# How Do Habits Form? Experimental Evidence from Health Screenings*

Damon Jones
University of Chicago

David Molitor
University of Illinois and NBER

Julian Reif
University of Illinois and NBER

February 2025

**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. We document evidence of strong habit formation from initial exposure: completing the first screening raised subsequent screening rates by 32.4–36.0 percentage points (84%–90%) in the second and third years, with no evidence of decay. In contrast, completing the second screening had a minimal effect on screenings in the third year. This pattern contradicts an addiction mechanism, which predicts a stronger relationship for behaviors closer in time, and instead supports an experience-good model where consumers learn the value of screening through first exposure.

**Keywords:** habit formation, wellness programs, health screenings, incentives

*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 preregistered with the American Economics Association RCT Registry (AEARCTR-0001368) and was approved by the Institutional Review Boards of the National Bureau of Economic Research, University of Illinois Urbana-Champaign, and University of Chicago. We gratefully acknowledge co-investigator Laura Payne for vital contributions to the study, Lauren Geary for project management, Michele Guerra for programmatic support, Illinois Human Resources for institutional support, and Nina Kerkebane for research assistance. We are also thankful for comments from Christopher Cronin, Tatyana Deryugina, Riley League, Sarah Miller, Lena Song, and seminar participants at the 114$^{th}$ Annual Conference on Taxation, the Bozeman Applied Economics Summer Conference, the Chicago Booth Junior Health Economics Summit, the Mid-Midwest Applied Microeconomics Workshop, the Midwest Health Economics Conference, and Toulouse School of Economics. The findings and conclusions expressed are solely those of the authors and do not necessarily represent the views of the National Institutes of Health, any of our funders, or the University of Illinois. An earlier version of this paper was titled, "Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study."

# 1 Introduction

Habits, which emerge when past consumption increases current consumption, shape many social, economic, and health behaviors (Becker, 1992). Identifying habit formation and its underlying mechanisms involves two key challenges. The first is distinguishing habits from other sources of behavioral persistence. Serial correlation in consumption may reflect persistent economic conditions, fixed preferences, or inattention rather than habit formation. The second challenge is uncovering underlying mechanisms governing the behavioral dynamics relevant to optimal taxation and incentive design (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001).

A fundamental question is whether habits are driven more by recent past consumption or by initial consumption. For example, addictive habits like cigarette smoking depend strongly on recent consumption, whereas experiential learning—like discovering one's taste for mangoes—can establish long-run habits through initial exposure alone. These mechanisms can be differentiated by observing behavior across at least three periods with exogenous variation in the initial and subsequent time periods, allowing comparison of their effects on later behavior. However, such data are rarely available.

This paper estimates habit formation in annual biometric health screenings and characterizes its underlying mechanisms using a large-scale randomized controlled trial. We randomly assigned individual-level financial incentives to 3,275 university employees ("Wave 1") each year for three years (2016–2018), paying those who completed health screenings. The three incentives were rerandomized and announced at the start of each program year, allowing us to estimate both contemporaneous ("direct") effects of the incentives on screening completion and their persistence effects up to two years later. Using multiple successive years of rerandomized incentives, we analyze how recent versus initial incentives influence current behavior, shedding light on underlying mechanisms. We validate our findings among a second group of 1,524 employees ("Wave 2") who received randomized incentives for two years (2017–2018). The second wave provides an opportunity to replicate the 1-year habit estimates from Wave 1 and the staggered and overlapping nature of the two waves allows us to compare, within the same year, how incentives affected behavior across two distinct

cohorts—one with prior exposure to incentives and another without.

We find that all financial incentives—the initial, the second, and the third—had large immediate effects, boosting contemporaneous screening rates by 12.4–20.3 percentage points. The initial incentive also produced persistent effects in both waves, boosting future screening rates by 4.5–8.9 percentage points (10–24%) one year later and 4.4 percentage points (12%) two years later. By contrast, the second incentive showed a small and statistically insignificant persistence effect. In other words, screening in the third year was affected by the initial incentives in year one, but not by the more recent incentive in year two.

We interpret these results through the framework of habit formation, defined as a positive causal effect of past screening completion on current screening completion.[1] To identify this effect, we use randomly assigned incentives as instruments for past screening completion. Our estimates indicate that completing the first screening raised the likelihood of screening completion one year later by 32.4–36.0 percentage points (84–89%) and two years later by 33.1 percentage points (90%). The consistency of these estimates across waves underscores the robustness of the effect, while its stability over time suggests the effect does not decay, at least within a two-year time period. We apply the same approach to estimate the effect of completing the second screening on the likelihood of completion in the third year, and instead find only a small and statistically insignificant effect of 6.3 percentage points.

These patterns help us to distinguish between different mechanisms. We focus on two broad classes of models: "reinforcement", where habits are primarily driven by recent consumption, and "initial exposure", where they depend mainly on initial consumption. The reinforcement class includes frameworks such as rational addiction and switching costs (Becker and Murphy, 1988), while the initial exposure class encompasses experience-good and learning models. Among these, addiction and experience-good models are particularly relevant, as they are commonly cited as drivers of persistence in health behaviors (Dupas, 2014; Carrera et al., 2020; Hussam et al., 2022). We derive a standard addiction model, demonstrating that it predicts a stronger influence of recent consumption on future behavior compared to more distant consumption. By contrast, the experience-good model predicts that habits emerge

[1]Our definition of habit formation is consistent with Becker (1992) and Royer, Stehr and Sydnor (2015). It encompasses a variety of mechanisms, including addiction, switching costs, learning, and taste discovery.

when consumers learn a good's value through initial exposure (Ackerberg, 2003; Dupas, 2014; Banerjee et al., 2021). In this case, early consumption plays a decisive role, while subsequent consumption has a limited impact on future behavior.

We find that habit formation in our setting is driven by initial screening completion. Notably, this pattern is inconsistent with addiction, where more recent behavior should exert the strongest influence. Instead, our findings align with an experience-good model where the initial exposure alone raises future demand. The staggered timing of our two waves strengthens this interpretation: the 2017 incentive produced no persistence when it was Wave 1's second exposure but strong persistence when it was Wave 2's initial exposure, indicating that the timing of exposure—rather than an idiosyncratic event in 2017—determines persistence.

Habits influence the elasticity of demand and thus should be factored into predictive models and the design of optimal incentives. To illustrate, we examine the implications of our findings for the cost-effectiveness of the financial incentives in our program, focusing on the initial (2016) incentive. Ignoring habit formation, one would conclude that the incentive increased completion rates by 12.4 percentage points for a single year. However, after accounting for habit formation, the total effect of this first incentive over the two-year program period is an increase of 21.3 percentage points—72% higher than the one-year estimate. Notably, this habit formation effect is confined to the first "dose" of the intervention. While the second dose increased contemporaneous completion rates by 20.3 percentage points, it had no detectable effect on future screenings. In other words, habit formation increased the cost-effectiveness of the first incentive but not the second incentive. This finding contrasts with predictions from a standard addiction model, where habit formation would increase the cost-effectiveness of both incentives. These results highlight the importance of identifying habit formation and understanding its underlying mechanisms.

Our study builds on a large literature documenting habit formation in diverse settings, from smoking to voting to social media. Some studies attribute habit formation to addiction, where current consumption increases future demand (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Allcott and Rogers, 2014; Fujiwara, Meng and Vogl, 2016; Allcott, Gentzkow and Song, 2022). Others have attributed it to inertia (Handel, 2013), environmental cues (Giuntella, Saccardo and Sadoff, 2024), start-up costs (Carrera et al.,

2020), switching costs (Polyakova, 2016), or learning through experience (Ackerberg, 2003; Dupas, 2014; Banerjee et al., 2021). However, the empirical evidence in these studies—which achieve identification based on structural assumptions or a one-time treatment—generally cannot distinguish between these different mechanisms (Hussam et al., 2022). Ackerberg (2001) and Osborne (2011) note that information on past consumption can help identify mechanisms underlying persistence in consumer behavior, but doing so requires structural assumptions in the absence of randomized variation.

Our paper advances this literature by introducing the first randomized framework to both identify habit formation and distinguish two potential underlying mechanisms: addiction and learning through experience. A key innovation is using multiple rounds of rerandomized incentives to measure the different persistence effects of repeated treatments. This research design enables us to rule out inattention as a driver of persistence, thereby supporting habit formation as the underlying channel, and to test whether initial incentives have a greater long-term impact than subsequent incentives—a key distinction for ruling out addiction in favor of learning through experience. Our experimental framework provides a useful template that could be applied to study other models of behavioral persistence. For example, the effects of switching costs and learning on brand choice is a longstanding topic in industrial organization (Ackerberg, 2003; Farrell, 2007). In a typical switching cost model, current consumption depends on the immediately preceding period's consumption, whereas in an experience-good model, it depends on earlier consumption when learning is fast. Employing multiple rounds of rerandomized incentives, as demonstrated in our study, allows one to distinguish between switching costs and learning, without relying on structural methods.

We also contribute to the literature on habit formation in health behaviors, which has predominantly focused on high-frequency activities such as exercising.[2] 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

[2]Even outside the realm of health, few studies have examined habit formation in infrequent behaviors. A prominent exception is Fujiwara, Meng and Vogl (2016), who study habit formation in voting.

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 on 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 literature 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 persistence effects of much larger incentives ($75–$200) on health screenings and shows that these persistence effects reflect habit formation. Unlike Schneider et al. (2023), we find that incentives for initial participation have lasting effects that are economically large and statistically significant.

The remainder of the paper is organized as follows. Section 2 presents a model of addiction and shows how its predictions differ from those produced by alternative mechanisms. Section 3 provides background on our empirical setting and describes our data. Section 4 outlines our empirical strategy. Section 5 describes our results, and Section 6 concludes.

# 2 Model

Addiction is a well-studied framework for understanding habit formation and has been applied to a number of different health-related behaviors, including smoking, exercise, and handwashing (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Carrera et al., 2020; Hussam et al., 2022). We present a basic model of addiction and derive the habit formation estimating equation described in Section 4. The model yields a clear testable prediction: current demand is influenced more strongly by recent consumption than by distant past consumption. We then describe how these predictions differ from those generated by alternative mechanisms such as an experience-good model.

## 2.1 Addiction

In each period $t \leq T$, consumers choose consumption of an addictive good $a_t$, sold at price $p_t$, and a composite good $c_t$, taken as numeraire. Consumption of these goods, along with unobserved factors $x_t$, produce flow utility $U(a_t, S_t, c_t, x_t)$. To capture habit formation, utility depends on a stock of past consumption $S_t$, which evolves according to the following law of motion:

$$S_{t+1} = (1-d)\, S_t + a_t, \tag{1}$$

where $d \in (0, 1]$ is the rate of depreciation.

In each period $t$, the consumer's objective is to maximize utility:

$$\max_{a_t, c_t} U(a_t, S_t, c_t, x_t), \tag{2}$$

where $S_t$ is given. For simplicity, we assume consumers are myopic. Introducing forward-looking behavior does not alter the key prediction regarding the relationships between current and past consumption.[3]

Consumers have income $w$ to allocate between the numeraire good $c_t$ and the addictive

[3]If consumers were forward-looking, demand would additionally depend on expectations of future consumption. In our setting, financial incentives were rerandomized annually and announced at the start of each program year. Because future incentives were unknown to participants, they cannot be used as instruments for future expectations.

good $a_t$, leading to the budget constraint:

$$w = c_t + p_t a_t. \tag{3}$$

To solve the model, we assume utility is concave and quadratic. This assumption, common in the addiction literature, delivers linear first-order conditions that facilitate empirical estimation (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Reif, 2019). Solving the consumer's optimization problem (2) subject to the budget constraint (3) yields the following demand equation:

$$a_t = k + \alpha S_t + \pi p_t + e_t, \tag{4}$$

where $k$ is a constant, $\alpha > 0$, $\pi < 0$, and the error term $e_t$ captures the effect of $x_t$ on consumption.[4] A full derivation is provided in Appendix A.

Now, consider a three-period model ($T = 3$) with initial stock $S_0 = 0$, which reflects the empirical setting examined in Section 4. Substituting the law of motion (1) into the demand equation (4) and back-substituting, we obtain our habit formation estimating equation:

$$a_3 = k + \alpha(1 - d)a_1 + \alpha a_2 + \pi p_3 + e_3. \tag{5}$$

We address the endogeneity of past consumption using randomized incentives, as described in Section 4.2.

Equation (5) shows that third-period demand depends on both first- and second-period consumption. The coefficient on second-period consumption ($\alpha$) is larger than the coefficient on first-period consumption ($\alpha(1 - d)$), since $(1 - d) < 1$ for $d > 0$. This yields the testable prediction that current demand is more strongly related to recent consumption than to earlier consumption.[5] In the special case where depreciation $d = 1$, the stock fully resets every period and $S_{t+1} = a_t$. The demand equation then simplifies to the myopic addiction

[4] Stable demand requires that $\alpha/d < 1$, ensuring that the influence of past consumption does not grow without bound. Conditions for $\alpha > 0$ are discussed in Appendix A.

[5] Alternatively, third-period demand can be expressed solely in terms of current and past exogenous variables. This alternative formulation yields the "reduced-form" persistence model described in Section 4.1 and generates analogous testable predictions.

model of Becker, Grossman and Murphy (1994):

$$a_3 = k + \alpha a_2 + p_3 + e_3. \tag{6}$$

In this case, only consumption in the immediately preceding period affects current demand; earlier consumption has no direct effect.

## 2.2 Alternative Models

Habit formation can arise from mechanisms beyond addiction, resulting in different consumption dynamics. One prominent channel considered in industrial organization is learning through experience, where consumers learn their preferences by sampling them directly (Erdem and Keane, 1996; Ackerberg, 2003; Israel, 2005). For example, a consumer trying mango for the first time can only discover through tasting it whether she enjoys it—a discovery that might lead to a lasting consumption habit. Unlike addiction, in this "experience good" model, current mango demand depends on initial rather than recent consumption.[6]

More formally, we define the experience-good model as the special case where the stock of past consumption is represented by a binary indicator:

$$S_t = \begin{cases} 1 & \text{if } a_\tau > 0 \text{ for any } \tau < t, \\ 0 & \text{otherwise.} \end{cases}$$

This formulation yields the same demand equation (4) as before. Considering a three-period model with an initial stock of $S_0 = 0$ yields the following habit formation estimating equation:

$$a_3 = k + \alpha \mathbf{1}(a_1 > 0) + \alpha \mathbf{1}(a_2 > 0 \mid a_1 = 0) + \pi p_3 + e_3. \tag{7}$$

A key prediction of this model is that the coefficients on first- and second-period consumption are identical. The regressor for second-period consumption is defined conditional on

[6] One could also consider models where an individual learns or discovers a preference over time. If learning occurs relatively quickly and the rate of depreciation rate is slow, it will still produce the pattern that current demand depends more on earlier consumption than more recent consumption.

first-period consumption being zero, ensuring that we measure only the effect of first-time consumption occurring in period 2.

Now, consider an alternative specification where the estimating equation includes unconditional consumption indicators, making it more directly comparable to equation (5). In this case, if most individuals consume in the first period, the coefficient on $a_1$ would naturally be larger than that on $a_2$, reflecting the fact that early consumption events are more common and therefore exert a greater impact on observed demand.

Several studies link the experience-good model to health behaviors. Crawford and Shum (2005) and Dickstein (2021) model pharmaceutical demand and drug adherence under experiential learning about drug efficacy. Dupas (2014) subsidizes the purchase of antimalarial bed nets and attributes the subsequent increase in long-run demand to learning about the product's true value. Similarly, Banerjee et al. (2021) find that a one-time health insurance subsidy led to sustained enrollment, which they attribute to learning through experience. Related to experience goods is the concept of start-up costs, which have been cited as drivers of habit formation in exercise (Charness and Gneezy, 2009; Carrera et al., 2020). In these models, current demand is shaped by initial consumption but does not necessarily depend on recent consumption.

To our knowledge, no prior study has implemented a research design to distinguish addiction from an experience good model. Hussam et al. (2022), for instance, interpret persistence in handwashing as an addiction but acknowledge that it may also reflect learning.[7] Similarly, Banerjee et al. (2021) attribute persistence in health insurance enrollment to learning from initial exposure, but it is also possible that persistence is driven instead by more frequent or more recent interaction with the health system.

Our analysis demonstrates that addiction and experience-good models can be empirically distinguished when consumption data span three or more periods, including the initial period in which learning might plausibly occur. Specifically, if habit formation is driven by addiction, third-period demand will be more strongly influenced by second-period consumption than first-period consumption. Conversely, an experience-good model predicts the opposite

[7]They note that increased handwashing may cause people to "learn that handwashing leads to improvements in health, and therefore [to] update their beliefs on the returns to the behavior" (Hussam et al., 2022).

pattern, where earlier consumption has a greater effect on current demand.

Other mechanisms, such as inertia and switching costs, can also drive habit formation. Handel (2013) and Polyakova (2016) argue that these mechanisms contribute to persistence in insurance plan enrollment. Like addiction, inertia and switching cost models predict that current demand should be most strongly related prior consumption from the most recent period. However, we believe these mechanisms are not relevant to our setting. Our subjects are required to make an active choice in each period, which rules out inertia as an explanation for our persistence. Furthermore, there is no clear source of switching costs in our context.

# 3 Background

The Illinois Workplace Wellness Study is a large-scale randomized controlled trial designed to evaluate the effects of workplace wellness programs on employee health, behavior, and productivity, and to examine how incentives shape participation (Jones, Molitor and Reif, 2019; Reif et al., 2020). Conducted at the University of Illinois at Urbana-Champaign, the study randomized 3,300 benefits-eligible employees to a "wellness" group eligible for a two-year workplace wellness program involving three annual health screenings from 2016 to 2018. The remaining 1,534 employees were assigned to a "non-wellness" group—ineligible for the wellness program but eligible for the 2017 and 2018 follow-up health screenings. We limit our analysis to subjects who remained continuously enrolled through the end of the intervention, resulting in a final sample size of 4,799.

Employees in the wellness group ($N = 3,275$), 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. 34 percent of Wave 1 members completed at least one activity during the two-year program.

Employees assigned to the non-wellness group ($N = 1,524$), 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.[8] 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.

# 4 Empirical Strategy

Our empirical strategy encompasses two objectives. The first is to estimate the causal effects of financial incentives on health screening completion. We define the "direct effect" of incentives as their impact on same-year screening completion and the "persistence effect" as their impact on future completion. The five treatments (incentives) embedded in the study design depicted in Figure 1 all produce measurable direct effects. We can measure persistence effects for three of these treatments: the 2016 incentives' effects on Wave 1's

[8]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.

2017–2018 screenings, and the 2017 incentives' effects on 2018 screenings for both Wave 1 and Wave 2. Our second objective is to estimate habit formation, for which we use an IV approach that assumes the persistence effect operates through past completion.[9]

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 and lead to habit formation. 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.

## 4.1 Persistence

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

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=1}^{T} \beta_\tau \text{INCENTIVE}_i^\tau + \gamma X_i + \epsilon_i. \tag{8}$$

The outcome, $\text{SCREEN}_i^t$, is an indicator equal to 1 if individual $i$ completed a health screening in event year $t \in \{1, 2, 3\}$, where event year 1 corresponds to 2016 for Wave 1 and 2017 for Wave 2. The focal explanatory variable is $\text{INCENTIVE}_i^\tau$, an indicator equal to 1 if individual $i$ was assigned a non-zero screening incentive in year $\tau$, with $\tau$ ranging from 1 to $T = 3$ for Wave 1 and from 1 to $T = 2$ for Wave 2 (Figure 1). We estimate the equation separately for each wave and screening year, with individuals as the unit of observation, resulting in five separate regressions: three for Wave 1, and two for Wave 2.

The focal parameter, $\beta_\tau$, represents the average treatment effect (ATE) of monetary

[9]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.

incentives assigned in year $\tau$ on screening completion. An individual receives the assigned incentive in year $\tau$ only upon completing a health screening that year. Thus, equation (8) identifies the direct effect of incentives on screening completion when screening is measured in the same year as the incentive ($t = \tau$) and a persistence effect when screening is measured in a future year ($t > \tau$). Since future incentives were unknown to participants, we expect $\beta_\tau = 0$ for $t < \tau$, providing a falsification test of our model.

Equation (8) includes multiple treatment variables. If the effect of an incentive depends on the receipt of past incentives, $\beta_\tau$ will capture a weighted average of ATEs across the counterfactual scenarios created by different combinations of incentives offered in previous years (Muralidharan, Romero and Wüthrich, 2023).[10] This weighted average is policy relevant because many real-world wellness programs and other similar interventions commonly offer incentives on an annual basis. In supplemental analyses, we extend equation (8) to include interaction terms.

In our baseline analysis, equation (8) 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.[11] Since randomization was performed at the individual level, we report conventional heteroskedasticity-robust standard errors (Abadie et al., 2023).

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

[10] Future incentives do not influence this average, as they were unknown at the time and could not affect current screening decisions.

[11] 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.

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.

## 4.2 Habit Formation

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

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=1}^{t-1} \theta_\tau \text{SCREEN}_i^\tau + \gamma X_i + \epsilon_i. \tag{9}$$

The focal explanatory variables, $\text{SCREEN}_i^\tau$, are indicators equal to 1 if individual $i$ completed a screening in previous event year $\tau < t$. The term $X_i$ represents a vector of individual-specific control variables. Our baseline analysis includes the contemporary financial incentive, $\text{INCENTIVE}_i^t$, as a control variable because it is a strong predictor of $\text{SCREEN}_i^t$. In supplemental analyses, we report a specification that includes post-Lasso controls. We estimate the equation separately for each wave and screening year $t > 1$, resulting in three separate regressions: two for Wave 1, and one for Wave 2.

Ordinary least squares (OLS) estimation of equation (9) 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 (9), 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 comple-

tion) 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 5.2. In models where there is only one endogenous regressor, we interpret $\theta_\tau$ as the average causal effect of screening completion in year $\tau$ among compliers, i.e., those induced to complete a health screening in year $\tau$ by a financial incentive. This interpretation requires the standard monotonicity assumption for local ATE (LATE) interpretations of 2SLS. For Wave 1, 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.

Equation (9) allows us to distinguish between addiction and experiential learning mechanisms by comparing the effects of the first and second incentives on the third screening for Wave 1. If habit formation follows the standard addiction model described by equation (5), then we expect that $\theta_2 > \theta_1$. If the stock fully resets every period, as in the model described by equation (6), the addiction model makes the stronger prediction that $\theta_2 > \theta_1 = 0$. By contrast, an experiential learning model where demand depends primarily on initial consumption predicts $\theta_2 < \theta_1$.

# 5 Results

## 5.1 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 (8), 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.[12] 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,

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

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 but was the first dose for Wave 2. We consider this possibility in further detail in Section 5.2.

## 5.2 Habit Formation

### 5.2.1 Main Estimates

Table 3 presents habit formation estimates from equation (9). 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.

Columns (4)–(6) address this bias by instrumenting for past screening completion with randomly assigned past incentives. Estimates in the first row show that completing the first health screening increased the likelihood of completing the second screening by 36.0 (84%) and 32.4 (89%) percentage points in Waves 1 and 2, respectively, and raised screening completion in the third year 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 similar estimates, as shown in columns (4)–(6) of Table A.5.

These IV results corroborate the conclusions of the reduced-form analysis in Section 5.1: habit formation produced by the initial health screening was equally strong in both waves, showed no decay after two years, and was greater than the habit formation produced by the second screening. We describe statistical tests to substantiate these conclusions and their implications for underlying mechanisms in the following section.

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 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.[13]

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

[13]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.

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 the possibility that offering a financial incentive prompts individuals to seek information about biometric screenings, thereby increasing the likelihood of future screenings regardless of whether they completed the initial incentivized screening. Although we consider this possibility unlikely, we lack evidence to rule it out definitively. However, even if the exclusion restriction were to fail, the persistence estimates presented in Table 2 would remain valid.

### 5.2.2 Mechanisms

Habit formation—a positive effect of past consumption on current consumption—can arise from different microeconomic foundations. Section 2 showed that we can distinguish two prominent models of habit formation—addiction and experiential learning—by comparing the effects of the first and second screening completions on screening completion in the third year of the study. Under addiction, the effect of the second screening is stronger, while under experiential learning we expect the initial screening effect to dominate.

The habit formation estimates presented in Table 3 provide evidence against addiction and instead support experiential learning through two key patterns. First, the initial screening's effect is stable across waves and persists without decay over two years, consistent with experiential learning through first exposure. Second, the stronger habit formation effects from the first screening, compared to the second, contradict the addiction model's prediction that more recent consumption should have the larger influence on current behavior.

To substantiate these patterns, we conduct statistical tests comparing the habit formation effects reported in Tables 3 and A.4. First, we test whether the effects of completing the first screening are equal across three scenarios: its impact on the second screening in Wave 1 (column (4)), the third screening in Wave 1 (column (5)), and the second screening in Wave 2 (column (6)). The results show that we cannot reject the null hypothesis of joint equality

across these three estimates, with $p > 0.85$ regardless of whether post-Lasso controls are included. These results indicate that the initial screening establishes a habit that persists consistently over two years and that the magnitude of this effect remains uniform across both waves in the first year.

Second, we test whether the habit formation effects from first and second screenings are equal. These tests lead us to reject equality. The 95% confidence interval for the second screening effect in column (5), $[-8.0, 20.5]$, does not contain any of the three first-year point estimates in columns (4)–(6), nor do any of the first-year confidence intervals contain the second-year point estimate. A formal test also rejects joint equality between the first- and second-year estimates in column (5), with $p < 0.1$.[14]

We further investigate the experiential learning mechanism using our rerandomized research design. As shown in column (5) of Table 3, the 2017 incentive had no detectable effect on 2018 screening completion among Wave 1 individuals, who were first offered screenings in 2016. Under the experiential learning hypothesis, the effect of the 2017 incentive should be larger for individuals assigned a low incentive in 2016, since they are more likely to complete their first screening in 2017. Column (4) of Table A.8 replicates our earlier estimate using the full Wave 1 sample. Columns (5) and (6) break down the estimate based on whether individuals were assigned a low or high incentive in 2016. The results do show a stronger effect for those who were assigned a low ($0) incentive in 2016: 13.6 percentage points compared to only 2.8 percentage points for the high incentive group. However, we cannot reject the equality of the estimates due to limited statistical power. We therefore interpret these results as providing additional suggestive evidence in favor of the experiential learning model. Columns (1)–(3) present similar estimates for the corresponding reduced-form specifications.

## 5.3 Discussion

Overlooking habit formation significantly underestimates the effectiveness of financial incentives. For example, consider the initial (2016) incentive offered to Wave 1. Without

[14]For greater power, we also use stacked data to estimate common effects in pooled samples. We reject equality of first and second screening effects for pooled 2018 screenings (columns (5)–(6), $p = 0.03$ with or without post-Lasso controls) and for all screenings (columns (4)–(6), $p = 0.03$ with post-Lasso controls, $p = 0.06$ 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 column (1) of 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.[15] 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 increased the effectiveness of our health screening incentives, this effect applies only to 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. This finding contrasts with predictions from a standard addiction model, where habit formation would increase the effectiveness of both doses.

Our results also shed light on the role of biased beliefs in shaping health behavior. We find evidence of habit formation, driven by a one-shot mechanism such as an experiential learning model. This pattern is consistent with individuals initially holding negatively biased beliefs—for example, overestimating the costs or underestimating the benefits of screening participation—which are revised positively after the first screening. However, different patterns may arise in settings where biased beliefs take a different form. For instance, if individuals hold positively biased beliefs, such as overconfidence in their ability to maintain health, initial participation in a health behavior may lead to a decline in that behavior over time as expectations are tempered.

# 6 Conclusion

We show that even infrequent health behaviors like annual health screenings can become strongly habitual. Completing a health screening for the first time raises the probability of

[15]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).

completing a future screening by over 30 percentage points. However, completing a second screening does not further reinforce this habit. This pattern contradicts the standard model of addiction, but is consistent with an experience-good model where consumers learn the value of health screenings.

The key to disentangling these two mechanisms is our novel experimental research design, which includes multiple rounds of rerandomized incentives that allow us to separately identify the effects of different past screening decisions. By contrast, prior work has generally relied on panel data coupled with quasi-experimental variation, a one-time experimental treatment, or structural modeling (Becker, Grossman and Murphy, 1994; Ackerberg, 2001; Gruber and Köszegi, 2001).

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.[16]

[16]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.

**Ackerberg, Daniel A.** 2003. "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination." *International Economic Review*, 44(3): 1007–1040.

**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.

**Banerjee, Abhijit, Amy Finkelstein, Rema Hanna, Benjamin A. Olken, Arianna Ornaghi, and Sudarno Sumarto.** 2021. "The Challenges of Universal Health Insurance in Developing Countries: Experimental Evidence from Indonesia's National Health Insurance." *American Economic Review*, 111(9): 3035–63.

**Becker, Gary S.** 1992. "Habits, Addictions, and Traditions." *Kyklos*, 45(3): 327–345.

**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.** 1994. "An empirical analysis of cigarette addiction." *The American Economic Review*, 84(3): 396–418.

**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.

**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.

**Crawford, Gregory S., and Matthew Shum.** 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica*, 73(4): 1137–1173.

**Dickstein, Michael J.** 2021. "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice." Working paper.

**Dupas, Pascaline.** 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment." *Econometrica*, 82(1): 197–228.

**Erdem, Tülin, and Michael P. Keane.** 1996. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science*, 15(1): 1–20.

**Farrell, J.** 2007. "Coordination and Lock-In: Competition with Switching Costs and Network Effects." *Handbook of Industrial Organization*, 3.

**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.

**Handel, Benjamin R.** 2013. "Adverse selection and inertia in health insurance markets: When nudging hurts." *American Economic Review*, 103(7): 2643–2682.

**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.

**Israel, Mark.** 2005. "Services as experience goods: An empirical examination of consumer learning in automobile insurance." *American Economic Review*, 95(5): 1444–1463.

**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.

**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.

**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.

**Polyakova, Maria.** 2016. "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D." *American Economic Journal: Applied Economics*, 8(3): 165–195.

**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.

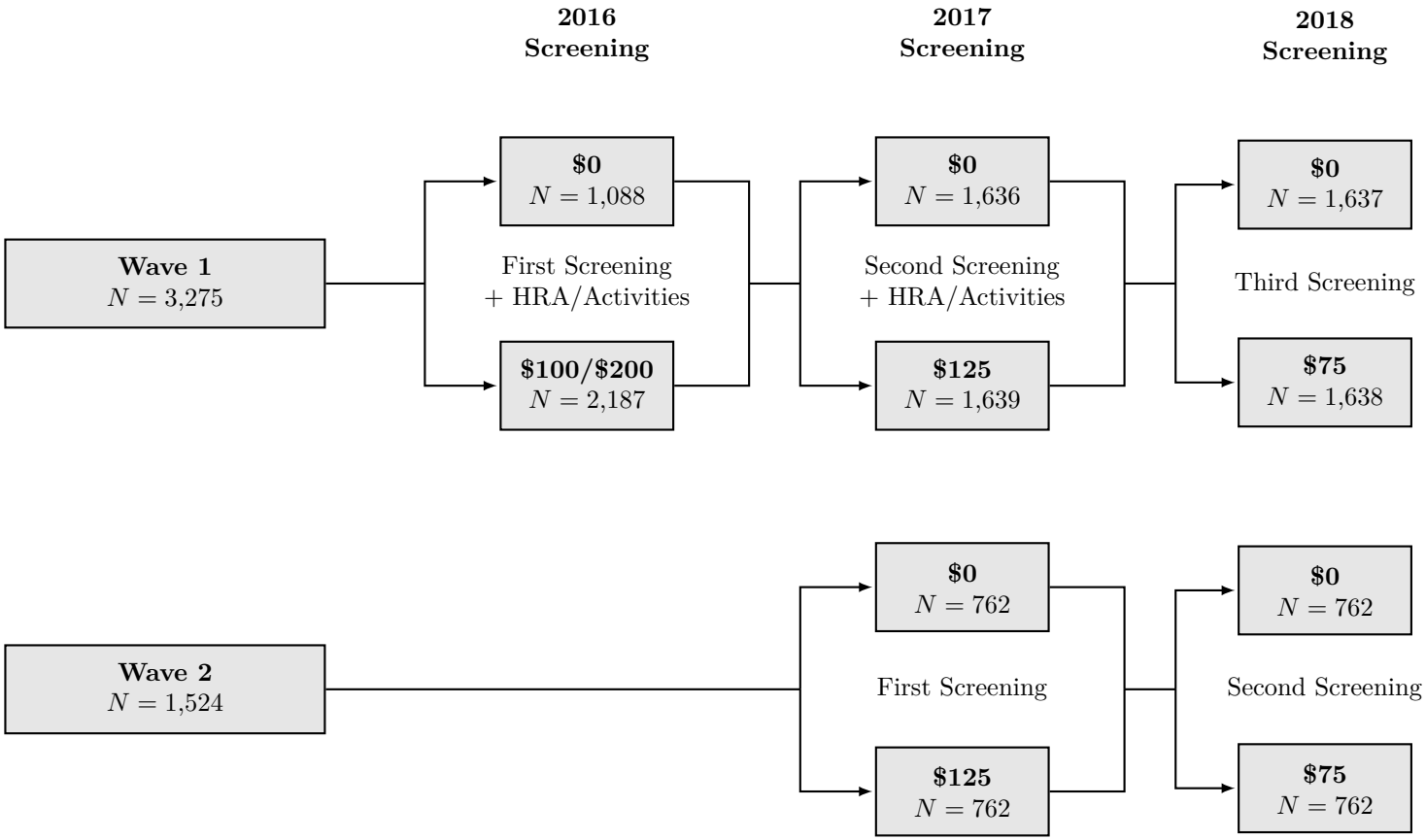
**Schneider, Florian H., Pol Campos-Mercade, Stephan Meier, Devin Pope, Erik Wengström, and Armando N. Meier.** 2023. "Financial incentives for vaccination do not have negative unintended consequences." *Nature*, 613(7944): 526–533.

**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.

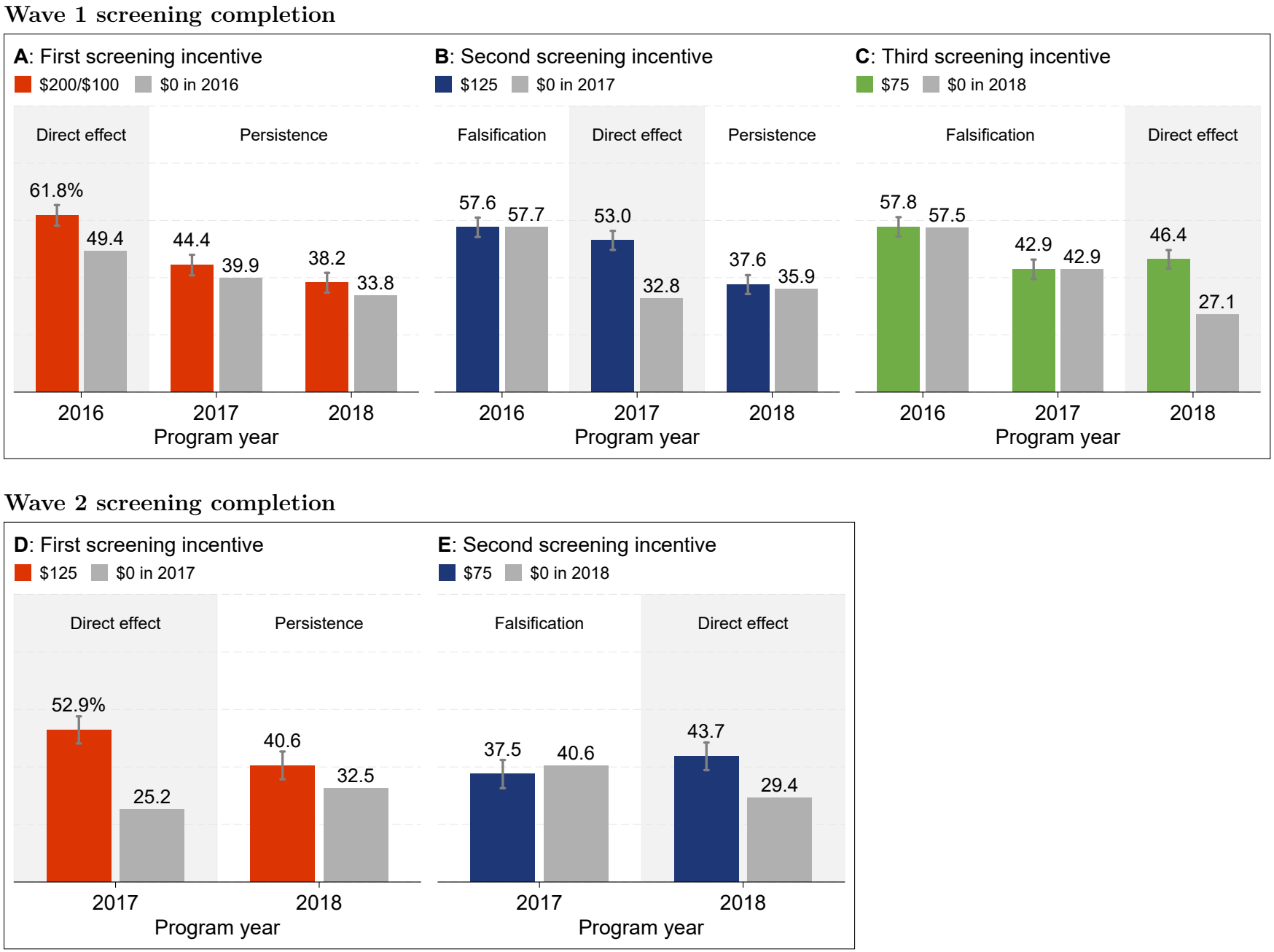
**Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein.** 2008. "Financial incentive-based approaches for weight loss: A randomized trial." *Journal of the American Medical Association*, 300: 2631–2637.

**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




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 $\beta_\tau$ from Equation (8). 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 $\theta_\tau$ from Equation (9). 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

# A Addiction Model Derivation

This section provides the full derivation of the demand equation (4) for the addiction model presented in Section 2. Following Reif (2019), we solve the consumer's optimization problem under the assumption that utility is concave and quadratic:

$$\begin{aligned} U\left(a_t, S_t, c_t, x_t\right) = &-\frac{1}{2}\left(u_{aa}a_t^2 + u_{ss}S_t^2 + u_{cc}c_t^2 + u_{xx}x_t^2\right) + u_{as}a_tS_t + u_{ac}a_tc_t + u_{ax}a_tx_t \\ &+ u_{sc}S_tc_t + u_{sx}S_tx_t + u_{cx}c_tx_t + u_aa_t + u_sS_t + u_cc_t + u_xx_t. \end{aligned} \tag{A.1}$$

Addictive behavior is characterized by the parameter $u_{as} > 0$, which captures the strength of intertemporal complementarity.

Maximizing utility with respect to $c_t$ and subject to the budget constraint (3) yields:

$$c_t = \frac{-\lambda + u_c + u_{ac}a_t + u_{sc}S_t + u_{cx}x_t}{u_{cc}},$$

where $\lambda$ is the marginal utility of wealth. Substituting this expression back into the utility function (A.1) reduces the consumer's objective function to a maximization problem in $a_t$ only:

$$\max_{a_t} U^*\left(a_t, S_t, x_t\right), \tag{A.2}$$

where:

$$\begin{aligned} U^*\left(a_t, S_t, x_t\right) = &-\frac{1}{2}\left(b_{aa}a_t^2 + b_{ss}S_t^2 + b_{xx}x_t^2\right) + b_{as}a_tS_t + b_{ax}a_tx_t \\ &+ b_{sx}S_tx_t + b_aa_t + b_sS_t + b_xx_t + b_k. \end{aligned} \tag{A.3}$$

Each parameter in the utility function (A.3) can be expressed in terms of the parameters from (A.1). The parameter $b_{as}$, which captures the effect of the stock of past consumption on the marginal utility of current consumption, is given by:

$$b_{as} = \frac{u_{ac}u_{sc} + u_{as}u_{cc}}{u_{cc}}.$$

A sufficient condition for $b_{as} > 0$ is $u_{ac}u_{sc} \geq 0$. This condition holds, for example, if utility is additively separable in $a_t$ and $c_t$. As in other studies of addiction, we assume $b_{as} > 0$.

Maximizing (A.2) subject to the budget constraint (3) yields the demand equation (4) from the main text:

$$a_t = k + \alpha S_t + \pi p_t + e_t,$$

where:

$$k = \frac{b_a}{b_{aa}}$$
$$\alpha = \frac{b_{as}}{b_{aa}} > 0$$
$$\pi = \frac{-\lambda}{b_{aa}} < 0$$

The error term $e_t$ captures the effect of (potentially unobservable) $x_t$ on demand:

$$e_t = \frac{b_{ax}}{b_{aa}} x_t.$$

Steady-state consumption of the addictive good is defined as:

$$a^* = S^* d.$$

Plugging this into the demand equation (4) yields:

$$a^* = \frac{E\left[\pi p_t + e_t\right]}{1 - \alpha/d},$$

where the expectation is taken with respect to the random variables $p_t$ and $x_t$ (the latter is included in $e_t$). Stable demand requires $1 - \alpha/d > 0$, or equivalently, $\alpha/d < 1$. Highly addictive goods, characterized by a large $\alpha$, are more likely to result in unstable demand.

**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 $\beta_\tau$ from Equation (8). 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 $\beta_\tau$ and its interactions from a version of Equation (8) 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 $\beta_\tau$ and its interactions from a version of Equation (8) 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 $\theta_\tau$ from Equation (9). 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 $\theta_\tau$ from Equation (9). 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 Completion on 2018 Health Screening Completion, 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 $\beta_\tau$ from Equation (8), while columns (3)–(4) report IV estimates of $\theta_\tau$ from Equation (9). 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 Completion on 2018 Health Screening Completion, 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 $\beta_\tau$ from Equation (8), while columns (3)–(4) report IV estimates of $\theta_\tau$ from Equation (9). 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.

**Table A.8:** Effect of 2017 Financial Incentives and 2017 Health Screening Completion on 2018 Health Screening Completion, by 2016 Financial Incentive

| | Reduced form | | | IV | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 2017 incentive ($125) | 0.012 | 0.025 | 0.006 | | | |
| | (0.017) | (0.028) | (0.020) | | | |
| Completed second screening | | | | 0.061 | 0.136 | 0.028 |
| | | | | (0.079) | (0.141) | (0.095) |
| Sample restrictions: | | | | | | |
| None | X | | | X | | |
| 2016 low incentive only | | X | | | X | |
| 2016 high incentive only | | | X | | | X |
| First-stage $F$ (second screening) | | | | 143.1 | 40.3 | 103.3 |
| Mean outcome | 0.368 | 0.338 | 0.382 | 0.368 | 0.338 | 0.382 |
| Sample size | 3,275 | 1,088 | 2,187 | 3,275 | 1,088 | 2,187 |

Notes: This table reports estimates of $\beta_\tau$ from Equation (8). The dependent variable is an indicator for whether a screening was completed in 2018. Column (1) replicates column (3) from Table 2. Column (2) omits the 2016 incentive regressor. Column (3) limits the sample to individuals who were assigned the low incentive ($0) in 2016. Column (4) limits the sample to individuals receiving a high incentive in 2016. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.