NBER WORKING PAPER SERIES

INCENTIVES AND HABIT FORMATION IN HEALTH SCREENINGS: EVIDENCE FROM THE ILLINOIS WORKPLACE WELLNESS STUDY

Damon Jones
David Molitor
Julian Reif

Working Paper 32745
http://www.nber.org/papers/w32745

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Damon Jones, David Molitor, and Julian Reif. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study
Damon Jones, David Molitor, and Julian Reif
NBER Working Paper No. 32745
July 2024
JEL No. H0,I1

ABSTRACT

We study habit formation in annual biometric health screenings using a field experiment that randomly assigned financial incentives to 4,799 employees over three years. Completing the first screening raised subsequent screenings by 32.4-36.0 percentage points (84%-90%) annually. Habit formation was similar whether employees were offered screenings as part of a comprehensive wellness program or just screenings alone, suggesting such habits can develop without frequent interactions. We rule out inattention as an explanation, using a subsample assigned more salient incentives. The long-run effect stems from the initial decision to participate, indicating a habit formation process with a one-shot mechanism.

Damon Jones
Harris School of Public Policy
University of Chicago
1307 E 60th Street
Chicago, IL 60637
and NBER
damonjones@uchicago.edu

Julian Reif
Department of Finance
University of Illinois Urbana-Champaign
515 E. Gregory Street
Urbana, IL 61820
and NBER
jreif@illinois.edu

David Molitor
University of Illinois Urbana-Champaign
340 Wohlers Hall
1206 S. Sixth Street
Urbana, IL 61820
and NBER
dmolitor@illinois.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/1368
Study website is available at
https://www.nber.org/programs-projects/projects-and-centers/illinois-workplace-wellness

1 Introduction

Smoking, poor diet, lack of exercise, and underuse of preventive care have all been linked to poor health outcomes. Changing these behaviors is difficult, even when the potential benefits to the individual may be large. For example, while evidence suggests that lifestyle modification programs can effectively prevent diabetes, many people at risk for the disease appear to be uninterested in such programs (Baicker, Mullainathan and Schwartzstein, 2015; Pryor and Volpp, 2018). Researchers have attributed these behavioral challenges to factors such as present bias, inattention, false beliefs, and poor access to health care (Royer, Stehr and Sydnor, 2015; Simon, Soni and Cawley, 2017; Bernheim, DellaVigna and Laibson, 2019). Financial incentives are often proposed as a solution, but questions remain about whether and how they create long-lasting habits, especially for infrequent behaviors such as annual checkups.

In this paper, we study habit formation in the decision to complete an annual biometric health screening program. We randomly assigned individual-level financial incentives to a group of 3,275 university employees ("Wave 1") every year for three years (2016–2018), and to a second group of 1,524 employees ("Wave 2") for two years (2017–2018). Our design yields five distinct experiments, one per wave and year of randomization, with incentive amounts ranging from \$75 to \$200. All five experiments allow us to estimate the contemporaneous ("direct") effect of the incentives on health screening participation. Three of the experiments enable us to further estimate whether the effects persist one year after randomization, and in one experiment we can also test for persistence after two years.

Together, these multiple experiments provide a robust framework for investigating whether and how financial incentives create long-lasting habits for infrequent health behaviors. Since the incentives were rerandomized and announced at the start of each program year, we can directly attribute changes in future screening behavior to the lasting impact of the initial incentive. The multiple waves in our study also enable us to test the replicability of these persistence effects across different program conditions. Additionally, the annual rounds of rerandomized incentives allow us to distinguish how current behavior is influenced by recent versus older incentives, a distinction that helps characterize the mechanisms underlying

habit formation in our setting.

We find that the incentives raised contemporaneous screening completion rates by 12.4 percentage points (22%) to 27.6 percentage points (71%) across the five experiments. The first-year incentives also increased future screenings in both waves. We estimate intent-to-treat effects of 4.5 percentage points (10%) and 8.9 percentage points (24%) on screening behavior one year later, and an intent-to-treat effect of 4.4 percentage points (12%) on behavior two years later. Instrumenting for screening completion using the randomly assigned incentives, we estimate that completing the first screening increases screenings one and two years later by 32.4–36.0 percentage points each year (84%–90%).

We define this last result, a causal effect of past screening completion on current screening completion, as habit formation.[1] Our instrumental variables (IV) analysis identifies this effect by using randomly assigned incentives as instruments for past participation. We address two potential threats to the required exclusion restriction necessary for our IV approach. First, we consider whether future incentives might correlate with initial incentives. However, balance tests confirm that the incentives were successfully rerandomized each year, ruling out this concern. Second, we address the threat of inattention, where subjects might have mistakenly believed that their initial incentives would remain the same in subsequent years. Our findings show that the effect of the first-year incentives on third-year participation is similar, regardless of whether subjects were assigned high or low incentives in the second year. This result indicates that our habit formation findings are not due to inattention.

To explore the mechanisms underlying habit formation, we perform two investigations. First, we compare the estimated degree of habit formation from the initial incentives across the two waves. Wave 1 individuals were offered three annual screenings during 2016–2018 and were also eligible for a comprehensive wellness program offered throughout the year between screenings.[2] Wave 2 individuals, who were offered screenings only in 2017 and 2018, faced a different sequence of financial incentives and were ineligible for the wellness program. Consequently, they received fewer communications from our research team and none from entities affiliated with the wellness program. Despite these differences, the one-

[1] Our definition of habit formation is consistent with Royer, Stehr and Sydnor (2015) and encompasses a variety of mechanisms, including addiction, learning, and taste discovery.

[2] This wellness program is studied in Jones, Molitor and Reif (2019) and Reif et al. (2020).

year IV estimate of 36.0 percentage points (84%) for Wave 1 is statistically indistinguishable from the Wave 2 estimate of 32.4 percentage points (89%). This finding suggests that habit formation in our setting arises without the need for more frequent communication or additional participation in related activities to support or build the habit.

Our second investigation focuses on data from Wave 1 individuals, who were eligible for three annual screenings. We compare the estimated degree of habit formation from the first and second incentives offered to this group and find that completing the 2016 screening increased 2018 completion rates by 33.1 percentage points, whereas completing the 2017 screening had a small and statistically insignificant effect on 2018 rates. This result indicates that habit formation in our setting is driven by initial screening completion rather than subsequent screenings. This pattern rules out reinforcement mechanisms like addiction, where recent consumption has the greatest effect on current behavior. Rather, it is consistent with mechanisms that follow a one-shot structure, such as information acquisition or taste discovery, where the initial effect has a lasting impact with minimal decay.

Our study builds on a large literature documenting habit formation in diverse settings, from smoking to voting to social media (Becker, Grossman and Murphy, 1991; Gruber and Köszegi, 2001; Fujiwara, Meng and Vogl, 2016; Allcott, Gentzkow and Song, 2022).[3] We use a randomized design, implemented separately for two different waves, to provide causal evidence of habit formation in the important domain of health screenings. A novel feature of our design is using multiple rounds of rerandomized incentives to characterize the effects of a series of treatment doses. While many prior studies of habit formation attribute their findings to a reinforcement model, such as the addiction model of Becker and Murphy (1988), our results instead support a one-shot model, such as learning, with no evidence of decay over the two-year follow-up period. This finding is particularly policy relevant, as the one-shot model implies that the cost-effectiveness of an intervention will substantially decrease after the first round. In line with this implication, our results show that while financial incentives always increase contemporaneous behavior, only the first incentive has positive spillovers on future behavior.

[3] Additional settings include energy conservation (Allcott and Rogers, 2014), political engagement (Bursztyn et al., 2021), recycling (Vollaard and van Soest, 2024), and sleep (Giuntella, Saccardo and Sadoff, 2024).

We also contribute to the literature on habit formation in health behaviors, which has predominantly focused on high-frequency activities such as exercising. A pioneering study by Charness and Gneezy (2009) finds that financial incentives for attending the gym for one month increased gym attendance for at least seven weeks after the intervention ended. More recent studies on exercise have either failed to find habit formation effects from simple incentives (Royer, Stehr and Sydnor, 2015; Patel et al., 2016; Carrera et al., 2018; Rohde and Verbeke, 2017) or found short-run effects that faded in subsequent months (Volpp et al., 2008; Acland and Levy, 2015; Carrera et al., 2020). However, more sophisticated incentives, such as commitment or time-bundled contracts, have produced long-term effects for gym attendance and walking (Royer, Stehr and Sydnor, 2015; Aggarwal, Dizon-Ross and Zucker, 2022). Financial incentives have also produced lasting effects up to three months post-intervention in children's healthy eating (Loewenstein, Price and Volpp, 2016) and up to nine months in daily hand washing (Hussam et al., 2022), though the effect was diminishing in the latter case. Our study adds to this work by providing evidence on habit formation for biometric health screenings, a low-frequency (annual) behavior. Our findings have important implications for public health goals, such as increasing annual vaccinations, as they show that even infrequent health behaviors can become habitual.

Finally, our study contributes to the extensive literature on using financial incentives to encourage positive changes in health behaviors. Similar to our study, this literature has explored low-frequency activities, including vaccinations (Stone et al., 2002; Campos-Mercade et al., 2021), cancer and cardiovascular screenings (Stone et al., 2002; Alsan, Garrick and Graziani, 2019), and workplace health screenings (Jones, Molitor and Reif, 2019; Song and Baicker, 2019). However, little is known about habit formation for these activities, which typically occur once per year, at most. One notable exception is Schneider et al. (2023), who find that a $24 financial incentive for the initial COVID-19 vaccine dose in Sweden did not significantly affect the likelihood of taking subsequent doses. In contrast, our study examines the effects of much larger incentives ($75–$200) on health screenings. Unlike Schneider et al. (2023), we find that incentives for initial participation have lasting effects that are economically large and statistically significant.

2 Background

The Illinois Workplace Wellness Study is a randomized controlled trial designed to evaluate the effects of workplace wellness programs on employee health, behavior, and productivity (Jones, Molitor and Reif, 2019; Reif et al., 2020). Conducted at the University of Illinois at Urbana-Champaign, the study randomly assigned 3,300 benefits-eligible employees to a "wellness" group, which was eligible to participate in a two-year workplace wellness program from 2016 to 2018. The remaining 1,534 employees were assigned to a "non-wellness" group. We limit our analysis to subjects who were continuously enrolled in the study through the end of the intervention, resulting in a final sample size of 4,799.

Employees in the wellness group, referred to as Wave 1 members, were invited for on-campus biometric health screenings in the fall of 2016, 2017, and 2018. During these annual screenings, clinicians measured each employee's height, weight, waist circumference, and blood pressure, and administered a fingerstick blood test to check for cholesterol, triglyceride, and glucose levels. Employees received their results within minutes and reviewed them with a health coach.

Wave 1 members who completed a health screening in 2016 or 2017 were also invited to complete an online health risk assessment (HRA), a questionnaire designed to assess a person's health habits. Completing the HRA made them eligible to choose from various wellness activities offered throughout the academic year, including a self-paced walking program, weight management classes, and a tobacco cessation program.

Employees assigned to the non-wellness group, referred to as Wave 2 members, were invited to complete health screenings in 2017 and 2018 so that researchers could compare their biometric health outcomes with those from the Wave 1 group (Reif et al., 2020). Unlike Wave 1 members, they were not eligible to participate in the initial (2016) health screening and were never invited to complete an online HRA or sign up for wellness activities.

Figure 1 presents the study's experimental design. Financial incentives were assigned randomly several weeks before each screening sign-up period. The annual rerandomization ensured that incentives were independent of previous incentives or employee outcomes, including past program participation. In 2016, Wave 1 employees were equally likely to receive

an incentive of \$0, \$100, or \$200.[4] In 2017 and 2018, Wave 1 and Wave 2 employees were assigned with equal probability to receive either \$0 or \$125 (in 2017) and either \$0 or \$75 (in 2018) upon successfully completing a screening. The incentives in each year were assigned at the individual level using stratified random sampling, as detailed in Jones, Molitor and Reif (2019).

The study's structure creates five distinct experiments, one per wave and year of randomization (Figure 1). All five experiments allow us to measure the "direct" effect of the incentive on contemporaneous screening rates. Three of them also allow us to estimate the persistence effects of monetary incentives on subsequent screening rates: (1) the effects of the 2016 incentives on Wave 1 screenings in 2017–2018; (2) the effect of the 2017 incentive on Wave 1 screenings in 2018; and (3) the effect of the 2017 incentive on Wave 2 screenings in 2018.

It is uncertain whether financial incentives to complete a screening will increase or decrease future screening completion. If individuals prefer to only undergo screenings every few years—perhaps because they feel that more frequent screenings do not generate information of sufficient value—then an incentive that boosts current screening rates might reduce future rates due to intertemporal substitution. On the other hand, screenings could exhibit intertemporal complementarity, resulting in a positive persistence effect. For instance, regular screenings allow individuals to monitor changes in biometrics over time. Completing an initial screening could also increase future screenings by sparking an individual's interest in health tracking or by making the logistics of scheduling and completing screenings more familiar and manageable.

3 Empirical Strategy

Our empirical strategy encompasses two objectives. The first is to estimate the causal effects of financial incentives on health screening completion. Using reduced-form analysis, we define the "direct effect" of incentives as their impact on same-year screening completion

[4] The 2016 incentives were paid only to employees who completed both a health screening and an online HRA. Of 1,900 participants who completed a health screening, 1,848 also completed an online HRA. In 2017, HRA completion was not required to receive the assigned incentive. In 2018, the HRA was not offered.

and the "persistence effect" as their impact on future completion. Our second objective is to estimate habit formation, for which we use an IV approach that assumes the persistence effect operates through past completion. We perform our analyses separately for each wave and screening year, which results in five sets of estimates (Figure 1).[5]

3.1 Persistence

We estimate the effects of past, current, and future financial incentives on screening completion using the following regression model:

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=\text{BASEYR}}^{2018} \beta_\tau^t \text{INCENTIVE}_i^\tau + \gamma X_i + \epsilon_i. \tag{1}$$

The outcome variable, SCREEN_i^t, equals 1 if individual i completed a health screening in year t, and 0 otherwise. The focal explanatory variable, INCENTIVE_i^τ, is an indicator equal to 1 if the individual was randomly assigned a non-zero screening incentive in year τ. The range for τ extends from BASEYR (defined as 2016 for Wave 1 and 2017 for Wave 2) to 2018.

The focal parameter, β_τ^t, represents the average treatment effect (ATE) of monetary incentives assigned in year τ on screening completion rates in year t. An individual receives the assigned incentive in year τ only upon completing a health screening that year. Thus, Equation (1) identifies the direct effect of incentives on screening completion for $t = \tau$ and a persistence effect for $t > \tau$. Given the annual rerandomization of incentives, disclosed shortly before screenings began, we expect $\beta_\tau^t = 0$ for $t < \tau$, providing a falsification test of our model.

Equation (1) includes multiple treatment variables. If the effect of an incentive varies with the receipt of past incentives, β_τ^t will capture a weighted average of ATEs under the various counterfactual scenarios created by the combinations of incentives offered in previous years (Muralidharan, Romero and Wüthrich, 2023). Only incentives offered up to year t are relevant for this average, as those offered after year t were not yet known and could not

[5] Our preregistered analysis plan specified estimating the effects of incentives on screening completion. We did not prespecify the specific models or mechanisms for habit formation examined in this paper.

affect screening decisions in year t. This weighted average is policy relevant as it is common for real-world wellness programs (and other interventions) to offer incentives on an annual basis. In supplemental analyses, we report a version of Equation (1) in which we include interaction terms.

In our baseline analysis, Equation (1) does not include additional control variables, X_i. Because incentives were randomly assigned, controls are not necessary to remove bias in the focal estimate but may increase precision. In supplemental analyses, we adopt a "post-Lasso" control specification, selecting controls via the Lasso double-selection method of Belloni, Chernozhukov and Hansen (2014). The set of potential control variables includes baseline demographics, health survey responses, health behaviors, and claims-based measures of medical spending and usage, along with all their pairwise interactions.[6] Since randomization was performed at the individual level and our strata are large, we report conventional heteroskedastic-robust standard errors (Abadie et al., 2023; de Chaisemartin and Ramirez-Cuellar, 2024).

Our experimental framework relies on the assumption that incentives are randomly assigned to participants. To validate this assumption, we test whether other screening incentives, baseline demographics, and survey variables jointly predict incentive amounts. Table 1 reports the averages of these variables (one per row) across the different treatment arms (one per column). We conduct joint balance tests for each of the five experiments in our study, represented by pairs of adjacent columns (columns (1)–(2), (3)–(4), etc.). The p-values from these tests, all 0.29 or greater, indicate that these variables collectively do not predict assigned incentives and support the null hypothesis of randomly assigned incentives in each experiment.

[6]For missing values, we impute means/modes and generate a variable that indicates missing values. The missing-value indicators are also included in the set of potential controls. Health behavior measures include participation in an annual running event and usage of campus recreational facilities. See Jones, Molitor and Reif (2019) for a detailed description of these variables.

3.2 Habit Formation

We define habit formation as the causal effect of past screening completion on current screening completion. We estimate habit formation using the following regression model:

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=\text{BASEYR}}^{t-1} \theta_\tau^t \text{SCREEN}_i^\tau + \gamma X_i + \epsilon_i. \tag{2}$$

The focal explanatory variables, SCREEN_i^τ, are indicators equal to 1 if individual i completed a screening in year τ, where $\tau < t$. The term X_i represents a vector of individual-specific control variables. Our baseline analysis includes the contemporary financial incentive, INCENTIVE_i^t, as a control variable because it is a strong predictor of SCREEN_i^t. In supplemental analyses, we report a specification that includes post-Lasso controls.

Ordinary least squares (OLS) estimation of Equation (2) will produce biased estimates if past screening completion is correlated with unobserved determinants of current screening completion. This bias is likely to be positive, as many factors that increase screening propensity in one period, such as higher health consciousness or a more proactive attitude toward preventive care, are likely to persist and increase the probability of future screening completion. However, the bias could be negative if people typically wait one or more years between screenings.

To address the endogeneity issue between current and past screening behavior, we perform an IV estimation of Equation (2), using the randomly assigned monetary incentives from prior years as instruments for past screening completion. When the outcome is the second screening for a given wave, there is a single endogenous regressor (initial screening completion) and one instrument (the initial incentive). When the outcome is the third screening, which happens only for Wave 1 in 2018, the model has two endogenous regressors, and we instrument for both using their respective screening incentives as instruments. We perform IV estimation via two-stage least squares (2SLS) and report first-stage F-statistics using the method of Sanderson and Windmeijer (2016), which allows us to test for weak identification in each endogenous regressor separately.

The IV analysis relies on the exclusion restriction that monetary incentives affect future screening completion solely through their influence on concurrent screening completion. We

address potential violations of this restriction in Section 4.2. In models where there is only one endogenous regressor, we interpret θ_τ^t as the average causal effect of screening completion in year τ on completion in year t among compliers, i.e., those induced to complete a health screening in year τ by a financial incentive. This interpretation requires the standard monotonicity assumptions for local ATE (LATE) interpretations of 2SLS. For the Wave 1 group, we have two endogenous regressors when the outcome variable is screening completion in 2018. In this case, our estimates can still be interpreted as capturing ATEs provided that we assume treatment effect homogeneity. Allowing for treatment effect heterogeneity requires imposing a monotonicity assumption in a context with two endogenous variables and non-mutually exclusive treatments, for which there is no widely accepted standard. For thoroughness, our appendix presents estimates of the coefficients on the two endogenous regressors obtained using two separate regressions, which preserves a standard LATE interpretation.

4 Results

4.1 Persistence

Figure 2 reports annual health screening completion rates by incentive level and experiment wave. The direct effect of an incentive can be assessed by comparing the completion rates of groups that received high and low incentives in the year the incentive was initially introduced. The persistence effects of incentives given before 2018 (shown in Panels A, B, and D) can be gauged by comparing completion rates in these groups in subsequent years. For the second and third incentives (in Panels B, C, and E), completion rates before the incentives were assigned provide a falsification test for whether incentives were successfully rerandomized each year.

Panel A of Figure 2 shows that for Wave 1 subjects, the initial (2016) financial incentive increased completion of that year's health screening from 49.4% to 61.8%. The panel also indicates that this initial incentive led to positive persistence effects, as those who were offered it were about 4.5 and 4.4 percentage points more likely to complete a health screening in

2017 and 2018, respectively, than those who were not. Error bars confirm the statistical significance of these effects ($p < 0.05$). Panel D shows that the initial incentive offered to individuals in Wave 2 (in 2017) produced similar results, affirming the replicability of a persistent impact of an initial incentive. In contrast, Panel B shows that the second incentive offered to Wave 1 subjects had a substantial direct effect but resulted in little to no persistence effect. Finally, the falsification tests in Panels B, C, and E all reveal similar screening rates between high- and low-incentive groups in years before the incentive assignment, consistent with incentives being rerandomized annually.

Table 2 displays regression estimates from Equation (1), analyzing the impact of financial incentives on current, future, and past screening completion rates. Each column is a separate regression, where the outcome is screening completion for a given wave and year. Rows report the effects of the incentives offered in each of the three years of the program. The direct effects on current completion, highlighted in gray, are shown along the diagonals in columns (1)–(3) for Wave 1 and columns (4)–(5) for Wave 2. The persistence effects on future completion are in bold, while the effects of incentives on past completion (in plain text) serve as falsification tests for the randomization process. The post-Lasso estimates are reported in Table A.1.

The values in the first row of Table 2 corroborate the patterns shown in Panel A of Figure 2: the 2016 incentive increased screening completion rates in that year by 12.4 percentage points and increased future completion rates in 2017 and 2018 by 4.5 and 4.4 percentage points, respectively. Across the five distinct experiments in the study, the direct effects ranged from 12.4 percentage points (an increase of 22%) to 27.6 percentage points (an increase of 71%), with all the effects being statistically significant ($p < 0.01$). First-year incentives increased future screening completion in both waves by 4.4 to 8.9 percentage points ($p < 0.01$), revealing strong persistence effects. By contrast, the second incentive offered to Wave 1 had a small and statistically insignificant persistence effect of 1.2 percentage points (column (3), row 2). Finally, all falsification estimates, reported below the direct effect estimates in columns (1), (2), and (4), are small and statistically insignificant, as expected.

Table A.2 presents estimates from a generalized model incorporating interactions between

the incentives assigned in different years.[7] We focus on the 2018 screening outcome, for which the number of previous incentive rounds is greatest. Column (1) of Table A.2 reproduces the baseline estimates from the third column of Table 2, while columns (2)–(5) report different combinations of interactions. Although including these interactions reduces statistical power, the estimates of the main effects remain largely unchanged. We do not detect any significant interaction effects, though we note that our lack of statistical power means we cannot rule out the possibility of meaningful interaction effects.

While all three persistence effects of the first incentives offered to the Wave 1 and Wave 2 groups are large and statistically significant, the effect of the second (2017) incentive on 2018 screening completion in the Wave 1 group is small and insignificant (see Panel B of Figure 2 and column (3), row 2 of Table 2). We consider three possible reasons for this discrepancy. First, there may have been something unique about the 2017 screening that influenced our results. For instance, perhaps people had unusually poor screening experiences, resulting in a diminished persistence effect. If that were the case, we would expect Wave 2 subjects, who received the same screening treatment in 2017 as Wave 1 subjects, to exhibit similarly weak persistence effects. However, the 2017 incentive produced a positive persistence effect on 2018 screening completion for Wave 2 subjects (see Panel D of Figure 2 and column (5), row 2 of Table 2).

Another possibility is that the 2017 incentive's direct effect on that year's screening completion may have been particularly weak for Wave 1 subjects, in which case we would expect the absolute magnitude of persistence to also be small. Comparing the direct effects in columns (2) and (4) of Table 2 provides some support for this hypothesis, though the differences in these effect sizes are small. In addition, the direct effect of the 2017 incentive was larger than the direct effect of the 2016 incentive, which did result in persistence effects. These patterns suggest that a lack of persistence of the second incentive among Wave 1 was not a consequence of weak direct effects.

A third possibility is that only the 2016 incentive matters for persistence within the window we examine. Unlike Wave 2 subjects, Wave 1 subjects were allowed to participate in the 2016 health screenings. Thus, the 2017 incentive represented a second "dose" for Wave 1

[7] The post-Lasso estimates, reported in Table A.3, are similar.

but was the first dose for Wave 2. We consider this possibility in further detail in Section 4.2.

4.2 Habit Formation

4.2.1 Main Estimates

Table 3 presents habit formation estimates from Equation (2). Columns (1)–(3) report OLS estimates of the effect of past screening completion on current completion, showing that all four estimates are positive and statistically significant. However, these estimates are prone to upward bias because an individual's decision to complete a health screening is positively correlated over time, and thus should not be interpreted causally.

Columns (4)–(6) address this bias by instrumenting for past screening completion with past incentives. Results in the first row show that completing the initial health screening increased the likelihood of completing a screening the next year by 36.0 (84%) and 32.4 (89%) percentage points in both waves, and raised screening completion two years later by 33.1 percentage points (90%) in Wave 1. These three estimates are statistically significant ($p < 0.01$) and similar in both absolute and relative terms. In contrast, the second row of column (5) indicates that completing the second screening raised screening rates one year later by a small and statistically insignificant 6.3 percentage points (17%). Post-Lasso estimates in Table A.4 provide similar results. Additionally, instrumenting for the two endogenous regressors separately using distinct regressions yields estimates similar to column (5), as shown in columns (5) and (6) of Table A.5.

These results reinforce our initial findings from Section 4.1: habit formation produced by the initial health screening was strong in both waves, showed minimal decay after two years, and was much greater than the habit formation produced by the second screening. Formally, we fail to reject the null hypothesis that the three habit formation estimates from the first screenings reported in the first row of columns (4)–(6) of Table 3 are equal ($p > 0.8$ with and without post-Lasso controls). For the third screening offered to Wave 1 (column (5)), we reject equality of the first and second screening effects ($p < 0.1$). We also jointly reject the equality of the three first-screening effects and the second-screening effect ($p = 0.06$ without controls, $p = 0.03$ with post-Lasso controls).

Our exclusion restriction assumes that the effect of financial incentives on future screening completion operates through an increase in prior screening completion. Because we rerandomized incentives each year, correlation in assigned incentive amounts over time is not a threat to validity. However, it is possible that subjects may not have fully understood or paid attention to how their financial incentives changed over time. For example, if subjects assigned to the high-incentive group in 2016 and then the low-incentive group in 2017 mistakenly thought they would still receive a high incentive in 2017, our exclusion restriction would not hold. To investigate that possibility, we estimate the effect of completing the 2016 screening on 2018 completion for the subsample of subjects who were assigned to the $0 group in 2017. If inattention drove our estimates, we would expect a smaller treatment effect estimate for this subsample: confused subjects who attended the 2017 screening expecting a high incentive would have learned they were mistaken, thereby reducing their turnout for the 2018 screening.[8]

We report the results of this investigation in Table A.6. Column (1) shows that the 2016 incentive raised the 2018 completion rate by 4.4 percentage points in the full sample. When we limit the regression to those individuals who were assigned the $0 incentive in 2017, the point estimate rises to 5.4 percentage points, indicating that confusion regarding payment is not driving our main estimates. Columns (3) and (4) report the corresponding IV estimates. Here, the point estimate from the subsample is slightly smaller than the estimate from the full sample, but the difference is not statistically significant. We obtain similar results if we include post-Lasso controls (Table A.7). Thus, we conclude that inattention does not explain our estimates.

Another potential threat to the exclusion restriction is that offering a financial incentive might cause individuals to seek out information about biometric screenings, increasing the likelihood of future screenings regardless of whether they had completed the initial incentivized screening. Although we consider this possibility unlikely, we lack evidence to rule it out definitively. Nevertheless, even if the exclusion restriction fails, the persistence estimates presented in Table 2 remain valid.

[8]The 2017 screening completion rate was 33.8% for the 1,091 people who were assigned to the high-incentive group in 2016 and the low-incentive group in 2017.

4.2.2 Channels

Habit formation can arise in various ways. One of the most well-studied channels is addiction, where past consumption reinforces current consumption through intertemporal complementarity (Becker and Murphy, 1988). In this model, current consumption is more influenced by recent past consumption than by distant past consumption due to depreciation.[9]

Alternative channels, such as learning or taste discovery, do not require reinforcement to produce habit formation. For example, consider consumers who avoid purchasing a good because they are unsure whether they will like it. If they are exogenously induced to consume the good by being offered a free sample and discover they like it, they may begin regularly purchasing it, thus creating a new habit. In this case, there are no intertemporal links in consumption once the initial consumption hurdle has been overcome. Further exogenous changes to consumption will have no long-run effects because they do not alter the consumers' knowledge.[10]

We investigate these potential channels using estimates from Wave 1, which was invited to a health screening for three consecutive years. Formally, we expect screening completion in the third year to be more strongly related to second-year completion rates than to first-year completion rates if habit formation operates through an addiction channel, i.e., that $\theta^{2018}_{2017} > \theta^{2018}_{2016}$ in Equation (2). We expect the opposite if habit formation operates by learning from consumption. Our estimates, presented in column (5) of Table 3, reject the addiction hypothesis and support alternatives such as learning. These results are consistent with subjects incurring a one-time cost of learning how to navigate the screening process or discovering whether they enjoy health screenings.

[9]This result holds whether or not individuals are forward-looking (Reif, 2019). Laibson (2001) generalizes the Becker-Murphy model by incorporating context-dependent cues. However, due to the lack of experimental variation in context within our study, we cannot provide relevant evidence for or against this cues model.

[10]The notion of learning about one's preferences by consuming an "experience good" was first emphasized by Nelson (1970). Ackerberg (2001) argues that the learning effects of yogurt advertising can be identified by comparing its effects on purchases for first-time versus repeat customers. Osborne (2011) notes that learning can be empirically distinguished from addiction ("switching costs") by comparing the effects of a first versus previous purchase on current consumption.

4.3 Discussion

Overlooking habit formation significantly underestimates the effectiveness of financial incentives. For example, consider the initial (2016) incentive offered to Wave 1. Without accounting for habit formation, one would conclude that this incentive increased completion rates by 12.4 percentage points (or 406 people) for one year, as shown in Table 2. However, the total effect over the two-year program period is an increase of 21.3 percentage points (or 698 people), which is 72% higher.[11] These figures only reflect the effect of habit formation over the two years of our intervention. Since we observed no significant decay in the effect during this period, it is reasonable to believe that the full effect would have been even greater had the program extended beyond two years.

While habit formation substantially improved the effectiveness of our health screening intervention, this effect appears to arise only with the initial exposure. The second dose of incentives had no detectable effect on future screenings, suggesting that while initial incentives may be more effective than previously thought, the benefits of subsequent doses are unlikely to be underestimated by ignoring habit formation.

5 Conclusion

We provide evidence of the long-term effects of financial incentives on health screenings. Financial rewards not only boost immediate participation but also increase the likelihood of future participation, with effects lasting at least two years. This prolonged effect operates through the initial screening—a second screening did not produce significant persistence. These findings have practical implications for health policy. In our setting, the results suggest that initial monetary incentives can lead to lasting changes in health-related behaviors.

A strength of our setting is that multiple experiments and rerandomization allow us to replicate our contemporaneous effects five times and our one-year persistence results twice. However, as we only have one wave lasting more than two years, we can estimate our longer-term persistence and test our habit formation models only once. Future work could extend

[11] Equivalently, each person induced by incentives to complete a screening in the first year is 36% more likely to complete a screening in the second year and 33% more likely in the third year (Table 3).

our analysis by examining persistence effects over a longer horizon, investigating whether the persistence of an initial incentive eventually decays. Finally, future work might combine evidence such as ours with benefit estimates to inform the optimal size, sequencing, and allocation of incentives over time under a fixed budget. Our evidence supports the front-loading of incentives, but this effect must be traded off against their contemporaneous effects over time.

While our study focuses on individual-level behaviors, we also acknowledge that health outcomes are influenced by system-level forces, such as health care markets, regulations, and social norms, as well as individual-level decisions, such as diet and exercise. Our findings do not address whether interventions at the system or individual level are more effective, or whether they serve as complements or substitutes. These questions are beyond the scope of this study, but would be interesting avenues for future work.[12]

[12]See Chater and Loewenstein (2023) for a broader discussion.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2023. "When should you adjust standard errors for clustering?" *Quarterly Journal of Economics*, 138(1): 1–35.

Ackerberg, Daniel A. 2001. "Empirically distinguishing informative and prestige effects of advertising." *RAND Journal of Economics*, 316–333.

Acland, Dan, and Matthew R. Levy. 2015. "Naiveté, projection bias, and habit formation in gym attendance." *Management Science*, 61(1): 146–160.

Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel Zucker. 2022. "Designing incentives for impatient people: An RCT promoting exercise to manage diabetes." Working paper.

Allcott, Hunt, and Todd Rogers. 2014. "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation." *American Economic Review*, 104(10): 3003–3037.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2022. "Digital addiction." *American Economic Review*, 112(7): 2424–2463.

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does diversity matter for health? Experimental evidence from Oakland." *American Economic Review*, 109(12): 4071–4111.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral hazard in health insurance." *Quarterly Journal of Economics*, 130(4): 1623–1667.

Becker, Gary S., and Kevin M. Murphy. 1988. "A theory of rational addiction." *Journal of Political Economy*, 96(4): 675–700.

Becker, Gary S., Michael Grossman, and Kevin M Murphy. 1991. "Rational addiction and the effect of price on consumption." *American Economic Review*, 81(2): 237–241.

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

Bernheim, B. Douglas, Stefano DellaVigna, and David Laibson, ed. 2019. *Handbook of Behavioral Economics - Foundations and Applications 2.* Elsevier.

Bursztyn, Leonardo, Davide Cantoni, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2021. "Persistent political engagement: Social interactions and the dynamics of protest movements." *American Economic Review: Insights*, 3(2): 233–250.

Campos-Mercade, Pol, Armando N. Meier, Florian H. Schneider, Stephan Meier, Devin Pope, and Erik Wengström. 2021. "Monetary incentives increase COVID-19 vaccinations." *Science*, 374(6569): 879–882.

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor. 2018. "Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members." *Journal of Health Economics*, 58: 202–214.

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor. 2020. "The structure of health incentives: Evidence from a field experiment." *Management Science*, 66(5): 1890–1908.

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

Chater, Nick, and George Loewenstein. 2023. "The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray." *Behavioral and Brain Sciences*, 46: e147.

de Chaisemartin, Clément, and Jaime Ramirez-Cuellar. 2024. "At what level should one cluster standard errors in paired and small-strata experiments?" *American Economic Journal: Applied Economics*, 16(1): 193–212.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit formation in voting: Evidence from rainy elections." *American Economic Journal: Applied Economics*, 8(4): 160–188.

Giuntella, Osea, Silvia Saccardo, and Sally Sadoff. 2024. "Sleep: Educational impact and habit formation." NBER Working Paper No. 32550.

Gruber, Jonathan, and Botond Köszegi. 2001. "Is addiction 'rational'? Theory and evidence." *Quarterly Journal of Economics*, 116(4): 1261–1303.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2022. "Rational habit formation: Experimental evidence from handwashing in India." *American Economic Journal: Applied Economics*, 14(1): 1–41.

Jones, Damon, David Molitor, and Julian Reif. 2019. "What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study." *Quarterly Journal of Economics*, 134(4): 1747–1791.

Laibson, David. 2001. "A cue-theory of consumption." *Quarterly Journal of Economics*, 116(1): 81–119.

Loewenstein, George, Joseph Price, and Kevin Volpp. 2016. "Habit formation in children: Evidence from incentives for healthy eating." *Journal of Health Economics*, 45: 47–54.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. 2023. "Factorial designs, model selection, and (incorrect) inference in randomized experiments." *Review of Economics and Statistics*, 1–44.

Nelson, Phillip. 1970. "Information and consumer behavior." *Journal of Political Economy*, 78(2): 311–329.

Osborne, Matthew. 2011. "Consumer learning, switching costs, and heterogeneity: A structural examination." *Quantitative Marketing and Economics*, 9: 25–70.

Patel, Mitesh S., David A. Asch, Roy Rosin, Dylan S. Small, Scarlett L. Bellamy, Jack Heuer, Susan Sproat, Chris Hyson, Nancy Haff, Samantha M. Lee, et al. 2016. "Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial." *Annals of Internal Medicine*, 164(6): 385–394.

Pryor, Katherine, and Kevin Volpp. 2018. "Deployment of preventive interventions: Time for a paradigm shift." *New England Journal of Medicine*, 378(19): 1761–1763.

Reif, Julian. 2019. "A model of addiction and social interactions." *Economic Inquiry*, 57(2): 759–773.

Reif, Julian, David Chan, Damon Jones, Laura Payne, and David Molitor. 2020.

"Effects of a workplace wellness program on employee health, health beliefs, and medical use: A randomized clinical trial." *JAMA Internal Medicine*, 180(7): 952–960.

Rohde, Kirsten I.M., and Willem Verbeke. 2017. "We like to see you in the gym: A field experiment on financial incentives for short and long term gym attendance." *Journal of Economic Behavior and Organization*, 134: 388–407.

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: 51–84.

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." *Journal of Econometrics*, 190(2): 212–221.

Schneider, Florian H., Pol Campos-Mercade, Stephan Meier, Devin Pope, Erik Wengström, and Armando N. Meier. 2023. "Financial incentives for vaccination do not have negative unintended consequences." *Nature*, 613(7944): 526–533.

Simon, Kosali, Aparna Soni, and John Cawley. 2017. "The impact of health insurance on preventive care and health behaviors: evidence from the first two years of the ACA Medicaid expansions." *Journal of Policy Analysis and Management*, 36(2): 390–417.

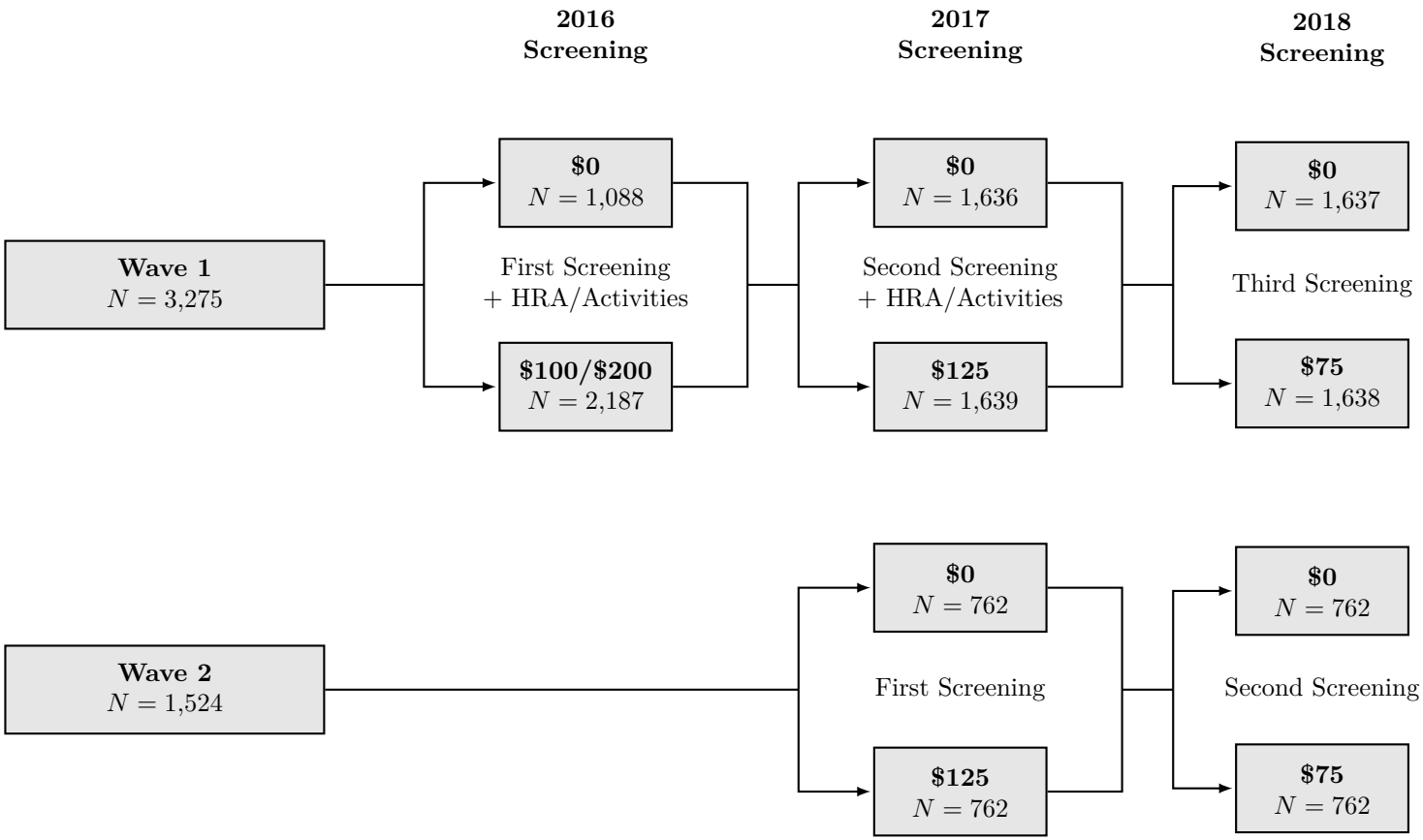
Song, Zirui, and Katherine Baicker. 2019. "Effect of a workplace wellness program on employee health and economic outcomes: A randomized clinical trial." *JAMA*, 321: 1491–1501.

Stone, Erin G., Sally C. Morton, Marlies E. Hulscher, Margaret A. Maglione, Elizabeth A. Roth, Jeremy M. Grimshaw, Brian S. Mittman, Lisa V. Rubenstein, Laurence Z. Rubenstein, and Paul G. Shekelle. 2002. "Interventions that increase use of adult immunization and cancer screening services: A meta-analysis." *Annals of Internal Medicine*, 136(9): 641–651.

Vollaard, Ben, and Daan van Soest. 2024. "Punishment to promote prosocial behavior: A field experiment." *Journal of Environmental Economics and Management*, 124: 102899.

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: 2631–2637.

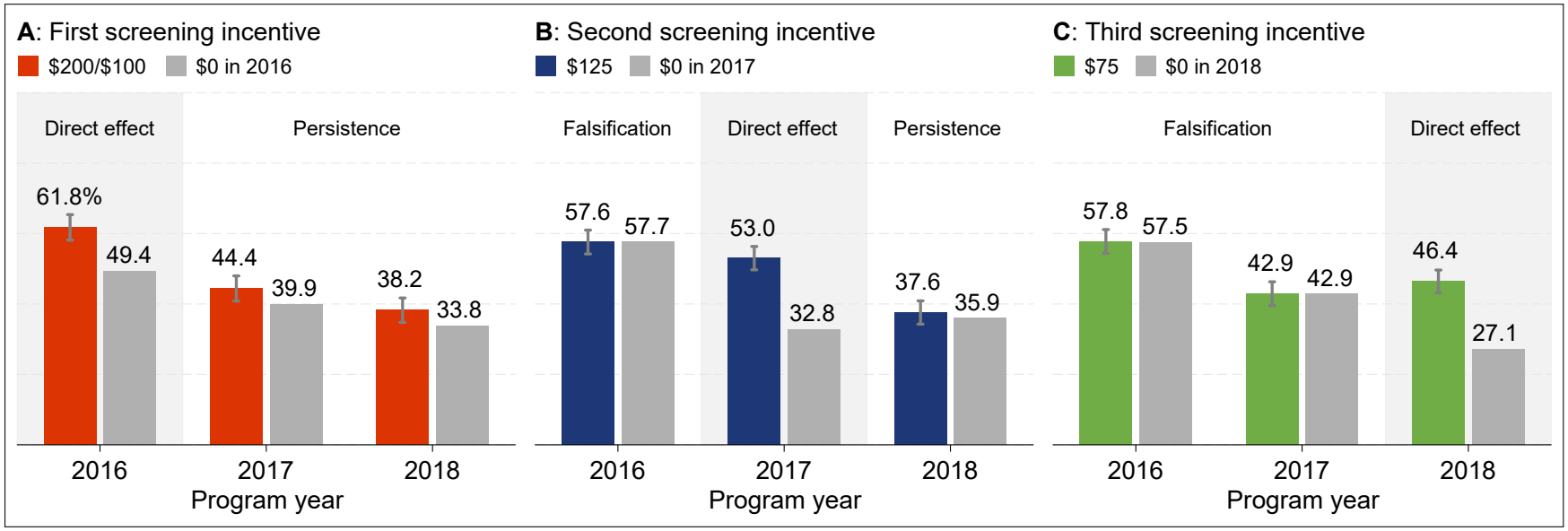
Figure 1: Experimental Design of the Illinois Workplace Wellness Study



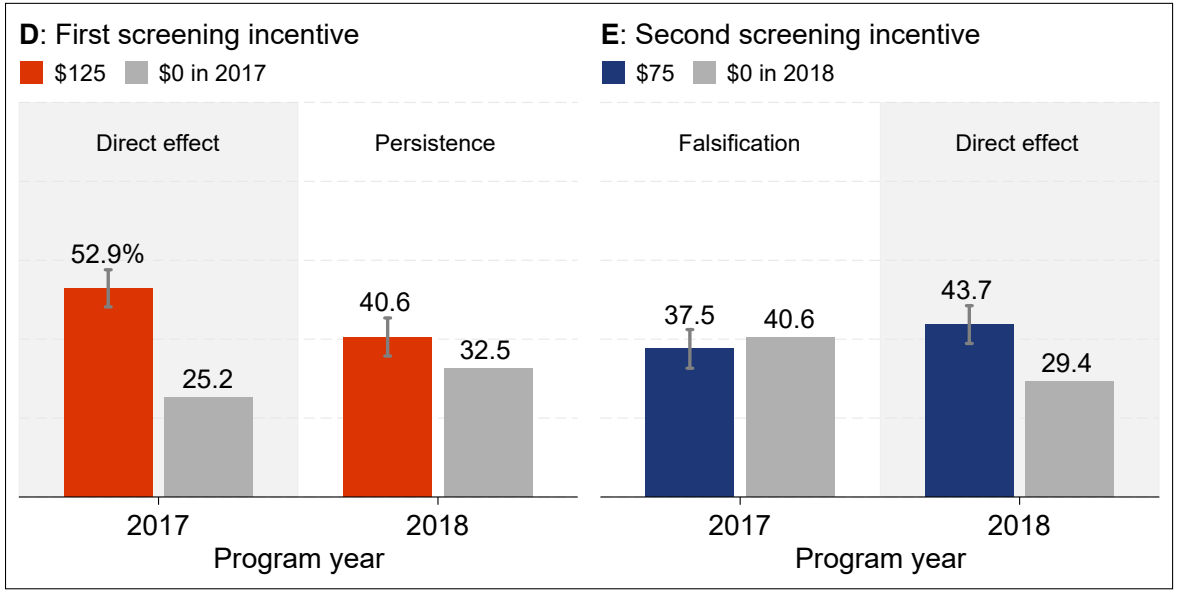
Notes: The figure depicts treatment assignments over time for subjects continuously enrolled through the end of the intervention. In 2016, the 4,799 subjects were randomly assigned to either Wave 1 or Wave 2. Wave 1 was invited to complete a biometric health screening in 2016, 2017, and 2018, while Wave 2 was invited only in 2017 and 2018. Subjects were randomly assigned to either a control ($0 incentive) or treatment (> $0 incentive) group a few weeks before each screening. Incentive assignments are uncorrelated across years.

Figure 2: Health Screening Completion Rates, by Incentive Groups

Wave 1 screening completion



Wave 2 screening completion



Notes: The figure panels report raw screening completion rates, by year and incentive level. Each panel reflects a specific wave and year of the incentive. Error bars show 95% confidence intervals of differences in screening rates between high- and low-incentive groups. Comparisons made one to two years after the assignment of incentives are labeled "persistence," while those made one to two years before the assignment are labeled "falsification."

Table 1: Balance Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Wave 1 screening incentives						Wave 2 screening incentives			
	First (2016)		Second (2017)		Third (2018)		First (2017)		Second (2018)	
	\$200/\$100	\$0	\$125	\$0	\$75	\$0	\$125	\$0	\$75	\$0
A. Screening Incentive Variables										
First-year screening incentive	1.00	0.00	0.67	0.67	0.67	0.67	1.00	0.00	0.47	0.53
Second-year screening incentive	0.50	0.50	1.00	0.00	0.51	0.49	0.47	0.53	1.00	0.00
Third-year screening incentive	0.50	0.50	0.51	0.49	1.00	0.00				
B. Stratification Variables										
Male [admin]	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.42	0.43	0.42
Age 50+ [admin]	0.32	0.33	0.33	0.32	0.32	0.33	0.32	0.32	0.32	0.32
Age 37–49 [admin]	0.33	0.33	0.33	0.33	0.33	0.33	0.34	0.34	0.34	0.34
White [admin]	0.83	0.84	0.84	0.83	0.83	0.84	0.84	0.84	0.86	0.82
Salary Q1 (bottom quartile) [admin]	0.24	0.24	0.24	0.24	0.25	0.24	0.24	0.25	0.25	0.24
Salary Q2 [admin]	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.25	0.26
Salary Q3 [admin]	0.25	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.24	0.26
Faculty [admin]	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.19	0.20	0.20
Academic staff [admin]	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
C. 2016 Health Survey Variables										
Ever screened [survey]	0.89	0.90	0.88	0.90	0.89	0.89	0.88	0.89	0.88	0.89
Physically active [survey]	0.39	0.37	0.38	0.39	0.38	0.39	0.35	0.37	0.36	0.36
Trying to be active [survey]	0.82	0.80	0.81	0.81	0.81	0.81	0.83	0.81	0.81	0.84
Current smoker (cigarettes) [survey]	0.07	0.06	0.07	0.06	0.06	0.07	0.06	0.09	0.07	0.07
Current smoker (other) [survey]	0.09	0.07	0.09	0.08	0.08	0.09	0.07	0.10	0.09	0.08
Former smoker [survey]	0.19	0.20	0.18	0.21	0.18	0.21	0.20	0.19	0.19	0.20
Drinker [survey]	0.64	0.65	0.64	0.65	0.66	0.64	0.64	0.67	0.65	0.66
Heavy drinker [survey]	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.06
Chronic condition [survey]	0.72	0.74	0.73	0.72	0.72	0.73	0.74	0.72	0.73	0.73
Excellent or v. good health [survey]	0.59	0.61	0.60	0.61	0.61	0.59	0.59	0.59	0.57	0.60
Not poor health [survey]	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Physical problems [survey]	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.38	0.40
Lots of energy [survey]	0.33	0.33	0.33	0.33	0.35	0.30	0.31	0.31	0.29	0.33
Bad emotional health [survey]	0.29	0.29	0.29	0.28	0.28	0.30	0.30	0.32	0.31	0.30
Overweight [survey]	0.52	0.56	0.54	0.53	0.53	0.53	0.55	0.54	0.52	0.56
High BP/cholesterol/glucose [survey]	0.29	0.31	0.30	0.29	0.29	0.30	0.32	0.29	0.31	0.30
Sedentary [survey]	0.55	0.53	0.55	0.54	0.56	0.53	0.56	0.53	0.53	0.56
Sample size	2,187	1,088	1,639	1,636	1,638	1,637	762	762	762	762
Joint balance test (p-value)		0.61		0.98		0.56		0.81		0.29

Notes: The table reports group means. The joint balance test row reports the p-value from testing whether assignment to a positive screening incentive in the specified year (column label) is predicted by the row variables, excluding the incentive variable itself.

Table 2: Effect of Financial Incentives on Health Screening Completion

	(1)	(2)	(3)	(4)	(5)
	Wave 1 screening completion			Wave 2 screening completion	
	2016	2017	2018	2017	2018
2016 incentive ($200/$100)	0.124**	**0.045***	**0.044***		
	(0.018)	**(0.018)**	**(0.017)**		
2017 incentive ($125)	–0.001	0.203**	**0.012**	0.276**	**0.089****
	(0.017)	(0.017)	**(0.017)**	(0.024)	**(0.024)**
2018 incentive ($75)	0.002	–0.004	0.192**	–0.013	0.149**
	(0.017)	(0.017)	(0.017)	(0.024)	(0.024)
Mean outcome	0.576	0.429	0.368	0.390	0.365
Sample size	3,275	3,275	3,275	1,524	1,524

Notes: This table reports estimates of β_τ^t from Equation (1). The dependent variable is an indicator for whether a screening was completed in the year specified in the column header. Gray-shaded estimates indicate the contemporaneous effect of the incentive on screening completion. Bold-faced estimates indicate the persistence effect of the incentive on future screening completion, and plain-text estimates report a falsification test of an unannounced incentive on past screening completion. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table 3: Effect of Past Screening Completion on Current Screening Completion

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS estimates			IV estimates		
	Wave 1 (2016–2018)		Wave 2 (2017–2018)	Wave 1 (2016–2018)		Wave 2 (2017–2018)
	Second screening	Third screening	Second screening	Second screening	Third screening	Second screening
Completed first screening	0.467**	0.183**	0.552**	0.360**	0.331**	0.324**
	(0.014)	(0.016)	(0.022)	(0.128)	(0.127)	(0.075)
Completed second screening		0.430**			0.063	
		(0.017)			(0.073)	
First-stage F (first screening)				45.6	48.1	131.6
First-stage F (second screening)					182.5	
Mean outcome	0.429	0.368	0.365	0.429	0.368	0.365
Sample size	3,275	3,275	1,524	3,275	3,275	1,524

Notes: This table reports OLS and IV estimates of θ_τ^t from Equation (2). Dependent variable is an indicator for whether the second (or third) screening was completed. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Online Appendix

"Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study"

Damon Jones, David Molitor, and Julian Reif

Table A.1: Effect of Financial Incentives on Health Screening Completion, Post-Lasso Controls

	(1)	(2)	(3)	(4)	(5)
	Wave 1 screening completion			Wave 2 screening completion	
	2016	2017	2018	2017	2018
2016 incentive ($200/$100)	0.133**	**0.046****	**0.045****		
	(0.018)	**(0.017)**	**(0.017)**		
2017 incentive ($125)	–0.003	0.200**	**0.014**	0.276**	**0.090****
	(0.016)	(0.016)	**(0.016)**	(0.023)	**(0.024)**
2018 incentive ($75)	–0.005	–0.007	0.188**	–0.002	0.154**
	(0.016)	(0.016)	(0.016)	(0.024)	(0.024)
Number of controls	53	40	42	40	42
Mean outcome	0.576	0.429	0.368	0.390	0.365
Sample size	3,275	3,275	3,275	1,524	1,524

Notes: This table reports estimates of β_τ^t from Equation (1). The dependent variable is an indicator for whether a screening was completed in the year specified in the column header. Each regression controls for a set of covariates selected by Lasso to predict the dependent variable in the full sample. "Number of controls" reports the number of selected covariates. Gray-shaded estimates indicate the contemporaneous effect of the incentive on screening completion. Bold-faced estimates indicate the persistence effect of the incentive on future screening completion, and plain-text estimates report a falsification test of an unannounced incentive on past screening completion. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.2: Effect of Financial Incentives on 2018 Health Screening Completion for Wave 1

	(1)	(2)	(3)	(4)	(5)
2016 incentive ($200/$100)	0.044*	0.054*	0.046*	0.044*	0.032
	(0.017)	(0.024)	(0.023)	(0.017)	(0.032)
2017 incentive ($125)	0.012	0.025	0.012	0.036	0.016
	(0.017)	(0.028)	(0.017)	(0.022)	(0.037)
2018 incentive ($75)	0.192**	0.193**	0.195**	0.216**	0.186**
	(0.017)	(0.017)	(0.028)	(0.023)	(0.039)
2016 incentive × 2017 incentive		–0.019			0.029
		(0.035)			(0.046)
2016 incentive × 2018 incentive			–0.003		0.045
			(0.035)		(0.049)
2017 incentive × 2018 incentive				–0.047	0.018
				(0.033)	(0.056)
2016 incentive × 2017 incentive × 2018 incentive					–0.096
					(0.069)
Mean outcome	0.368	0.368	0.368	0.368	0.368
Sample size	3,275	3,275	3,275	3,275	3,275

Notes: This table reports estimates of β_τ^t and its interactions from a version of Equation (1) that incorporates interactions between the different assigned treatments. The dependent variable is an indicator for whether a screening was completed in 2018. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.3: Effect of Financial Incentives on 2018 Health Screening Completion for Wave 1, Post-Lasso Controls

	(1)	(2)	(3)	(4)	(5)
2016 incentive ($200/$100)	0.045**	0.046	0.043	0.045**	0.013
	(0.017)	(0.024)	(0.023)	(0.017)	(0.031)
2017 incentive ($125)	0.014	0.014	0.014	0.035	–0.006
	(0.016)	(0.027)	(0.016)	(0.022)	(0.036)
2018 incentive ($75)	0.188**	0.188**	0.185**	0.209**	0.164**
	(0.016)	(0.016)	(0.028)	(0.023)	(0.039)
2016 incentive × 2017 incentive		–0.001			0.062
		(0.034)			(0.045)
2016 incentive × 2018 incentive			0.004		0.067
			(0.034)		(0.048)
2017 incentive × 2018 incentive				–0.042	0.042
				(0.032)	(0.055)
2016 incentive × 2017 incentive × 2018 incentive					–0.126
					(0.068)
Number of controls	42	42	42	42	42
Mean outcome	0.368	0.368	0.368	0.368	0.368
Sample size	3,275	3,275	3,275	3,275	3,275

Notes: This table reports estimates of β_τ^t and its interactions from a version of Equation (1) that incorporates interactions between the different assigned treatments. The dependent variable is an indicator for whether a screening was completed in 2018. Each regression controls for a set of covariates selected by Lasso to predict the dependent variable in the full sample. "Number of controls" reports the number of selected covariates. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.4: Effect of Past Screening Completion on Current Screening Completion, Post-Lasso Controls

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS estimates			IV estimates		
	Wave 1 (2016–2018)		Wave 2 (2017–2018)	Wave 1 (2016–2018)		Wave 2 (2017–2018)
	Second screening	Third screening	Second screening	Second screening	Third screening	Second screening
Completed first screening	0.431**	0.169**	0.524**	0.364**	0.330**	0.318**
	(0.015)	(0.016)	(0.023)	(0.124)	(0.123)	(0.071)
Completed second screening		0.412**			0.064	
		(0.018)			(0.071)	
Number of controls	40	42	42	40	42	42
First-stage F (first screening)				49.9	52.4	142.6
First-stage F (second screening)					189.8	
Mean outcome	0.429	0.368	0.365	0.429	0.368	0.365
Sample size	3,275	3,275	1,524	3,275	3,275	1,524

Notes: This table reports OLS and IV estimates of θ_τ^t from Equation (2). Dependent variable is an indicator for whether the second (or third) screening was completed. Each regression controls for a set of covariates selected by Lasso to predict the dependent variable in the full sample. "Number of controls" reports the number of selected covariates. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.5: Effect of Past Screening Completion on 2018 Screening Completion

	OLS estimates			IV estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
Completed first screening	0.183**	0.383**		0.331**	0.354**	
	(0.016)	(0.015)		(0.127)	(0.129)	
Completed second screening	0.430**		0.515**	0.063		0.061
	(0.017)		(0.015)	(0.073)		(0.079)
First-stage F (first screening)				48.1	45.6	
First-stage F (second screening)				182.5		143.1
Mean outcome	0.368	0.368	0.368	0.368	0.368	0.368
Sample size	3,275	3,275	3,275	3,275	3,275	3,275

Notes: This table reports OLS and IV estimates of θ_τ^t from Equation (2). The dependent variable is an indicator for whether the 2018 screening was completed. Columns (1) and (4) replicate columns (2) and (5) from Table 3. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.6: Effect of 2016 Financial Incentives and 2016 Health Screening Participation on 2018 Health Screening Participation, for Subjects Assigned a $0 Incentive in 2017

	(1)	(2)	(3)	(4)
	Reduced form		IV	
	Full sample	Subsample	Full sample	Subsample
2016 incentive ($200/$100)	0.044*	0.054*		
	(0.017)	(0.024)		
Completed first screening			0.354**	0.340*
			(0.129)	(0.143)
First-stage F (first screening)			45.6	37.3
Mean outcome	0.368	0.359	0.368	0.359
Sample size	3,275	1,636	3,275	1,636

Notes: Columns (1)–(2) report estimates of β_τ^t from Equation (1), while columns (3)–(4) report IV estimates of θ_τ^t from Equation (2). The dependent variable is an indicator for completing the 2018 health screening. Columns (1) and (3) include the full Wave 1 sample. Columns (2) and (4) limit the sample to Wave 1 subjects who were assigned the $0 incentive in 2017. The IV specification instruments for the endogenous regressor "Completed first screening" using an indicator for the 2016 financial incentive. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table A.7: Effect of 2016 Financial Incentives and 2016 Health Screening Participation on 2018 Health Screening Participation, for Subjects Assigned a \$0 Incentive in 2017, Post-Lasso Controls

	(1)	(2)	(3)	(4)
	Reduced form		IV	
	Full sample	Subsample	Full sample	Subsample
2016 incentive (\$200/\$100)	0.045**	0.045		
	(0.017)	(0.024)		
Completed first screening			0.355**	0.292*
			(0.123)	(0.144)
Number of controls	42	42	42	42
First-stage F (first screening)			49.9	36.8
Mean outcome	0.368	0.359	0.368	0.359
Sample size	3,275	1,636	3,275	1,636

Notes: Columns (1)–(2) report estimates of β_τ^t from Equation (1), while columns (3)–(4) report IV estimates of θ_τ^t from Equation (2). The dependent variable is an indicator for completing the 2018 health screening. Columns (1) and (3) include the full Wave 1 sample. Columns (2) and (4) limit the sample to Wave 1 subjects who were assigned the \$0 incentive in 2017. The IV specification instruments for the endogenous regressor "Completed first screening" using an indicator for the 2016 financial incentive. Each regression controls for a set of covariates selected by Lasso to predict the dependent variable in the full sample. "Number of controls" reports the number of selected covariates. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.