related outcomes and that originate from similar data sources.

[17] We have made our generalized Stata code module publicly available for other interested researchers to use. It can be installed by typing "`ssc install wyoung, replace`" at the Stata prompt. We provide additional documentation of this multiple testing adjustment in Appendix C.

4 First-Year Results

4.1 Participation

We begin by summarizing the effect of incentives on participation. Figure 2 reports that 56.0 percent of participants in the treatment group completed both the health screening and online HRA, which together comprise the first major step of our workplace wellness program. These participants earned their assigned rewards ($0, $100, or $200), and were subsequently allowed to sign up for wellness activities; the remaining 44 percent were excluded. In the fall, 39.5 percent of the treatment group registered for an activity, and 27.4 percent completed enough of the activity to earn their assigned activity reward. Registration and completion rates were slightly lower for the spring wellness activity. By way of comparison, a survey of employers with workplace wellness programs found that less than 50 percent of their eligible employees complete health screenings, and that most firms have wellness activity participation rates of less than 20 percent (Mattke et al., 2013).

Figure 3 reports participation rates for different levels of incentives, first for the screening and HRA stage and then for the fall activities.[18] The first set of three dark bars in Figure 3a show how screening participation varies as a function of the screening incentive. Increasing the screening incentive from $0 to $100 boosts participation from 46.9 percent to 58.5 percent. This difference is statistically significant at the 5 percent level. Increasing the screening incentive to $200 increases turnout further, to 62.5 percent. The second set of two dark bars in Figure 3a shows screening participation as a function of the wellness activity incentives. Increasing the activity incentive from $25 to $75 increases turnout from 53.6 percent to 58.4 percent, indicating that at least some participants were forward looking: they understood that they needed to first complete the screening and HRA in order to later be eligible to sign up for a wellness activity.

Table 2 provides formal statistical testing of the patterns described above for health

[18] We report the results for spring activities, which are very similar to those for the fall, in Appendix A.

screening participation. Panel A reports estimates of equations (1) and (2), using the completion of the screening and HRA as the outcome variable. Columns (1) and (2) pool together groups B and C, while columns (3) and (4) estimate the effects separately for groups B and C. The omitted group in each specification is group A25: members who were assigned a $0 screening incentive and a $25 wellness incentive.

As reported in Panel A, the baseline participation rate for the screening and HRA in the omitted group is 44.5 percent (see column (1) or (3)). Column (3) of Panel A shows that the screening/HRA completion rates of treatment groups B and C are larger than those of group A by 11.6 ($p < 0.001$) and 15.6 ($p < 0.001$) percentage points, respectively. In addition, the difference between group B and C is marginally significant ($p = 0.05$). We also estimate that a $75 wellness incentive increases screening and HRA completion by 4.9 percentage points relative to a $25 wellness incentive ($p < 0.01$). Comparing columns (1) and (3) to columns (2) and (4), respectively, shows that controlling for baseline stratification variables has very little effect on the point estimates.

We find consistently positive, but marginally diminishing, effects of monetary rewards on screening and HRA participation. The optimal reward amount depends on the marginal cost and marginal benefit associated with additional participation. Using our participation results, it is straightforward to provide some basic estimates of marginal cost using data on the field costs of our study.

The lightly shaded bars in Figure 3a report the realized average variable costs for treatment groups with different monetary rewards. The average variable costs are equal to the average monetary incentives paid to the group plus the costs of providing the health screening, the HRA, and the wellness activities.[19] We calculate the marginal cost of the additional participation induced by each reward by dividing the increase in average cost for each group

[19]Our variable cost measure does not account for paid time off or the fixed costs of managing the iThrive intervention. The health screening and HRA cost $78.22 per participant. This includes the costs of purchasing a fingerstick blood test, hiring nurses to administer the test, and licensing the HRA. The wellness activities cost an average of $26.07 per enrollee per semester. Employees who declined to participate in the health screening are assigned a variable cost of $0.

by the corresponding increase in participation.[20] The results of those calculations are plotted in Figure 3c. The marginal cost is increasing in the share of employees participating and is largest (at $1,750) for group C, whose members received $200 if they completed a screening and HRA. All else equal, this estimate implies that the optimal screening incentive is less than $200 if the marginal benefit associated with additional participation in group C is less than $1,750. Interestingly, the marginal cost of using activity incentives to increase screening participation lines up closely with that of the screening incentives.

We repeat this exercise for fall activity participation in Figures 3b and 3d. Here, a different pattern emerges. Screening incentives have only a small effect on fall activity completion, and, as a result, generate a relatively steep marginal cost curve. On the other hand, wellness activity incentives have a sizeable effect on activity completion, and exhibit a much flatter marginal cost.[21]

Panels B and C of Table 2 report that the screening incentives for groups B and C increase the completion probability for the fall or spring wellness activity by about 4-5 percentage points ($0.004 \leq p \leq 0.03$). Finally, the $75 wellness incentive, as compared to a $25 incentive, generates a 12 percentage point increase in the likelihood of completing a fall or spring wellness activity ($p < 0.001$). This last effect is sizeable when compared to a baseline completion rate of 18.2 percent in the fall and 13.7 percent in the spring for group A (see column (1) or (3)).

Overall, we find that financial incentives have a significant, but diminishing, effect on health screening participation. This suggests that when screening incentives are large, further increases in reward sizes will result in larger transfers to existing participants but little change in total participation. By contrast, we find that screening incentives have little effect on subsequent wellness activity participation, while wellness incentives have a relatively

[20] For the $25 activity incentive and $0 screening incentive groups, the marginal cost is calculated relative to a baseline of 0 percent participation and $0 average variable cost. Thus, the marginal cost for these two groups is simply the group's average variable cost divided by its participation rate.

[21] We find qualitatively similar patterns for spring activity participation, which we present in Appendix Figure A.1.

large effect. For this reason, the back-loaded wellness activity incentives are arguably more cost-effective than the upfront screening incentives: they are about as effective as screening incentives in increasing screening participation—as evidenced by similar marginal cost curves (Figure 3c)—and at the same time are more efficient at increasing wellness activity completion—i.e., they have a flatter marginal cost curve (Figure 3d).

4.2 Selection

4.2.1 Average Selection

Next, we characterize the types of workers most likely to participate in our wellness program. We focus on medical spending and health behaviors, which are primary targets of wellness programs, and on salary, which is useful for understanding the redistribution effects of these programs. Selection results for the full set of pre-specified observables are presented in Appendix Tables A.1a through A.1d.

Table 3 reports our main selection results, as estimated by equation (3). We test for selection at three different, sequential points in the study: completing the health screening and HRA; completing a fall wellness activity; and completing a spring wellness activity. Column (1) reports the mean of the selection variable of interest for employees assigned to one of our study's treatment groups. Columns (3)-(5) report the difference in means between those employees who successfully completed the participation outcome of interest and those who did not. We also report family-wise p-values in brackets that account for the number of selection variables in each "family."[22]

Column (3) of the first row of Table 3 reports that employees who completed the screening and HRA spent, on average, $115.4 per month less on health care in the 13 months prior to our study than employees who did not participate. This pattern of advantageous selection

[22]The seven families of outcome variables are defined in Section 3.4. The family-wise p-values reported in Table 3 account for all the variables in the family, including ones that are not reported in the main text. An expanded version of Table 3 that reports estimates for all pre-specified outcomes is provided in Appendix Tables A.1a through A.1d.

is strongly significant using conventional inference ($p = 0.027$), and remains marginally significant even after adjusting for the five outcomes in this family (family-wise $p = 0.082$). The magnitude is also economically significant, representing about 24 percent of the \$479 in average monthly spending (column (1)). Columns (4) and (5) present further evidence of advantageous selection into the fall and spring wellness activities, although in these cases the magnitude of selection falls by half and becomes statistically insignificant.

In contrast, the second row of Table 3 reports that employees participating in our wellness program were *more* likely to have non-zero medical spending at baseline than non-participants, by about 5 percentage points (family-wise $p \leq 0.02$), for all three participation outcomes. When combined with our results from the first row on average spending, this suggests that our wellness program is more attractive to employees with moderate spending than to employees in either tail of the spending distribution.

We investigate these results further in Figure 4, which displays the empirical distributions of prior spending for those employees who participated in screening and for those who did not. We perform two tests of the equality of the spending distributions across these two samples: Pearson's chi-squared test and the non-parametric Kolmogorov-Smirnov test.[23] Both tests strongly reject the null hypothesis that these two samples were drawn from the same distribution (Chi-squared $p < 0.001$; Kolmogorov-Smirnov $p = 0.006$). More specifically, Figure 4 reveals a "tail-trimming" effect: participating (screened) employees are less likely to be high spenders ($> \$2,338$ per month), but they are also less likely to be low spenders (\$0 per month). Because medical spending is highly skewed to the right, the overall effect on the mean among participants is negative, which explains the advantageous selection effect reported in the first row of Table 3.

Panel B of Table 3 reports selection estimates for income. The first row reports that the average annual salary of participants is lower than that of non-participants, significantly so for the fall and spring wellness activities (family-wise $p \leq 0.012$). This initially suggests that

[23] These tests were not specified in our pre-analysis plan.

participants are disproportionately lower-income. Yet, the second row of Panel B reports that the share of screening participants in the first (bottom) quartile of income is actually 6.9 percentage points *lower* than the share among non-participants (family-wise $p < 0.001$). Columns (4) and (5) also report negative, albeit smaller, selection effects for the fall and spring wellness activities. We again delve deeper by comparing the entire empirical distributions of income for participants and non-participants in Figure 5. We can reject that these two samples came from the same distribution ($p \leq 0.002$). As in Figure 4, we again find a tail-trimming effect: participating employees are less likely to come from either tail of the income distribution.

Lastly, we test for differences in baseline health behaviors as measured by our administrative data variables. The first row of Panel C in Table 3 reports that the share of screening participants who had previously participated in one of the IL Marathon/5K/10K running events is 8.9 percentage points larger than the share among non-participants (family-wise $p < 0.001$), a sizeable difference that represents over 75 percent of the mean participation rate of 11.8 percent (column (1)). This selection effect is even larger for the fall and spring wellness activities. The second row of Panel C reports that participants also visited the campus gym facilities more frequently, although these selection effects are only statistically significant for screening and HRA completion (family-wise $p = 0.013$).

Prior studies have raised concerns that the benefits of wellness programs accrue primarily to higher-income employees with lower health risks (Horwitz, Kelly and DiNardo, 2013). Our results are broadly consistent with these concerns: participating employees are less likely to have very high medical spending, less likely to be in the bottom quartile of income, and more likely to engage in healthy activities such as running or visiting the gym. At the same time, participating employees are also less likely to have very low medical spending or have very high incomes, which suggests a more nuanced story.

4.2.2 Marginal Selection

Our study design allows us to characterize not only how participants differ from non-participants on average, but also how the *marginal* participant varies as we increase incentives. As reported previously in Table 3, screening participants had *lower* baseline medical spending than non-participants, on average. Figure 6a (orange bars) shows how this pattern of selection varies by screening incentive size. For example, participants in the treatment groups with \$100 and \$200 screening incentives spent, on average, \$79 *more* per month ($p = 0.06$) than participants in the treatment group with a \$0 screening incentive. At low levels of screening incentives, wellness programs attract below-average spenders, but as incentive levels increase, the marginal participants have spending levels that are higher than the average participant. Thus, over the range of incentives we offer, increasing the size of the screening incentive reduces the average amount of selection. By contrast, Figure 6a (blue bars) illustrates a different pattern for wellness activity incentives: as we increase activity incentives, the marginal participant has significantly *lower* spending ($p = 0.03$).

While we have less power for other outcomes, we find similar selection patterns when using pre-intervention health behaviors as a proxy for health status in Figures 6b and 6c.[24] As we increase screening incentives, the marginal participant is potentially less likely to have participated in a prior marathon or have used the campus gym. Conversely, increasing wellness activity incentives potentially draws in marginal participants with a higher propensity for gym use. Thus, the selection patterns are potentially heterogeneous across type of incentive. As was the case when we examined the marginal cost of increasing participation, the type of incentive matters when examining selection into wellness programs.

[24] Marginal selection patterns with respect to income and non-zero health spending are provided in Appendix A.

4.2.3 Health Care Cost-Savings via Selection

The selection patterns we have uncovered may provide, by themselves, a potential motive for firms to offer wellness programs. We have shown that wellness participants have lower medical spending on average than non-participants. If wellness programs differentially increase the recruitment or retention of these types of employees, then the accompanying reduction in health care costs will save firms money.[25]

A simple back-of-the-envelope calculation demonstrates this possibility. In our setting, 39 percent ($= 4,834/12,459$) of eligible employees enrolled into our study, and 56 percent of the treatment group completed a screening and health assessment (Figure 2). Participating employees spent on average $138.2 per month less than non-participants in the post-period (Table 6, column 4), which translates into an annual spending difference of $1,658. When combined with average program costs of $271 per participant, this implies that the employer would need to increase the share of employees who are similar to wellness participants by 4.3 (e.g. $0.39 \times 0.56 \times 271/(1658 - 271)$) percentage points in order for the resulting reduction in medical spending to offset the entire cost of the wellness program. To be clear, this calculation does not imply that adoption of workplace wellness programs is socially beneficial. But, it does provide a profit-maximizing rationale for firms to adopt wellness programs, even in the absence of any direct effects on health, productivity, or medical spending. We estimate changes in employee retention in Sections 4.3.1 and 5, allowing us to assess how realistic this rationale may be in practice.

[25] Wellness participants differ from non-participants along other dimensions as well (e.g., health behaviors). Because it is difficult in many cases to sign, let alone quantify, a firm's preferences over these other dimensions, we focus our cost-savings discussion on the medical spending consequences.

4.3 Causal Effects

4.3.1 Intent-to-Treat

We estimate the causal, intent-to-treat (ITT) effect of our intervention on three domains of outcomes: medical spending, employment and productivity, and health behaviors. Table 4 reports estimates of equation (4) for all administratively measured outcomes, as well as a select set of outcomes from the one-year follow-up survey. An expanded version of this table reporting 41 administrative and survey outcomes is provided in Appendix Tables A.2a through A.2f.

We report ITT estimates using three specifications. The first includes no control variables. Our second specification includes fixed effects for the 69 strata used for stratified random assignment at baseline. Because the probability of treatment assignment was constant across strata, these controls are included not to reduce bias, but to improve the precision of the treatment effect estimates (Bruhn and McKenzie, 2009). Our third specification includes a set of baseline outcomes and covariates chosen via Lasso, as described in Section 3.3.

Medical spending We do not detect statistically significant effects of treatment on average medical spending over the first 12 months (August 2016–July 2017) of the wellness intervention in any of our specifications. Column (2) of the first row of Table 4 shows that the difference in average spending between treatment and control was only $10.6 per month. The point estimate increases slightly when using either of our control strategies (columns (3) or (4)) but remains small and statistically indistinguishable from zero. The post-Lasso specification generates a significant improvement in precision, with a standard error about 24 percent smaller than that of either the no-control or strata fixed effects specifications. In the rest of Panel A, we continue to find small and insignificant results for different subcategories of spending, as well as the probability of any spending over this 12-month period.

Panels (a) and (b) of Figure 7—which reproduce the basic results for total and non-zero spending presented in Panel A, column (2) of Table 4—reveal no significant differences in

average spending or probability of any spending between treatment and control. However, these results do not rule out mean-preserving treatment effects that alter other moments of the distribution. We investigate this possibility in Panel (c) of Figure 7, which displays the empirical distributions of spending for the treatment and control groups, but we fail to observe any clear differences between these two groups. This is confirmed formally by Pearson's chi-squared test and the Kolmogorov-Smirnov test, which both fail to reject the null hypothesis that the control and treatment samples were drawn from the same spending distribution ($p = 0.828$ and $p = 0.521$, respectively).

Finally, we investigate the potential for spending treatment effects to vary by treatment arm. Those results, which are available in Appendix Tables A.4a and A.4b, show no evidence of meaningful differences in spending effects across treatment arms.

Employment and productivity Next, we estimate the effect of treatment on various employment and productivity outcomes. As reported in Panel B of Table 4, we do not detect statistically significant effects after 12 months of the wellness intervention on any of our administratively measured outcomes: annual salary, the probability of job promotion or job termination, and sick leave taken. Among self-reported employment and productivity outcomes measured by the one-year follow-up survey, we find no statistically significant effects on most measures, including being happier at work than last year or feeling very productive at work. The only exception is that individuals in the treatment group are 5.7 percentage points (7.2 percent) more likely (family-wise $p < 0.001$) to believe that management places a priority on health and safety (column (2), Table 4). Appendix Tables A.2c and A.2d report ITT estimates for all administrative and pre-specified survey productivity measures.

Column (1) of Table 4 reports that 17.6 percent of our sample had received a promotion and 11.3 percent had ceased employment by the end of the first year, suggesting that our null estimates are not due to stickiness in career progression.[26] A more serious concern

[26]There is even less stickiness in the 30-month estimates reported in Section 5. As discussed in that section, our precision allows us to reject small increases in productivity during the first 30 months following randomization.

is whether our productivity measures are sufficiently meaningful and/or precise to draw conclusions. Following Baker, Gibbs and Holmstrom (1994), we perform a cross-validation exercise, comparing each of our administrative measures of employment and productivity to each of our survey measures of work and productivity. As reported in Table 5, we find a strong degree of concordance between the independently-measured administrative and survey variables. The eighth row of column (3) reports that individuals who self-report receiving "a promotion or more responsibility at work" are 22.5 percent more likely to have an official title change in our administrative data, and column (2) reports that they are 22.9 percent more likely to have received a promotion, which we define as having both a job title change *and* a non-zero salary raise.[27] More generally, our administrative measure of promotion is positively correlated with self-reported job satisfaction and happiness at work, and negatively correlated with self-reported job search. Likewise, the first row of column (5) reports that survey respondents who indicated they had taken any sick days were recorded in the administrative data as taking 3.2 more sick days than respondents who had not indicated taking sick days. The high overall agreement between our survey and administrative variables both increases our confidence in their accuracy and validates their relevance as measures of productivity.

Health behaviors Finally, we investigate health behaviors, which may respond more quickly to a wellness intervention than medical spending and productivity outcomes. Our main results are reported in Panel C of Table 4. We find very small and statistically insignificant effects of treatment assignment on participation in any running event of the April 2017 Illinois Marathon (i.e. 5K, 10K, and half/full marathons). Similarly, we do not find meaningful effects on the average number of days per month that an employee visits a campus recreation facility. However, we do find that individuals in the treatment group are nearly 4 percentage points more likely (family-wise $p = 0.001$) to report having a previous health

[27] As discussed in Section 2.3, less than five percent of employees with job title changes did not also have a salary raise. We obtain a similar causal effect estimate if we look only at job title changes rather than our constructed promotion measure (see Appendix Table A.2c).

screening. This provides evidence that our program provided biometric health information to a significant number of employees who report not previously being screened, and did not completely crowd out screenings that would have otherwise occurred within the first year of our study.

Discussion Across all 41 outcome variables we examine, we only find two statistically significant effects of our intervention: an increase in the number of employees who ever received a health screening, and an increase in the number who believe that management places a priority on health and safety.[28] The next section addresses the precision of our estimates by quantifying what effects we can rule out. But first, we mention two caveats.

First, these results only include one year of data. While we do not find significant effects for most of the outcomes we examine, it is possible that longer-run effects may emerge in the second or third year following the intervention. We return to this question below in Section 5. Second, our analysis assumes that the control group was unaffected by the intervention. The research team's contact with the control group was confined to the communication procedures employed for the 2016 and 2017 online surveys. Although we never shared details of the intervention with the control group, some of them may have learned about it from their colleagues. To evaluate how often this occurred, we asked study participants on the 2017 follow-up survey whether they ever talked about the iThrive workplace wellness program with any of their coworkers. Only 3 percent of the control group responded affirmatively, compared to 44 percent of the treatment group.

4.3.2 Comparison to Prior Studies

We now compare our estimates to the prior literature, which has focused on medical spending and absenteeism. This exercise employs a spending estimate derived from a data sample that winsorizes (top-codes) medical spending at the one percent level (see Column 3 of Table 7).

[28] We show in the appendix that these two effects are driven by the health screening component of our intervention rather than the wellness activity component.

We do this to reduce the influence of a small number of extreme outliers on the precision of our estimate, as has been done in prior studies (e.g. Clemens and Gottlieb, 2014).[29]

Figure 8 illustrates how our estimates compare to the prior literature.[30] The top-left figure in Panel (a) plots the distribution of the intent-to-treat (ITT) point estimates for medical spending from 22 prior workplace wellness studies. The figure also plots our ITT point estimate for total medical spending from Table 4, and shows that our 95-percent confidence interval rules out 20 of these 22 estimates. For ease of comparison, all effects are expressed as percent changes. The bottom-left figure in Panel (a) plots the distribution of treatment-on-the-treated (TOT) estimates for health spending from 33 prior studies, along with the IV estimates from our study. In this case, our 95-percent confidence interval rules out 23 of the 33 studies. Overall, our confidence intervals rule out 43 of 55 (78 percent) prior ITT and TOT point estimates for health spending.[31] The two figures in Panel (b) repeat this exercise for absenteeism, and show that our estimates rule out 53 of 60 (90 percent) prior ITT and TOT point estimates for absenteeism. Across both sets of outcomes, we rule out 96 of 115 (83 percent) prior estimates. If we restrict our comparison to just the studies that lasted 12 months or less, we rule out 61 of 72 (85 percent) prior estimates, and if we restrict our comparison to only the set of RCTs, we rule out 21 of 22 (95 percent) prior estimates.

We can also combine our spending and absenteeism estimates with our cost data to calculate a return on investment (ROI) for workplace wellness programs. The 99 percent confidence intervals for the ROI associated with our intervention rule out the widely cited savings estimates reported in the meta-analysis of Baicker, Cutler and Song (2010).[32]

[29] Winsorizing can introduce bias if there are heterogeneous treatment effects in the tails of the spending distribution. However, Figure 7c provides evidence of a consistently null treatment effect throughout the spending distribution. This evidence is further supported by Table 7, which shows that the point estimate of the medical spending treatment effect changes little after winsorization. For completeness, Appendix Figure A.3 illustrates the stability of the point estimate across a wide range of winsorization levels.

[30] Appendix B provides the sources and calculations underlying the point estimates reported in Figure 8.

[31] If we do not winsorize medical spending, we rule out 41 of 55 (75 percent) prior health studies.

[32] The first year of the iThrive program cost \$152 ($= \$271 \times 0.56$) per person assigned to treatment. This is a conservative estimate because it does not account for paid time off or the fixed costs of managing iThrive. Focusing on the first year of our intervention and assuming that the cost of a sick day equals \$240, we calculate that the lower bounds of the 99 percent confidence intervals for annual medical and absenteeism costs are -\$398 ($= (17.0 - 2.577 \times 19.5) \times 12$) and -\$91 ($= (0.138 - 2.577 \times 0.200) \times 240$), which imply ROI

4.3.3 IV versus OLS

Across a variety of outcomes, we find very little evidence that our workplace wellness intervention had any effect in its first year. As shown above, our results differ from many prior studies that find significant reductions in health expenditures and absenteeism. One possible reason for this discrepancy is the presence of advantageous selection bias in these other studies, which are generally not randomized controlled trials. A second possibility is that there is something unique about our setting. We investigate these competing explanations by performing a typical observational (OLS) analysis and comparing its results to those of our experimental estimates.[33] Specifically, we estimate

$$Y_i = \alpha + \gamma P_i + \Gamma X_i + \varepsilon_i, \tag{5}$$

where Y_i is the outcome variable as in (4), P_i is an indicator for participating in the screening and HRA, and X_i is a vector of variables that control for potentially non-random selection into participation.

We estimate two variants of equation (5). The first is an instrumental variables (IV) specification that includes observations for individuals in the treatment or control groups, and uses treatment assignment as an instrument for completing the screening and HRA. The second variant estimates equation (5) using OLS, restricted to individuals in the treatment group. For each of these two variants, we estimate three specifications similar to those used for the ITT analysis described above (no controls, strata fixed effects, and post-Lasso).[34]

lower bounds of 2.61 and 0.60, respectively. By comparison, Baicker, Cutler and Song (2010) found that spending fell by $3.27, and absenteeism costs fell by $2.73, for every dollar spent on wellness programs.

[33] This observational analysis was not specified in our pre-analysis plan.

[34] To select controls for the post-Lasso IV specification, we follow the "triple" selection strategy proposed in Chernozhukov, Hansen and Spindler (2015). This strategy first estimates three Lasso regressions of (1) the (endogenous) focal independent variable on all potential controls and instruments; (2) the focal independent variable on all potential controls; and (3) the outcome on all potential controls. It then forms a 2SLS estimator using instruments selected in step (1) and all controls selected in any of the steps (1)-(3). When the instrument is randomly assigned, as it is in our setting, the set of controls selected in steps (1)-(2) above will be the same, in expectation. Thus, we form our 2SLS estimator using treatment assignment as the instrument and controls selected in Lasso steps (2) or (3) of this algorithm.

This generates six estimates for each outcome variable. Table 6 reports the results for our primary outcomes of interest. The results for all pre-specified administrative and survey outcomes are reported in Appendix Tables A.3e-A.3f. We expect OLS and IV to be most similar for the post-Lasso specification, which controls for lagged outcomes and hundreds of other possible variables.

As in our previous ITT analysis, the IV estimates reported in columns (1)-(3) are small and indistinguishable from zero for nearly every outcome. By contrast, the observational estimates reported in columns (4)-(6) are frequently large and statistically significant. Moreover, the IV estimate rules out the OLS estimate for several key outcomes. Based on our most precise and well-controlled specification (post-Lasso), the OLS monthly spending estimate of −$105.1 (row 1, column (6)) lies outside the 95 percent confidence interval of the IV estimate of $50.4 with a standard error of $59.4 (row 1, column (3)). For participation in the 2017 IL Marathon/10K/5K, the OLS estimate of 0.024 lies outside the 99 percent confidence interval of the corresponding IV estimate of -0.011 (standard error = 0.011). For campus gym visits, the OLS estimate of 2.160 lies just inside the 95 percent confidence interval of the corresponding IV estimate of 0.757 (standard error = 0.656). Under the assumption that the IV (RCT) estimates are unbiased, these difference imply that even after conditioning on a rich set of controls, participants selected into our workplace wellness program on the basis of lower-than-average contemporaneous spending and higher-than-average health activity. This is consistent with the evidence presented in Section 3.2 that pre-existing spending is lower, and pre-existing behaviors are healthier, among participants than among non-participants.

In addition, the observational estimates presented in columns (4)-(6) are in line with estimates from previous observational studies, which suggests that our setting is not particularly unique. In the spirit of LaLonde (1986), these estimates demonstrate that even well-controlled observational analyses can suffer from significant selection bias in our setting, suggesting that similar biases are present in other wellness program settings as well.

5 Longer-Run Results

The first year of our intervention concluded in July 2017. We continued to offer the iThrive wellness program to the treatment group in the subsequent year. We maintained the same basic structure as in the first year, but offered smaller incentives—a design choice influenced both by a smaller budget and the diminishing effect of incentives on participation that we observed during the first year.[35] In particular, the second year of iThrive again included a health screening, an online health assessment, and a set of wellness activities offered in both the fall and spring semesters. iThrive officially ended in the fall of 2018 with a third and final health screening.

This section reports estimates of the causal, intent-to-treat effect of our two-year intervention on longer-run outcomes using data that extend up to two-and-a-half years (30 months) after the intervention began. We note that our study design entailed offering follow-up health screenings to the treatment *and* control groups in 2017 and 2018, one and two years after the intervention began, respectively. This means the control group received a partial treatment, which potentially attenuates treatment effect estimates beyond 12 months for outcomes affected by screening in the short run. However, the scope for attenuation of our longer-run estimates is limited. Control group participants were eligible only to receive a health screening; they were ineligible for both the health risk assessment and the wellness activities. Moreover, we know from our estimates above that the full intervention—screening, health risk assessment, and wellness activities—had little effect on most outcomes during the first 12 months.

Table 8 summarizes our key treatment effect estimates after 24 to 30 months, with the time horizon varying across outcomes based on data availability.[36] Overall, the longer-run estimates are qualitatively similar to those from the one-year analysis. Notably, we continue to find no effects on job promotion, despite a mean 30-month promotion rate of 36.0

[35] Appendix Figure A.4 illustrates the structure of incentives and treatments offered in the second year of the wellness program.

[36] Longer-run results for our full set of outcomes are presented in Appendix Tables A.7a–A.7g.

percent. The 30-month effect on job termination, which at 12 months was insignificant at -1.2 percentage points, is now very close to zero (0.2 percentage points), despite a mean 30-month termination rate of 20.4 percent. Our 95-percent confidence interval for job termination rules out a positive retention effect of 2.4 percentage points (12.0 percent) for iThrive. For perspective, this upper bound is well below the 4.3 percentage points needed to generate the screening savings discussed in Section 4.2.3.

Although we previously found that individuals in the treatment group were more likely to believe management places a priority on health and safety after the first year, the two-year estimate is attenuated and is no longer statistically significant in our preferred (post-Lasso) specification. We continue to find that individuals in the treatment group are more likely to report having a previous health screening, and this effect remains significant.

The point estimate for 29-month total medical spending is lower than the first-year estimates and the standard error has increased. The reduction in precision is likely caused by outliers, as described previously in Section 4.3.2. As with our 12-month estimates, we reduce the influence of outliers by winsorizing at the one percent level. Spending estimates at various levels of winsorization are presented in Table 9. For one percent winsorization (column (3)), we estimate an intent-to-treat effect of \$5.4 with a 95 percent confidence interval of $[-33.4, 44.2]$. This is very similar to the winsorized 12-month estimate of \$17.0 and 95 percent confidence interval of $[-21.2, 55.2]$ (column (3) of Table 7).

Increasing the length of the follow-up window raises concerns about the potential for differential attrition between the control and treatment groups. However, Appendix Table A.8 shows that health insurance enrollment is nearly identical in the control and treatment groups over both the 12-month and 30-month post-periods. In addition, rates of job exit, which measure sample attrition for outcomes derived from University administrative data, and rates of one-year follow-up survey completion are also similar for both the treatment and control groups. The only case in which attrition is statistically significantly different involves two-year follow-up survey completion, which only affects two of the major outcomes

presented in Table 8. Both the treatment and control groups nonetheless completed the follow-up surveys at high and broadly similar rates, although outcomes derived from the two-year survey should potentially be given less weight than those from other data sources.

6 Conclusion

This paper presents a first set of findings from the Illinois Workplace Wellness Study. We find a large but diminishing effect of incentives on wellness participation. At large incentive levels, further increases have little effect on participation and thus primarily just increase compensation for inframarginal participants. We also find that employees who chose to participate in our wellness program were less likely to be in the bottom quartile of the income distribution, and already had lower medical spending and healthier behaviors than non-participants prior to our intervention. This suggests that workplace wellness programs shift costs onto low-income employees with high health care spending and poor health habits. Moreover, the large magnitude of our selection on prior spending suggests that a potential value of wellness programs to firms may be their potential to attract and retain workers with low health care costs. All else equal, reducing the share of non-participating employees by 4.3 percentage points would lower total medical spending in our setting by an amount sufficient to pay for our entire program. In our setting, however, we estimate null retention effects, suggesting this effect (if it exists) would likely need to operate through a recruitment channel, which we cannot estimate using our study design.

After more than two years, we find no significant effects of our wellness program on medical spending, health behaviors, or employee productivity, with two exceptions. First, after 12 months, treated employees are more likely to believe that the employer places a priority on worker health and safety. Second, treated employees are more likely to report having received a health screening, an effect which persists for at least 24 months. Our null results are economically meaningful: we can rule out 83 percent of the medical spending

and absenteeism estimates from the prior literature, along with the average ROIs calculated by Baicker, Cutler and Song (2010) in a widely cited meta-analysis. Our OLS estimate is consistent with results from the prior literature, but ruled out by our IV estimate, suggesting that non-RCT studies in this literature suffer from selection bias.

Well-designed studies have found that monetary incentives can successfully promote exercise (e.g., Charness and Gneezy, 2009), and there is ample evidence that exercise improves health (e.g., Warburton, Nicol and Bredin, 2006). Indeed, this is a primary rationale behind using public policies to encourage wellness programs. However, our study shows that it is difficult to achieve these benefits with large-scale interventions, a point that is echoed by Cawley and Price (2013). One reason for this may be that those who benefit the most (e.g., smokers and those with high health care costs) decline to participate, even when offered large monetary incentives. This suggests these individuals may face large barriers to participation. Future research that leads to a better understanding of those barriers could help wellness programs better target these individuals.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" *NBER Working Paper #24003*.

Abraham, Jean, and Katie M White. 2017. "Tracking The Changing Landscape Of Corporate Wellness Companies." *Health Affairs*, 36(2): 222–228.

Aldana, Steven G. 2001. "Financial impact of health promotion programs: a comprehensive review of the literature." *American Journal of Health Promotion*, 15(5): 296–320.

Baicker, Katherine, David Cutler, and Zirui Song. 2010. "Workplace wellness programs can generate savings." *Health Affairs*, 29(2): 304–311.

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994. "The Internal Economics of the Firm: Evidence from Personnel Data." *The Quarterly Journal of Economics*, 109(4): 881–919.

Baxter, Siyan, Kristy Sanderson, Alison J Venn, C Leigh Blizzard, and Andrew J Palmer. 2014. "The relationship between return on investment and quality of study methodology in workplace health promotion programs." *American Journal of Health Promotion*, 28(6): 347–363.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies*, 81(2): 608–650.

Bhattacharya, Jayanta, and William B Vogt. 2014. "Employment and adverse selection in health insurance." Vol. 17, 79–104.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 1(4): 200–232.

Burd, Steven A. 2009. "How Safeway Is Cutting Health-Care Costs." *The Wall Street Journal*. http://www.wsj.com/articles/SB124476804026308603.

Cawley, John. 2014. "The Affordable Care Act permits greater financial rewards for weight loss: a good idea in principle, but many practical concerns remain." *Journal of Policy Analysis and Management*, 33(3): 810–820.

Cawley, John, and Joshua A Price. 2013. "A case study of a workplace wellness program that offers financial incentives for weight loss." *Journal of Health Economics*, 32(5): 794–803.

Chapman, Larry S. 2012. "Meta-evaluation of worksite health promotion economic return studies: 2012 update." *American Journal of Health Promotion*, 26(4): 1–12.

Charness, Gary, and Uri Gneezy. 2009. "Incentives to exercise." *Econometrica*, 77(3): 909–931.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. "Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments." *American Economic Review*, 105(5): 486–90.

Chiappori, Pierre-André, and Bernard Salanié. 2013. "Asymmetric information in insurance markets: Predictions and tests." In *Handbook of insurance*. 397–422. Springer.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104(4): 1320–1349.

Cutler, David M, and Richard J Zeckhauser. 2000. "The anatomy of health insurance." *Handbook of health economics*, 1: 563–643.

Geruso, Michael, and Timothy J. Layton. 2017. "Selection in Health Insurance Markets and Its Policy Remedies." *Journal of Economic Perspectives*, 31(4): 23–50.

Gowrisankaran, Gautam, Karen Norberg, Steven Kymes, Michael E Chernew, Dustin Stwalley, Leah Kemper, and William Peck. 2013. "A hospital system's wellness program linked to health plan enrollment cut hospitalizations but not overall costs." *Health Affairs*, 32(3): 477–485.

Gubler, Timothy, Ian Larkin, and Lamar Pierce. 2017. "Doing well by making well: The impact of corporate wellness programs on employee productivity." *Management Science*.

Haisley, Emily, Kevin G Volpp, Thomas Pellathy, and George Loewenstein. 2012. "The impact of alternative incentive schemes on completion of health risk assessments." *American Journal of Health Promotion*, 26(3): 184–188.

Handel, Benjamin, and Jonathan Kolstad. 2017. "Wearable Technologies and Health Behaviors: New Data and New Methods to Understand Population Health." *American Economic Review: Papers and Proceedings*, 107(5): 481–85.

Harrison, Glenn W, and John A List. 2004. "Field experiments." *Journal of Economic literature*, 42(4): 1009–1055.

Horwitz, Jill R, Brenna D Kelly, and John E DiNardo. 2013. "Wellness incentives in the workplace: cost savings through cost shifting to unhealthy workers." *Health Affairs*,

32(3): 468–476.

Jaspen, Bruce. 2015. "Employers Boost Wellness Spending 17% From Yoga to Risk Assessments." *Forbes Online*. http://www.forbes.com/sites/brucejapsen/2015/03/26/employers-boost-wellness-spending-17-from-yoga-to-risk-assessments/#6a37ebf2350f.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, 76(4): 604–620.

Lazear, Edward P. 2000. "Performance pay and productivity." *American Economic Review*, 90(5): 1346–1361.

Lazear, Edward P., and Paul Oyer. 2012. "Personnel Economics." *The Handbook of Organizational Economics*, 479–519. Princeton University Press.

Liu, Tim, Christos Makridis, Paige Ouimet, and Elena Simintzi. 2017. "Is Cash Still King: Why Firms Offer Non-Wage Compensation and the Implications for Shareholder Value." Available at SSRN: https://ssrn.com/abstract=3088067.

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. 1995. *Microeconomic theory*. Vol. 1, Oxford university press New York.

Mattke, Soeren, Christopher Schnyer, and Kristin R Van Busum. 2012. "A review of the US workplace wellness market." *The RAND Corporation*, Occasional Paper Series, https://www.dol.gov/sites/default/files/ebsa/researchers/analysis/health-and-welfare/workplacewellnessmarketreview2012.pdf.

Mattke, Soeren, Hangsheng Liu, John Caloyeras, Christina Y Huang, Kristin R Van Busum, Dmitry Khodyakov, and Victoria Shier. 2013. "Workplace wellness programs study: Final report." *RAND Health Quarterly*, 3(2).

McIntyre, Adrianna, Nicholas Bagley, Austin Frakt, and Aaron Carroll. 2017. "The dubious empirical and legal foundations of workplace wellness programs." *Health Matrix*, 27: 59.

Meenan, Richard T, Thomas M Vogt, Andrew E Williams, Victor J Stevens, Cheryl L Albright, and Claudio Nigg. 2010. "Economic evaluation of a worksite obesity prevention and intervention trial among hotel workers in Hawaii." *Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine*, 52(Suppl 1): S8.

Oyer, Paul. 2008. "Salary or benefits?" In *Work, Earnings and Other Aspects of the Employment Relation*. 429–467. Emerald Group Publishing Limited.

Pelletier, Kenneth R. 2011. "A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VIII 2008 to 2010." *Journal of occupational and environmental medicine*, 53(11): 1310–1331.

Reynolds, Chelsea. 2010. "Myth Surrounds Reform's 'Safeway Amendment." *Covering Health*. http://healthjournalism.org/blog/2010/01/myth-surrounds-reforms-safeway-amendment/.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *The quarterly journal of economics*, 629–649.

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company." *American Economic Journal: Applied Economics*, 7(3): 51–84.

Ryde, Gemma C, Nicholas D Gilson, Nicola W Burton, and Wendy J Brown. 2013. "Recruitment rates in workplace physical activity interventions: characteristics for success." *American Journal of Health Promotion*, 27(5): e101–e112.

Salop, Joanne, and Steven Salop. 1976. "Self-selection and turnover in the labor market." *The Quarterly Journal of Economics*, 619–627.

Saunders, Rob, Madhu Vulimiri, Mark Japinga, William Bleser, and Charlene Wong. 2018. "Are Carrots Good for Your Health? Current Evidence on Health Behavior Incentives in the Medicaid Program." Duke Margolis Center for Health Policy.https://healthpolicy.duke.edu/sites/default/files/atoms/files/duke_healthybehaviorincentives_6.1.pdf.

Soler, Robin E, Kimberly D Leeks, Sima Razi, David P Hopkins, Matt Griffith, Adam Aten, Sajal K Chattopadhyay, Susan C Smith, Nancy Habarta, Ron Z Goetzel, et al. 2010. "A systematic review of selected interventions for worksite health promotion: the assessment of health risks with feedback." *American Journal of Preventive Medicine*, 38(2): S237–S262.

Spence, Michael. 1973. "Job market signaling." *The Quarterly Journal of Economics*, 87(3): 355–374.

Terry, Paul E, Jinnet Briggs Fowles, Min Xi, and Lisa Harvey. 2011. "The ACTIVATE study: results from a group-randomized controlled trial comparing a traditional worksite health promotion program with an activated consumer program." *American Journal of Health Promotion*, 26(2): e64–e73.

The Kaiser Family Foundation and Health Research and Educational Trust. 2016*a*. "Employer Health Benefits: 2016 Annual Survey." http://files.kff.org/attachment/Report-Employer-Health-Benefits-2016-Annual-Survey.

The Kaiser Family Foundation and Health Research and Educational Trust. 2016*b*. "Workplace Wellness Programs Characteristics and Requirements." http://files.kff.org/attachment/Issue-Brief-Workplace-Wellness-Programs-Characteristics-and-Requirements.

The Kaiser Family Foundation and Health Research and Educational Trust. 2017. "Employer Health Benefits: 2017 Annual Survey." http://files.kff.org/attachment/Report-Employer-Health-Benefits-Annual-Survey-2017.

Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov. 2016. "Using Double-Lasso Regression for Principled Variable Selection."

Volpp, Kevin G, David A Asch, Robert Galvin, and George Loewenstein. 2011. "Redesigning employee health incentives—lessons from behavioral economics." *New England Journal of Medicine*, 365(5): 388–390.

Volpp, Kevin G, Leslie K John, Andrea B Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008. "Financial incentive–based approaches for weight loss: a randomized trial." *Journal of the American Medical Association*, 300(22): 2631–2637.

Warburton, Darren ER, Crystal Whitney Nicol, and Shannon SD Bredin. 2006. "Health benefits of physical activity: the evidence." *Canadian Medical Association Journal*, 174(6): 801–809.

Westfall, Peter H, and S Stanley Young. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279, John Wiley & Sons.

Wilson, Charles. 1977. "A model of insurance markets with incomplete information." *Journal of Economic theory*, 16(2): 167–207.

Figure 1: Experimental Design of the Illinois Workplace Wellness Study

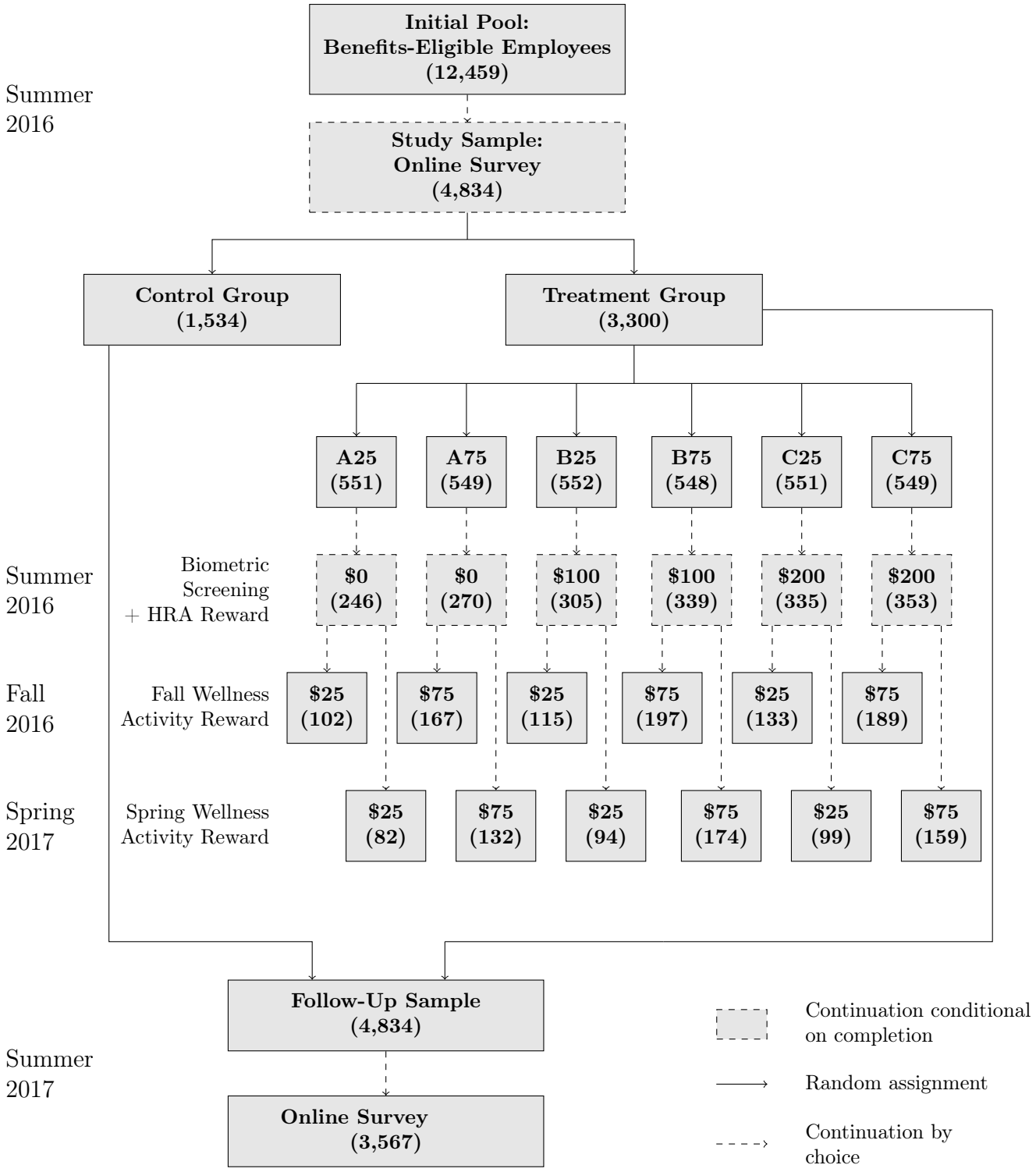
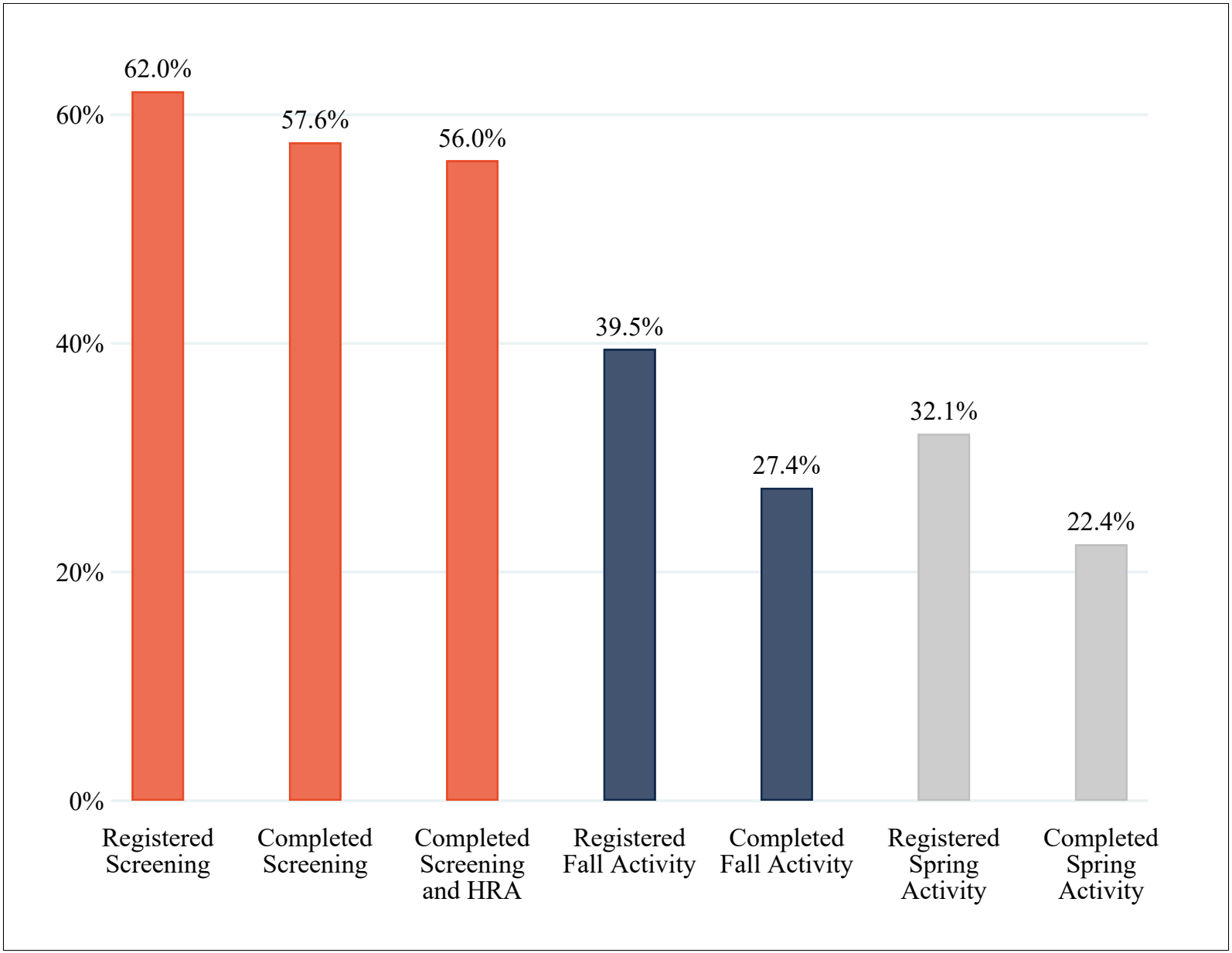
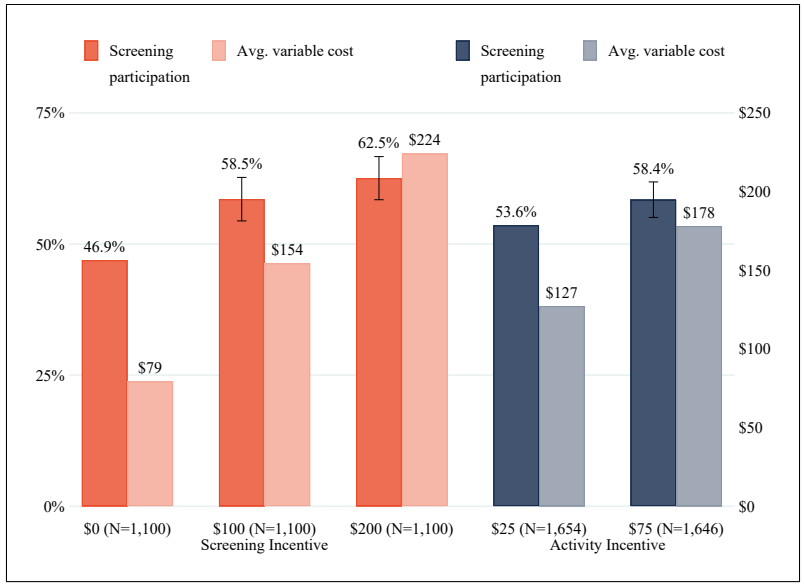


Figure 2: Employee participation rates in the first year of the workplace wellness program

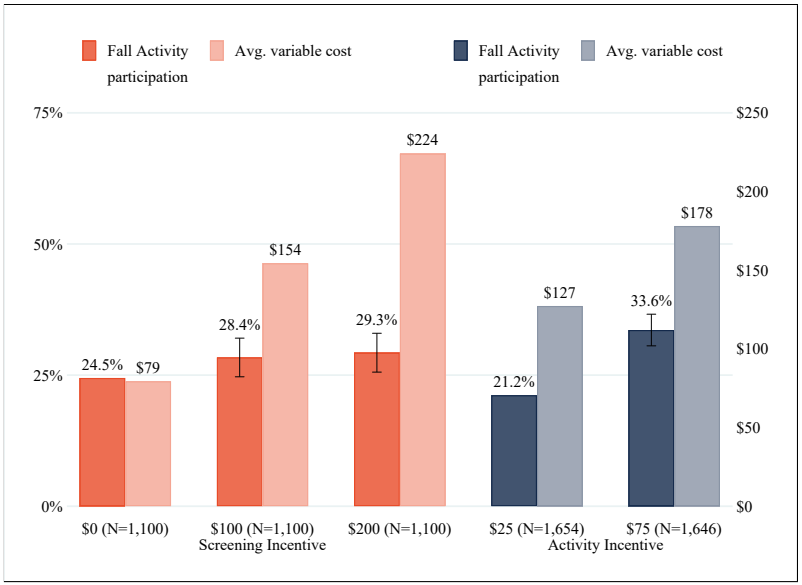


Notes: Participation rates are measured as a fraction of the treatment group ($N = 3,300$).

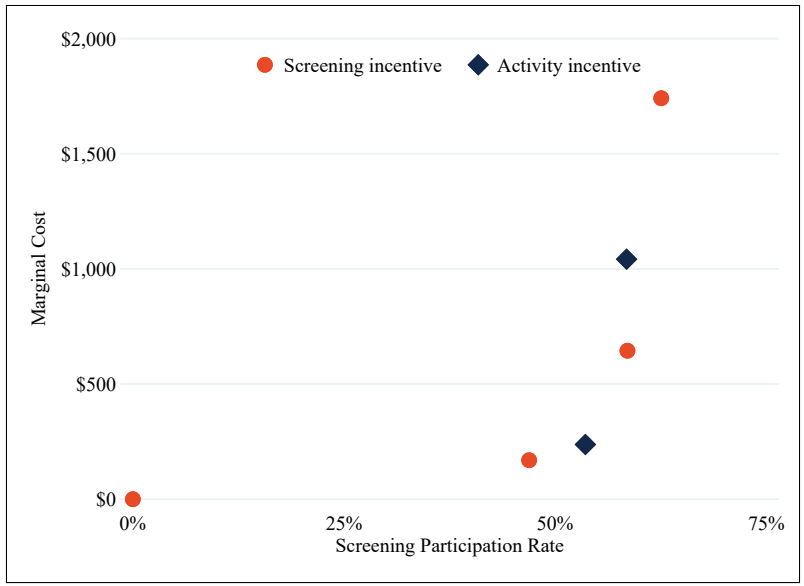
Figure 3: Marginal cost of inducing additional participation into health screening/HRA and fall activities



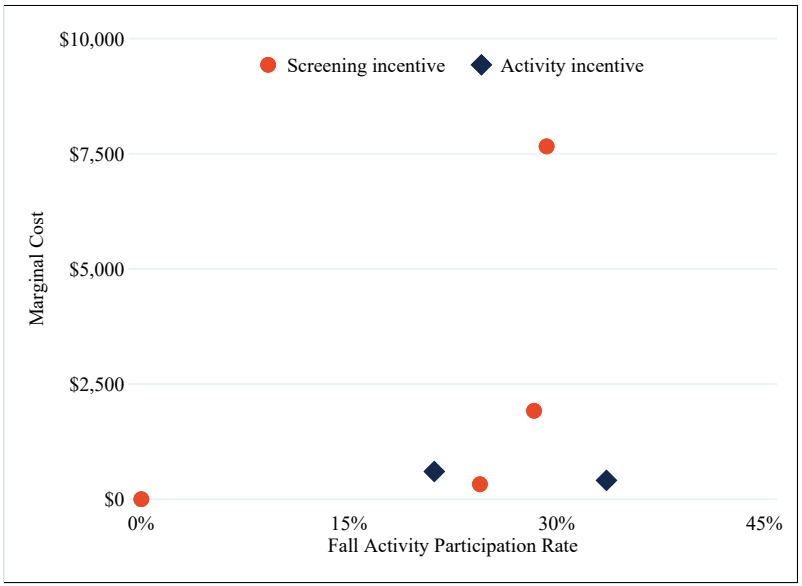
(a) Screening participation and costs



(b) Fall activities participation and costs



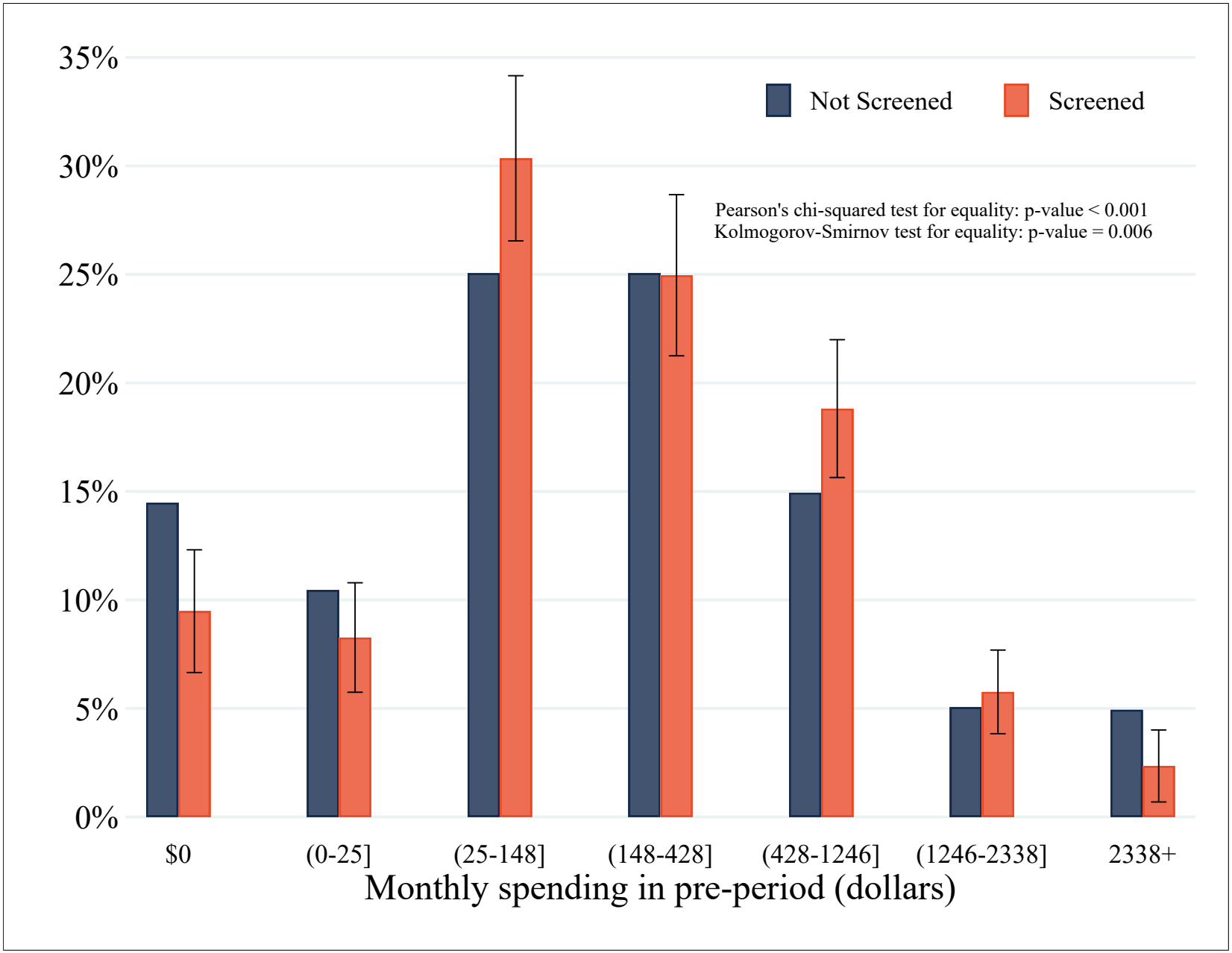
(c) Marginal cost of additional screening participation



(d) Marginal cost of additional fall activity participation

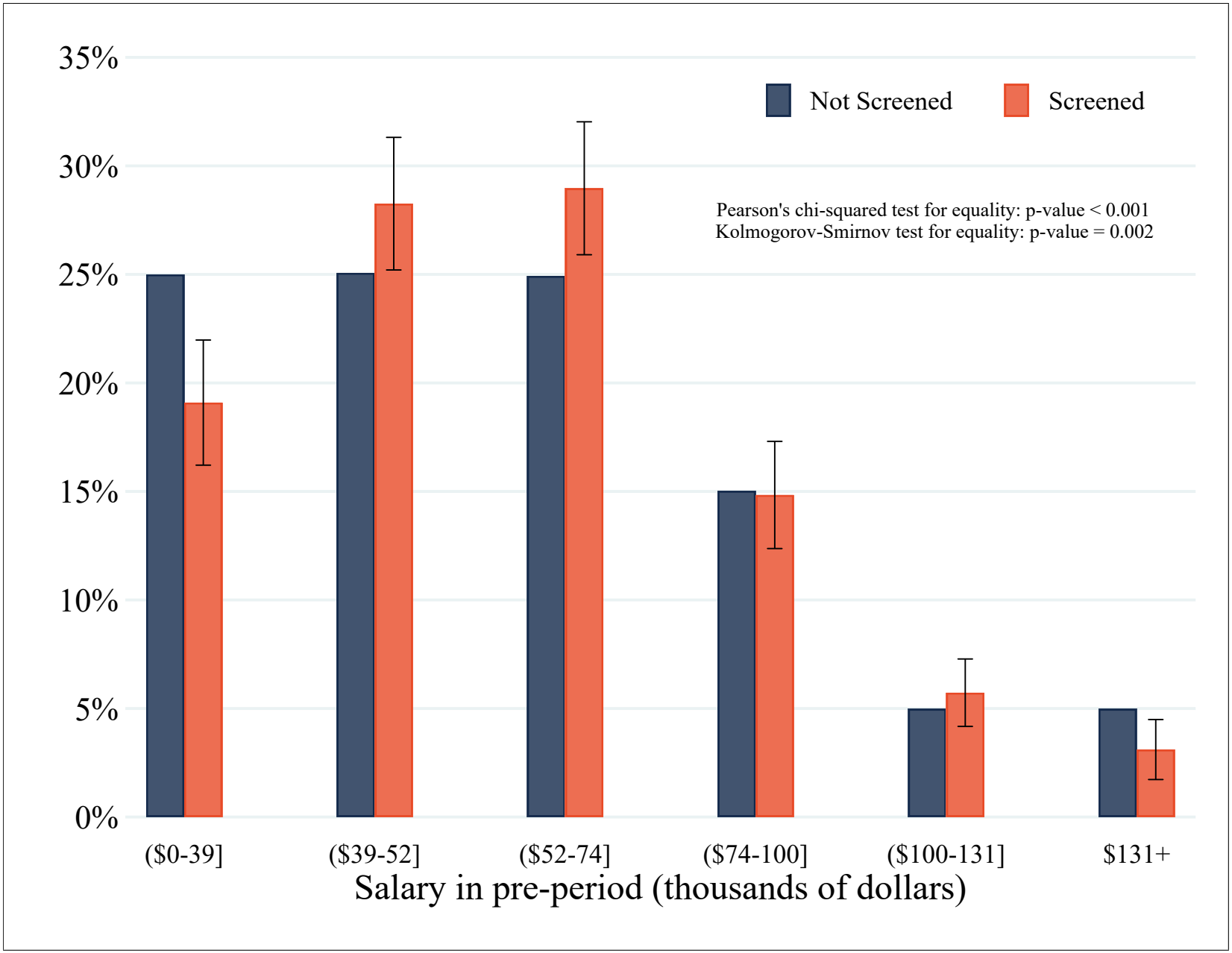
Notes: Panels (a) and (b) plot health screening participation rates (PR) and average variable costs (AVC) as a function of screening and activity incentives, separately for screening and fall activity particpation. Vertical bars display 95% confidence intervals on the difference in means relative to the lowest reward group. AVC includes costs of the health screening, HRA, and wellness activities. Panels (c) and (d) plot the implied marginal costs (MC), calculated as MC=$\frac{\Delta \text{AVC}}{\Delta \text{PR}}$. The MC of the control group (PR=0 percent) is set equal to 0.

Figure 4: Pre-intervention medical spending among treatment group, by participation status



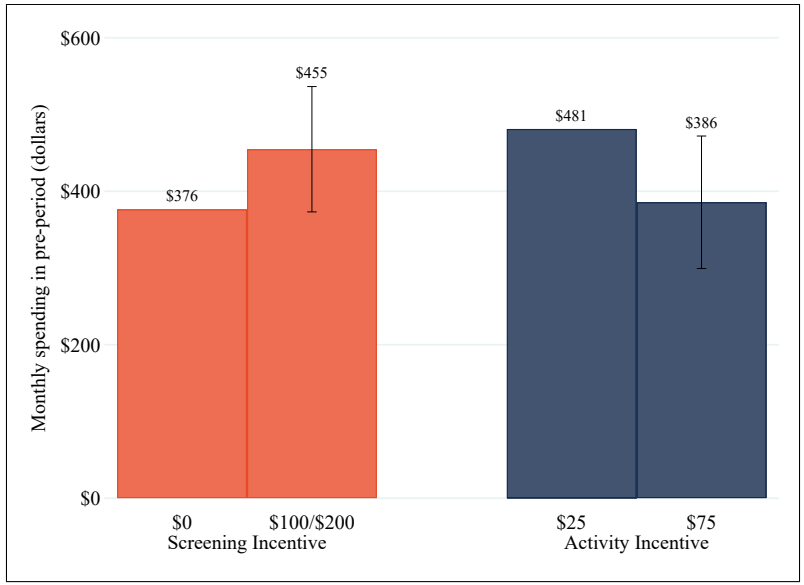
Notes: Data are from claims covering the period July 2015–July 2016 ($N = 2,188$). The first two bins ($0 and $(0-25]$) include 25 percent of those not screened. The remaining five bins were defined to include 25, 25, 15, 5, and 5 percent of those not screened, respectively. The null hypothesis of the Pearson's chi-squared and the non-parametric Kolmogorov-Smirnov tests is that the two samples are drawn from the same distribution.

Figure 5: Pre-intervention salary among treatment group, by participation status

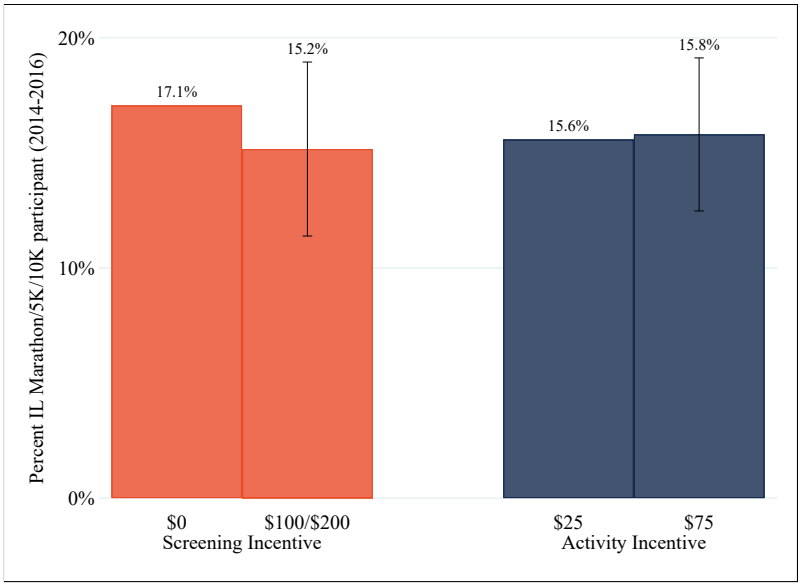


Notes: Salary was measured on June 1, 2016 ($N = 3,257$). The six bins were defined to include 25, 25, 25, 15, 5, and 5 percent of employees not screened, respectively. The null hypothesis of the Pearson's chi-squared and the non-parametric Kolmogorov-Smirnov tests is that the two samples are drawn from the same distribution.

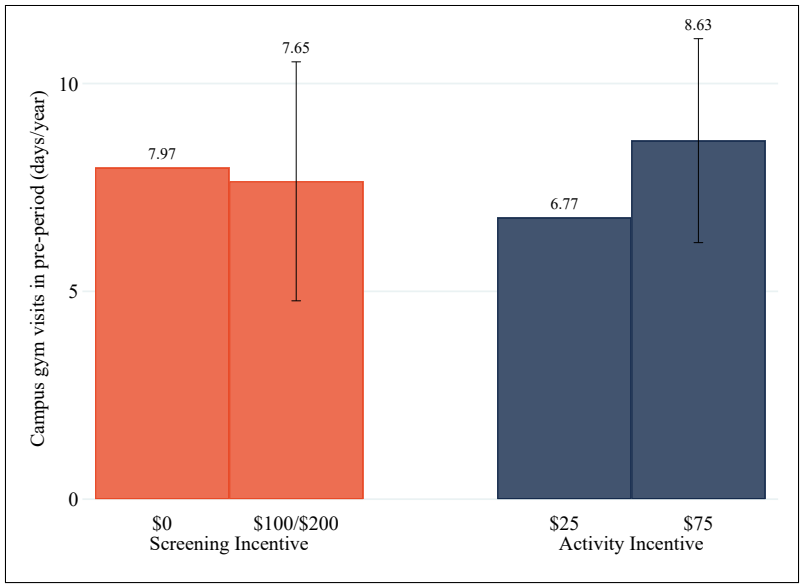
Figure 6: Marginal Selection on Medical Spending and Health Behaviors



(a) Total spending (dollars/month)



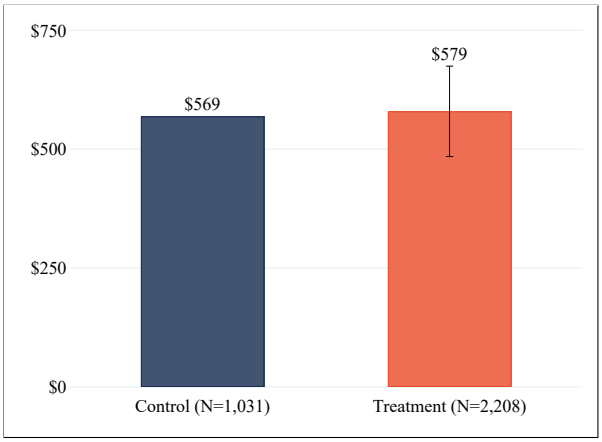
(b) IL Marathon/10K/5K participant (2014-2016)



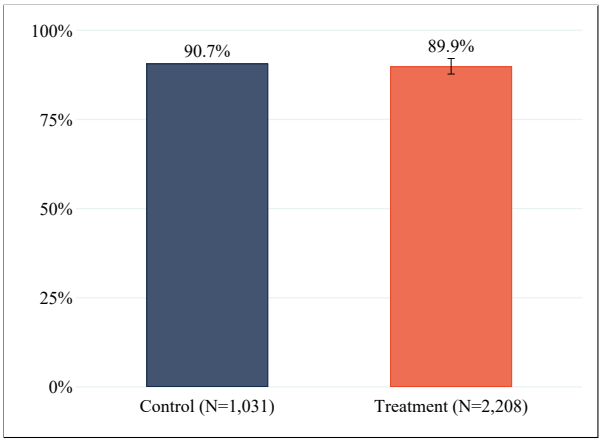
(c) Gym visits (days/year)

Notes: Each panel presents average characteristics of members of different treatment arms, conditional on having completed the screenings/HRA. The $100 and $200 treatment groups are combined. Vertical bars represent 95% confidence intervals on the difference in means between each pair of treatment groups.

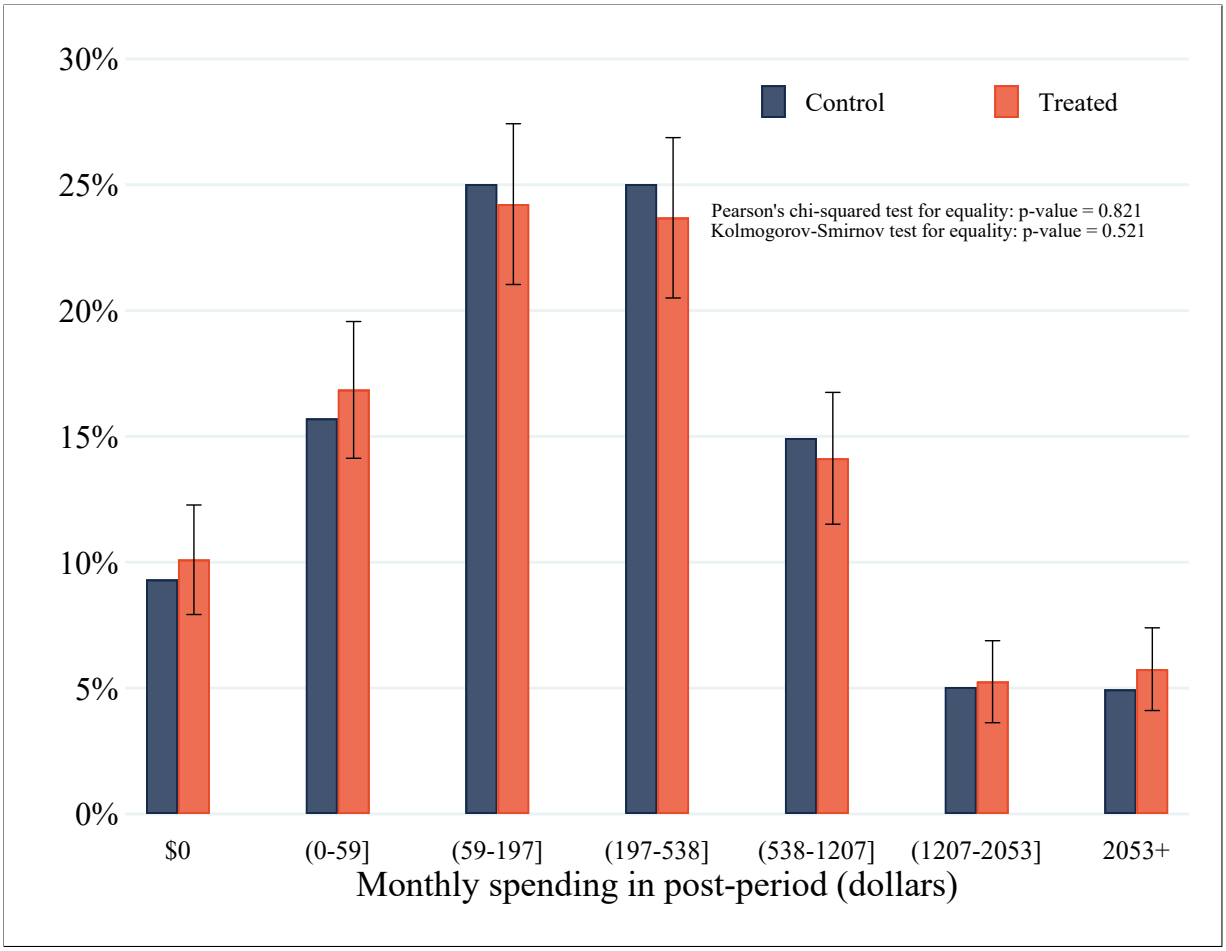
Figure 7: Post-intervention medical spending by treatment status



(a) Average monthly spending (dollars)



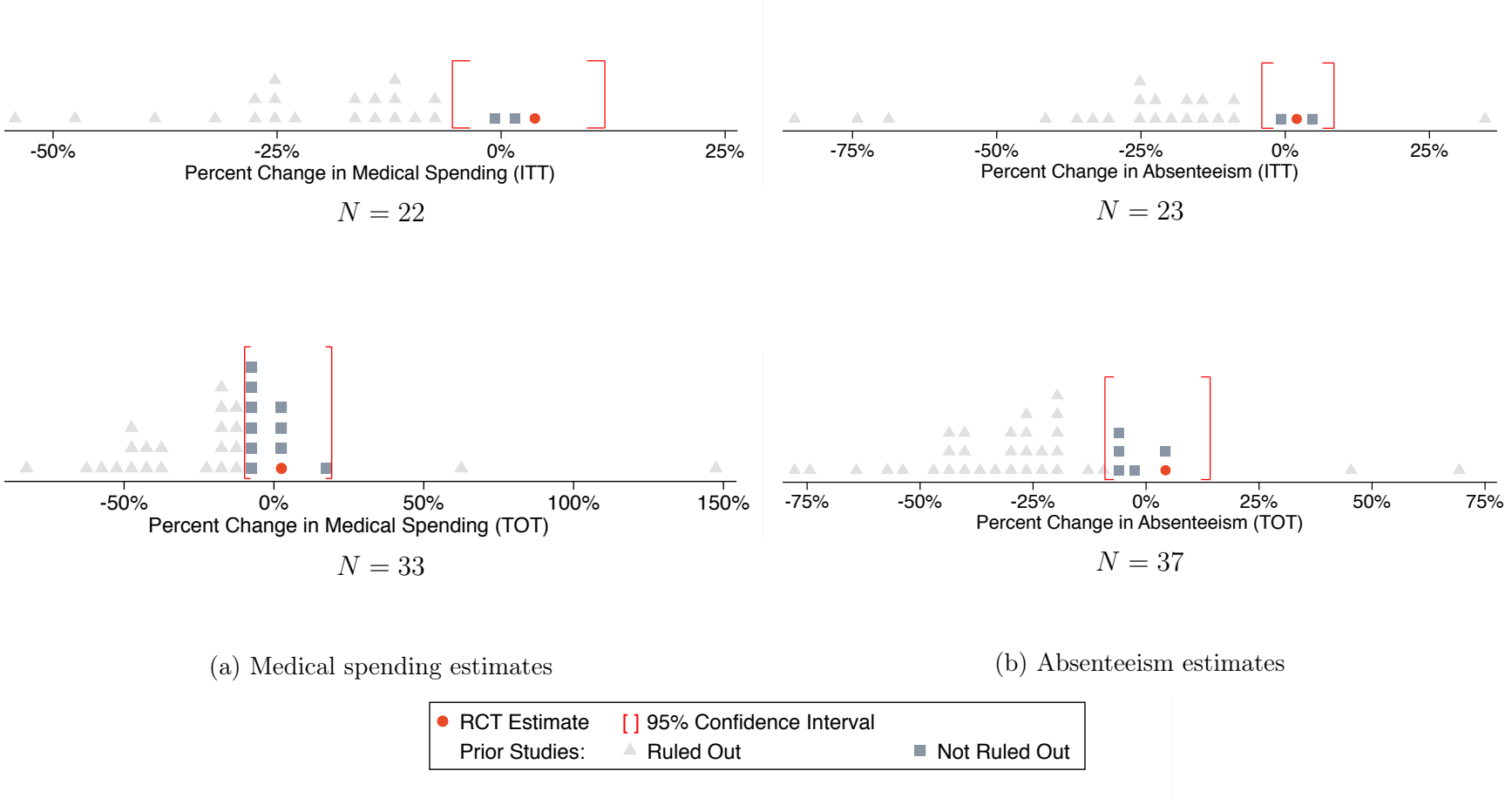
(b) Probability of any spending



(c) Histogram of average monthly spending, by quantile of control group spending ($N = 3,238$)

Notes: Results based on health care claims over the 12-month period August 2016–July 2017. The null hypothesis of the Pearson's chi-squared and the non-parametric Kolmogorov-Smirnov tests is that the two samples are drawn from the same distribution.

Figure 8: Comparison of experimental estimates to prior studies



(a) Medical spending estimates

(b) Absenteeism estimates

Notes: Each figure shows the distribution of N point estimates from prior workplace wellness studies. Panel (a) plots intent-to-treat (ITT) and treatment-on-the-treated (TOT) estimates for medical spending. Panel (b) plots corresponding estimates for absenteeism. The point estimates from our own study ("RCT Estimate"), and their associated confidence intervals, are taken from Table 7, Column 3, for medical spending, and Table 4, Column 4 and Table 6, Column 3 for absenteeism. Our RCT estimates and confidence intervals are plotted in order to demonstrate the share of prior study point estimates we are able to rule out. Appendix Table B.1 provides the full details of this meta-analysis.

Table 1a: Means of Study Variables at Baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Enrolled in Study								
	Not in Study	Control	A25	A75	B25	B75	C25	C75	p-value	Sample size
A. Stratification Variables										
Male	0.536	0.426	0.423	0.434	0.429	0.427	0.421	0.432	1.000	12,459
Age 50+	0.430	0.323	0.332	0.322	0.326	0.325	0.328	0.326	1.000	12,459
Age 37-49	0.362	0.340	0.330	0.333	0.330	0.336	0.330	0.335	0.999	12,459
White	0.774	0.841	0.828	0.847	0.835	0.832	0.842	0.831	0.971	12,459
Salary Q1 (bottom quartile)	0.234	0.244	0.243	0.239	0.246	0.237	0.241	0.244	1.000	12,459
Salary Q2	0.189	0.255	0.254	0.259	0.255	0.261	0.258	0.266	0.999	12,459
Salary Q3	0.197	0.249	0.252	0.260	0.250	0.248	0.250	0.240	0.996	12,459
Faculty	0.298	0.196	0.198	0.202	0.199	0.203	0.198	0.204	1.000	12,459
Academic Staff	0.324	0.443	0.439	0.439	0.438	0.434	0.436	0.435	1.000	12,459
B. 2016 Survey Variables										
Ever screened		0.885	0.895	0.900	0.891	0.876	0.887	0.902	0.817	4,834
Physically active		0.359	0.350	0.397	0.399	0.392	0.370	0.381	0.387	4,834
Trying to be active		0.822	0.799	0.791	0.799	0.843	0.797	0.827	0.161	4,834
Current smoker (cigarettes)		0.072	0.051	0.060	0.062	0.075	0.071	0.075	0.513	4,833
Current smoker (other)		0.085	0.075	0.062	0.089	0.089	0.096	0.100	0.224	4,833
Former smoker		0.198	0.216	0.186	0.185	0.204	0.211	0.171	0.481	4,833
Drinker		0.657	0.641	0.658	0.636	0.625	0.656	0.656	0.836	4,830
Heavy drinker		0.050	0.051	0.035	0.054	0.044	0.056	0.055	0.553	4,829
Chronic condition		0.729	0.751	0.729	0.712	0.741	0.701	0.721	0.562	4,834
Excellent or v. good health		0.586	0.613	0.619	0.612	0.604	0.563	0.603	0.433	4,834
Not poor health		0.989	0.982	0.991	0.993	0.987	0.995	0.989	0.509	4,834
Physical problems		0.392	0.387	0.395	0.380	0.392	0.401	0.375	0.979	4,834
Lots of energy		0.310	0.339	0.324	0.346	0.327	0.323	0.321	0.790	4,834
Bad emotional health		0.308	0.247	0.326	0.292	0.288	0.279	0.299	0.078	4,834
Overweight		0.545	0.577	0.530	0.507	0.518	0.552	0.514	0.202	4,834
High BP/cholesterol/glucose		0.308	0.328	0.281	0.292	0.266	0.290	0.313	0.273	4,834
Sedentary		0.545	0.569	0.499	0.538	0.571	0.530	0.545	0.239	4,833
Pharmaceutical drug utilization		0.723	0.736	0.710	0.710	0.670	0.708	0.701	0.286	4,830
Physician/ER utilization		0.772	0.797	0.734	0.774	0.712	0.715	0.760	0.003	4,833
Hospital utilization		0.038	0.036	0.020	0.024	0.022	0.034	0.026	0.168	4,833
Any sick days in past year		0.618	0.628	0.622	0.580	0.607	0.583	0.581	0.325	4,828
Worked 50+ hours/week		0.187	0.162	0.168	0.192	0.175	0.176	0.164	0.711	4,831
Very satisfied with job		0.396	0.385	0.426	0.408	0.389	0.435	0.408	0.534	4,832
Very or somewhat satisfied with job		0.836	0.858	0.829	0.841	0.847	0.842	0.852	0.818	4,832
Management priority on health/safety		0.771	0.797	0.780	0.746	0.781	0.791	0.796	0.399	4,831
Sample size	7,625	1,534	551	549	552	548	551	549		
Joint balance test for panel A (p-value)									1.000	4,834
Joint balance test for panel B (p-value)									0.165	4,817

Notes: Columns (1)-(8) report unweighted means for different, nonoverlapping subsets of university employees. Column (9) reports the p-value from a joint test of equality of the seven coefficients reported in Columns (2)-(8). We also estimate a seemingly unrelated regression model to test whether the variables listed in a particular panel predict enrollment into any of the seven control or treatment groups. The joint balance test row reports the p-value from jointly testing whether all regression coefficients across all seven study groups are equal to 0.

Table 1b: Means of Study Variables at Baseline, Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Enrolled in Study								
	Not in Study	Control	A25	A75	B25	B75	C25	C75	*p*-value	Sample size
C. Health Claims Variables (2015-2016)										
Total spending (dollars/month)	579	506	452	393	486	458	500	494	0.572	8,096
Office spending	54	67	61	53	54	49	79	50	0.331	8,096
Hospital spending	345	283	242	231	281	239	263	300	0.707	8,096
Drug spending	105	103	97	75	113	124	94	103	0.842	8,096
Non-zero medical spending	0.888	0.899	0.911	0.886	0.901	0.862	0.867	0.886	0.282	8,096
D. Health Behavior and Productivity Variables										
Sick leave (days/year)	5.89	6.05	6.53	5.82	5.69	6.36	6.24	6.13	0.394	12,459
Annual salary (dollars)	73,927	61,528	62,774	60,579	60,906	62,719	61,042	62,407	0.875	12,221
IL Marathon/10K/5K (2014-2016)	0.072	0.107	0.120	0.120	0.118	0.111	0.102	0.137	0.597	12,459
Campus gym visits (days/year)	6.14	7.36	5.44	8.68	7.68	5.69	5.34	7.86	0.119	12,459
Sample size	7,625	1,534	551	549	552	548	551	549		
Joint balance test for panel C (*p*-value)									0.207	3,223
Joint balance test for panel D (*p*-value)									0.438	4,770

Notes: Columns (1)-(8) report unweighted means for different, nonoverlapping subsets of university employees. Column (9) reports the *p*-value from a joint test of equality of the seven coefficients reported in Columns (2)-(8). We also estimate a seemingly unrelated regression model to test whether the variables listed in a particular panel predict enrollment into any of the seven control or treatment groups. The joint balance test row reports the *p*-value from jointly testing whether all regression coefficients across all seven study groups are equal to 0.

Table 2: Wellness Program Participation by Treatment Group

	(1)	(2)	(3)	(4)
A. Summer 2016 Screening and HRA Completion				
Group B* or C* (B25, B75, C25, C75)	0.136***	0.137***		
	(0.018)	(0.018)		
Group B* (B25, B75)			0.116***	0.117***
			(0.021)	(0.021)
Group C* (C25, C75)			0.156***	0.157***
			(0.021)	(0.021)
Group *75 (A75, B75, C75)	0.049***	0.050***	0.049***	0.049***
	(0.017)	(0.017)	(0.017)	(0.017)
Constant	0.445***	0.444***	0.445***	0.444***
	(0.017)	(0.017)	(0.017)	(0.017)
N	3,300	3,300	3,300	3,300
Strata FE	No	Yes	No	Yes
F Test	0.000	0.000	0.000	0.000
B. Fall 2016 Activity Completion				
Group B* or C* (B25, B75, C25, C75)	0.044***	0.044***		
	(0.016)	(0.016)		
Group B* (B25, B75)			0.039**	0.039**
			(0.019)	(0.018)
Group C* (C25, C75)			0.048**	0.049***
			(0.019)	(0.019)
Group *75 (A75, B75, C75)	0.124***	0.126***	0.124***	0.126***
	(0.015)	(0.015)	(0.015)	(0.015)
Constant	0.182***	0.182***	0.182***	0.182***
	(0.014)	(0.014)	(0.014)	(0.014)
N	3,300	3,300	3,300	3,300
Strata FE	No	Yes	No	Yes
F Test	0.000	0.000	0.000	0.000
C. Spring 2017 Activity Completion				
Group B* or C* (B25, B75, C25, C75)	0.045***	0.045***		
	(0.015)	(0.015)		
Group B* (B25, B75)			0.049***	0.049***
			(0.017)	(0.017)
Group C* (C25, C75)			0.040**	0.040**
			(0.017)	(0.017)
Group *75 (A75, B75, C75)	0.116***	0.117***	0.116***	0.117***
	(0.014)	(0.014)	(0.014)	(0.014)
Constant	0.137***	0.136***	0.137***	0.136***
	(0.013)	(0.013)	(0.013)	(0.013)
N	3,300	3,300	3,300	3,300
Strata FE	No	Yes	No	Yes
F Test	0.000	0.000	0.000	0.000

Notes: This table reports rates of completion for the three components of the wellness program tied to completion incentives. Each column in each panel reports estimates from a separate regression estimated over individuals in one of the six treatment groups (A25, A75, B25, B75, C25, and C75). The outcome in each regression is an indicator for completing the program component indicated by the panel, and the independent variables are indicators for inclusion in the specified treatment groups. The regressions reported in Columns (2) and (4) are the same as those reported in Columns (1) and (3), respectively, but with the addition of strata fixed effects. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table 3: Selection on Medical Spending, Income, and Health Behaviors

	(1)	(2)	(3)	(4)	(5)
Selection Variable	Mean	*N*	Completed Screening and HRA	Completed Fall Activity	Completed Spring Activity
A. Baseline Medical Spending					
Total spending (dollars/month) [admin]	479	2,188	-115.4**	-60.6	-62.5
			(52.2)	(43.6)	(44.3)
			[0.082]	[0.405]	[0.273]
Non-zero medical spending [admin]	0.885	2,188	0.050***	0.049***	0.046***
			(0.014)	(0.014)	(0.014)
			[0.008]	[0.005]	[0.020]
B. Baseline Income					
Annual salary (dollars) [admin]	61,736	3,257	-782.7	-3363.9***	-3429.1***
			(1248.3)	(1191.6)	(1251.8)
			[0.519]	[0.009]	[0.012]
Salary Q1 (bottom quartile) [admin]	0.242	3,300	-0.069***	-0.022	-0.036**
			(0.015)	(0.016)	(0.017)
			[0.000]	[0.398]	[0.121]
C. Baseline Health Behaviors					
IL Marathon/10K/5K (2014-2016) [admin]	0.118	3,300	0.089***	0.111***	0.090***
			(0.011)	(0.014)	(0.016)
			[0.000]	[0.000]	[0.000]
Campus gym visits (days/year) [admin]	6.780	3,300	2.178**	1.006	1.629
			(0.885)	(1.024)	(1.132)
			[0.013]	[0.328]	[0.153]

Notes: Column (1) reports the mean among subjects assigned to treatment. Columns (3)-(5) report the difference in means between those who completed the participation outcome and those who did not. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome (selection) variables in each family and are estimated using 10,000 bootstraps.

Table 4: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
A. Medical Spending				
Total spending (dollars/month) [admin]	576.1	10.6	20.7	34.7
		(48.5)	(48.7)	(36.9)
		[0.938]	[0.927]	[0.861]
	N=3,239	N=3,239	N=3,239	N=3,152
Drug spending [admin]	132.0	-8.5	-5.4	-6.1
		(26.5)	(25.7)	(12.0)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152
Office spending [admin]	69.5	-6.1	-5.7	-2.0
		(10.0)	(9.8)	(4.4)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152
Hospital spending [admin]	312.9	22.0	28.7	24.6
		(30.9)	(32.2)	(28.2)
		[0.938]	[0.865]	[0.869]
	N=3,239	N=3,239	N=3,239	N=3,152
Non-zero medical spending [admin]	0.902	-0.008	-0.007	0.002
		(0.011)	(0.011)	(0.010)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152
B. Employment and Productivity				
Annual salary (share of baseline salary) [admin]	0.059	-0.000	-0.002	-0.001
		(0.005)	(0.005)	(0.004)
		[0.969]	[0.875]	[0.922]
	N=4,146	N=4,146	N=4,146	N=4,130
Job promotion [admin]	0.176	-0.003	-0.006	-0.004
		(0.013)	(0.013)	(0.012)
		[0.952]	[0.875]	[0.922]
	N=4,146	N=4,146	N=4,146	N=4,130
Job terminated [admin]	0.113	-0.013	-0.014	-0.012
		(0.010)	(0.010)	(0.009)
		[0.630]	[0.509]	[0.571]
	N=4,834	N=4,834	N=4,834	N=4,753
Sick leave (days/year) [admin]	6.341	0.186	0.249	0.138
		(0.230)	(0.208)	(0.200)
		[0.816]	[0.563]	[0.880]
	N=4,782	N=4,782	N=4,782	N=4,712
Management priority on health/safety [survey]	0.790	0.057***	0.057***	0.050***
		(0.015)	(0.015)	(0.014)
		[0.001]	[0.001]	[0.003]
	N=3,566	N=3,566	N=3,566	N=3,514
C. Health Status and Behaviors				
IL Marathon/10K/5K 2017 [admin]	0.066	0.002	0.002	-0.005
		(0.008)	(0.008)	(0.006)
		[0.975]	[0.962]	[0.471]
	N=4,834	N=4,834	N=4,834	N=4,817
Campus gym visits (days/year) [admin]	5.839	-0.062	-0.068	0.401
		(0.733)	(0.721)	(0.360)
		[0.975]	[0.962]	[0.471]
	N=4,834	N=4,834	N=4,834	N=4,817
Ever screened [survey]	0.942	0.039***	0.042***	0.036***
		(0.009)	(0.009)	(0.008)
		[0.001]	[0.000]	[0.000]
	N=3,567	N=3,567	N=3,567	N=3,557

Notes: Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is specified by the table row. The focal independent variable is an indicator for inclusion in the treatment group, and the control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict the dependent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome variables in each family. See Appendix Tables A.2a-A.2f for results for all outcomes, categorized by family.

Table 5: Associations Between Administrative and Survey Measures of Productivity

	(1) Annual salary (share of baseline salary)	(2) Job promotion	(3) Job title change	(4) Job terminated	(5) Sick leave (days/year)
Any sick days in past year [survey]	-0.009*	0.005	0.009	-0.026***	3.242***
	(0.005)	(0.014)	(0.014)	(0.007)	(0.240)
	N=3,322	N=3,322	N=3,322	N=3,565	N=3,565
Worked 50+ hours/week [survey]	0.006	-0.032*	-0.036**	0.011	-3.278***
	(0.007)	(0.018)	(0.018)	(0.010)	(0.295)
	N=3,323	N=3,323	N=3,323	N=3,566	N=3,566
Very satisfied with job [survey]	0.011**	0.021	0.021	-0.013*	-1.004***
	(0.005)	(0.014)	(0.014)	(0.007)	(0.250)
	N=3,320	N=3,320	N=3,320	N=3,564	N=3,564
Very or somewhat satisfied with job [survey]	0.026***	0.050***	0.043**	-0.026**	-1.440***
	(0.005)	(0.017)	(0.017)	(0.011)	(0.334)
	N=3,320	N=3,320	N=3,320	N=3,564	N=3,564
Management priority on health/safety [survey]	0.013**	0.048***	0.046***	-0.023**	-0.184
	(0.005)	(0.015)	(0.016)	(0.010)	(0.299)
	N=3,322	N=3,322	N=3,322	N=3,566	N=3,566
Happier at work than last year [survey]	0.026***	0.085***	0.089***	-0.004	-0.894***
	(0.005)	(0.013)	(0.013)	(0.007)	(0.249)
	N=3,319	N=3,319	N=3,319	N=3,562	N=3,562
Feel very productive at work [survey]	0.004	0.017	0.014	-0.011	0.130
	(0.005)	(0.014)	(0.014)	(0.007)	(0.245)
	N=3,323	N=3,323	N=3,323	N=3,567	N=3,567
Received promotion [survey]	0.050***	0.229***	0.225***	-0.013*	0.007
	(0.005)	(0.013)	(0.013)	(0.007)	(0.246)
	N=3,319	N=3,319	N=3,319	N=3,562	N=3,562
Job search very likely [survey]	0.003	-0.049**	-0.046**	0.166***	-1.522***
	(0.007)	(0.019)	(0.020)	(0.018)	(0.347)
	N=3,319	N=3,319	N=3,319	N=3,561	N=3,561
Job search somewhat/very likely [survey]	0.004	-0.042***	-0.039***	0.086***	-0.502*
	(0.005)	(0.014)	(0.014)	(0.009)	(0.262)
	N=3,319	N=3,319	N=3,319	N=3,561	N=3,561
Sample size (outcome mean)	3,309	3,309	3,309	3,550	3,550
Outcome mean	0.061	0.184	0.192	0.045	6.473

Notes: Each row and column reports estimates from a separate regression, where observations include all individuals who completed the one-year follow-up survey. The outcome in each regression is specified by the table column. The independent variable is the survey response, which is always an indicator variable. (The Presenteeism survey variable is omitted from this table.) The outcome mean is calculated for the sample of observations that have non-missing values for all survey variables listed in this table. Regressions are unweighted. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes.

Table 6: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
A. Medical Spending						
Total spending (dollars/month) [admin]	17.3	33.8	52.0	-136.8**	-161.1**	-103.4*
	(79.0)	(78.7)	(59.4)	(68.6)	(66.0)	(61.9)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Drug spending [admin]	-13.8	-8.8	-12.8	-26.3	-34.5	-7.3
	(43.2)	(41.5)	(20.4)	(27.2)	(26.8)	(12.0)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Office spending [admin]	-9.9	-9.4	-3.2	12.2	9.5	8.8*
	(16.2)	(15.9)	(6.8)	(7.5)	(7.2)	(5.1)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Hospital spending [admin]	35.8	46.9	45.5	-117.5**	-126.0**	-83.0
	(50.4)	(52.1)	(45.7)	(55.7)	(52.7)	(51.8)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Non-zero medical spending [admin]	-0.013	-0.012	0.004	0.061***	0.043***	0.036***
	(0.018)	(0.018)	(0.016)	(0.014)	(0.013)	(0.012)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
B. Employment and Productivity						
Annual salary (share of baseline salary) [admin]	-0.000	-0.003	-0.003	0.004	0.005	0.006
	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)
	N=4,146	N=4,146	N=4,130	N=2,840	N=2,840	N=2,828
Job promotion [admin]	-0.006	-0.010	-0.009	0.019	0.015	0.009
	(0.022)	(0.021)	(0.021)	(0.014)	(0.014)	(0.015)
	N=4,146	N=4,146	N=4,130	N=2,840	N=2,840	N=2,828
Job terminated [admin]	-0.022	-0.024	-0.023	-0.080***	-0.078***	-0.063***
	(0.018)	(0.017)	(0.017)	(0.011)	(0.011)	(0.011)
	N=4834.000	N=4834.000	N=4753.000	N=3,300	N=3,300	N=3,244
Sick leave (days/year) [admin]	0.322	0.432	0.224	0.275	0.039	-0.068
	(0.398)	(0.358)	(0.344)	(0.272)	(0.253)	(0.251)
	N=4,782	N=4,782	N=4,712	N=3,264	N=3,264	N=3,216
Management priority on health/safety [survey]	0.087***	0.087***	0.077***	-0.004	-0.012	-0.007
	(0.023)	(0.023)	(0.021)	(0.017)	(0.017)	(0.016)
	N=3,566	N=3,566	N=3,514	N=2,410	N=2,410	N=2,376
C. Health Status and Behaviors						
IL Marathon/10K/5K 2017 [admin]	0.003	0.003	-0.011	0.059***	0.054***	0.024***
	(0.014)	(0.013)	(0.011)	(0.008)	(0.008)	(0.006)
	N=4834.000	N=4834.000	N=4,817	N=3,300	N=3,300	N=3,287
Campus gym visits (days/year) [admin]	-0.110	-0.121	0.757	3.527***	3.849***	2.160***
	(1.309)	(1.276)	(0.656)	(0.813)	(0.804)	(0.425)
	N=4834.000	N=4834.000	N=4,817	N=3,300	N=3,300	N=3,287
Ever screened [survey]	0.060***	0.065***	0.056***	0.073***	0.074***	0.061***
	(0.014)	(0.013)	(0.012)	(0.011)	(0.010)	(0.009)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404

Notes: Each row and column reports estimates from a separate regression. The outcome in each regression is specified by the table row, and the (endogenous) focal independent variable is an indicator for completing the screening and HRA. For the IV specifications (columns (1)-(3)), the instrument is an indicator for inclusion in the treatment group, and observations include individuals in the control or treatment groups. For the OLS specifications (columns (4)-(6)), there is no instrument and observations are restricted to individuals in the treatment group. The control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict either the dependent variable or the focal independent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table 7: First-Year Winsorized Medical Spending Treatment Effects

	(1)	(2)	(3)	(4)	(5)
A. ITT Estimates (Post-Lasso)					
Total spending (dollars/month) [admin]	34.7 (36.9) [-37.7, 107.0]	15.8 (23.2) [-29.8, 61.3]	17.0 (19.5) [-21.2, 55.2]	15.4 (13.5) [-11.1, 41.8]	10.8 (9.7) [-8.3, 29.9]
N	3,152	3,152	3,152	3,152	3,152
Winsorization (percent)	0	0.5	1	2.5	5
B. IV Estimates (Post-Lasso)					
Total spending (dollars/month) [admin]	52.0 (59.4) [-64.4, 168.4]	20.1 (38.0) [-54.4, 94.5]	21.8 (31.7) [-40.4, 83.9]	21.8 (21.9) [-21.0, 64.7]	16.6 (15.7) [-14.2, 47.4]
N	3,152	3,152	3,152	3,152	3,152
Winsorization (percent)	0	0.5	1	2.5	5

Notes: Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is winsorized (top-coded) average monthly medical spending over the first 12 months of the intervention, winsorized at the level indicated in each column. Regressions are weighted by the number of months of coverage. In Panel A (ITT), the focal independent variable is an indicator for inclusion in the treatment group, and all regressions include the same controls as the ITT post-Lasso specification reported in row 1 and column (4) of Table 4. In Panel B (IV), the (endogenous) focal independent variable is an indicator for completing the screening and HRA, the instrument is an indicator for inclusion in the treatment group, and all regressions include the same controls as the IV post-Lasso specification reported in row 1 and column (3) of Table 6. Column (1) replicates the (non-winsorized) ITT and IV post-Lasso results reported in Table 4 and Table 6. Robust standard errors are reported in parentheses, and 95% confidence intervals are reported in brackets. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table 8: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
A. Medical Spending					
Total spending (dollars/month) [admin]	29	628.1	-65.7	-56.8	-29.6
			(56.1)	(55.6)	(45.5)
			[0.693]	[0.786]	[0.683]
		N=3,302	N=3,302	N=3,302	N=3,155
Drug spending [admin]	29	145.1	-24.0	-22.4	-20.2
			(27.0)	(26.8)	(15.5)
			[0.847]	[0.872]	[0.620]
		N=3,302	N=3,302	N=3,302	N=3,155
Office spending [admin]	29	72.5	-6.6	-6.1	-4.9
			(8.5)	(8.4)	(5.4)
			[0.847]	[0.872]	[0.683]
		N=3,302	N=3,302	N=3,302	N=3,155
Hospital spending [admin]	29	337.9	-26.0	-19.8	-11.8
			(33.5)	(33.5)	(29.8)
			[0.847]	[0.872]	[0.703]
		N=3,302	N=3,302	N=3,302	N=3,155
Non-zero medical spending [admin]	29	0.949	0.003	0.004	0.008
			(0.008)	(0.008)	(0.007)
			[0.847]	[0.872]	[0.673]
		N=3,302	N=3,302	N=3,302	N=3,155
B. Employment and Productivity					
Annual salary (share of baseline salary) [admin]	30	0.121	-0.001	0.001	-0.001
			(0.007)	(0.006)	(0.006)
			[1.000]	[0.996]	[0.998]
		N=3,635	N=3,635	N=3,635	N=3,619
Job promotion [admin]	30	0.360	0.006	0.004	0.006
			(0.017)	(0.017)	(0.016)
			[0.996]	[0.996]	[0.996]
		N=3,635	N=3,635	N=3,635	N=3,619
Job terminated [admin]	30	0.204	0.002	0.000	0.002
			(0.012)	(0.012)	(0.012)
			[1.000]	[0.996]	[0.998]
		N=4,834	N=4,834	N=4,834	N=4,753
Sick leave (days/year) [admin]	30	6.066	0.013	0.059	0.018
			(0.204)	(0.175)	(0.169)
			[1.000]	[0.996]	[0.998]
		N=4,782	N=4,782	N=4,782	N=4,712
Management priority on health/safety [survey]	24	0.784	0.028*	0.028*	0.024
			(0.016)	(0.016)	(0.015)
			[0.539]	[0.534]	[0.632]
		N=3,018	N=3,018	N=3,018	N=2,976
C. Health Status and Behaviors					
IL Marathon/10K/5K 2018 [admin]	30	0.052	0.006	0.006	0.001
			(0.007)	(0.007)	(0.006)
			[0.625]	[0.594]	[0.995]
		N=4,834	N=4,834	N=4,834	N=4,817
Campus gym visits (days/year) [admin]	30	5.047	-0.342	-0.337	0.001
			(0.660)	(0.649)	(0.367)
			[0.625]	[0.595]	[0.998]
		N=4,834	N=4,834	N=4,834	N=4,817
Ever screened [survey]	24	0.962	0.029***	0.031***	0.027***
			(0.008)	(0.008)	(0.007)
			[0.006]	[0.003]	[0.005]
		N=3,020	N=3,020	N=3,020	N=3,010

Notes: Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is specified by the table row. The focal independent variable is an indicator for inclusion in the treatment group, and the control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict the dependent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome variables in each family. See Appendix Tables A.7a-A.7f for results for all outcomes, categorized by family.

Table 9: Longer-Run Winsorized Medical Spending Treatment Effects

	(1)	(2)	(3)	(4)	(5)
A. ITT Estimates (Post-Lasso)					
Total spending (dollars/month) [admin]	-29.6	-1.4	5.4	11.8	10.2
	(45.5)	(24.2)	(19.8)	(13.7)	(9.8)
	[-118.9, 59.7]	[-48.9, 46.0]	[-33.4, 44.2]	[-15.1, 38.7]	[-9.0, 29.4]
N	3,155	3,155	3,155	3,155	3,155
Winsorization (percent)	0	0.5	1	2.5	5

Notes: Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is winsorized (top-coded) average monthly medical spending over the first 29 months of the intervention, winsorized at the level indicated in each column. Regressions are weighted by the number of months of coverage. In Panel A (ITT), the focal independent variable is an indicator for inclusion in the treatment group, and all regressions include the same controls as the ITT post-Lasso specification reported in row 1 and column (5) of Table 8. Column (1) replicates the (non-winsorized) ITT result reported in Table 8. Robust standard errors are reported in parentheses, and 95% confidence intervals are reported in brackets. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Appendix A: For Online Publication Only

Appendix Figure A.1 reports how participation in the spring wellness activities varies as a function of rewards. It also reports the marginal cost of the additional participation induced by each reward. An increase in the size of screening incentives has modest, positive effects on participation rates for spring wellness activities; an increase in the size of participation incentives has a large, positive effect.

Panels (a)-(c) of Appendix Figure A.2 shows how selection on prior probability of nonzero medical spending, prior annual salary, and prior annual salary in the first quartile varies as a function of the monetary incentives assigned to study participants. Panel (a) shows that, at larger incentive levels, participants are slightly less likely to have non-zero medical spending in the prior year. Panels (b) and (c) show little effect of the size of incentives on selection with respect to annual salary.

Appendix Tables A.1a–A.1d provide selection results for the full set of pre-specified variables shown in Tables 1a and 1b using equation (3). Appendix Tables A.2a–A.2f report the causal, intent-to-treat (ITT) effect of our intervention on all pre-specified variables. In addition, Appendix Table A.2g provides results for different measures of medical utilization. Appendix Tables A.3a–A.3g provide the corresponding IV and OLS estimates of equation (5) for all pre-specified variables.

Appendix Tables A.4a and A.4b report intent-to-treat estimates for medical spending from a model that allows the treatment effect to vary by treatment group. We do not find statistically significant treatment effects for any treatment group in any of these specifications.

As discussed in Section 4.3.1, we find two statistically significant effects of our intervention: an increase in the number of employees who ever received a health screening, and an increase in employees who believe that management places a priority on health and safety. Because our monetary incentives were varied independently across the health screening and wellness activity components of our study, these incentives can be used as instruments for participation in those components. Appendix Table A.5 reports estimates of those IV regressions. For both outcomes, the effects are driven by the health screening component of our intervention.

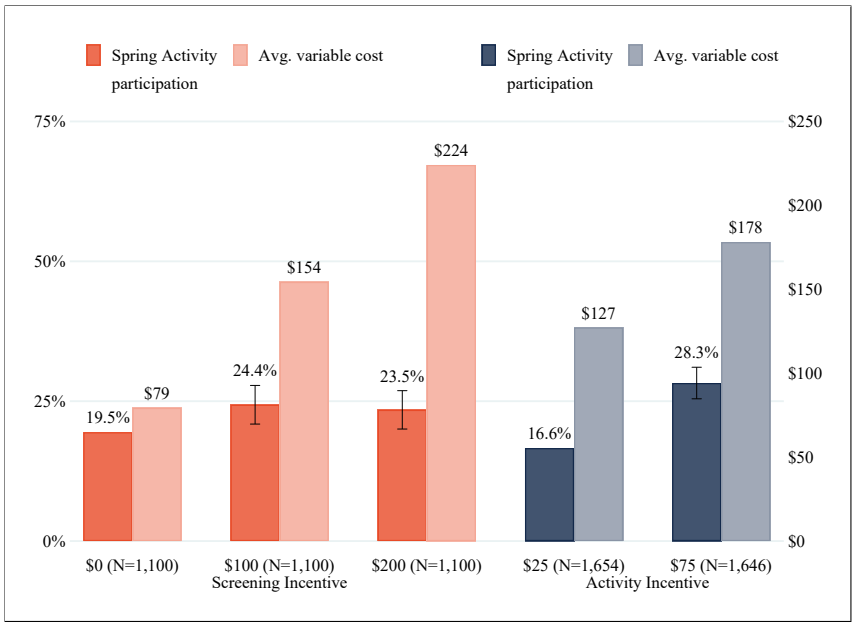
Appendix Table A.6 reports ITT and IV estimates of winsorized 12-month medical spending, for both "no control" and post-Lasso specifications.

Appendix Tables A.7a-A.7f report the causal, intent-to-treat (ITT) effect of our intervention on longer-run versions of all pre-specified variables. In addition, Appendix Table A.7g provides results for different longer-run measures of medical utilization. The time horizon of each longer-run outcome ranges from 24–30 months, as reported in column (1) of the table.

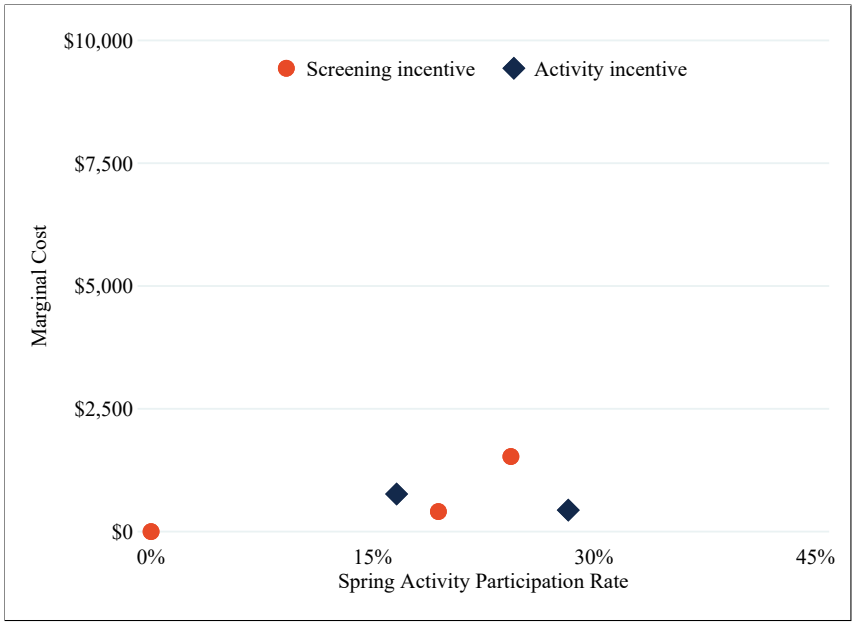
Appendix Table A.8 reports results of tests for differential attrition between the control and treatment groups. We perform these tests separately for our three main sources of data: health insurance claims, university administrative data, and online surveys. The fractions of control and treatment subjects included in the health insurance and university administrative datasets are very similar. The control group's survey completion rate is slightly higher than the treatment group's completion rate in both 2017 and 2018.

Finally, Appendix Table A.9 provides the definition, data source, and time period for every variable presented in the paper.

Figure A.1: Marginal cost of inducing additional participation into spring wellness activities



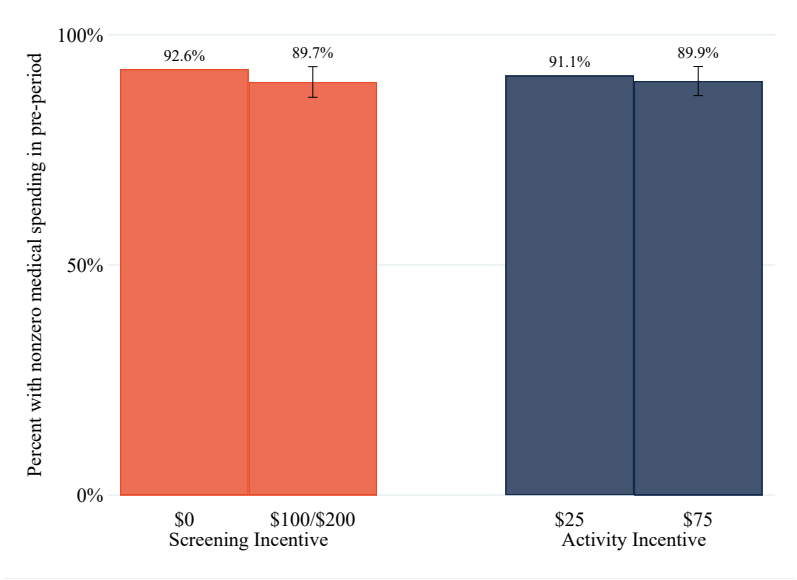
(a) Spring activities participation and costs



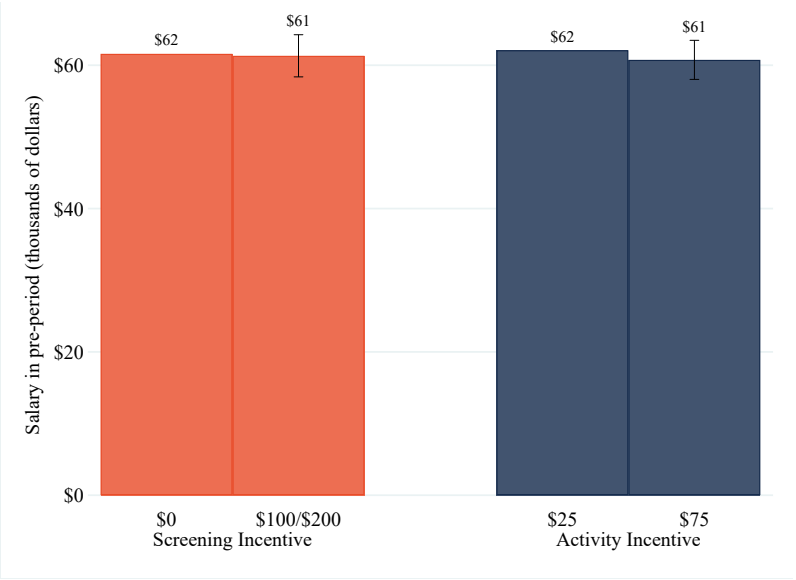
(b) Marginal cost of additional spring participation

Notes: Panel (a) plots health screening participation rates (PR) and average variable costs (AVC) as a function of screening and activity incentives. Vertical bars display 95% confidence intervals on the difference in means relative to the lowest reward group. AVC includes costs of the health screening, HRA, and wellness activities. Panel (b) plots the implied marginal costs (MC), calculated as MC=$\frac{\Delta \text{AVC}}{\Delta \text{PR}}$. The MC of the control group (PR=0 percent) is set equal to 0. We omit the MC for group C because its marginal PR is negative.

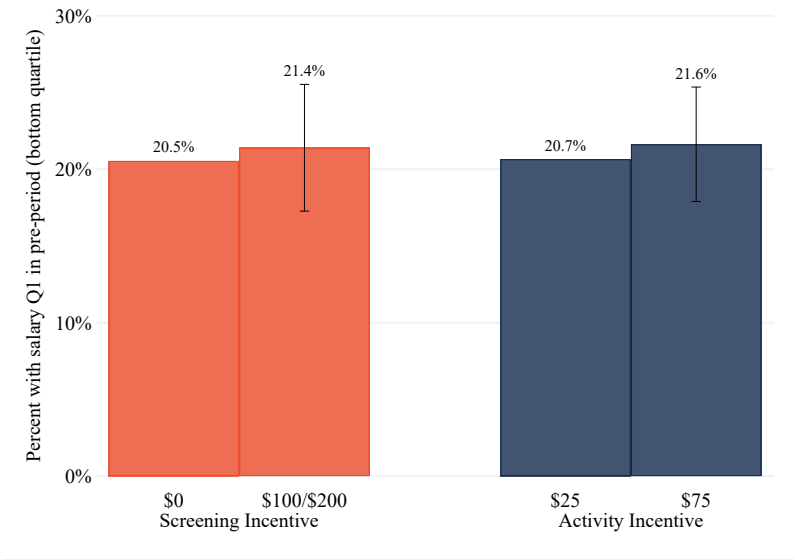
Figure A.2: Marginal Selection on Non-zero Spending and Income



(a) Non-zero Medical Spending



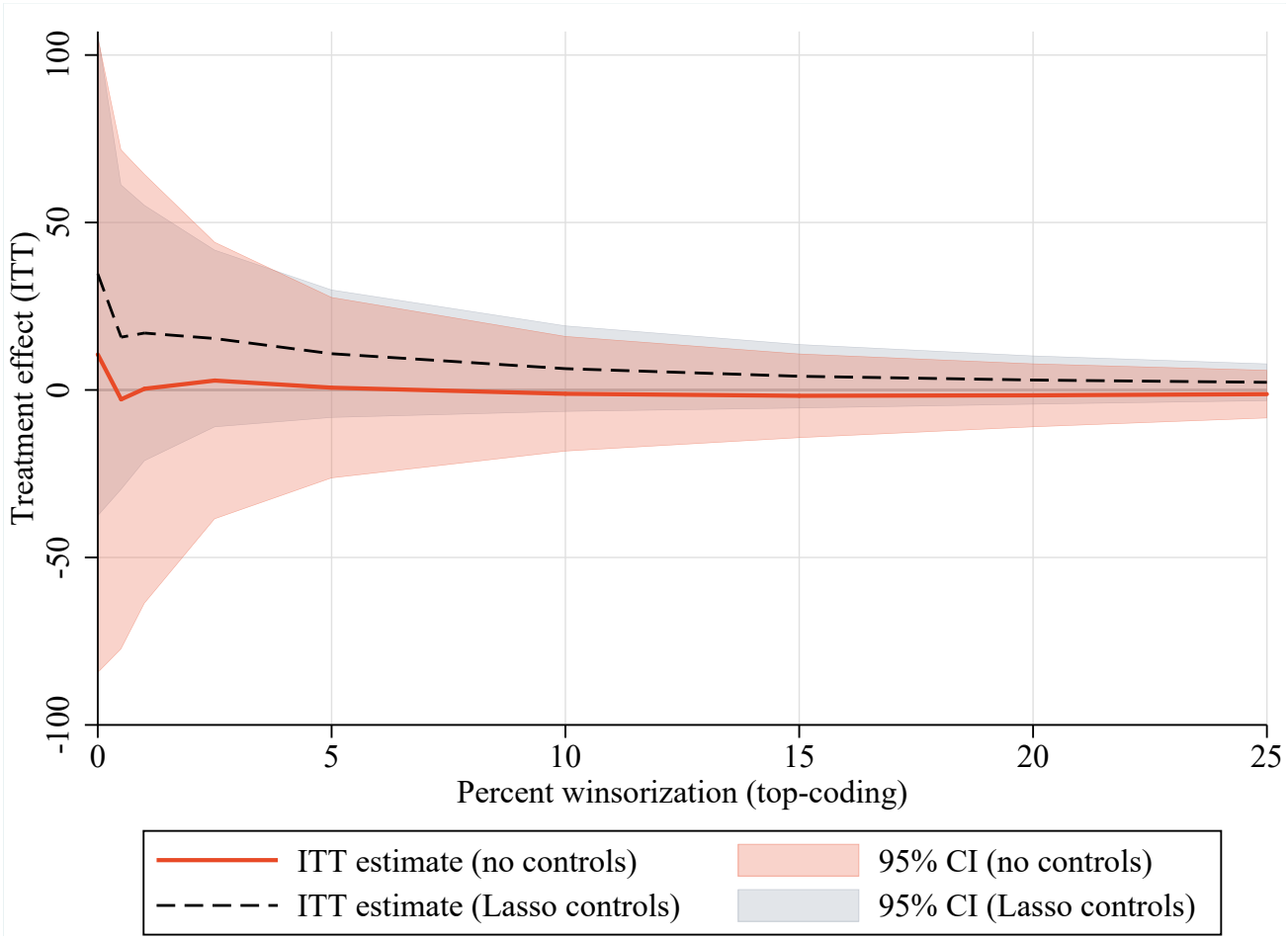
(b) Salary annual (thousands of dollars)



(c) Salary Q1 (bottom quartile)

Notes: Each panel presents average characteristics of members of different treatment arms, conditional on having completed the screenings/HRA. The $100 and $200 treatment groups are combined. Vertical bars represent 95% confidence intervals on the difference in means between each pair of treatment groups.

Figure A.3: Pre-intervention medical spending among treatment group, by participation status



Notes: The figure reports how intent-to-treat (ITT) medical spending effect estimates vary by the degree of winsorization (top-coding) of medical spending, calculated as the average monthly health care spending over the first 12 months of the wellness program (August 2016–July 2017). Each ITT estimate is estimated from a separate regression of medical spending (winsorized at the level indicated by the horizontal axis) on an indicator for inclusion in the treatment group. Observations include individuals in the control or treatment groups, and regressions are weighted by the number of months of medical coverage. The solid orange line reports estimates from a specification that includes no controls. The dashed black line reports estimates from a specification that includes the same controls as the ITT post-Lasso specification reported in row 1 and column (4) of Table 4. Shaded regions indicate 95% confidence intervals based on robust standard errors. The values of the ITT point estimates and confidence intervals for selected levels of winsorization are reported in Panel A (no controls) and Panel B (post-Lasso controls) of Table A.6.

Figure A.4: Experimental Design of the Illinois Workplace Wellness Study in Year 2

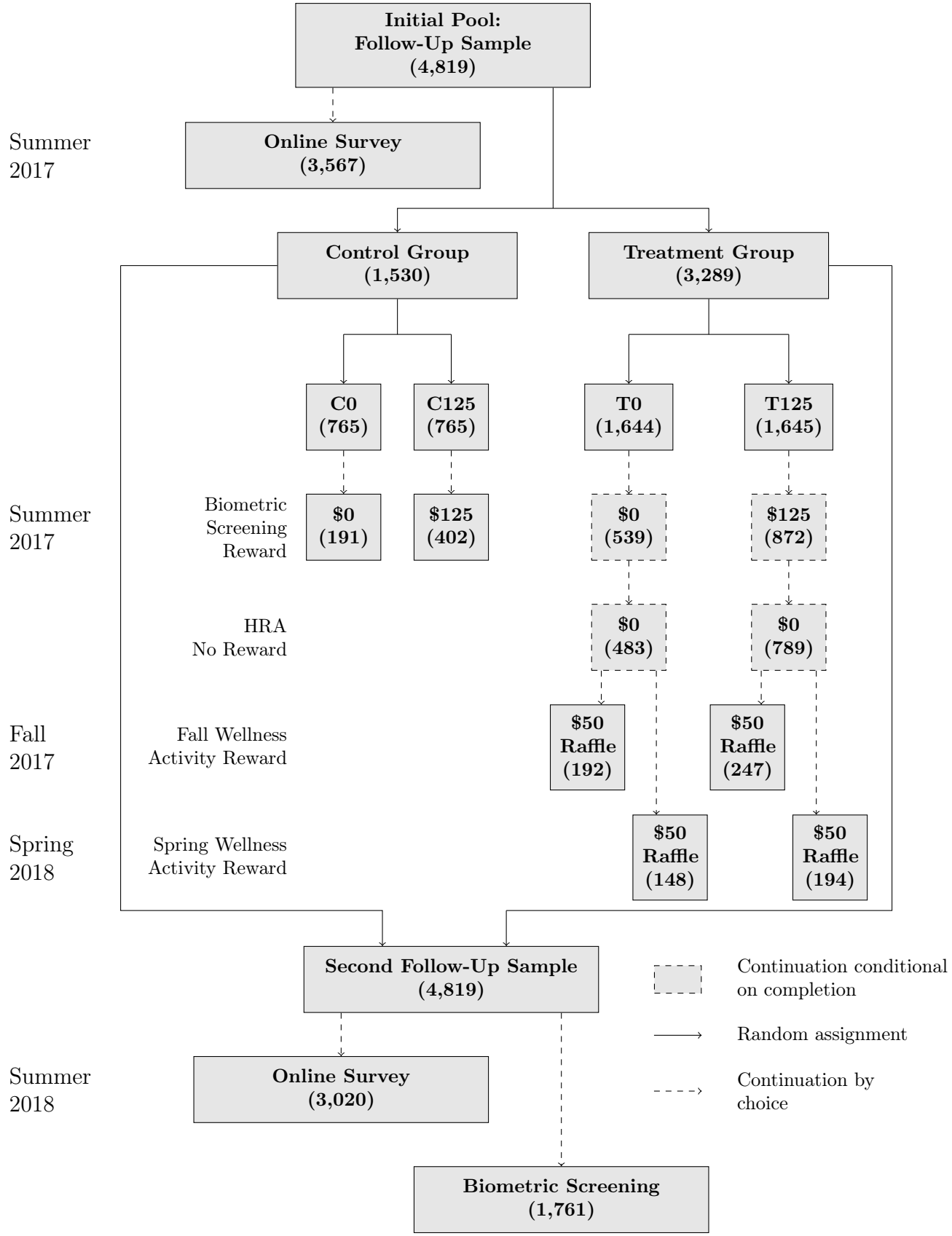


Table A.1a: Selection on Strata Variables

	(1)	(2)	(3)	(4)	(5)
Selection Variable	Mean	N	Completed Screening and HRA	Completed Fall Activity	Completed Spring Activity
Male [admin]	0.428	3,300	-0.058***	-0.114***	-0.149***
			(0.017)	(0.019)	(0.020)
			[0.005]	[0.000]	[0.000]
Age 50+ [admin]	0.327	3,300	-0.027	-0.015	-0.020
			(0.016)	(0.018)	(0.019)
			[0.270]	[0.399]	[0.473]
Age 37-49 [admin]	0.332	3,300	0.008	0.026	0.017
			(0.017)	(0.019)	(0.020)
			[0.850]	[0.398]	[0.473]
White [admin]	0.836	3,300	-0.001	0.046***	0.036**
			(0.013)	(0.014)	(0.015)
			[0.962]	[0.005]	[0.072]
Salary Q1 (bottom quartile) [admin]	0.242	3,300	-0.069***	-0.022	-0.036**
			(0.015)	(0.016)	(0.017)
			[0.000]	[0.398]	[0.121]
Salary Q2 [admin]	0.259	3,300	0.038**	0.028	0.058***
			(0.015)	(0.017)	(0.019)
			[0.052]	[0.346]	[0.012]
Salary Q3 [admin]	0.250	3,300	0.044***	0.043**	0.040**
			(0.015)	(0.017)	(0.019)
			[0.019]	[0.067]	[0.121]
Faculty [admin]	0.201	3,300	-0.051***	-0.098***	-0.097***
			(0.014)	(0.014)	(0.015)
			[0.002]	[0.000]	[0.000]
Academic Staff [admin]	0.437	3,300	0.077***	0.077***	0.086***
			(0.017)	(0.019)	(0.021)
			[0.000]	[0.001]	[0.000]

Notes: Column (1) reports the mean among subjects assigned to treatment. Columns (3)-(5) report the difference in means between those who completed the participation outcome and those who did not. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome (selection) variables in each family and are estimated using 10,000 bootstraps.

Table A.1b: Selection on Health Care Utilization Variables

	(1)	(2)	(3)	(4)	(5)
Selection Variable	Mean	N	Completed Screening and HRA	Completed Fall Activity	Completed Spring Activity
Total spending (dollars/month) [admin]	479	2,188	-115.4**	-60.6	-62.5
			(52.2)	(43.6)	(44.3)
			[0.082]	[0.405]	[0.273]
Office spending [admin]	59	2,188	2.4	-5.6	-12.4**
			(7.2)	(6.5)	(6.2)
			[0.742]	[0.641]	[0.146]
Hospital spending [admin]	268	2,188	-103.7**	-47.3*	-62.8**
			(40.3)	(28.3)	(27.5)
			[0.046]	[0.297]	[0.104]
Drug spending [admin]	104	2,188	-14.7	-4.3	14.5
			(20.6)	(25.5)	(28.9)
			[0.732]	[0.872]	[0.635]
Non-zero medical spending [admin]	0.885	2,188	0.050***	0.049***	0.046***
			(0.014)	(0.014)	(0.014)
			[0.008]	[0.005]	[0.020]
Pharmaceutical drug utilization [survey]	0.706	3,297	-0.001	0.029*	0.040**
			(0.016)	(0.018)	(0.019)
			[0.929]	[0.183]	[0.059]
Physician/ER utilization [survey]	0.748	3,300	0.050***	0.070***	0.061***
			(0.015)	(0.016)	(0.017)
			[0.003]	[0.000]	[0.002]
Hospital utilization [survey]	0.027	3,299	-0.012**	-0.005	-0.012**
			(0.006)	(0.006)	(0.006)
			[0.072]	[0.400]	[0.059]

Notes: Column (1) reports the mean among subjects assigned to treatment. Columns (3)-(5) report the difference in means between those who completed the participation outcome and those who did not. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome (selection) variables in each family and are estimated using 10,000 bootstraps.

Table A.1c: Selection on Employment and Productivity Variables

	(1)	(2)	(3)	(4)	(5)
Selection Variable	Mean	N	Completed Screening and HRA	Completed Fall Activity	Completed Spring Activity
Sick leave (days/year) [admin]	6.274	3,296	0.473*	0.705**	0.617**
			(0.267)	(0.290)	(0.312)
			[0.144]	[0.015]	[0.048]
Annual salary (dollars) [admin]	61,736	3,257	-782.7	-3363.9***	-3429.1***
			(1248.3)	(1191.6)	(1251.8)
			[0.519]	[0.009]	[0.012]
Any sick days in past year [survey]	0.600	3,296	0.043**	0.057***	0.051**
			(0.017)	(0.019)	(0.020)
			[0.049]	[0.008]	[0.046]
Worked 50+ hours/week [survey]	0.173	3,297	-0.058***	-0.065***	-0.064***
			(0.013)	(0.014)	(0.014)
			[0.000]	[0.000]	[0.000]
Very satisfied with job [survey]	0.408	3,299	0.002	0.002	0.002
			(0.017)	(0.019)	(0.021)
			[0.899]	[0.921]	[0.911]
Very or somewhat satisfied with job [survey]	0.845	3,299	0.023*	0.043***	0.029**
			(0.013)	(0.013)	(0.014)
			[0.193]	[0.005]	[0.092]
Management priority on health/safety [survey]	0.782	3,299	0.012	0.033**	0.035**
			(0.015)	(0.016)	(0.017)
			[0.618]	[0.062]	[0.092]

Notes: Column (1) reports the mean among subjects assigned to treatment. Columns (3)-(5) report the difference in means between those who completed the participation outcome and those who did not. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome (selection) variables in each family and are estimated using 10,000 bootstraps.

Table A.1d: Selection on Health and Behavior Variables

	(1)	(2)	(3)	(4)	(5)
Selection Variable	Mean	*N*	Completed Screening and HRA	Completed Fall Activity	Completed Spring Activity
IL Marathon/10K/5K (2014-2016) [admin]	0.118	3,300	0.089***	0.111***	0.090***
			(0.011)	(0.014)	(0.016)
			[0.000]	[0.000]	[0.000]
Campus gym visits (days/year) [admin]	6.780	3,300	2.178**	1.006	1.629
			(0.885)	(1.024)	(1.132)
			[0.013]	[0.328]	[0.153]
Ever screened [survey]	0.892	3,300	0.033***	0.042***	0.035***
			(0.011)	(0.011)	(0.012)
			[0.029]	[0.002]	[0.046]
Physically active [survey]	0.382	3,300	-0.015	0.013	0.040*
			(0.017)	(0.019)	(0.020)
			[0.909]	[0.964]	[0.445]
Trying to be active [survey]	0.809	3,300	0.045***	0.033**	0.030*
			(0.014)	(0.015)	(0.016)
			[0.014]	[0.293]	[0.445]
Current smoker (cigarettes) [survey]	0.065	3,299	-0.041***	-0.047***	-0.053***
			(0.009)	(0.008)	(0.008)
			[0.000]	[0.000]	[0.000]
Current smoker (other) [survey]	0.085	3,299	-0.034***	-0.046***	-0.066***
			(0.010)	(0.010)	(0.009)
			[0.011]	[0.000]	[0.000]
Former smoker [survey]	0.196	3,299	-0.009	-0.004	-0.019
			(0.014)	(0.015)	(0.016)
			[0.909]	[0.964]	[0.770]
Drinker [survey]	0.645	3,296	0.026	0.021	0.009
			(0.017)	(0.019)	(0.020)
			[0.707]	[0.889]	[0.929]
Heavy drinker [survey]	0.049	3,295	-0.010	-0.005	-0.006
			(0.008)	(0.008)	(0.009)
			[0.798]	[0.964]	[0.929]
Chronic condition [survey]	0.726	3,300	0.024	0.038**	0.023
			(0.016)	(0.017)	(0.018)
			[0.707]	[0.293]	[0.770]
Excellent or v. good health [survey]	0.602	3,300	-0.022	0.032*	0.060***
			(0.017)	(0.019)	(0.020)
			[0.798]	[0.626]	[0.045]
Not poor health [survey]	0.989	3,300	0.003	0.005	0.007*
			(0.004)	(0.004)	(0.003)
			[0.909]	[0.703]	[0.445]
Physical problems [survey]	0.388	3,300	0.022	-0.015	-0.027
			(0.017)	(0.019)	(0.020)
			[0.798]	[0.964]	[0.750]
Lots of energy [survey]	0.330	3,300	-0.031*	0.006	0.014
			(0.017)	(0.018)	(0.020)
			[0.502]	[0.964]	[0.929]
Bad emotional health [survey]	0.288	3,300	0.001	-0.019	-0.041**
			(0.016)	(0.018)	(0.018)
			[0.944]	[0.889]	[0.280]
Overweight [survey]	0.533	3,300	0.057***	0.015	-0.008
			(0.017)	(0.019)	(0.021)
			[0.015]	[0.964]	[0.929]
High BP/cholesterol/glucose [survey]	0.295	3,300	-0.007	-0.022	-0.034*
			(0.016)	(0.018)	(0.019)
			[0.909]	[0.866]	[0.445]
Sedentary [survey]	0.542	3,299	0.117***	0.115***	0.110***
			(0.017)	(0.019)	(0.020)
			[0.000]	[0.000]	[0.000]

Notes: Column (1) reports the mean among subjects assigned to treatment. Columns (3)-(5) report the difference in means between those who completed the participation outcome and those who did not. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome (selection) variables in each family and are estimated using 10,000 bootstraps.

Table A.2a: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
A. Medical Spending [admin]				
Total spending (dollars/month) [admin]	576.1	10.6	20.7	34.7
		(48.5)	(48.7)	(36.9)
		[0.938]	[0.927]	[0.861]
	N=3,239	N=3,239	N=3,239	N=3,152
Drug spending [admin]	132.0	-8.5	-5.4	-6.1
		(26.5)	(25.7)	(12.0)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152
Office spending [admin]	69.5	-6.1	-5.7	-2.0
		(10.0)	(9.8)	(4.4)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152
Hospital spending [admin]	312.9	22.0	28.7	24.6
		(30.9)	(32.2)	(28.2)
		[0.938]	[0.865]	[0.869]
	N=3,239	N=3,239	N=3,239	N=3,152
Non-zero medical spending [admin]	0.902	-0.008	-0.007	0.002
		(0.011)	(0.011)	(0.010)
		[0.938]	[0.927]	[0.947]
	N=3,239	N=3,239	N=3,239	N=3,152

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is specified by the table row. The focal independent variable is an indicator for inclusion in the treatment group, and the control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict the dependent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome variables in the table.

Table A.2b: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
A. Medical Spending [survey]				
Pharmaceutical drug utilization [survey]	0.725	-0.011	-0.009	-0.002
		(0.016)	(0.015)	(0.014)
		[0.851]	[0.864]	[0.895]
	N=3,567	N=3,567	N=3,567	N=2,433
Physician/ER utilization [survey]	0.745	0.003	0.002	0.018
		(0.016)	(0.015)	(0.017)
		[0.863]	[0.919]	[0.632]
	N=3,567	N=3,567	N=3,567	N=2,433
Hospital utilization [survey]	0.026	0.003	0.004	0.006
		(0.006)	(0.006)	(0.007)
		[0.851]	[0.864]	[0.632]
	N=3,567	N=3,567	N=3,567	N=2,433

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.2c: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
B. Employment and Productivity [admin]				
Annual salary (share of baseline salary) [admin]	0.059	-0.000	-0.002	-0.001
		(0.005)	(0.005)	(0.004)
		[0.969]	[0.875]	[0.922]
	N=4,146	N=4,146	N=4,146	N=4,130
Job promotion [admin]	0.176	-0.003	-0.006	-0.004
		(0.013)	(0.013)	(0.012)
		[0.952]	[0.875]	[0.922]
	N=4,146	N=4,146	N=4,146	N=4,130
Job title change [admin]	0.184	-0.006	-0.009	-0.008
		(0.013)	(0.013)	(0.013)
		[0.882]	[0.753]	[0.880]
	N=4,146	N=4,146	N=4,146	N=4,130
Job terminated [admin]	0.113	-0.013	-0.014	-0.012
		(0.010)	(0.010)	(0.009)
		[0.630]	[0.509]	[0.571]
	N=4,834	N=4,834	N=4,834	N=4,753
Sick leave (days/year) [admin]	6.341	0.186	0.249	0.138
		(0.230)	(0.208)	(0.200)
		[0.816]	[0.563]	[0.880]
	N=4,782	N=4,782	N=4,782	N=4,712

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.2d: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
B. Employment and Productivity [survey]				
Any sick days in past year [survey]	0.576	0.005	0.007	0.012
		(0.018)	(0.017)	(0.016)
		[0.997]	[0.994]	[0.961]
	N=3,565	N=3,565	N=3,565	N=3,514
Worked 50+ hours/week [survey]	0.150	-0.004	-0.008	0.005
		(0.013)	(0.012)	(0.010)
		[0.997]	[0.991]	[0.961]
	N=3,566	N=3,566	N=3,566	N=3,515
Very satisfied with job [survey]	0.387	-0.025	-0.028	-0.029*
		(0.017)	(0.017)	(0.015)
		[0.749]	[0.631]	[0.376]
	N=3,564	N=3,564	N=3,564	N=3,512
Very or somewhat satisfied with job [survey]	0.835	-0.004	-0.006	-0.013
		(0.013)	(0.013)	(0.012)
		[0.997]	[0.994]	[0.882]
	N=3,564	N=3,564	N=3,564	N=3,512
Management priority on health/safety [survey]	0.790	0.057***	0.057***	0.050***
		(0.015)	(0.015)	(0.014)
		[0.001]	[0.001]	[0.003]
	N=3,566	N=3,566	N=3,566	N=3,514
Happier at work than last year [survey]	0.542	0.009	0.005	-0.003
		(0.018)	(0.018)	(0.018)
		[0.995]	[0.994]	[0.978]
	N=3,562	N=3,562	N=3,562	N=3,510
Presenteeism [survey]	23.900	-0.023	-0.050	-0.151
		(0.261)	(0.259)	(0.238)
		[0.997]	[0.994]	[0.961]
	N=3,567	N=3,567	N=3,567	N=3,515
Feel very productive at work [survey]	0.449	-0.018	-0.013	-0.021
		(0.018)	(0.018)	(0.017)
		[0.930]	[0.991]	[0.868]
	N=3,567	N=3,567	N=3,567	N=3,515
Received promotion [survey]	0.472	0.008	0.000	0.002
		(0.018)	(0.018)	(0.018)
		[0.995]	[0.994]	[0.978]
	N=3,562	N=3,562	N=3,562	N=3,511
Job search very likely [survey]	0.139	0.031**	0.026**	0.027**
		(0.012)	(0.012)	(0.011)
		[0.095]	[0.208]	[0.143]
	N=3,561	N=3,561	N=3,561	N=3,511
Job search somewhat/very likely [survey]	0.337	0.019	0.012	0.013
		(0.017)	(0.017)	(0.016)
		[0.908]	[0.991]	[0.961]
	N=3,561	N=3,561	N=3,561	N=3,511

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.2e: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
C. Health Status and Behaviors [admin]				
IL Marathon/10K/5K 2017 [admin]	0.066	0.002	0.002	-0.005
		(0.008)	(0.008)	(0.006)
		[0.975]	[0.962]	[0.471]
	N=4,834	N=4,834	N=4,834	N=4,817
Campus gym visits (days/year) [admin]	5.839	-0.062	-0.068	0.401
		(0.733)	(0.721)	(0.360)
		[0.975]	[0.962]	[0.471]
	N=4,834	N=4,834	N=4,834	N=4,817

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.2f: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
C. Health Status and Behaviors [survey]				
Ever screened [survey]	0.942	0.039***	0.042***	0.036***
		(0.009)	(0.009)	(0.008)
		[0.001]	[0.000]	[0.000]
	N=3,567	N=3,567	N=3,567	N=3,557
Physically active [survey]	0.381	0.015	0.016	-0.009
		(0.017)	(0.017)	(0.012)
		[0.991]	[0.981]	[0.977]
	N=3,567	N=3,567	N=3,567	N=3,557
Trying to be active [survey]	0.825	0.005	0.007	0.017
		(0.014)	(0.014)	(0.012)
		[1.000]	[0.996]	[0.723]
	N=3,567	N=3,567	N=3,567	N=3,557
Current smoker (cigarettes) [survey]	0.060	-0.023**	-0.022**	-0.009*
		(0.009)	(0.009)	(0.005)
		[0.139]	[0.159]	[0.589]
	N=3,566	N=3,566	N=3,566	N=3,556
Drinker [survey]	0.672	-0.012	-0.013	-0.003
		(0.017)	(0.016)	(0.013)
		[0.998]	[0.983]	[0.992]
	N=3,565	N=3,565	N=3,565	N=3,555
Heavy drinker [survey]	0.047	-0.003	-0.002	0.003
		(0.008)	(0.008)	(0.007)
		[1.000]	[0.999]	[0.992]
	N=3,563	N=3,563	N=3,563	N=3,553
Chronic condition [survey]	0.735	-0.004	0.003	0.001
		(0.016)	(0.015)	(0.012)
		[1.000]	[0.999]	[0.997]
	N=3,565	N=3,565	N=3,565	N=3,555
Excellent or v. good health [survey]	0.564	-0.004	-0.007	-0.024
		(0.018)	(0.017)	(0.015)
		[1.000]	[0.996]	[0.689]
	N=3,567	N=3,567	N=3,567	N=3,557
Not poor health [survey]	0.990	-0.004	-0.005	-0.005*
		(0.003)	(0.003)	(0.003)
		[0.952]	[0.863]	[0.675]
	N=3,567	N=3,567	N=3,567	N=3,557
Physical problems [survey]	0.403	-0.007	-0.003	0.001
		(0.018)	(0.017)	(0.015)
		[1.000]	[0.999]	[0.997]
	N=3,567	N=3,567	N=3,567	N=3,557
Lots of energy [survey]	0.309	0.040**	0.039**	0.027*
		(0.016)	(0.016)	(0.014)
		[0.176]	[0.166]	[0.530]
	N=3,566	N=3,566	N=3,566	N=3,556
Bad emotional health [survey]	0.311	0.017	0.015	0.021
		(0.016)	(0.016)	(0.015)
		[0.977]	[0.981]	[0.723]
	N=3,566	N=3,566	N=3,566	N=3,556
Overweight [survey]	0.562	0.009	0.018	0.027**
		(0.018)	(0.017)	(0.011)
		[0.999]	[0.980]	[0.162]
	N=3,567	N=3,567	N=3,567	N=3,557
High BP/cholesterol/glucose [survey]	0.324	0.005	0.015	0.020
		(0.017)	(0.016)	(0.013)
		[1.000]	[0.981]	[0.699]
	N=3,567	N=3,567	N=3,567	N=3,557
Sedentary [survey]	0.560	0.001	-0.002	-0.008
		(0.018)	(0.017)	(0.013)
		[1.000]	[0.999]	[0.977]
	N=3,565	N=3,565	N=3,565	N=3,555

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.2g: First-Year Treatment Effects (ITT)

Outcome Variable	(1) Mean	(2) No Controls	(3) Strata FEs	(4) Post-Lasso
D. Medical Utilization (Quantity) [admin]				
Time to first claim <= 1 month [admin]	0.578	-0.029	-0.027	-0.011
		(0.019)	(0.018)	(0.015)
		[0.509]	[0.549]	[0.947]
	N=3,163	N=3,163	N=3,163	N=3,145
Time to first claim <= 2 months [admin]	0.689	-0.004	-0.001	0.012
		(0.018)	(0.017)	(0.014)
		[0.944]	[0.991]	[0.935]
	N=3,166	N=3,166	N=3,166	N=3,145
Time to first claim <= 3 months [admin]	0.758	0.008	0.010	0.022
		(0.016)	(0.016)	(0.014)
		[0.944]	[0.957]	[0.524]
	N=3,167	N=3,167	N=3,167	N=3,145
Time to first claim <= 6 months [admin]	0.842	-0.009	-0.007	-0.001
		(0.014)	(0.013)	(0.012)
		[0.944]	[0.970]	[0.976]
	N=3,176	N=3,176	N=3,176	N=3,147
Time to first claim <= 12 months [admin]	0.902	-0.008	-0.007	0.002
		(0.011)	(0.011)	(0.010)
		[0.944]	[0.957]	[0.976]
	N=3,239	N=3,239	N=3,239	N=3,152
Pharmaceutical events (days/month) [admin]	0.821	-0.022	-0.008	0.010
		(0.038)	(0.036)	(0.018)
		[0.944]	[0.991]	[0.953]
	N=3,239	N=3,239	N=3,239	N=3,152
Physician office visits (days/month) [admin]	0.308	0.032	0.031	0.010
		(0.025)	(0.026)	(0.018)
		[0.703]	[0.727]	[0.953]
	N=3,239	N=3,239	N=3,239	N=3,152
Hospital stays (days/month) [admin]	0.490	-0.011	-0.006	0.027
		(0.024)	(0.024)	(0.017)
		[0.944]	[0.991]	[0.524]
	N=3,239	N=3,239	N=3,239	N=3,152

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.2a for additional details.

Table A.3a: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
A. Medical Spending [admin]						
Total spending (dollars/month) [admin]	17.3	33.8	52.0	-136.8**	-161.1**	-103.4*
	(79.0)	(78.7)	(59.4)	(68.6)	(66.0)	(61.9)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Drug spending [admin]	-13.8	-8.8	-12.8	-26.3	-34.5	-7.3
	(43.2)	(41.5)	(20.4)	(27.2)	(26.8)	(12.0)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Office spending [admin]	-9.9	-9.4	-3.2	12.2	9.5	8.8*
	(16.2)	(15.9)	(6.8)	(7.5)	(7.2)	(5.1)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Hospital spending [admin]	35.8	46.9	45.5	-117.5**	-126.0**	-83.0
	(50.4)	(52.1)	(45.7)	(55.7)	(52.7)	(51.8)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Non-zero medical spending [admin]	-0.013	-0.012	0.004	0.061***	0.043***	0.036***
	(0.018)	(0.018)	(0.016)	(0.014)	(0.013)	(0.012)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140

Notes: Each row and column reports estimates from a separate regression. The outcome in each regression is specified by the table row, and the (endogenous) focal independent variable is an indicator for completing the screening and HRA. For the IV specifications (columns (1)-(3)), the instrument is an indicator for inclusion in the treatment group, and observations include individuals in the control or treatment groups. For the OLS specifications (columns (4)-(6)), there is no instrument and observations are restricted to individuals in the treatment group. The control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict either the dependent variable or the focal independent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table A.3b: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
A. Medical Spending [survey]						
Pharmaceutical drug utilization [survey]	-0.017	-0.013	0.000	0.022	0.018	0.018
	(0.024)	(0.023)	(0.020)	(0.019)	(0.019)	(0.019)
	N=3,567	N=3,567	N=2,433	N=2,410	N=2,410	N=1,641
Physician/ER utilization [survey]	0.004	0.002	0.025	0.024	0.020	0.016
	(0.024)	(0.023)	(0.025)	(0.019)	(0.019)	(0.022)
	N=3,567	N=3,567	N=2,433	N=2,410	N=2,410	N=1,641
Hospital utilization [survey]	0.005	0.006	0.012	-0.009	-0.010	-0.015
	(0.008)	(0.008)	(0.010)	(0.007)	(0.008)	(0.010)
	N=3,567	N=3,567	N=2,433	N=2,410	N=2,410	N=1,641

Notes: See notes to Appendix Table A.3a.

Table A.3c: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
B. Employment and Productivity [admin]						
Annual salary (share of baseline salary) [admin]	-0.000	-0.003	-0.003	0.004	0.005	0.006
	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)
	N=4,146	N=4,146	N=4,130	N=2,840	N=2,840	N=2,828
Job promotion [admin]	-0.006	-0.010	-0.009	0.019	0.015	0.009
	(0.022)	(0.021)	(0.021)	(0.014)	(0.014)	(0.015)
	N=4,146	N=4,146	N=4,130	N=2,840	N=2,840	N=2,828
Job title change [admin]	-0.011	-0.015	-0.015	0.015	0.012	0.006
	(0.022)	(0.022)	(0.022)	(0.015)	(0.015)	(0.015)
	N=4,146	N=4,146	N=4,130	N=2,840	N=2,840	N=2,828
Job terminated [admin]	-0.022	-0.024	-0.023	-0.080***	-0.078***	-0.063***
	(0.018)	(0.017)	(0.017)	(0.011)	(0.011)	(0.011)
	N=4834.000	N=4834.000	N=4753.000	N=3,300	N=3,300	N=3,244
Sick leave (days/year) [admin]	0.322	0.432	0.224	0.275	0.039	-0.068
	(0.398)	(0.358)	(0.344)	(0.272)	(0.253)	(0.251)
	N=4,782	N=4,782	N=4,712	N=3,264	N=3,264	N=3,216

Notes: See notes to Appendix Table A.3a.

Table A.3d: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
B. Employment and Productivity [survey]						
Any sick days in past year [survey]	0.007	0.011	0.021	0.004	-0.004	-0.020
	(0.027)	(0.026)	(0.024)	(0.021)	(0.021)	(0.019)
	N=3,565	N=3,565	N=3,514	N=2,409	N=2,409	N=2,376
Worked 50+ hours/week [survey]	-0.006	-0.013	0.008	-0.037**	-0.034**	-0.009
	(0.020)	(0.018)	(0.015)	(0.016)	(0.015)	(0.012)
	N=3,566	N=3,566	N=3,515	N=2,409	N=2,409	N=2,376
Very satisfied with job [survey]	-0.038	-0.042	-0.043*	-0.017	-0.018	-0.012
	(0.027)	(0.026)	(0.023)	(0.021)	(0.021)	(0.018)
	N=3,564	N=3,564	N=3,512	N=2,407	N=2,407	N=2,373
Very or somewhat satisfied with job [survey]	-0.006	-0.009	-0.020	0.003	0.001	0.005
	(0.020)	(0.020)	(0.018)	(0.016)	(0.016)	(0.015)
	N=3,564	N=3,564	N=3,512	N=2,407	N=2,407	N=2,373
Management priority on health/safety [survey]	0.087***	0.087***	0.077***	-0.004	-0.012	-0.007
	(0.023)	(0.023)	(0.021)	(0.017)	(0.017)	(0.016)
	N=3,566	N=3,566	N=3,514	N=2,410	N=2,410	N=2,376
Happier at work than last year [survey]	0.014	0.008	-0.004	0.022	0.023	0.013
	(0.027)	(0.027)	(0.027)	(0.021)	(0.022)	(0.021)
	N=3,562	N=3,562	N=3,510	N=2,408	N=2,408	N=2,374
Presenteeism [survey]	-0.035	-0.076	-0.226	-0.378	-0.304	-0.334
	(0.397)	(0.391)	(0.361)	(0.312)	(0.314)	(0.289)
	N=3,567	N=3,567	N=3,515	N=2,410	N=2,410	N=2,376
Feel very productive at work [survey]	-0.027	-0.020	-0.030	-0.040*	-0.043**	-0.036*
	(0.027)	(0.027)	(0.026)	(0.021)	(0.021)	(0.020)
	N=3,567	N=3,567	N=3,515	N=2,410	N=2,410	N=2,376
Received promotion [survey]	0.012	0.000	0.001	0.032	0.039*	0.024
	(0.027)	(0.027)	(0.027)	(0.021)	(0.021)	(0.021)
	N=3,562	N=3,562	N=3,511	N=2,408	N=2,408	N=2,375
Job search very likely [survey]	0.047**	0.040**	0.039**	-0.011	-0.013	-0.001
	(0.018)	(0.018)	(0.017)	(0.015)	(0.015)	(0.014)
	N=3,561	N=3,561	N=3,511	N=2,406	N=2,406	N=2,374
Job search somewhat/very likely [survey]	0.028	0.019	0.018	-0.030	-0.033*	-0.023
	(0.026)	(0.025)	(0.024)	(0.021)	(0.020)	(0.019)
	N=3,561	N=3,561	N=3,511	N=2,406	N=2,406	N=2,374

Notes: See notes to Appendix Table A.3a.

Table A.3e: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
C. Health Status and Behaviors [admin]						
IL Marathon/10K/5K 2017 [admin]	0.003	0.003	-0.011	0.059***	0.054***	0.024***
	(0.014)	(0.013)	(0.011)	(0.008)	(0.008)	(0.006)
	N=4834.000	N=4834.000	N=4,817	N=3,300	N=3,300	N=3,287
Campus gym visits (days/year) [admin]	-0.110	-0.121	0.757	3.527***	3.849***	2.160***
	(1.309)	(1.276)	(0.656)	(0.813)	(0.804)	(0.425)
	N=4834.000	N=4834.000	N=4,817	N=3,300	N=3,300	N=3,287

Notes: See notes to Appendix Table A.3a.

Table A.3f: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
C. Health Status and Behaviors [survey]						
Ever screened [survey]	0.060***	0.065***	0.056***	0.073***	0.074***	0.061***
	(0.014)	(0.013)	(0.012)	(0.011)	(0.010)	(0.009)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Physically active [survey]	0.023	0.025	-0.016	0.020	0.032	0.027*
	(0.026)	(0.026)	(0.019)	(0.021)	(0.020)	(0.015)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Trying to be active [survey]	0.008	0.010	0.028	0.052***	0.049***	0.036**
	(0.021)	(0.020)	(0.018)	(0.017)	(0.017)	(0.015)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Current smoker (cigarettes) [survey]	-0.035**	-0.034**	-0.014*	-0.033***	-0.032***	-0.005
	(0.014)	(0.013)	(0.008)	(0.010)	(0.010)	(0.006)
	N=3,566	N=3,566	N=3,556	N=2,410	N=2,410	N=2,404
Drinker [survey]	-0.018	-0.020	-0.007	0.010	0.015	-0.010
	(0.025)	(0.025)	(0.021)	(0.020)	(0.020)	(0.017)
	N=3,565	N=3,565	N=3,555	N=2,409	N=2,409	N=2,403
Heavy drinker [survey]	-0.004	-0.002	0.005	-0.003	-0.003	0.001
	(0.012)	(0.012)	(0.010)	(0.009)	(0.009)	(0.008)
	N=3,563	N=3,563	N=3,553	N=2,408	N=2,408	N=2,402
Chronic condition [survey]	-0.005	0.005	0.000	0.033*	0.037**	0.016
	(0.024)	(0.023)	(0.018)	(0.019)	(0.019)	(0.014)
	N=3,565	N=3,565	N=3,555	N=2,409	N=2,409	N=2,403
Excellent or v. good health [survey]	-0.007	-0.011	-0.034	-0.015	-0.018	0.005
	(0.027)	(0.026)	(0.023)	(0.021)	(0.021)	(0.018)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Not poor health [survey]	-0.006	-0.007	-0.008	0.009*	0.008	0.009*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Physical problems [survey]	-0.010	-0.005	0.000	0.025	0.026	0.011
	(0.027)	(0.026)	(0.024)	(0.021)	(0.021)	(0.020)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Lots of energy [survey]	0.060**	0.060**	0.036*	-0.030	-0.026	-0.013
	(0.025)	(0.024)	(0.022)	(0.020)	(0.020)	(0.018)
	N=3,566	N=3,566	N=3,556	N=2,410	N=2,410	N=2,404
Bad emotional health [survey]	0.026	0.022	0.035	-0.003	-0.005	0.003
	(0.025)	(0.025)	(0.023)	(0.020)	(0.020)	(0.019)
	N=3,566	N=3,566	N=3,556	N=2,410	N=2,410	N=2,404
Overweight [survey]	0.014	0.027	0.041**	0.031	0.029	-0.005
	(0.027)	(0.026)	(0.016)	(0.021)	(0.021)	(0.015)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
High BP/cholesterol/glucose [survey]	0.008	0.023	0.028	0.030	0.033*	0.032*
	(0.025)	(0.025)	(0.020)	(0.020)	(0.020)	(0.017)
	N=3,567	N=3,567	N=3,557	N=2,410	N=2,410	N=2,404
Sedentary [survey]	0.002	-0.003	-0.012	0.074***	0.056***	-0.003
	(0.027)	(0.026)	(0.020)	(0.021)	(0.021)	(0.016)
	N=3,565	N=3,565	N=3,555	N=2,408	N=2,408	N=2,402

Notes: See notes to Appendix Table A.3a.

Table A.3g: First-Year Treatment Effects: Experimental vs. Observational Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Experimental (IV)			Observational (OLS)		
Outcome Variable	No Controls	Strata FEs	Post-Lasso	No Controls	Strata FEs	Post-Lasso
D. Medical Utilization (Quantity) [admin]						
Time to first claim <= 1 month [admin]	-0.048	-0.045	-0.015	0.035	0.010	0.027
	(0.031)	(0.029)	(0.024)	(0.022)	(0.021)	(0.018)
	N=3,163	N=3,163	N=3,145	N=2,149	N=2,149	N=2,134
Time to first claim <= 2 months [admin]	-0.007	-0.002	0.019	0.054***	0.030	0.032*
	(0.029)	(0.027)	(0.023)	(0.021)	(0.020)	(0.017)
	N=3,166	N=3,166	N=3,145	N=2,152	N=2,152	N=2,134
Time to first claim <= 3 months [admin]	0.013	0.016	0.034	0.060***	0.035*	0.041**
	(0.027)	(0.025)	(0.023)	(0.019)	(0.019)	(0.017)
	N=3,167	N=3,167	N=3,145	N=2,153	N=2,153	N=2,134
Time to first claim <= 6 months [admin]	-0.014	-0.011	0.002	0.070***	0.050***	0.051***
	(0.023)	(0.022)	(0.020)	(0.017)	(0.016)	(0.015)
	N=3,176	N=3,176	N=3,147	N=2,160	N=2,160	N=2,136
Time to first claim <= 12 months [admin]	-0.013	-0.012	0.004	0.061***	0.043***	0.036***
	(0.018)	(0.018)	(0.016)	(0.014)	(0.013)	(0.012)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Pharmaceutical events (days/month) [admin]	-0.035	-0.013	0.015	-0.106**	-0.136***	-0.043**
	(0.061)	(0.058)	(0.030)	(0.045)	(0.043)	(0.022)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Physician office visits (days/month) [admin]	0.052	0.051	0.016	0.058*	0.044	0.042**
	(0.041)	(0.041)	(0.029)	(0.032)	(0.034)	(0.021)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140
Hospital stays (days/month) [admin]	-0.018	-0.009	0.041	-0.019	-0.038	-0.027
	(0.039)	(0.038)	(0.028)	(0.030)	(0.030)	(0.023)
	N=3,239	N=3,239	N=3,152	N=2,208	N=2,208	N=2,140

Notes: See notes to Appendix Table A.3a.

Table A.4a: First-Year Treatment Effects (ITT) by Treatment Group: Total Health Care Spending

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Group (any)	10.62	20.70				
	(48.49)	(48.67)				
Group A* (A25, A75)			16.72	33.36		
			(61.96)	(62.34)		
Group B* (B25, B75)			81.86	89.19		
			(95.64)	(95.16)		
Group C* (C25, C75)			9.83	16.71		
			(59.46)	(59.35)		
Group *75 (A75, B75, C75)			-52.64	-53.31		
			(60.11)	(60.21)		
Group A25					19.88	28.60
					(66.16)	(66.31)
Group B25					159.19	171.40
					(129.55)	(129.16)
Group C25					-68.27	-58.43
					(58.59)	(57.21)
Group A75					-39.18	-14.88
					(69.71)	(69.46)
Group B75					-51.01	-49.68
					(70.29)	(70.57)
Group C75					42.45	45.72
					(81.02)	(80.89)
Constant	568.83***	561.94***	568.83***	562.05***	568.83***	562.04***
	(38.00)	(37.53)	(38.02)	(37.55)	(38.03)	(37.56)
N	3,239	3,239	3,239	3,239	3,239	3,239
Strata FE	No	Yes	No	Yes	No	Yes
F Test	0.83	0.67	0.90	0.90	0.52	0.54

Notes: Each column reports estimates from a separate regression estimated over individuals in the treatment and control groups in the claims sample. The outcome in each regression is average monthly health care spending over the first 12 months of the wellness program (August 2016 - July 2017), and regressions are weighted by the number of months of coverage. The independent variables are indicators for inclusion in the specified treatment groups. Regressions reported in columns (2), (4), and (6) are the same as those reported in columns (1), (3), and (5) respectively, but with the addition of strata fixed effects. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table A.4b: First-Year Treatment Effects (ITT) by Treatment Group: Any Health Care Spending

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Group (any)	-0.008	-0.007				
	(0.011)	(0.011)				
Group A* (A25, A75)			0.004	0.004		
			(0.015)	(0.015)		
Group B* (B25, B75)			-0.012	-0.013		
			(0.016)	(0.015)		
Group C* (C25, C75)			-0.001	-0.006		
			(0.016)	(0.015)		
Group *75 (A75, B75, C75)			-0.009	-0.005		
			(0.013)	(0.012)		
Group A25					0.017	0.012
					(0.017)	(0.016)
Group B25					-0.015	-0.014
					(0.018)	(0.018)
Group C25					-0.011	-0.012
					(0.018)	(0.017)
Group A75					-0.018	-0.009
					(0.019)	(0.018)
Group B75					-0.019	-0.017
					(0.019)	(0.018)
Group C75					-0.000	-0.003
					(0.018)	(0.018)
Constant	0.907***	0.907***	0.907***	0.907***	0.907***	0.907***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
N	3,239	3,239	3,239	3,239	3,239	3,239
Strata FE	No	Yes	No	Yes	No	Yes
F Test	0.477	0.495	0.699	0.745	0.566	0.778

Notes: Each column reports estimates from a separate regression estimated over individuals in the treatment and control groups in the claims sample. The outcome in each regression is an indicator for positive health care spending over the first 12 months of the wellness program (August 2016 - July 2017). The independent variables are indicators for inclusion in the specified treatment groups. Regressions reported in columns (2), (4), and (6) are the same as those reported in columns (1), (3), and (5) respectively, but with the addition of strata fixed effects. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table A.5: First-Year IV Treatment Effects: Screening and Wellness Participation

	(1)	(2)	(3)	(4)
	Ever screened	Ever screened	Management priority on health/safety	Management priority on health/safety
Completed Screening and HRA	0.097***	0.098***	0.124**	0.117**
	(0.026)	(0.026)	(0.050)	(0.050)
Completed Fall and Spring Wellness Activities	-0.098	-0.087	-0.122	-0.103
	(0.061)	(0.059)	(0.119)	(0.118)
N	3,567	3,567	3,566	3,566
Strata FE	No	Yes	No	Yes
First-stage F-statistic	12.580	12.814	12.580	12.814

Notes: Each column reports estimates from a separate regression. The outcome variable is specified by the column heading. We instrument for both regressors using six indicators for inclusion in the six treatment groups. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level.

Table A.6: First-Year Winsorized Medical Spending Treatment Effects

	(1)	(2)	(3)	(4)	(5)
A. ITT Estimates (No Controls)					
Total spending (dollars/month) [admin]	10.6	-2.8	0.4	2.8	0.7
	(48.5)	(38.0)	(32.7)	(21.1)	(13.8)
	[-84.5, 105.7]	[-77.3, 71.8]	[-63.7, 64.5]	[-38.5, 44.2]	[-26.3, 27.7]
N	3,239	3,239	3,239	3,239	3,239
Winsorization (percent)	0	0.5	1	2.5	5
B. ITT Estimates (Post-Lasso)					
Total spending (dollars/month) [admin]	34.7	15.8	17.0	15.4	10.8
	(36.9)	(23.2)	(19.5)	(13.5)	(9.7)
	[-37.7, 107.0]	[-29.8, 61.3]	[-21.2, 55.2]	[-11.1, 41.8]	[-8.3, 29.9]
N	3,152	3,152	3,152	3,152	3,152
Winsorization (percent)	0	0.5	1	2.5	5
C. IV Estimates (No Controls)					
Total spending (dollars/month) [admin]	17.3	-4.6	0.6	4.6	1.1
	(79.0)	(61.9)	(53.3)	(34.4)	(22.5)
	[-137.6, 172.2]	[-126.0, 116.9]	[-103.8, 105.0]	[-62.8, 72.0]	[-42.9, 45.1]
N	3,239	3,239	3,239	3,239	3,239
Winsorization (percent)	0	0.5	1	2.5	5
D. IV Estimates (Post-Lasso)					
Total spending (dollars/month) [admin]	52.0	20.1	21.8	21.8	16.6
	(59.4)	(38.0)	(31.7)	(21.9)	(15.7)
	[-64.4, 168.4]	[-54.4, 94.5]	[-40.4, 83.9]	[-21.0, 64.7]	[-14.2, 47.4]
N	3,152	3,152	3,152	3,152	3,152
Winsorization (percent)	0	0.5	1	2.5	5

Notes: Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is winsorized (top-coded) average monthly health care spending over the first 12 months of the wellness program (August 2016 - July 2017), winsorized at the level indicated in each column. Regressions are weighted by the number of months of coverage. In Panels A and B (ITT), the focal independent variable is an indicator for inclusion in the treatment group. The specifications reported in Panel A do not include controls, while those reported in Panel B include the same controls as the ITT post-Lasso specification reported in row 1 and column (4) of Table 4. In Panels C and D (IV), the (endogenous) focal independent variable is an indicator for completing the screening and HRA and the instrument is an indicator for inclusion in the treatment group. The specifications reported in Panel C do not include controls, while those reported in Panel D include the same controls as the IV post-Lasso specification reported in row 1 and column (3) of Table 6. There is no winsorization of the outcome in column (1), and thus the ITT and IV estimates are identical to the total spending effects of the corresponding No Controls and Post-Lasso specifications reported in Table 4 and Table 6. Robust standard errors are reported in parentheses, and 95% confidence intervals are reported in brackets. A */**/*** indicates significance at the 10/5/1% level using conventional inference.

Table A.7a: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
A. Medical Spending [admin]					
Total spending (dollars/month) [admin]	29	628.1	-65.7	-56.8	-29.6
			(56.1)	(55.6)	(45.5)
			[0.693]	[0.786]	[0.683]
		N=3,302	N=3,302	N=3,302	N=3,155
Drug spending [admin]	29	145.1	-24.0	-22.4	-20.2
			(27.0)	(26.8)	(15.5)
			[0.847]	[0.872]	[0.620]
		N=3,302	N=3,302	N=3,302	N=3,155
Office spending [admin]	29	72.5	-6.6	-6.1	-4.9
			(8.5)	(8.4)	(5.4)
			[0.847]	[0.872]	[0.683]
		N=3,302	N=3,302	N=3,302	N=3,155
Hospital spending [admin]	29	337.9	-26.0	-19.8	-11.8
			(33.5)	(33.5)	(29.8)
			[0.847]	[0.872]	[0.703]
		N=3,302	N=3,302	N=3,302	N=3,155
Non-zero medical spending [admin]	29	0.949	0.003	0.004	0.008
			(0.008)	(0.008)	(0.007)
			[0.847]	[0.872]	[0.673]
		N=3,302	N=3,302	N=3,302	N=3,155

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. Each row and column reports estimates from a separate regression, where observations include individuals in the control or treatment groups. The outcome in each regression is specified by the table row. The focal independent variable is an indicator for inclusion in the treatment group, and the control strategy is specified by the column. Post-Lasso controls include covariates selected by Lasso to predict the dependent variable. The set of potential predictors include baseline values of all available variables in the same family of outcomes, strata variables, and the baseline (2016) survey variables reported in Table 1a, as well as all two-way interactions between these predictors. Robust standard errors are reported in parentheses. A */**/*** indicates significance at the 10/5/1% level using conventional inference, i.e., not adjusting for multiple outcomes. Family-wise p-values, reported in brackets, adjust for the number of outcome variables in the table.

Table A.7b: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
A. Medical Spending [survey]					
Pharmaceutical drug utilization [survey]	24	0.729	0.010	0.012	0.023
			(0.017)	(0.017)	(0.016)
			[0.911]	[0.845]	[0.393]
		N=3,019	N=3,019	N=3,019	N=2,104
Physician/ER utilization [survey]	24	0.765	-0.004	-0.007	-0.009
			(0.016)	(0.016)	(0.018)
			[0.911]	[0.874]	[0.695]
		N=3,019	N=3,019	N=3,019	N=2,105
Hospital utilization [survey]	24	0.027	0.003	0.003	0.006
			(0.006)	(0.006)	(0.007)
			[0.911]	[0.874]	[0.695]
		N=3,019	N=3,019	N=3,019	N=2,105

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.7c: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
B. Employment and Productivity [admin]					
Annual salary (share of baseline salary) [admin]	30	0.121	-0.001	0.001	-0.001
			(0.007)	(0.006)	(0.006)
			[1.000]	[0.996]	[0.998]
		N=3,635	N=3,635	N=3,635	N=3,619
Job promotion [admin]	30	0.360	0.006	0.004	0.006
			(0.017)	(0.017)	(0.016)
			[0.996]	[0.996]	[0.996]
		N=3,635	N=3,635	N=3,635	N=3,619
Job title change [admin]	30	0.377	-0.002	-0.004	-0.004
			(0.017)	(0.017)	(0.017)
			[1.000]	[0.996]	[0.998]
		N=3,635	N=3,635	N=3,635	N=3,619
Job terminated [admin]	30	0.204	0.002	0.000	0.002
			(0.012)	(0.012)	(0.012)
			[1.000]	[0.996]	[0.998]
		N=4,834	N=4,834	N=4,834	N=4,753
Sick leave (days/year) [admin]	30	6.066	0.013	0.059	0.018
			(0.204)	(0.175)	(0.169)
			[1.000]	[0.996]	[0.998]
		N=4,782	N=4,782	N=4,782	N=4,712

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.7d: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
B. Employment and Productivity [survey]					
Any sick days in past year [survey]	24	0.591	-0.009	-0.008	-0.007
			(0.019)	(0.019)	(0.017)
			[0.996]	[0.987]	[0.995]
		N=3,018	N=3,018	N=3,018	N=2,976
Worked 50+ hours/week [survey]	24	0.139	-0.034**	-0.032**	-0.023**
			(0.014)	(0.013)	(0.011)
			[0.136]	[0.154]	[0.383]
		N=3,018	N=3,018	N=3,018	N=2,976
Very satisfied with job [survey]	24	0.406	-0.010	-0.012	-0.006
			(0.019)	(0.019)	(0.017)
			[0.996]	[0.983]	[0.995]
		N=3,017	N=3,017	N=3,017	N=2,975
Very or somewhat satisfied with job [survey]	24	0.833	0.004	0.003	-0.002
			(0.015)	(0.015)	(0.013)
			[0.996]	[0.987]	[0.995]
		N=3,017	N=3,017	N=3,017	N=2,975
Management priority on health/safety [survey]	24	0.784	0.028*	0.028*	0.024
			(0.016)	(0.016)	(0.015)
			[0.539]	[0.534]	[0.632]
		N=3,018	N=3,018	N=3,018	N=2,976
Happier at work than last year [survey]	24	0.541	0.006	0.005	-0.001
			(0.019)	(0.019)	(0.019)
			[0.996]	[0.987]	[0.995]
		N=3,012	N=3,012	N=3,012	N=2,970
Presenteeism [survey]	24	23.899	0.316	0.292	0.258
			(0.282)	(0.281)	(0.258)
			[0.896]	[0.923]	[0.924]
		N=3,020	N=3,020	N=3,020	N=2,978
Feel very productive at work [survey]	24	0.437	-0.005	-0.007	-0.008
			(0.019)	(0.019)	(0.019)
			[0.996]	[0.987]	[0.995]
		N=3,017	N=3,017	N=3,017	N=2,975
Received promotion [survey]	24	0.487	-0.008	-0.013	-0.017
			(0.019)	(0.019)	(0.019)
			[0.996]	[0.983]	[0.930]
		N=3,017	N=3,017	N=3,017	N=2,975
Job search very likely [survey]	24	0.119	0.018	0.016	0.017
			(0.012)	(0.012)	(0.012)
			[0.742]	[0.797]	[0.740]
		N=3,016	N=3,016	N=3,016	N=2,974
Job search somewhat/very likely [survey]	24	0.296	0.015	0.018	0.019
			(0.018)	(0.017)	(0.016)
			[0.962]	[0.923]	[0.872]
		N=3,016	N=3,016	N=3,016	N=2,974

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.7e: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
C. Health Status and Behaviors [admin]					
IL Marathon/10K/5K 2018 [admin]	30	0.052	0.006	0.006	0.001
			(0.007)	(0.007)	(0.006)
			[0.625]	[0.594]	[0.995]
		N=4,834	N=4,834	N=4,834	N=4,817
Campus gym visits (days/year) [admin]	30	5.047	-0.342	-0.337	0.001
			(0.660)	(0.649)	(0.367)
			[0.625]	[0.595]	[0.998]
		N=4,834	N=4,834	N=4,834	N=4,817

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.7f: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
C. Health Status and Behaviors [survey]					
Ever screened [survey]	24	0.962	0.029***	0.031***	0.027***
			(0.008)	(0.008)	(0.007)
			[0.006]	[0.003]	[0.005]
		N=3,020	N=3,020	N=3,020	N=3,010
Physically active [survey]	24	0.373	0.039**	0.044**	0.030**
			(0.019)	(0.018)	(0.014)
			[0.359]	[0.170]	[0.363]
		N=3,020	N=3,020	N=3,020	N=3,010
Trying to be active [survey]	24	0.809	-0.004	-0.001	0.003
			(0.015)	(0.015)	(0.013)
			[1.000]	[1.000]	[0.999]
		N=3,020	N=3,020	N=3,020	N=3,010
Current smoker (cigarettes) [survey]	24	0.057	-0.021**	-0.023**	-0.013**
			(0.010)	(0.009)	(0.006)
			[0.290]	[0.180]	[0.387]
		N=3,017	N=3,017	N=3,017	N=3,007
Drinker [survey]	24	0.665	-0.009	-0.011	0.003
			(0.018)	(0.018)	(0.014)
			[1.000]	[1.000]	[0.999]
		N=3,017	N=3,017	N=3,017	N=3,007
Heavy drinker [survey]	24	0.048	0.005	0.005	0.007
			(0.008)	(0.008)	(0.007)
			[1.000]	[1.000]	[0.980]
		N=3,017	N=3,017	N=3,017	N=3,007
Chronic condition [survey]	24	0.758	-0.003	0.001	-0.006
			(0.017)	(0.016)	(0.013)
			[1.000]	[1.000]	[0.999]
		N=3,020	N=3,020	N=3,020	N=3,010
Excellent or v. good health [survey]	24	0.569	0.009	0.008	-0.005
			(0.019)	(0.019)	(0.017)
			[1.000]	[1.000]	[0.999]
		N=3,020	N=3,020	N=3,020	N=3,010
Not poor health [survey]	24	0.992	-0.006**	-0.006*	-0.006**
			(0.003)	(0.003)	(0.003)
			[0.359]	[0.462]	[0.456]
		N=3,020	N=3,020	N=3,020	N=3,010
Physical problems [survey]	24	0.406	-0.011	-0.009	-0.011
			(0.019)	(0.019)	(0.017)
			[1.000]	[1.000]	[0.998]
		N=3,020	N=3,020	N=3,020	N=3,010
Lots of energy [survey]	24	0.316	0.030*	0.032*	0.027*
			(0.018)	(0.018)	(0.016)
			[0.650]	[0.524]	[0.625]
		N=3,019	N=3,019	N=3,019	N=3,009
Bad emotional health [survey]	24	0.296	-0.005	-0.008	-0.004
			(0.018)	(0.018)	(0.016)
			[1.000]	[1.000]	[0.999]
		N=3,018	N=3,018	N=3,018	N=3,008
Overweight [survey]	24	0.564	0.005	0.008	0.006
			(0.019)	(0.019)	(0.013)
			[1.000]	[1.000]	[0.999]
		N=3,019	N=3,019	N=3,019	N=3,010
High BP/cholesterol/glucose [survey]	24	0.330	-0.011	-0.002	0.004
			(0.018)	(0.018)	(0.015)
			[1.000]	[1.000]	[0.999]
		N=3,019	N=3,019	N=3,019	N=3,009
Sedentary [survey]	24	0.553	0.008	0.009	-0.007
			(0.019)	(0.019)	(0.015)
			[1.000]	[1.000]	[0.999]
		N=3,018	N=3,018	N=3,018	N=3,008

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.7g: Longer-Run Treatment Effects (ITT)

Outcome Variable	(1) Horizon (months)	(2) Mean	(3) No Controls	(4) Strata FEs	(5) Post-Lasso
D. Medical Utilization (Quantity) [admin]					
Time to first claim <= 29 months [admin]	29	0.949	0.003	0.004	0.008
			(0.008)	(0.008)	(0.007)
			[0.886]	[0.933]	[0.714]
		N=3,302	N=3,302	N=3,302	N=3,155
Pharmaceutical events (days/month) [admin]	29	0.828	-0.024	-0.013	0.003
			(0.037)	(0.035)	(0.018)
			[0.886]	[0.933]	[0.958]
		N=3,302	N=3,302	N=3,302	N=3,155
Physician office visits (days/month) [admin]	29	0.322	0.012	0.013	-0.004
			(0.024)	(0.024)	(0.017)
			[0.886]	[0.933]	[0.958]
		N=3,302	N=3,302	N=3,302	N=3,155
Hospital stays (days/month) [admin]	29	0.500	-0.019	-0.016	0.014
			(0.024)	(0.024)	(0.017)
			[0.886]	[0.932]	[0.782]
		N=3,302	N=3,302	N=3,302	N=3,155

Notes: The outcomes in this table constitute a single family of outcomes for calculating family-wise p-values. See notes to Appendix Table A.7a for additional details.

Table A.8: Tests of Differential Attrition Between Control and Treatment Groups

	(1) Control	(2) Treatment	(3) Difference	(4) p-value
Health insurance enrollment (first 12 months)	0.634	0.638	0.004	0.768
Health insurance enrollment (first 30 months)	0.599	0.600	0.001	0.951
Job terminated (first 12 months)	0.122	0.109	-0.013	0.210
Job terminated (first 30 months)	0.202	0.204	0.002	0.862
2017 survey	0.754	0.730	-0.024	0.075
2018 survey	0.646	0.615	-0.031	0.036
Sample size	1,534	3,300		

Notes: This table reports health insurance enrollment rates and job termination rates for the first 12 and the first 30 months following randomization, and completion rates for the 2017 and 2018 online surveys. An individual's insurance enrollment is defined as the number of covered months divided by the length of the sample period (either 12 or 30 months). Columns (1)-(2) report unweighted means for the control and treatment groups. Column (3) reports the difference between the two means and column (4) reports the p-value from a joint test of equality of the two means.

Table A.9: Variable Definitions

Variable Name	Data Source	Survey Question(s)	Formula	Time Period
Male	Human resources data (C)	N/A	Sex = Male	May 30, 2016
Age 50+	Human resources data (C)	N/A	50 $\leq$ Age	May 30, 2016
Age 37-49	Human resources data (C)	N/A	37 $\leq$ Age $\leq$ 49	May 30, 2016
White	Human resources data (C)	N/A	Race = White	May 30, 2016
Salary Q1 (bottom quartile)	Human resources data (C)	N/A	Salary $\leq$ 25th percentile	Pre-period: May 30, 2016 Post-period: August 15, 2017
Salary Q2	Human resources data (C)	N/A	25th pctile $\leq$ Salary $\leq$ 50th pctile	Pre-period: May 30, 2016 Post-period: August 15, 2017
Salary Q3	Human resources data (C)	N/A	50th pctile $\leq$ Salary $\leq$ 75th pctile	Pre-period: May 30, 2016 Post-period: August 15, 2017
Faculty	Human resources data (C)	N/A	Employment Class = Faculty	May 30, 2016
Academic Staff	Human resources data (C)	N/A	Employment Class = Academic Staff	May 30, 2016
Annual salary	Human resources data (C)	N/A	N/A	Pre-period: May 30, 2016 Post-period: August 15, 2017
Annual salary (share of baseline salary)	Human resources data (C)	N/A	(Current salary / May 30, 2016 Salary) - 1	Pre-period: N/A Post-period: August 15, 2017
Job title change	Human resources data (C)	N/A	In case of multiple titles, only measures title changes associated with primary job.	Pre-period: N/A Post-period: August 15, 2017
Job promotion	Human resources data (C)	N/A	Job title change = 1 & Salary raise > 0	Pre-period: N/A Post-period: August 15, 2017
Job terminated	Human resources data (C)	N/A	TerminationDate<= August 15, 2017	Pre-period: N/A Post-period: August 15, 2017

Table A.9: Variable Definitions

Variable Name	Data Source	Survey Question(s)	Formula	Time Period
Sick leave (days/year)	Human resources data (C)	N/A	Sick days are measured monthly for CS employees, and biannually (August 15th and May 15th) for AP and Faculty employees. Number of sick days is normalized by fraction of year employed.	Pre-period:[37] 8/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17
Ever screened	2016 Online survey (A) 2017 Online survey (G)	A1 (G1) Have you ever had your cholesterol checked? A2 (G2) Have you ever had a blood test for high blood sugar or diabetes, other than during pregnancy? A3 (G3) Have you ever had a blood test for high blood sugar or diabetes? A4 (G4) In the last 12 months, have you had a Pap test or Pap smear? A5 (G5) In the last 12 months, have you had a mammogram? A8 (G8) In the last 12 months, have you had a sigmoidoscopy or a colonoscopy? A9 (G9) In the last 12 months, have you had a blood test to check for prostate cancer?	Any of A1-A5, A8-A9 (G1-G5, G8-G9) = "Yes"	(A) July 2016 (G) July 2017
Physically active	2016 Online survey (A) 2017 Online survey (G)	A11 (G11) Compared with most people your age, would you say you are more physically active, less physically active, or about the same?	A11 (G11) ="More active"	(A) July 2016 (G) July 2017
Trying to be active	2016 Online survey (A) 2017 Online survey (G)	A12 (G12) In the last 12 months, have you been told by a doctor or health professional to increase your physical activity or exercise? A13 (G13) Are you currently trying to increase your physical activity or exercise?	A12 (G12) = "Yes" or A13 (G13) = "Yes"	(A) July 2016 (G) July 2017
Current smoker (cigarettes)	2016 Online survey (A) 2017 Online survey (G)	A16 (G16) Have you smoked at least 100 cigarettes in your entire life? A17 (G17) Do you now smoke cigarettes every day, some days, or not at all?	A16 (G16) = "Yes" and A17 (G17) = "Every day" or "Some days"	(A) July 2016 (G) July 2017

[37] Pre- and Post-period are offset by 15 days for AP and Faculty employees (see description in Formula).

Table A.9: Variable Definitions

Variable Name	Data Source	Survey Question(s)	Formula	Time Period
Current smoker (other)	2016 Online survey (A) 2017 Online survey (G)	A22 (G22) Do you now smoke or use any other type of tobacco product, such as pipes, cigars, or chewing tobacco, every day, some days, or not at all? A23 (G23) Do you now use e-cigarettes (also known as vape-pens, hookah-pens, e-hookahs, or e-vaporizers) every day, some days, or not at all?	A22 (G22) & A23 (G23) != "Not at all"	(A) July 2016 (G) July 2017
Former smoker	2016 Online survey (A) 2017 Online survey (G)	A16 (G16) Have you smoked at least 100 cigarettes in your entire life? A17 (G17) Do you now smoke cigarettes every day, some days, or not at all?	A16 (G16) = "Yes" and A17 (G17) = "Not at all"	(A) July 2016 (G) July 2017
Drinker	2016 Online survey (A) 2017 Online survey (G)	A24 (G24) In the last 7 days, on how many days did you drink any type of alcoholic beverage?	A24 (G24) != 0	(A) July 2016 (G) July 2017
Heavy drinker	2016 Online survey (A) 2017 Online survey (G)	A25 (G25) In the last 7 days, on the days when you did drink alcohol, how many drinks did you usually have per day? One ?drink? is a 12 ounce can of beer, a 5 ounce glass of wine, or a 1.5 ounce shot of liquor.	A25 (G25) $\geq$ 4 if female A25 (G25) $\geq$ 5 if male	(A) July 2016 (G) July 2017
Chronic condition	2016 Online survey (A) 2017 Online survey (G)	A27 (G27) Have you ever been told by a doctor or other health professional that you have any of the following? Mark all that apply.	At least one box is checked	(A) July 2016 (G) July 2017
Excellent or v. good health	2016 Online survey (A) 2017 Online survey (G)	A28 (G28) Overall, how would you rate your health during the past 4 weeks?	A28 (G28) = "Excellent" or "Very good"	(A) July 2016 (G) July 2017
Not poor health	2016 Online survey (A) 2017 Online survey (G)	A28 (G28) Overall, how would you rate your health during the past 4 weeks?	A28 (G28) != "Poor"	(A) July 2016 (G) July 2017
Physical problems	2016 Online survey (A) 2017 Online survey (G)	A29 (G29) During the past 4 weeks, how much did physical health problems limit your usual physical activities (such as walking or climbing stairs)? A30 (G30) During the past 4 weeks, how much difficulty did you have doing your daily work, both at home and away from home, because of your physical health? A31 (G31) How much bodily pain have you had during the past 4 weeks?	A29 (G29)="Somewhat?, "Quite a lot?, "Could not do physical activities? **or** A30 (G30) = "Some?, "Quite a lot?, "Could not do daily work? **or** A31 (G31) = "Mild?, "Moderate?, "Severe?, "Very severe?	(A) July 2016 (G) July 2017
Lots of energy	2016 Online survey (A) 2017 Online survey (G)	A32 (G32) During the past 4 weeks, how much energy did you have?	A32 (G32) = "An extraordinary amount", or "Quite a lot"	(A) July 2016 (G) July 2017

Table A.9: Variable Definitions

Variable Name	Data Source	Survey Question(s)	Formula	Time Period
Bad emotional health	2016 Online survey (A) 2017 Online survey (G)	A33 (G33) During the past 4 weeks, how much have you been bothered by emotional problems (such as feeling anxious, depressed or irritable)?	A33 (G33) ov= “Moderately”, “Quite a lot”, “Extremely”	(A) July 2016 (G) July 2017
Overweight	2016 Online survey (A) 2017 Online survey (G)	A39 (39) How would you describe your body weight?	A39 (G39) = “Overweight” or “Very overweight”	
High BP / cholesterol / glucose	2016 Online survey (A) 2017 Online survey (G)	A40 (G40) How would you describe your blood pressure level? That is, if we measured it right now, do you think your blood pressure level would be: A41 (G41) How would you describe your cholesterol level? That is, if we measured it right now, do you think your cholesterol level would be: A42 (G42) How would you describe your blood glucose level? That is, if we measured it right now, do you think your blood glucose level would be:	A40 or A41 or A42 (G40 or G41 or G42) = “High” or “Very high”	(A) July 2016 (G) July 2017
Sedentary	2016 Online survey (A) 2017 Online survey (G)	A53 (G63) On an average day, how often does your job involve standing or walking around?	A53 (G63) = “None at all” or “Some, but less than 1 hour”	(A) July 2016 (G) July 2017
Pharmaceutical drug utilization	2016 Online survey (A) 2017 Online survey (G)	A34 (G34) How many different prescription medications are you currently taking? A35 (G35) How many different over-the-counter medications are you currently taking?	A34 (G34) $>$ 0 or A35 (G35) $>$ 0	(A) July 2016 (G) July 2017
Physician/ER utilization	2016 Online survey (A) 2017 Online survey (G)	A36 (G36) In the last 6 months, how many times did you go to a doctor?s office, clinic, emergency room, or other healthcare provider to get care for yourself? Do not include dental visits. Your best estimate is fine.	A36 (G36) != “None”	(A) July 2016 (G) July 2017
Hospital utilization	2016 Online survey (A) 2017 Online survey (G)	A37 (G37) = In the last 6 months, how many different times were you a patient in a hospital at least overnight? Do not include hospital stays to deliver a baby. Your best estimate is fine.	A37 (G37) != “None”	(A) July 2016 (G) July 2017
Any sick days in past year	2016 Online survey (A) 2017 Online survey (G)	A45 (G46) In the last 12 months, about how many days of work have you missed because of disability or poor health? Your best estimate is fine.	A45 (G46) != 0	(A) July 2016 (G) July 2017
Worked 50+ hours/week	2016 Online survey (A) 2017 Online survey (G)	A44 (G45) About how many hours a week do you usually work at your current job or jobs?	A44 (G45) = “50 or more”	(A) July 2016 (G) July 2017

Table A.9: Variable Definitions

Variable Name	Data Source	Survey Question(s)	Formula	Time Period
Very satisfied with job	2016 Online survey (A) 2017 Online survey (G)	A46 (G53) How satisfied are you with your current job?	A46 (G53) = "Very satisfied"	(A) July 2016 (G) July 2017
Very or somewhat satisfied with job	2016 Online survey (A) 2017 Online survey (G)	A46 (G53) How satisfied are you with your current job?	A46 (G53) = "Very satisfied" or "Somewhat satisfied"	(A) July 2016 (G) July 2017
Management priority on health/safety	2016 Online survey (A) 2017 Online survey (G)	A52 (G62) How much of a priority do you think your unit's management places on the health and safety of workers?	A52 (G62) = "Very high priority" or "Some priority"	(A) July 2016 (G) July 2017
Happier at work than last year	2017 Online survey (G)	G54 Do you feel happier at work this year than you did last year?	G54 = Yes	July 2017
Presenteeism	2017 Online survey (G)	G47 Despite having disability or poor health, I was able to finish hard tasks in my work. G48 At work, I was able to focus on achieving my goals despite disability or poor health. G49 Despite having disability or poor health, I felt energetic enough to complete all my work. G50 Because of disability or poor health, the stresses of my job were much harder to handle. G51 My disability or poor health distracted me from taking pleasure in my work. G52 I felt hopeless about finishing certain work tasks, due to my disability or poor health.	Stanford Presenteeim Scale (SPS-6), using G47-G52	July 2017
Feel very productive at work	2017 Online survey (G)	G56 How productive do you feel at work?	G56 = "Very productive"	July 2017
Received promotion	2017 Online survey (G)	G57 During the last 12 months, have you been given a promotion or more responsibility at work?	G57 = "Yes"	July 2017
Job search very likely	2017 Online survey (G)	G64 Taking everything into consideration, how likely are you to make a genuine effort to find a job with a new employer (outside the university) within the next year?	G64 = "Very likely"	July 2017
Job search somewhat / very likely	2017 Online survey (G)	G64 Taking everything into consideration, how likely are you to make a genuine effort to find a job with a new employer (outside the university) within the next year?	G64 = "Very likely" or "Somewhat likely"	July 2017
Total spending (dollars/month)	Health Insurance Claims Data (B)	N/A	Monthly Average	Pre-period: 7/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17
Drug spending	Health Insurance Claims Data (B)	N/A	Monthly Average	Pre-period: 7/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17

Table A.9: Variable Definitions

Variable Name	**Data Source**	**Survey Question(s)**	**Formula**	**Time Period**
Office spending	Health Insurance Claims Data (B)	N/A	Monthly Average	Pre-period: 7/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17
Hospital spending	Health Insurance Claims Data (B)	N/A	Monthly Average	Pre-period: 7/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17
Non-zero medical spending	Health Insurance Claims Data (B)	N/A	Monthly Average	Pre-period: 7/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17
IL Marathon/10K/5K	Human Resources Data (C)	N/A	Pre-period: participated in *at least one* event during 2014 - 2016	Pre-period: April 2014 - April 2016 Post-period: April 2017
Campus gym visits (days/year)	Human Resources Data (C)	N/A	Number of visits to gym, measured by ID card swipe-in	Pre-period: 8/1/15 - 7/31/16 Post-period: 8/1/16 - 7/31/17

Notes: Variable definitions are based on our pre-analysis plan, which is available at http://www.socialscienceregistry.org/trials/1368.