NBER WORKING PAPER SERIES

# HOW DO HABITS FORM? EXPERIMENTAL EVIDENCE FROM HEALTH SCREENINGS

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, Revised July 2026

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 114th 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, the NBER Economics of Health Program Meeting, and the 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." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Do Habits Form? Experimental Evidence from Health Screenings
Damon Jones, David Molitor, and Julian Reif
NBER Working Paper No. 32745
July 2024, Revised July 2026
JEL No. H0, I1

## ABSTRACT

We develop a framework for distinguishing between two competing mechanisms of habit formation and apply it to a three-year randomized experiment on annual health screenings among 4,799 university employees. Annually re-randomized incentives generate exogenous variation in early, late, and repeated screening, allowing us to test whether habits are strengthened by additional consumption or instead form discretely once consumption crosses a threshold. We find strong habit formation from initial exposure: completing a first screening raised subsequent screening rates by 32.4–36.1 percentage points (90%–111%). In contrast, completing a second screening had no detectable additional effect. Our results reject reinforcement models, including addiction, which predict repeated participation should strengthen the habit, and instead support a threshold model in which a single exposure is sufficient to establish and sustain the habit. Observational estimates imply the opposite conclusion, showing the value of rigorous research design in identifying mechanisms of habit formation.

Damon Jones
The University of Chicago
Harris School of Public Policy
and NBER
damonjones@uchicago.edu

Julian Reif
University of Illinois Urbana-Champaign
Department of Finance
and NBER
jreif@illinois.edu

David Molitor
University of Illinois Urbana-Champaign
Department of Finance
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

Habits shape many social, economic, and health behaviors (Pollak, 1970; Becker, 1992). Empirical work has mostly focused on distinguishing habits from other sources of behavioral persistence, such as fixed preferences or persistent economic conditions. Yet for both theory and policy, the central question is not only whether habits exist, but also how they are established and sustained.

How habits form matters. Cigarette smoking might sustain itself through repeated consumption, whereas a lasting taste for a new food can form after a single experience. Both produce persistent behavior, but their implications for taxes and incentives differ (Becker, Grossman and Murphy, 1994; Hintermann and Lange, 2013). Consider annual vaccination: if a first visit establishes the habit, incentives should target initial take-up; if repetition sustains it, incentives should recur. Distinguishing these mechanisms requires exogenous variation in multiple periods of past consumption—for example, separate variation in initial and subsequent vaccination decisions—which few empirical settings provide.

This paper develops a framework to characterize and distinguish competing mechanisms of habit formation. We define a habit as a positive causal effect of past consumption on current consumption and organize habit-forming mechanisms into two classes. The first, which we call *reinforcement models*, includes rational addiction, switching costs, and related mechanisms in which recent consumption affects current behavior more strongly than distant consumption (Klemperer, 1987; Becker and Murphy, 1988). The second, *threshold models*, includes experience-good and learning mechanisms in which a habit is established once consumption exceeds a critical threshold (Dupas, 2014; Banerjee et al., 2021). In a two-period setting, these classes are observationally equivalent because initial consumption is also the most recent. We show that with three periods, separating the mechanisms is possible, provided there is exogenous variation in early, late, and repeated consumption. This variation identifies two distinguishing features: *recency*, whether recent actions matter more than earlier ones, and *frequency*, whether repetition further strengthens the habit. Reinforcement models predict both effects; threshold models predict neither.

We apply this framework to study habit formation in health screenings, drawing on a

large-scale randomized controlled trial among 4,799 university employees. Employees were randomly divided into two waves: 3,275 ("Wave 1") were offered financial incentives to complete an annual biometric health screening each year for three years (2016–2018), while the other 1,524 employees ("Wave 2") were invited to participate only in the final two years. Crucially, incentives were re-randomized at the start of each year, generating the exogenous variation in early, late, and repeated screening that our framework requires.[1]

The reduced-form effects of the incentives provide a first indication of the underlying mechanism. In Wave 1, the initial (2016) incentive had a large contemporaneous effect, boosting first-year screening rates by 12.4 percentage points. Screening rates remained 4.5 percentage points higher one year later and 4.4 percentage points higher two years later. The second-year (2017) incentive also produced a large contemporaneous effect but, unlike the initial incentive, generated little persistence. Wave 2 confirms the persistent effect of a first incentive in an independent sample: its initial (2017) incentive raised screening rates by 27.7 percentage points that year and by 8.0 percentage points one year later. In both waves, roughly one-third of the initial boost persisted, indicating substantial habit formation. The persistent effect of the initial incentive, together with the absence of persistence from the subsequent incentive, is suggestive of the underlying mechanism: this pattern is difficult to reconcile with reinforcement models, in which repeated exposure progressively strengthens habits. However, because the second-year incentive in Wave 1 induced screening among both first-time and repeat participants, the reduced-form estimates do not separately identify the effects of initial and repeated screening.

To separate these effects, we use the randomized incentives as instruments for screening participation in two complementary ways. Our first approach estimates parametric benchmarks: for the reinforcement class, an addiction model in which past consumption accumulates into a geometrically depreciating habit stock; for the threshold class, a zero-threshold (experience-good) model in which any prior screening suffices to establish the habit. By imposing structure, these specifications deliver powerful tests of their respective mechanisms.

These parametric estimates reveal strong habit formation from initial exposure and reject

[1]Our preregistered analysis plan specified estimation of the contemporaneous effects of incentives on screening completion. The analysis of persistence and habit formation mechanisms was not prespecified.

the addiction model in favor of the zero-threshold model. The key distinguishing prediction of the addiction model is that more recent consumption should have a larger effect. Estimating this specification on Wave 1, we find the opposite: completing a screening in 2016 raised the probability of screening in 2018 by 33.1 percentage points, whereas completing one in 2017 raised it by a statistically insignificant 6.3 percentage points. Under the zero-threshold specification, by contrast, completing at least one prior screening raised subsequent screening rates by 32.4–36.1 percentage points, a 90–111 percent increase relative to the control-group screening rate. When we allow a second prior screening to have an additional effect beyond the first, the estimated incremental effect is small and statistically indistinguishable from zero, as the zero-threshold model predicts.

Our second approach relaxes these parametric assumptions and estimates a general recency–frequency specification that relates third-year screening to mutually exclusive histories of prior behavior: screened only in the first year, only in the second year, or in both years. This specification places minimal functional-form restrictions on the data while preserving the key distinction between the two classes: reinforcement models predict recency and frequency effects, whereas the zero-threshold model predicts neither.

The general specification yields the same verdict as the parametric tests. Screening only in the first year, only in the second year, or in both years raised third-year screening by amounts that are large, similar in magnitude, and statistically indistinguishable. This pattern is precisely the equality the zero-threshold model implies, and it is inconsistent with reinforcement models, which predict a larger effect of second-year screening (recency) and a larger effect of screening in both years (frequency). Taken together, the evidence rejects reinforcement in favor of a threshold mechanism of habit formation in annual screenings: a single screening is sufficient to establish and sustain the habit, and additional screenings provide no incremental reinforcement.

Our randomized design is essential for reaching this conclusion. Observational (OLS) estimates point to a very different conclusion: they exhibit strong recency and frequency effects, consistent with a reinforcement mechanism, and reject the key restrictions implied by the threshold model. This contrast likely reflects persistent unobserved differences in individuals' propensity to screen. Research design thus matters not only for distinguishing

habits from other sources of behavioral persistence, but also for identifying how those habits form. In this setting, standard observational evidence would recover the wrong model of habit formation.

Finally, we show how understanding the mechanisms of habit formation can inform the design of cost-effective interventions. A naive evaluation that ignores habit formation would credit the initial Wave 1 incentive with a 12.4 percentage point increase in screening for a single year. Accounting for the habit it created, however, raises the total effect to 21.3 percentage points over the three-year program period, 71% higher than the one-year estimate. Notably, this persistence is confined to the first "dose" of the intervention: the second dose increased contemporaneous completion rates by 20.3 percentage points but had no detectable effect on subsequent screenings. Habit formation in this setting thus raises the cost-effectiveness of the first incentive but not the second, contrary to the predictions of reinforcement models, in which repeated participation amplifies the long-run effect of both incentives.

A large literature in economics documents habit formation. Many studies explicitly or implicitly feature reinforcement models, a classic case being cigarette addiction (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001). This class of models has also been applied to other settings including gym attendance, energy conservation, voting, hand washing, and social media use.[2] Models of switching costs and inertia can likewise be viewed as special cases of reinforcement models.[3] Other studies instead emphasize mechanisms resembling threshold models, notably experience-good models in which agents learn about a good's benefits through initial consumption. These models have been applied to brand persistence in food consumption, pharmaceutical demand, and health insurance enrollment.[4] In the simplest case, a single exposure to an experience good can generate persistently higher subsequent consumption (Dupas, 2014).

In most of these studies, reinforcement or threshold mechanisms are assumed rather than empirically distinguished, although observed patterns may sometimes be consistent with

[2]These include Charness and Gneezy (2009), Allcott and Rogers (2014), Fujiwara, Meng and Vogl (2016), Hussam et al. (2022), and Allcott, Gentzkow and Song (2022).

[3]These include, for example, Farrell (2007), Handel (2013), and Polyakova (2016).

[4]These include Ackerberg (2001), Ackerberg (2003), Crawford and Shum (2005), Dickstein (2021), and Banerjee et al. (2021).

either class, as noted by Hussam et al. (2022). A notable exception, and the paper closest to ours, is Carrera et al. (2019), who study an employer-provided gym membership and develop a model that generates either reinforcement or threshold dynamics depending on parameter values. Empirically, however, they find limited persistence following an eight-week incentive program, suggesting that exercise habits rarely become self-sustaining.

Our approach to studying habit formation differs from the existing literature in two key respects. First, we organize a wide range of previously studied mechanisms into two broad classes and show how recency and frequency can be used to characterize and distinguish them empirically. Second, we recast habit estimation as an instrumental variables (IV) problem. This brings the standard IV apparatus to bear, making transparent the variation required to identify recency and frequency effects—namely, valid instruments for early, late, and repeated actions.[5] This perspective also highlights the additional conditions required to interpret multivariate IV estimates as local average treatment effects when multiple endogenous treatments are present (Bhuller and Sigstad, 2024). More broadly, the framework offers a general strategy for uncovering the mechanisms of habit formation that extends naturally to non-experimental settings outside health. Quasi-experimental shocks could reveal, for example, whether voting habits strengthen with repetition, helping determine whether mobilization efforts should focus on converting first-time voters or on repeatedly contacting the same individuals over time (Gerber, Green and Shachar, 2003; Fujiwara, Meng and Vogl, 2016). Exogenous variation in purchases over time could likewise distinguish switching costs from experience-good learning in brand choice, helping determine whether firms compete to attract first-time buyers or instead profit from consumer lock-in (Ackerberg, 2003; Farrell, 2007).

Our paper also contributes to work on financial incentives for health behaviors by bringing insights from the habit formation literature to low-frequency behaviors. Most studies of habit formation in health behaviors focus on high-frequency activities, such as gym attendance, exercise, healthy eating, and daily hand-washing.[6] By contrast, the literature

[5]Fully specified structural models also seek to identify the effect of past behavior on current behavior, and may generate other distinguishing predictions, such as different patterns of decay over time (Ackerberg, 2001; Osborne, 2011).

[6]These include Volpp et al. (2008), Charness and Gneezy (2009), Acland and Levy (2015), Royer, Stehr and Sydnor (2015), Loewenstein, Price and Volpp (2016), Patel et al. (2016), Rohde and Verbeke (2017),

on incentives for low-frequency behaviors such as vaccinations, cancer and cardiovascular screenings, and workplace health screenings has largely overlooked the possibility of habit formation.[7] An exception is Schneider et al. (2023), who find that a $24 financial incentive for an initial COVID-19 vaccine dose in Sweden did not significantly affect uptake of subsequent doses. We study much larger incentives ($75–$200) for another low-frequency (annual) behavior, biometric health screenings, and find economically large and statistically significant persistence effects, showing that even infrequent health behaviors can become habitual.[8]

The remainder of the paper is organized as follows. Section 2 presents a general model of habit formation and shows how its predictions differ depending on the underlying mechanism. Section 3 provides background on our empirical setting and describes our data. Section 4 outlines our empirical strategy. Section 5 presents our results, and Section 6 concludes.

# 2 Model

This section develops a framework for distinguishing competing mechanisms of habit formation. Following the economics literature, we define habit formation as a positive causal effect of past consumption on current consumption (Becker, 1992; Royer, Stehr and Sydnor, 2015). We derive testable restrictions that separate two broad classes—reinforcement and threshold models—using experimental variation in early, late, and repeated actions. The key distinction is whether additional or more recent past actions continue to strengthen current demand after a habit has been established. In reinforcement models, past actions accumulate into a stock that places greater weight on more recent behavior. In threshold models, the stock depends only on whether prior actions have exceeded a critical level; once that level is reached, additional actions—regardless of their timing—have no further effect on demand.

These classes encompass leading explanations for behavioral persistence. The reinforce-

Carrera et al. (2018), Beshears et al. (2020), Carrera et al. (2019), Aggarwal, Dizon-Ross and Zucker (2022), and Hussam et al. (2022).

[7] These include Stone et al. (2002), Alsan, Garrick and Graziani (2019), Jones, Molitor and Reif (2019), Song and Baicker (2019), and Campos-Mercade et al. (2021).

[8] Even outside the realm of health, few studies have examined habit formation in infrequent behaviors. Two prominent exceptions include Gerber, Green and Shachar (2003) and Fujiwara, Meng and Vogl (2016), who study habit formation in voting.

ment class includes the canonical addiction model of Becker and Murphy (1988) and, as we show below, nests switching-cost and inertia models as special cases (Handel, 2013; Polyakova, 2016). The threshold class captures experience-good and learning models, as well as models with fixed entry costs, where behavior changes discretely once a critical level is reached (Ackerberg, 2003; Dupas, 2014; Carrera et al., 2019; Banerjee et al., 2021).[9]

We proceed in three steps. First, we define the two general classes of mechanisms—reinforcement and threshold—through restrictions on the habit stock. Second, we derive predictions from one parametric benchmark in each class: a geometrically depreciating addiction model for reinforcement and a zero-threshold model for threshold behavior. Third, we show that both benchmarks can be tested in a non-parametric, recency–frequency specification that relates current behavior to three mutually exclusive histories of past behavior: early-only, late-only, and repeated consumption.

## 2.1 Setup

In each period $t \leq T$, consumers choose consumption of a habit-forming good $a_t$, sold at price $p_t$, and a composite numeraire good $c_t$. Flow utility is given by $U(a_t, S_t, c_t, z_t)$, where $z_t$ is an exogenous utility shifter. Past consumption of the habit-forming good affects current utility through the habit stock $S_t$, which depends on past consumption and the initial stock:

$$S_t = g(a_{t-1}, a_{t-2}, \ldots, a_1; S_0) \tag{1}$$

Each period, consumers maximize utility subject to their budget constraint:

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

$$\text{s.t.} \quad w_t = c_t + p_t a_t, \tag{3}$$

[9] While these threshold mechanisms differ in their underlying microfoundations (e.g., some emphasize learning while others emphasize fixed costs), they generate the same predictions for screening completion in our experimental setting and are therefore observationally equivalent here.

where income $w_t$ is divided between the two goods. For simplicity, we assume consumers are myopic: they ignore the effect of current consumption on future utility through $S_{t+1}$.[10]

For tractability, we treat $a_t$ as continuous and assume a concave, quadratic utility function—a common specification in the addiction literature (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Reif, 2019).[11] As shown in Appendix A.1, solving the consumer's optimization problem (2) subject to the budget constraint (3) yields the following demand equation:

$$a_t = \kappa + \theta S_t + \pi p_t + \delta z_t, \tag{4}$$

where $\kappa$ is a constant, $\pi$ captures price sensitivity, and $\delta$ reflects the effect of $z_t$ on demand. The parameter of interest, $\theta$, captures the effect of past consumption on current demand: $\theta > 0$ implies habit formation, while $\theta < 0$ indicates satiation or fatigue. Appendix A.1 derives conditions for $\theta > 0$, which we take to be the relevant case in our empirical setting. Identifying $\theta$ requires exogenous variation in the stock $S_t$. Our randomized incentives provide this variation by shifting prior consumption.[12]

## 2.2 Reinforcement Models

We define reinforcement models as those where the habit stock, $S_t$, increases with past actions and places greater weight on more recent actions than on more distant ones. Formally:

**Definition 1 (Reinforcement)**

$$\frac{\partial S_t}{\partial a_{t-1}} > \frac{\partial S_t}{\partial a_{t-k}} \geq \frac{\partial S_t}{\partial a_{t-j}} \geq 0, \quad \text{for all } j > k > 1$$

This definition characterizes the reinforcement class. To obtain a parametric benchmark with sharp empirical restrictions, we specialize this class to the standard geometric depreciation

[10] In forward-looking models, current demand may also depend on expected future consumption. In our setting, however, future incentives were not known in advance, so consumers could not condition current choices on future incentive realizations. We therefore abstract from forward-looking behavior.

[11] Appendix A.2 presents a generalized reinforcement model that relaxes the quadratic utility assumption, while Appendix A.3 solves a discrete-choice version of the model that yields a similar estimating equation.

[12] Equation (4) is a Frisch demand equation, which holds the marginal utility of income constant. Although the randomized incentives are monetary, they represent less than 1% of annual income for the median participant, so any associated income effects are likely negligible.

specification from the addiction literature (Becker and Murphy, 1988):

$$S_t = (1 - d)S_{t-1} + a_{t-1}, \tag{5}$$

where $d \in (0, 1]$ is the rate of depreciation. Stable demand requires $\theta/d < 1$, which ensures that the effect of past consumption on demand does not grow without bound. When depreciation is complete ($d = 1$), the habit stock depends only on consumption in the immediately preceding period. In a discrete choice setting, this special case corresponds to a switching-cost or inertia model in which consumers incur a cost whenever their current action differs from that of the previous period (see, e.g., Handel, 2013; Polyakova, 2016).

Both waves of our empirical setting (Section 3) begin with no prior screening exposure, so $S_0 = 0$. Consider first a two-period setting ($T = 2$), as in Wave 2. Combining equations (4) and (5) yields:

$$a_2 = \kappa + \theta a_1 + \pi p_2 + \delta z_2. \tag{6}$$

Extending to three periods ($T = 3$)—as in Wave 1 of our application—and substituting recursively yields:[13]

$$a_3 = \kappa + \theta a_2 + \theta(1 - d)a_1 + \pi p_3 + \delta z_3. \tag{7}$$

This benchmark delivers two key empirical implications. First, *recency* matters: more recent consumption has a larger effect on current demand than more distant consumption, because the habit stock depreciates over time ($d > 0$). Second, *frequency* matters: repeated consumption further increases current demand, because past actions accumulate in the habit stock (provided $d < 1$).

## 2.3 Threshold Models

We define threshold models as those where the habit stock depends only on whether cumulative past consumption exceeds a critical level. Once past consumption exceeds the threshold $\underline{a} \geq 0$, the stock remains fixed and does not depreciate. Formally:

[13] Alternatively, third-period demand can be expressed solely in terms of current and past exogenous variables, yielding the "reduced-form" persistence model described in Section 4.1.

**Definition 2 (Threshold)**

$$S_t = \begin{cases} 1 & \text{if } \sum_{\tau=1}^{t-1} a_\tau > \underline{a}, \\ 0 & \text{otherwise.} \end{cases}$$

In a two-period setting with initial stock $S_0 = 0$, the threshold model yields the same qualitative prediction as the addiction model: first-period consumption raises second-period demand. The models diverge, however, in settings with at least three periods. Substituting the law of motion into the demand equation yields:

$$a_3 = \kappa + \theta \mathbf{1}(a_1 + a_2 > \underline{a}) + \pi p_3 + \delta z_3. \tag{8}$$

Two testable implications distinguish threshold behavior from reinforcement. First, *recency* does not matter: conditional on exceeding the threshold, consumption in period 1 and period 2 have identical effects on current demand. Second, *frequency* does not matter once the threshold is reached: additional past consumption beyond the threshold has no incremental effect on current demand.

This definition characterizes the threshold class. For the empirical benchmark, we focus on the zero-threshold case, in which one prior action is sufficient to establish the habit. This benchmark corresponds to experience-good models with rapid learning, where consumers learn their preferences after initial consumption (Erdem and Keane, 1996; Ackerberg, 2003; Israel, 2005), and to fixed-entry-cost models, where a one-time cost paid upon initial participation lowers the effective cost of subsequent participation.

## 2.4 Unified Recency–Frequency Specification

The reinforcement and threshold models generate distinct predictions for how third-period demand depends on the recency and frequency of past consumption. This subsection translates those predictions into a non-parametric specification saturated in prior screening histories. To match our empirical setting, we present a discrete version of the model in the main text and provide a formal derivation in Appendix A.3. Appendix A.4 develops a continuous version that yields qualitatively similar results.

Define four mutually exclusive histories of past consumption: no prior consumption, early-only consumption, late-only consumption, and repeated consumption. Let the no-prior-consumption group be the omitted category. We denote these categories as:

$$D_0 \equiv \mathbf{1}\{a_1 = 0,\ a_2 = 0\},$$
$$D_1 \equiv \mathbf{1}\{a_1 = 1,\ a_2 = 0\},$$
$$D_2 \equiv \mathbf{1}\{a_1 = 0,\ a_2 = 1\},$$
$$D_3 \equiv \mathbf{1}\{a_1 = 1,\ a_2 = 1\},$$

The corresponding treatment effects compare third-period demand for each history to third-period demand for the no-prior-consumption group:

$$\beta_j \equiv \mathbb{E}[a_3 \mid D_j = 1] - \mathbb{E}[a_3 \mid D_0 = 1], \qquad j \in \{1, 2, 3\} \tag{9}$$

The three coefficients have a direct interpretation: $\beta_1$ measures the effect of early-only consumption, $\beta_2$ the effect of late-only consumption, and $\beta_3$ the effect of repeated consumption, each relative to no prior consumption.

The reinforcement and threshold model classes impose distinct restrictions on these coefficients. Under reinforcement, later consumption should matter more than earlier consumption because the habit stock depreciates. Thus, reinforcement predicts a recency effect: $\beta_2 > \beta_1$. Reinforcement also predicts a frequency effect: $\beta_3 \geq \beta_2$ and $\beta_3 > \beta_1$. The standard addiction benchmark, which assumes geometric decay of the habit stock, delivers the same recency restriction and a sharper frequency restriction:

$$\beta_2 > \beta_1 \quad \text{(recency)}, \tag{10}$$
$$\beta_3 = \beta_1 + \beta_2 \quad \text{(frequency)}. \tag{11}$$

Under a threshold model, behavior depends only on whether prior consumption exceeds the threshold. Timing is therefore irrelevant, so there is no recency effect ($\beta_1 = \beta_2$). Whether a frequency effect arises depends on the threshold. In the zero-threshold benchmark ($\underline{a} = 0$),

any prior consumption suffices to establish the habit, so early-only, late-only, and repeated consumption all have the same effect on third-period demand:[14]

$$\beta_1 = \beta_2 = \beta_3 \quad \text{(no recency and no frequency)}. \tag{12}$$

# 3 Institutional Setting

The Illinois Workplace Wellness Study is a large-scale randomized controlled trial conducted at the University of Illinois Urbana-Champaign to evaluate workplace wellness programs (Jones, Molitor and Reif, 2019; Reif et al., 2020). We use its annually rerandomized incentives for biometric health screenings to study habit formation. The habit we study is completion of the on-campus biometric health screening, which the university offered for the first time as part of this wellness program. In 2016, 4,834 benefits-eligible employees were randomly assigned to one of two study waves. Wave 1 ($N = 3,300$) was invited to complete annual health screenings in 2016, 2017, and 2018 and was eligible for the broader workplace wellness program. Wave 2 ($N = 1,534$) was invited to complete health screenings only in 2017 and 2018 and was not eligible for the broader wellness program. We limit our analysis to subjects who remained continuously enrolled through the end of the intervention, yielding a final sample size of 4,799 (3,275 in Wave 1 and 1,524 in Wave 2).

The biometric health screening is the focal behavior of our analysis. Participants invited to a screening completed it on campus during the fall. At each screening, clinicians measured height, weight, waist circumference, and blood pressure, and administered a fingerstick blood test to assess cholesterol, triglyceride, and glucose levels. Results were provided within minutes and reviewed with a health coach.

Wave 1 members who completed a 2016 or 2017 health screening were also invited to complete an online health risk assessment (HRA), which made them eligible for additional wellness activities, including a walking program, weight management classes, and a tobacco cessation program. Overall, 34 percent of Wave 1 members completed at least one activity.

[14] With $\underline{a} = 1$, only individuals with $a_1 = a_2 = 1$ cross the threshold, implying $\beta_3 > 0$ and $\beta_1 = \beta_2 = 0$. With $\underline{a} \geq 2$, no possible consumption history crosses the threshold, so $\beta_1 = \beta_2 = \beta_3 = 0$. We show in Section 5.2 that both alternative thresholds are inconsistent with our data.

Wave 2 members were not offered the HRA or these activities. These additional program components are not the focus of our analysis, which centers on screening completion, but they are part of the broader wellness intervention available to Wave 1.

Figure 1 summarizes the experimental design. Financial incentives for screening completion were assigned several weeks before each screening sign-up period using stratified individual-level randomization (Jones, Molitor and Reif, 2019). The key design feature is that screening incentives were rerandomized each year. As a result, each year's incentive assignment was independent of previous incentives, prior screening decisions, and other past outcomes. In 2016, Wave 1 employees were assigned with equal probability to incentives of \$0, \$100, or \$200.[15] In 2017 and 2018, employees in both waves were assigned with equal probability to receive \$0 or \$125 (in 2017) and \$0 or \$75 (in 2018), conditional on completing a screening. Together, these annual rerandomizations constitute five distinct randomized experiments: three for Wave 1 (2016, 2017, 2018) and two for Wave 2 (2017, 2018). This repeated rerandomization generates exogenous variation in early, late, and repeated screening, providing exactly the empirical variation required by the framework in Section 2.

# 4 Empirical Strategy

Our empirical strategy proceeds in two steps. Section 4.1 estimates reduced-form effects of incentives on screening completion, distinguishing the "direct effect" (their impact on same-year completion) from the "persistence effect" (their impact on future completion). Section 4.2 then uses instrumental variables (IV) to estimate habit formation—the causal effect of past screening completion on current screening—using the randomized incentives as instruments for past screening, operationalizing the model classes derived in Section 2.

Both steps rely on the random assignment of incentives. To assess this assumption, we test whether each year's assigned incentive can be predicted by other years' incentives, baseline characteristics, and survey variables. Table 1 reports mean values of these variables

[15]The 2016 incentives were paid only to employees who completed both a health screening and an online HRA. Of the 1,888 participants in the analysis sample who completed a health screening, 1,838 also completed an online HRA. In 2017, HRA completion was not required to receive the assigned incentive, and in 2018 the HRA was not offered.

across treatment arms.[16] Joint balance tests for each of the five experiments, represented by pairs of adjacent columns (columns (1)–(2), (3)–(4), etc.), yield $p$-values of 0.29 or greater, indicating no systematic differences across treatment arms and supporting the validity of random assignment.

## 4.1 Reduced Form

We estimate the effects of incentives on screening completion using the following reduced-form specification:

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=1}^{T} \delta_\tau \text{INCENTIVE}_i^\tau + \gamma X_i + \epsilon_i. \tag{13}$$

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 variables of interest are the incentives $\text{INCENTIVE}_i^\tau$, indicators equal to 1 if individual $i$ was assigned a non-zero screening incentive in year $\tau$. We pool the \$100 and \$200 arms of the 2016 incentive into a single indicator, since these two arms produced similar direct effects on 2016 screening (Jones, Molitor and Reif, 2019). We estimate equation (13) separately for each wave and screening year, yielding five separate regressions (Figure 1).

The coefficient $\delta_\tau$ captures the average treatment effect of incentives assigned in year $\tau$, with three distinct interpretations depending on the relative timing of $t$ and $\tau$. When $t = \tau$, $\delta_\tau$ identifies the direct effect of incentives on contemporaneous screening. When $t > \tau$, it captures the persistence effect on future screening. When $t < \tau$, $\delta_\tau$ provides a falsification test: because future incentives were unknown to participants at the time of decision-making, the coefficient should equal zero.

Because multiple incentives enter the regression simultaneously, each $\delta_\tau$ represents a weighted average of treatment effects across different histories of prior incentives (Muralidharan, Romero and Wüthrich, 2023). This weighted average is policy relevant, as many real-world programs offer incentives repeatedly over time. In supplemental analyses, we

[16] The survey variable "ever screened" is a broad self-reported measure that indicates yes to any of several preventive screening questions, and is not the same as completion of the on-campus biometric screening used in our analysis.

extend equation (13) to allow for interactions across years.

Random assignment ensures unbiased estimation without controls, so our baseline specification omits $X_i$. To improve precision, we also report supplemental "post-Lasso" specifications that select controls using the double-selection method of Belloni, Chernozhukov and Hansen (2014), following the implementation in Jones, Molitor and Reif (2019). The set of candidate control variables includes baseline demographics, health survey responses, health behaviors, and claims-based measures of medical spending and usage, along with their pairwise interactions.[17] We report heteroskedasticity-robust standard errors (Abadie et al., 2023).

## 4.2 Habit Formation

We now turn to estimating habit formation—the causal effect of past screening completion on current screening completion—using past incentives as instruments for past screening. Section 2 derived testable restrictions that distinguish reinforcement from threshold models. The remainder of this subsection operationalizes those restrictions: Sections 4.2.1 and 4.2.2 specify estimating equations for each model class, Section 4.2.3 presents a unified, recency–frequency specification that nests both, and Section 4.2.4 discusses identification.

### 4.2.1 Reinforcement Model

To test the reinforcement benchmark, we estimate the addiction specification from Section 2.2, regressing current screening on up to two lags of past screening:

$$\text{SCREEN}_i^t = \kappa + \theta_1 \text{SCREEN}_i^{t-1} + \theta_2 \text{SCREEN}_i^{t-2} + \gamma X_i + \epsilon_i. \tag{14}$$

The control vector $X_i$ contains the contemporaneous incentive indicator, $\text{INCENTIVE}_i^t$, given its strong predictive power for screening completion. We also estimate a supplemental post-Lasso specification that adds the controls selected in Section 4.1 for the same screening outcome.

[17] Missing values are imputed using means or modes, and missingness indicators are included among the candidate controls. See Jones, Molitor and Reif (2019) for a detailed description of these control variables.

Because the waves differ in the number of screening years, and hence in the available lags, we estimate equation (14) separately by wave, yielding three regressions: for Wave 1, we regress 2017 screening on its one available lag and 2018 screening on both lags; for Wave 2, which was first invited in 2017, we regress 2018 screening on its one available lag.

The focal parameters $\theta_1$ and $\theta_2$ measure the effects of screening one and two years earlier, respectively. In the single lag regression, the key question is simply whether past screening increases current screening ($\theta_1 > 0$), that is, whether habit formation is present. With two lags, the regression further provides a direct test of reinforcement because the addiction model predicts a recency effect, with more recent screening exerting a larger effect than earlier screening: $\theta_1 > \theta_2$.

### 4.2.2 Threshold Model

To test the threshold benchmark, we estimate the zero-threshold specification from Section 2.3. This model yields two testable predictions, which we examine in turn: additional screenings beyond the first have no incremental effect (no frequency effect), and the timing of the initial screening is irrelevant (no recency effect).

We test the absence of a frequency effect by regressing 2018 screening on indicators for any prior screening ($\text{EVER}_i$) and for screening in both prior years ($\text{TWICE}_i$):

$$\text{SCREEN}_i^{2018} = \kappa + \theta_3 \text{EVER}_i + \theta_4 \text{TWICE}_i + \gamma X_i + \epsilon_i. \tag{15}$$

We test the absence of a recency effect by regressing 2018 screening on indicators for screening in 2016 ($\text{SCREEN}_i^{2016}$) and first-time screening in 2017 ($\text{ONLY}_i^{2017}$):[18]

$$\text{SCREEN}_i^{2018} = \kappa + \theta_5 \text{SCREEN}_i^{2016} + \theta_6 \text{ONLY}_i^{2017} + \gamma X_i + \epsilon_i. \tag{16}$$

In both equations, the control vector $X_i$ contains the contemporaneous (2018) incentive indicator, $\text{INCENTIVE}_i^{2018}$. We also estimate supplemental post-Lasso versions of both

[18]Technically, this test is best interpreted as a joint test of recency and frequency. Those who screen in 2016 can repeat in 2017, and thus $\theta_5$ could also capture frequency effects. Even so, it remains a valid test of the zero-threshold model, which implies neither recency nor frequency effects.

equations.

We estimate each equation separately by wave and, to gain statistical power, on the pooled Waves 1 and 2 sample, where $X_i$ additionally includes wave fixed effects and their interaction with the 2018 incentive indicator. For Wave 2, whose participants have $\text{SCREEN}_i^{2016} = 0$ by design, both specifications collapse to the single-lag version of equation (14), which can detect habit formation but cannot distinguish reinforcement from threshold mechanisms.

The focal parameters are $\theta_4$, $\theta_5$, and $\theta_6$. The threshold model predicts no frequency effect ($\theta_4 = 0$): conditional on having been screened at least once, a second prior screening yields no additional boost to subsequent demand. It likewise predicts no recency effect ($\theta_5 = \theta_6$): a first-time screening has the same effect on subsequent demand regardless of whether it occurs in 2016 or 2017.

### 4.2.3 Recency and Frequency

Finally, we estimate the unified three-period specification from Section 2.4, which nests both the reinforcement and threshold classes, regressing 2018 screening on mutually exclusive indicators for prior screening histories:

$$\text{SCREEN}_i^{2018} = \kappa + \beta_1 \text{ONLY}_i^{2016} + \beta_2 \text{ONLY}_i^{2017} + \beta_3 \text{TWICE}_i + \gamma X_i + \epsilon_i. \tag{17}$$

We estimate equation (17) on the pooled Waves 1 and 2 sample. The control vector $X_i$ is the same as in the pooled threshold specifications: the contemporaneous (2018) incentive indicator, wave fixed effects, and their interaction. We also estimate a supplemental post-Lasso version.

The focal parameters $\beta_1$, $\beta_2$, and $\beta_3$ measure the effects of screening only in 2016, only in 2017, and in both prior years. Because Wave 2 participants have $\text{SCREEN}_i^{2016} = 0$ by design, $\beta_1$ and $\beta_3$ are identified entirely from Wave 1, while $\beta_2$ is identified using variation from either wave.

These parameters, estimated from a single regression, provide a unified test of the reinforcement and threshold mechanisms, as derived in Section 2.4. Reinforcement implies a recency effect ($\beta_2 > \beta_1$) and a frequency effect ($\beta_3 > \beta_1$ and $\beta_3 \geq \beta_2$), with the addiction

model further sharpening the frequency restriction to $\beta_3 = \beta_1 + \beta_2$. By contrast, the zero-threshold model implies $\beta_1 = \beta_2 = \beta_3$: any prior screening is sufficient, and neither recency nor frequency matters for current demand.

#### 4.2.4 Identification

Ordinary least squares (OLS) estimation of equations (14)–(17) will be biased if past screening completion is correlated with unobserved determinants of current screening. This bias is likely positive, since persistent factors such as health preferences increase screening propensity across periods, though it could be negative if people prefer to space screenings over time.

To address this endogeneity, we instrument past screening completion using the randomly assigned incentives from prior years and estimate all models by two-stage least squares (2SLS). The instrument set varies across specifications. In the reinforcement specification (equation (14)), the endogenous regressors are lagged screening indicators, and we use the corresponding year-specific incentives as excluded instruments. In the threshold and unified specifications (equations (15), (16), and (17)), the endogenous regressors are nonlinear functions of past screening histories (e.g., EVER, TWICE), and we use the full set of prior randomized incentives and their interactions as instruments. In pooled specifications, we further interact instruments with wave fixed effects to allow first-stage relationships to differ across waves, reflecting the fact that the incentive regime differs between Wave 1 and Wave 2. We report first-stage $F$-statistics using the method of Sanderson and Windmeijer (2016), which provides separate weak-identification tests for each endogenous regressor.

The main identifying assumption is that incentives affect future screening only through their effect on past screening completion. We discuss potential violations of this exclusion restriction in Section 5.2. In models with a single endogenous treatment, the 2SLS estimate can be interpreted as a local average treatment effect (LATE) among compliers, i.e., individuals whose screening decisions respond to incentives, under the standard monotonicity assumption.

In models with multiple endogenous treatments, a LATE interpretation of 2SLS estimates requires more than the standard monotonicity assumption. Bhuller and Sigstad (2024) show

that multivariate 2SLS coefficients identify positively weighted averages of individual treatment effects if and only if average conditional monotonicity and no cross effects hold. Average conditional monotonicity requires that, holding the other predicted treatment fixed, instrument variation does not on average reduce uptake of a given treatment. No cross effects rules out heterogeneous cross-responses to the instruments. In our setting, for example, although the 2016 incentive may affect 2017 screening through habit formation, no cross effects requires that this spillover occur in the same relative way across individuals. While no cross effects is a non-trivial assumption, Bhuller and Sigstad (2024) derive observable implications of these identifying restrictions, allowing researchers to assess whether individuals respond to the instruments in a manner consistent with a multivariate LATE interpretation.

We test these observable implications in our data and find no evidence that the identifying restrictions underlying the multivariate 2SLS estimates are violated. In twelve subsamples defined by pre-determined characteristics, individuals take up the screening histories the instruments predict for them (supporting average conditional monotonicity) and no others (joint cross-effect $p$-values 0.38–0.99); an outcome-based version of the test likewise fails to reject ($p$-values 0.28 and 0.17).[19] Moreover, complier characteristics are similar across the three prior-screening history categories, and the overidentified specifications pass tests of overidentifying restrictions ($p$-values 0.34–0.78), consistent with stable effects across complier populations. Finally, our conclusions do not hinge on multi-treatment 2SLS: the reduced-form estimates in Table 2 and single-endogenous-regressor specifications estimated separately (Appendix Table A.5) require only standard single-treatment LATE assumptions and exhibit the same pattern—large, persistent effects of initial screening and no detectable incremental

[19]Let $D_1$, $D_2$, and $D_3$ denote the three mutually exclusive screening-history indicators in equation (17), and let $P_1$, $P_2$, and $P_3$ denote predicted histories, obtained by projecting each indicator on the instrument set in the full sample. If average conditional monotonicity and no cross effects both hold, then within any subsample defined by pre-determined characteristics, the regressions

$$D_k = \alpha_k + \sum_{l=1}^{3} \eta_{kl} P_l + \phi_k X + \varepsilon_k, \qquad k = 1, 2, 3,$$

where $X$ denotes the specification's exogenous controls, must yield non-negative own-treatment coefficients ($\eta_{kk} \geq 0$) and zero cross-treatment coefficients ($\eta_{kl} = 0$ for $l \neq k$) (Bhuller and Sigstad, 2024, Proposition 4). The full sample recovers $\eta_{kk} = 1$ and $\eta_{kl} = 0$ by construction, so only subsamples are informative; ours split the sample in two on each of six pre-determined characteristics: sex, age (50 and over), race (white), faculty status, salary (below median), and self-reported prior screening. The outcome-based test replaces the dependent variables with $YD_k$ and $(1-Y)D_k$, where $Y$ denotes 2018 screening; it is informative even in the full sample, where we implement it ($p = 0.28$ and 0.17, respectively).

effect of repeated screening.

# 5 Results

## 5.1 Reduced Form Effects of Incentives

We begin with reduced-form evidence on how financial incentives affect health screening completion both contemporaneously (direct effects) and in subsequent years (persistence effects). The central pattern is that the first incentive individuals are offered generates substantial persistence, whereas subsequent incentives do not. This pattern provides the first indication that habit formation in our setting operates through initial exposure rather than repeated reinforcement.

Figure 2 illustrates this pattern graphically. For Wave 1, the initial 2016 incentive increased screening completion from 49.4% to 61.8% (panel A). This effect persisted: screening rates in the treated group remained 4.5 percentage points higher in 2017 and 4.4 percentage points higher in 2018. A similar pattern appears in Wave 2. The initial 2017 incentive increased screening completion by 27.7 percentage points in that year and by 8.0 percentage points in 2018. By contrast, the second incentive offered to Wave 1 in 2017 had a large contemporaneous effect but little subsequent persistence. The falsification comparisons in panels B, C, and E show no meaningful differences in screening rates prior to incentive assignment, consistent with successful annual rerandomization.

Table 2 confirms these patterns in regression form, reporting estimates of equation (13) for each wave and screening year. The diagonal entries show large direct effects of incentives on contemporaneous screening completion. The upper-diagonal (bold-faced) entries show that persistence is concentrated in first-time incentives. Across the two waves, first-year incentives increased future screening completion by 4.4 to 8.9 percentage points, whereas the second incentive offered to Wave 1 had a small and statistically insignificant persistence effect of 1.2 percentage points. Thus, the reduced-form evidence points to a sharp asymmetry: incentives have durable effects when they induce initial participation, but not when they induce repeated participation.[20]

[20]Post-Lasso specifications yield similar results (Table A.1). Tables A.2 and A.3 re-estimate equation (13)

## 5.2 Habit Formation

We next interpret the persistence documented in Section 5.1 through the lens of habit formation. We proceed in three steps: we first test the reinforcement model, then the threshold model, and finally the unified specification that nests both.

### 5.2.1 Reinforcement Model

Table 3 reports estimates of the reinforcement model described in Section 2.2. Panel A presents IV estimates that instrument for past screening completion with randomly assigned past incentives. Panel B reports the corresponding OLS estimates.

Columns (1) and (3) of Panel A report single-lag IV estimates from equation (14). For Wave 1, completing the 2016 screening raised the probability of completing the 2017 screening by 36.0 percentage points, a 90 percent increase relative to the control-group rate of 39.9 percent (Figure 2). For Wave 2, completing the 2017 screening raised the probability of completing the 2018 screening by 32.4 percentage points, essentially doubling the control-group rate of 32.5 percent. Both effects are statistically significant ($p < 0.01$) and similar in magnitude, a notable replication given that the waves differed in the timing and conditions of their first screening invitations.

Column (2) of Panel A exploits the three-period structure of Wave 1 to test the recency prediction of the reinforcement model implied by equation (7): more recent consumption should have a larger effect on current demand than more distant consumption. Specifically, completing the 2017 screening should have a larger effect on 2018 completion than completing the 2016 screening, i.e., $\theta_1 > \theta_2$.

The estimates strongly contradict this prediction. Completing the 2016 screening increases the probability of completing the 2018 screening by 33.1 percentage points ($p < 0.01$), whereas completing the 2017 screening increases it by only 6.3 percentage points and is statistically insignificant ($p = 0.39$). The more recent effect is not only smaller but also statistically distinguishable from the earlier effect: the 95 percent confidence interval for the

for 2018 screening completion in the Wave 1 sample, adding interactions that allow each incentive's effect to vary with assignments in other years (Muralidharan, Romero and Wüthrich, 2023). The interaction terms are statistically insignificant, though imprecisely estimated, and the main effects remain qualitatively stable, continuing to support a persistent effect of the 2016 incentive but not the 2017 incentive.

2017 effect, $[-8.0, 20.5]$, does not contain the 2016 point estimate, and a Wald test rejects equality of the two coefficients at the 10 percent level ($p = 0.088$). These results imply that habit formation is driven primarily by the initial screening rather than by reinforcement from repeated participation.

Panel B of Table 3 reports the corresponding OLS estimates. All are positive and statistically significant, but OLS is upward biased because unobserved determinants of screening participation—such as health preferences—are persistent over time. As a result, OLS yields the misleading conclusion that more recent screening has a stronger effect than earlier screening, a result directly contradicted by the IV estimates.

### 5.2.2 Threshold Model

Table 4 reports estimates of the threshold model described in Section 2.3. All specifications use 2018 screening completion as the outcome. Panel A presents IV estimates, and Panel B reports the corresponding OLS estimates.

Columns (1)–(3) of Panel A estimate a restricted version of equation (15) that includes only an indicator for having ever completed a screening prior to 2018. For Wave 1, Wave 2, and the pooled sample, having ever been screened increases the probability of completing the 2018 screening by 32.4 to 36.1 percentage points, an increase of 100–111 percent relative to the 32.5 percent screening rate among participants never assigned a prior-year incentive, with all estimates statistically significant at the 1 percent level. These results indicate strong habit formation following initial exposure but do not, on their own, distinguish between the reinforcement and threshold mechanisms. For Wave 2, which has only two screening opportunities, this specification coincides with the single-lag model in Table 3.

Columns (4) and (5) of Panel A relax the restriction that additional screenings have no further effect by separately estimating the incremental effect of completing a second screening (equation (15)), providing a direct test of frequency. In both the Wave 1 and pooled specifications, this incremental effect is small and statistically insignificant. Conditional on at least one prior screening, completing a second screening does not increase the probability of screening in 2018. This finding is consistent with the threshold model's no-frequency prediction that habit formation operates through initial exposure rather than repeated rein-

forcement.

Columns (6)–(8) test the model's no-recency implication by distinguishing between first-time screening in 2016 and first-time screening in 2017 (equation (16)). In the pooled sample (column 8), the estimated effects are similar in magnitude, and we cannot reject equality between them ($p = 0.41$), consistent with the model's predictions. Column (6), which uses only Wave 1 data, is less precise because most Wave 1 participants screened in 2017 had also been screened in 2016. Since Wave 2 has no 2016 screening by design, the specification in column (7) collapses to a single-lag model and recovers the same habit-formation parameter estimated in column (2). Column (6), which uses only Wave 1 data, is less precise because most participants who screened in 2017 had also screened in 2016, leaving few first-time screeners in that year. Panel B reports the corresponding OLS estimates. Unlike the IV estimates, OLS implies that both having ever been screened and having been screened twice have large and statistically significant effects on 2018 completion (columns (4) and (5)). The threshold model's no-frequency prediction—that additional screenings provide no incremental benefit conditional on prior screening—is therefore rejected under OLS. This contrast with the IV estimates underscores the importance of accounting for endogeneity in past screening behavior.

### 5.2.3 Recency and Frequency

Table 5 reports estimates from the unified, three-period model in equation (17), which relates 2018 screening completion to mutually exclusive prior screening histories. This specification nests the reinforcement and threshold models and allows us to test their restrictions directly within a common framework.

Column (1) reports IV estimates. All three coefficients are positive and statistically significant, indicating substantial habit formation regardless of whether prior screening occurred only in 2016, only in 2017, or in both years. The estimated effects are 44.0 percentage points for screening only in 2016, 31.6 percentage points for screening only in 2017, and 40.3 percentage points for screening in both years.

These estimates permit direct tests of the competing models. Reinforcement models predict that more recent and more frequent past screening should have larger effects on

current behavior, implying $\beta_3 \geq \beta_2 > \beta_1$. The IV estimates do not support this pattern: the coefficients are similar in magnitude and do not exhibit the ordering implied by recency or frequency effects. A sharper prediction arises from the canonical addiction model, which implies additivity, $\beta_3 = \beta_1 + \beta_2$. This restriction is also rejected ($p = 0.017$), indicating that repeated participation does not generate additive reinforcement effects. By contrast, the threshold model predicts that all three coefficients are equal ($\beta_1 = \beta_2 = \beta_3$). We fail to reject this restriction in the IV estimates ($p = 0.678$), consistent with habit formation driven by initial exposure rather than reinforcement.

Column (2) reports the corresponding OLS estimates. Although all three coefficients remain positive and statistically significant, the implied pattern differs sharply from the IV results. Under OLS, the threshold model is strongly rejected ($p < 0.001$), and the ordering of the coefficients follows the pattern predicted by reinforcement models ($\beta_3 \geq \beta_2 > \beta_1$). This pattern suggests that serial correlation in unobserved determinants of screening behavior biases OLS estimates upward, producing spurious evidence of reinforcement.

Overall, the unified, recency–frequency framework consolidates the evidence from the preceding analyses: once past screening completion is instrumented using randomized incentives, the data reject reinforcement and support a threshold model in which initial exposure increases subsequent participation.

## 5.3 Robustness

### 5.3.1 Alternative Explanations for Reduced Form Patterns

We consider alternative interpretations of the persistence patterns documented above. First, the weak persistence of the 2017 incentive in Wave 1 could reflect an unusually poor screening experience in that year. However, this explanation is difficult to reconcile with the positive persistence effect of the same 2017 screening on Wave 2 subjects. Second, the 2017 incentive might have had a weaker direct effect in Wave 1, mechanically limiting persistence. But the contemporaneous effect of the 2017 incentive in Wave 1 was not smaller than that of the 2016 incentive, which did generate substantial persistence. Taken together, these patterns support a conclusion that persistence in this setting is driven primarily by initial exposure.

### 5.3.2 Assessing the Exclusion Restriction

For persistence to reflect habit formation, the exclusion restriction requires that financial incentives affect future screening only through their effect on prior screening completion. A potential concern is that subjects may not have fully understood or paid attention to changes in their incentives across years. For example, individuals assigned a high incentive in 2016 may have expected this incentive to always persist. In that case, incentives could affect future screening through mistaken beliefs rather than through past screening completion, violating the exclusion restriction.

To investigate that possibility, we consider a null hypothesis under which persistence is driven entirely by inattention to changes in incentives across years, with beliefs corrected by experiencing the 2017 payment. Under this null, the effect of the 2016 incentive on 2018 completion should be smaller in the subsample assigned to the $0 incentive in 2017, since individuals who attended the 2017 screening expecting a high payment would have learned they were mistaken, thereby reducing their likelihood of returning in 2018.[21] We therefore estimate equation (13) for 2018 screening, omitting the 2017 incentive indicator, on the full Wave 1 sample and on this subsample.

Table A.6 reports the results. Column (1) shows that the 2016 incentive raised the 2018 completion rate by 4.4 percentage points in the full sample. Restricting the sample to those assigned the $0 incentive in 2017 (column (2)) *raises* the point estimate to 5.4 percentage points, providing no evidence that confusion about incentive payments was driving our persistence estimates. Estimates from specifications that include post-Lasso controls are similar (Table A.7). Thus, we conclude that inattention does not explain our results.

A related concern is that the Wave 1 persistence estimates may reflect not only screening completion, but also the bundled HRA and broader wellness activities for which Wave 1 members were eligible upon screening completion. Wave 2 provides a way to assess this interpretation: Wave 2 members were not offered the HRA or the additional wellness activities. Yet the Wave 2 first-screening incentive still generates a comparable persistence effect into 2018, suggesting that the habit-formation result is not driven by the Wave 1 bundle.

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

### 5.4 Discussion

In our setting, overlooking habit formation significantly underestimates the effectiveness of financial incentives. Consider the initial (2016) incentive offered to Wave 1. Absent habit formation, one would conclude that this incentive increased completion rates by 12.4 percentage points (or 407 additional screenings, had all 3,275 Wave 1 participants been incentivized) for one year, as shown in column (1) of Table 2. Accounting for persistence nearly doubles this estimate: the cumulative effect across all three screenings is a 21.3 percentage point increase (or 697 people), a 71% increase relative to the one-year estimate. These estimates capture only the habit-formation effect observed over the three-year program period. Because persistence shows little decay over this period, the long-run impact of the initial incentive would likely be larger if the program were extended.

However, this amplification is confined to initial exposure. A second incentive generates large contemporaneous responses but had no detectable effect on future screening behavior. As a result, while ignoring habit formation leads to substantial underestimation of the returns to initial incentives, it is unlikely to similarly bias estimates of repeated incentives. This pattern contrasts sharply with reinforcement models, which predict that habit formation should amplify the long-run effects of both initial and subsequent incentives.

## 6 Conclusion

We show that even infrequent health behaviors, such as annual health screenings, can become strongly habitual. Completing a health screening for the first time raises the probability of completing a future screening by 32.4–36.1 percentage points—an effect that replicates across two waves—but additional screenings do not further reinforce this habit. This pattern is inconsistent with reinforcement models in which more recent or more frequent past behavior progressively strengthens current demand. Instead, it supports a threshold model in which a single exposure is sufficient to establish and sustain a habit.

Our empirical design disentangles these mechanisms by combining multiple rounds of rerandomized incentives with an IV approach that treats past screening as endogenous. This design allows us to separately identify the effects of initial and subsequent screenings,

and thereby to test competing models of habit formation. Prior studies usually assume one mechanism rather than identify it: much of the existing literature imposes either a reinforcement or threshold-like mechanism, while our design provides direct tests between them. The setting also broadens the evidence on habit formation in health behavior, which has largely focused on high-frequency activities such as exercise, healthy eating, and hand washing. We show that even low-frequency behaviors can become habitual.

From a policy perspective, our results suggest that, in the case of annual health screenings, incentives aimed at inducing initial participation may have larger and more durable effects than repeated incentives targeting experienced individuals. At the same time, our findings identify the form of the habit more clearly than its precise microfoundations. One plausible interpretation is experiential learning: individuals may update beliefs about the costs or benefits of screening after participating once. However, other mechanisms, such as fixed start-up costs that do not depend on beliefs, could generate similar dynamics. Distinguishing among these microfoundations—for example by measuring beliefs directly and tracing how they evolve with experience—is an important direction for future research. Such evidence would clarify why a single exposure is sufficient to establish a lasting habit and would help determine when initial incentives can generate persistent behavior change.

# 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.

**Beshears, John, Hae Nim Lee, Katherine L. Milkman, Robert Mislavsky, and Jessica Wisdom.** 2020. "Creating Exercise Habits Using Incentives: The Trade-off Between Flexibility and Routinization." *Management Science*, 67(7): 4139–4171.

**Bhuller, Manudeep, and Henrik Sigstad.** 2024. "2SLS with multiple treatments." *Journal of Econometrics*, 242(1): 105785.

**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.** 2019. "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.

**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.

**Gerber, Alan S, Donald P Green, and Ron Shachar.** 2003. "Voting may be habit-forming: Evidence from a randomized field experiment." *American Journal of Political Science*, 47(3): 540–550.

**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.

**Hintermann, Beat, and Andreas Lange.** 2013. "Learning abatement costs: On the dynamics of the optimal regulation of experience goods." *Journal of environmental economics and management*, 66(3): 625–638.

**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.

**Klemperer, Paul.** 1987. "Markets with consumer switching costs." *The quarterly journal of economics*, 102(2): 375–394.

**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.

**Pollak, Robert A.** 1970. "Habit Formation and Dynamic Demand Functions." *Journal of Political Economy*, 78(4, Part 1): 745–763.

**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.

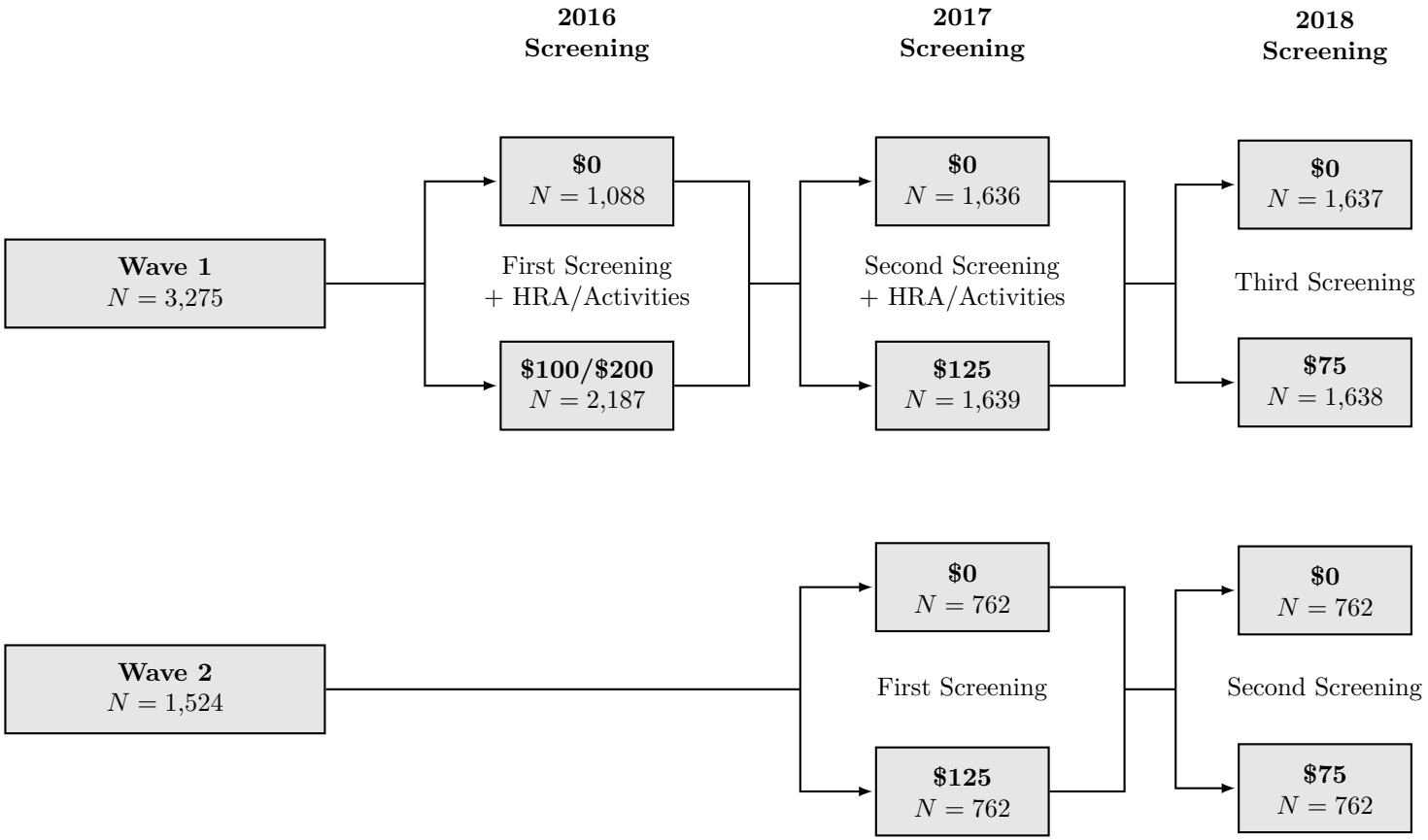
**Schneider, Florian H., Pol Campos-Mercade, Stephan Meier, Devin Pope, Erik Wengström, and Armando N. Meier.** 2023. "Financial incentives for vaccination do not have negative unintended consequences." *Nature*, 613(7944): 526–533.

**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.

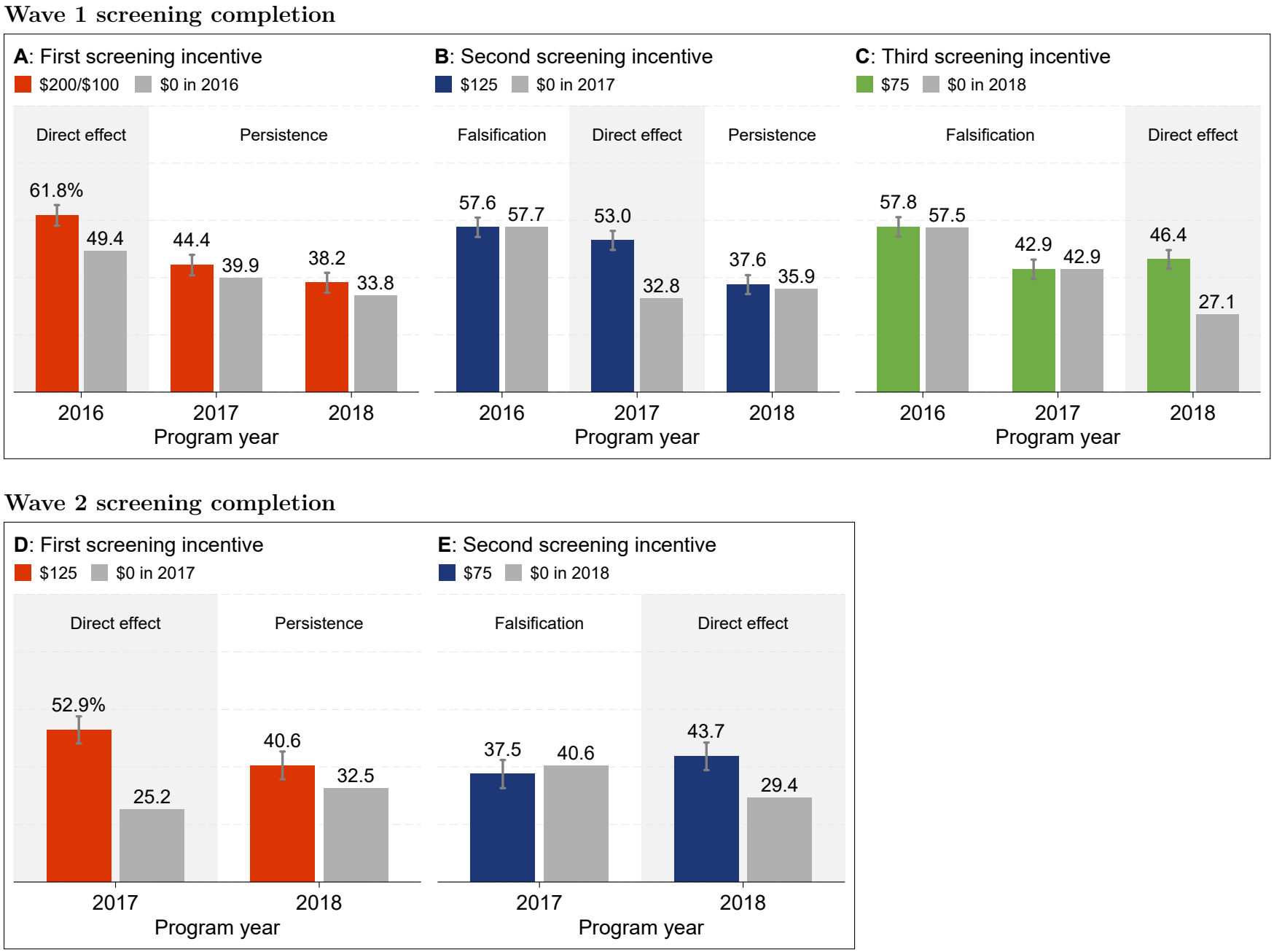
**Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein.** 2008. "Financial incentive-based approaches for weight loss: A randomized trial." *Journal of the American Medical Association*, 300: 2631–2637.

**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. Incentives were re-randomized annually.

**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 | | | Wave 2 | |
| | 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 $\delta_\tau$ from equation (13). 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 (direct) 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. Estimates from post-Lasso versions of these specifications are reported in Table A.1. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

**Table 3:** Reinforcement Model Estimates

| | (1) | (2) | (3) |
|---|---|---|---|
| | Wave 1 | | Wave 2 |
| | Screened 2017 | Screened 2018 | Screened 2018 |
| **A. IV estimates** | | | |
| Screened 2016 | $\theta_1$: 0.360** | $\theta_2$: 0.331** | |
| | (0.128) | (0.127) | |
| Screened 2017 | | $\theta_1$: 0.063 | $\theta_1$: 0.324** |
| | | (0.073) | (0.075) |
| First-stage $F$ (Screened 2016) | 45.6 | 48.1 | |
| First-stage $F$ (Screened 2017) | | 182.5 | 131.6 |
| Mean outcome | 0.429 | 0.368 | 0.365 |
| Sample size | 3,275 | 3,275 | 1,524 |
| **B. OLS estimates** | | | |
| Screened 2016 | $\theta_1$: 0.467** | $\theta_2$: 0.183** | |
| | (0.014) | (0.016) | |
| Screened 2017 | | $\theta_1$: 0.430** | $\theta_1$: 0.552** |
| | | (0.017) | (0.022) |
| Mean outcome | 0.429 | 0.368 | 0.365 |
| Sample size | 3,275 | 3,275 | 1,524 |

Notes: This table reports IV and OLS estimates of $\theta_1$ and $\theta_2$ from equation (14). The dependent variable is an indicator for whether the 2017 (or 2018) screening was completed. Reinforcement models like addiction predict $\theta_1 > \theta_2$. All regressions control for the contemporaneous incentive indicator. Estimates from post-Lasso versions of these specifications are reported in Table A.4. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

**Table 4:** Threshold Model Estimates

| | (1) Wave 1 | (2) Wave 2 | (3) Waves 1 & 2 | (4) Wave 1 | (5) Waves 1 & 2 | (6) Wave 1 | (7) Wave 2 | (8) Waves 1 & 2 |
|---|---|---|---|---|---|---|---|---|
| **A. IV estimates** | | | | | | | | |
| Screened ever | 0.361** | 0.324** | 0.334** | 0.432* | 0.337** | | | |
| | (0.130) | (0.075) | (0.065) | (0.216) | (0.071) | | | |
| Screened twice | | | | –0.065 | –0.011 | | | |
| | | | | (0.158) | (0.104) | | | |
| Screened 2016 | | | | | | 0.389** | | 0.419** |
| | | | | | | (0.133) | | (0.123) |
| Screened 2017 only | | | | | | 0.180 | 0.324** | 0.308** |
| | | | | | | (0.224) | (0.075) | (0.072) |
| First-stage $F$ (endog. 1) | 14.9 | 131.6 | 44.1 | 11.8 | 51.7 | 22.3 | | 60.0 |
| First-stage $F$ (endog. 2) | | | | 17.0 | 71.8 | 29.6 | 131.6 | 63.6 |
| Mean outcome | 0.368 | 0.365 | 0.367 | 0.368 | 0.367 | 0.368 | 0.365 | 0.367 |
| Sample size | 3,275 | 1,524 | 4,799 | 3,275 | 4,799 | 3,275 | 1,524 | 4,799 |
| **B. OLS estimates** | | | | | | | | |
| Screened ever | 0.444** | 0.552** | 0.480** | 0.229** | 0.362** | | | |
| | (0.013) | (0.022) | (0.011) | (0.017) | (0.014) | | | |
| Screened twice | | | | 0.384** | 0.310** | | | |
| | | | | (0.019) | (0.018) | | | |
| Screened 2016 | | | | | | 0.449** | | 0.464** |
| | | | | | | (0.014) | | (0.014) |
| Screened 2017 only | | | | | | 0.408** | 0.552** | 0.503** |
| | | | | | | (0.033) | (0.022) | (0.018) |
| Mean outcome | 0.368 | 0.365 | 0.367 | 0.368 | 0.367 | 0.368 | 0.365 | 0.367 |
| Sample size | 3,275 | 1,524 | 4,799 | 3,275 | 4,799 | 3,275 | 1,524 | 4,799 |

Notes: This table reports IV and OLS estimates of equations (15) and (16). The dependent variable is an indicator for whether the 2018 screening was completed. Columns (1)–(3) estimate the effect of having ever been screened prior to 2018. Columns (4)–(5) distinguish between being screened once and being screened twice. Columns (6)–(8) separately estimate the effects of screening in 2016 and screening in 2017 only. Under the zero-threshold model, screening demand increases after an initial screening, but subsequent screenings do not further increase demand. All regressions control for the contemporaneous (2018) incentive indicator; pooled specifications additionally include wave fixed effects and their interaction with the incentive. Estimates from post-Lasso versions of these specifications are reported in Table A.8. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level. The first-stage F statistics correspond to the first and second endogenous regressors in each specification.

**Table 5:** Recency and Frequency Estimates

| | (1) | (2) |
|---|---|---|
| | IV | OLS |
| Screened 2016 only | $\beta_1$: 0.440** | $\beta_1$: 0.188** |
| | (0.139) | (0.018) |
| Screened 2017 only | $\beta_2$: 0.316** | $\beta_2$: 0.503** |
| | (0.075) | (0.018) |
| Screened twice | $\beta_3$: 0.403** | $\beta_3$: 0.628** |
| | (0.133) | (0.015) |
| Tests of model restrictions ($p$-values) | | |
| Reinforcement: $\beta_1 + \beta_2 = \beta_3$ | 0.017 | 0.021 |
| Threshold: $\beta_1 = \beta_2 = \beta_3$ | 0.678 | 0.000 |
| First-stage $F$ (Screened 2016 only) | 92.2 | |
| First-stage $F$ (Screened 2017 only) | 70.3 | |
| First-stage $F$ (Screened twice) | 91.3 | |
| Mean outcome | 0.367 | 0.367 |
| Sample size | 4,799 | 4,799 |

Notes: This table reports IV and OLS estimates of $\beta_1$, $\beta_2$, and $\beta_3$ from equation (17). Recency effects arise when $\beta_2 > \beta_1$, and frequency effects arise when $\beta_3 \geq \beta_2$ and $\beta_3 > \beta_1$. The dependent variable is an indicator for whether the 2018 screening was completed. All regressions control for the contemporaneous (2018) incentive indicator, interacted with wave. The model restrictions correspond to a reinforcement model with geometric depreciation (e.g., addiction) and a zero-threshold model (e.g., experience goods). Estimates from post-Lasso versions of these specifications are reported in Table A.9. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

# Online Appendix

"How Do Habits Form? Experimental Evidence from Health Screenings"

Damon Jones, David Molitor, and Julian Reif

# A Habit Formation Model Derivations

## A.1 Deriving the Habit-Demand Equation

This subsection derives the demand equation (4) presented in Section 2.1 and shows how the habit stock enters current demand in reduced form. Following Reif (2019), we solve the consumer's optimization problem assuming that utility is concave and quadratic:

$$\begin{aligned} U(a_t, S_t, c_t, z_t) = &-\frac{1}{2}\Big(u_{aa}a_t^2 + u_{ss}S_t^2 + u_{cc}c_t^2 + u_{zz}z_t^2\Big) + u_{as}a_tS_t + u_{ac}a_tc_t + u_{az}a_tz_t \\ &+ u_{sc}S_tc_t + u_{sz}S_tz_t + u_{cz}c_tz_t + u_aa_t + u_sS_t + u_cc_t + u_zz_t \end{aligned} \tag{A.1}$$

Habit formation is driven by the parameter $u_{as}$, which captures the strength of intertemporal complementarity and is assumed to be positive. Assuming agents are myopic, we have the following utility maximization problem:

$$\begin{aligned} \max_{a_t, c_t} \quad & U(a_t, S_t, c_t, z_t) \\ \text{s.t.} \quad & w_t = c_t + p_ta_t \end{aligned} \tag{A.2}$$

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

$$c_t = \frac{-\lambda + u_c + u_{ac}a_t + u_{sc}S_t + u_{cz}z_t}{u_{cc}} \tag{A.3}$$

where $\lambda$ is the marginal utility of income. Substituting (A.3) into the utility function (A.1) yields a reduced-form utility that depends only on $a_t$, $S_t$, and $z_t$:

$$U^*(a_t, S_t, z_t) = -\frac{1}{2}\Big(b_{aa}a_t^2 + b_{ss}S_t^2 + b_{zz}z_t^2\Big) + b_{as}a_tS_t + b_{az}a_tz_t + b_{sz}S_tz_t + b_aa_t + b_sS_t + b_zz_t + b_k \tag{A.4}$$

where each $b$-parameter is a function of the original $u$-parameters. In particular,

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

The first-order condition for $a_t$, together with (A.3), yields demand equation (4):

$$a_t = \kappa + \theta S_t + \pi p_t + \delta z_t$$

where:

$$\kappa = \frac{b_a}{b_{aa}}$$
$$\theta = \frac{b_{as}}{b_{aa}} > 0$$
$$\pi = \frac{-\lambda}{b_{aa}} < 0$$
$$\delta = \frac{b_{az}}{b_{aa}}$$

The parameter $\theta$ captures the strength of habit formation by measuring the effect of the habit stock $S_t$ on current demand. A sufficient condition for $\theta > 0$ is $b_{as} > 0$, which holds if $u_{ac}u_{sc} \geq 0$; one simple special case is additive separability in $a_t$ and $c_t$.

## A.2 A General Reinforcement Model with Continuous $a_t$

We now relax the benchmark addiction model to show that its key recency and frequency implications do not depend on quadratic utility or geometric depreciation. We assume quasi-linear utility together with two maintained assumptions on preferences: current consumption and the habit stock are complements ($\partial^2 u/\partial a \partial S > 0$), and utility is concave in current consumption ($\partial^2 u/\partial a^2 < 0$).[1]

$$\max_{a_t, c_t} U(a_t, S_t, c_t, z_t) = u(a_t, S_t, z_t) + c_t$$
$$\text{s.t. } w_t = c_t + p_t a_t$$

The first-order conditions are:

$$\frac{\partial u}{\partial a} = p$$
$$c = w - pa$$

To derive the effect of past consumption on current consumption, differentiate the first-order condition with respect to $S_t$:

$$\frac{\partial^2 u}{\partial a^2}\frac{\partial a_t}{\partial S_t} + \frac{\partial^2 u}{\partial a \partial S} = 0$$

[1] Quasilinearity allows us to abstract from income effects and is closely related to the setup in the previous subsection, since that analysis likewise focused on a Frisch demand function that holds the marginal utility of income constant.

so that:

$$\begin{aligned}\frac{\partial a_t}{\partial S_t} &= -\frac{\partial^2 u/\partial a \partial S}{\partial^2 u/\partial a^2} \\ &> 0\end{aligned}$$

where the inequality follows from the maintained assumptions that $\partial^2 u/\partial a \partial S > 0$ and $\partial^2 u/\partial a^2 < 0$.

Next, differentiate the first-order condition with respect to $a_{t-k}$:

$$\frac{\partial^2 u}{\partial a^2}\frac{\partial a_t}{\partial a_{t-k}} + \frac{\partial^2 u}{\partial a \partial S}\frac{\partial S_t}{\partial a_{t-k}} = 0$$

which implies:

$$\begin{aligned}\frac{\partial a_t}{\partial a_{t-k}} &= -\frac{\partial^2 u/\partial a \partial S}{\partial^2 u/\partial a^2}\frac{\partial S_t}{\partial a_{t-k}} \\ &= \frac{\partial a_t}{\partial S_t}\frac{\partial S_t}{\partial a_{t-k}}\end{aligned}$$

Since $\frac{\partial a_t}{\partial S_t} > 0$, Definition 1 implies that more recent actions have a larger effect on current demand than more distant actions. In particular:

$$\frac{\partial a_t}{\partial a_{t-1}} > \frac{\partial a_t}{\partial a_{t-2}},$$

which is the continuous analogue of the recency prediction, $\beta_2 > \beta_1$, in the discrete three-period setting.

A natural continuous analogue of completing both prior actions is to increase both lagged actions by a common amount. Define:

$$\beta_3^* \equiv \frac{da_t}{d\varepsilon}, \qquad a_{t-1} \mapsto a_{t-1} + \varepsilon,\ a_{t-2} \mapsto a_{t-2} + \varepsilon.$$

By the chain rule:

$$\beta_3^* = \frac{\partial a_t}{\partial a_{t-1}} + \frac{\partial a_t}{\partial a_{t-2}}$$

Since $\frac{\partial a_t}{\partial a_{t-1}} > 0$ and $\frac{\partial a_t}{\partial a_{t-2}} \geq 0$, it follows that:

$$\begin{aligned}\beta_3^* &\geq \frac{\partial a_t}{\partial a_{t-1}}, \\ \beta_3^* &> \frac{\partial a_t}{\partial a_{t-2}},\end{aligned}$$

which are the continuous analogs of the frequency predictions, $\beta_3 \geq \beta_2$ and $\beta_3 > \beta_1$, in the discrete three-period setting.

## A.3 Binary Choice and the Recency/Frequency Restrictions

Sections 2.2 and 2.3 present the recency and frequency restrictions implied by the reinforcement and threshold models for continuous consumption. This subsection considers the binary choice case and derives the discrete testable restrictions presented in Section 2.4. Let $a_t \in \{0, 1\}$ be a binary choice made every period. We assume utility takes a quasilinear form and augment it with an additional random utility shock, $v_t$, associated with choosing $a_t = 1$:

$$\begin{aligned}\tilde{U}(a_t, S_t, c_t, z_t) =& u(a_t, S_t, z_t) + c_t - v_t a_t \\ =& u(a_t, S_t, z_t) + w_t - p_t a_t - v_t a_t\end{aligned}$$

where in the second line, we've substituted for consumption using the budget constraint $c_t + p_t a_t = w_t$. The random shock $v_t$ has a CDF $F_v(\cdot)$. In each period, a myopic agent will choose $a_t = 1$ if the decision yields higher utility:

$$\begin{aligned}a_t &= \mathbf{1}\Big(\tilde{U}(1, S_t, w - p_t, z_t) - \tilde{U}(0, S_t, w_t, z_t) > 0\Big) \\ &= \mathbf{1}(u(1, S_t, z_t) - u(0, S_t, z_t) - p_t > v_t) \\ &= \mathbf{1}(\triangle u(S_t) - p_t > v_t)\end{aligned}$$

where we've defined $\triangle u(S_t) \equiv u(1, S_t, z_t) - u(0, S_t, z_t)$. It follows that the probability of choosing $a_t = 1$ is:

$$\begin{aligned}\Pr(a_t = 1) &= \mathbb{E}[\mathbf{1}(\triangle u(S_t) - p_t > v_t)] \\ &= F_v(\triangle u(S_t) - p_t)\end{aligned} \tag{A.5}$$

We proceed in two steps. First, we derive qualitative restrictions under weak assumptions on preferences. We then impose additional structure to obtain a sharper additivity restriction.

As a first step, we assume that $a_t$ and $S_t$ are complementary ($\partial^2 u / \partial a \partial S > 0$), which in turn implies that $u(a_t, S_t, z_t)$ has an increasing differences property:

$$\triangle u(S_t') > \triangle u(S_t), \quad \text{for } S_t' > S_t$$

Following the main text, define mutually exclusive indicators for past consumption histories:

$$\begin{aligned}D_0 &\equiv \mathbf{1}\{a_1 = 0,\ a_2 = 0\} \\ D_1 &\equiv \mathbf{1}\{a_1 = 1,\ a_2 = 0\} \\ D_2 &\equiv \mathbf{1}\{a_1 = 0,\ a_2 = 1\} \\ D_3 &\equiv \mathbf{1}\{a_1 = 1,\ a_2 = 1\}.\end{aligned}$$

Define corresponding treatment effects as:

$$\begin{aligned}\beta_j &\equiv \mathbb{E}[a_3 \mid D_j = 1] - \mathbb{E}[a_3 \mid D_0 = 1] \\ &= F_v(\triangle u(S_3^j) - p_t) - F_v(\triangle u(0) - p_t)\end{aligned}, \qquad j \in \{1, 2, 3\} \tag{A.6}$$

where $S_3^j$ is the stock conditional on $D_j = 1$. Assume that, for each $j \in \{0, 1, 2, 3\}$, $(p_3, z_3)$

is identically distributed across the $\{D_j\}$ cells and that the utility shock $v_3$ has the same distribution $F_v$ across cells.[2]

Under the addiction model, the stock evolves according to:

$$S_t = (1-d)S_{t-1} + a_{t-1}$$

The corresponding treatment effects satisfy:

$$\begin{aligned}
\beta_1 =& F_v(\triangle u(1-d) - p_t) - F_v(\triangle u(0) - p_t) \\
\beta_2 =& F_v(\triangle u(1) - p_t) - F_v(\triangle u(0) - p_t) \\
\beta_3 =& F_v(\triangle u(2-d) - p_t) - F_v(\triangle u(0) - p_t)
\end{aligned}$$

Given the monotonicity of $F_v(\cdot)$, the addiction model implies $\beta_3 \geq \beta_2 > \beta_1$, which embeds both the recency prediction ($\beta_2 > \beta_1$) and the frequency predictions ($\beta_3 \geq \beta_2$ and $\beta_3 > \beta_1$).

Under the zero-threshold model ($\underline{a} = 0$), the stock evolves according to:

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

The corresponding treatment effects satisfy:

$$\beta_1 = \beta_2 = \beta_3 = F_v(\triangle u(1) - p_t) - F_v(\triangle u(0) - p_t)$$

Thus, the threshold model exhibits neither frequency nor recency effects.

If we place additional structure on the $u(\cdot)$ and $F_v(\cdot)$ functions, we obtain a sharper prediction under the addiction model. We make the following two assumptions:

$$\begin{aligned}
&\text{Linear } \triangle u : \triangle u(S) = \triangle u(0) + kS \\
&\text{Linear } F_v : F_v(x + \triangle u(0) - p_t) = F_v(\triangle u(0) - p_t) + bx
\end{aligned}$$

Under these assumptions, the addiction model implies:

$$\begin{aligned}
\beta_1 &= F_v(\triangle u(1-d) - p_t) - F_v(\triangle u(0) - p_t) \\
&= (1-d)bk \\
\beta_2 &= F_v(\triangle u(1) - p_t) - F_v(\triangle u(0) - p_t) \\
&= bk \\
\beta_3 &= F_v(\triangle u(2-d) - p_t) - F_v(\triangle u(0) - p_t) \\
&= (2-d)bk \\
&= \beta_1 + \beta_2
\end{aligned}$$

[2]For identification, it is sufficient that $(p_3, z_3)$ be uncorrelated with the instruments $Z$. The assumption of identical distributions across $\{D_j\}$ cells is imposed only for notational simplicity when deriving the algebraic relationships in this section.

Thus, under the additional linearity assumptions, the addiction model satisfies

$$\beta_3 = \beta_1 + \beta_2$$

which is the additivity restriction stated in Section 2.4.

## A.4 Continuous Analogues of the Recency/Frequency Model

Section 2.4 presents the unified recency-frequency framework for binary consumption. This subsection derives the continuous analogues of the addiction and zero-threshold models and shows that they imply the same qualitative restrictions as the binary case. We begin from the addiction demand equation (7):

$$a_3 = \kappa + \theta a_2 + \theta(1-d)a_1 + \pi p_3 + \delta z_3,$$

where $d \in (0,1]$ is the depreciation rate. Define the mutually exclusive and exhaustive indicators:

$$\begin{aligned} D_0 &\equiv \mathbf{1}\{a_1 = 0,\ a_2 = 0\}, \\ D_1 &\equiv \mathbf{1}\{a_1 > 0,\ a_2 = 0\}, \\ D_2 &\equiv \mathbf{1}\{a_1 = 0,\ a_2 > 0\}, \\ D_3 &\equiv \mathbf{1}\{a_1 > 0,\ a_2 > 0\}. \end{aligned}$$

Consider the saturated "unified" projection:

$$a_3 = \kappa + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \pi p_3 + \delta z_3 + u_3,$$

and assume that, for each $j \in \{0,1,2,3\}$, $(p_3, z_3)$ is identically distributed across the $\{D_j\}$ cells and $\mathbb{E}[u_3 \mid D_j] = 0$.[3] Because the specification is saturated in $\{D_j\}$, each coefficient equals a conditional mean difference:

$$\beta_j = \mathbb{E}[a_3 \mid D_j] - \mathbb{E}[a_3 \mid D_0], \qquad j \in \{1,2,3\}. \tag{A.7}$$

Let $\eta \equiv \kappa + \pi\mathbb{E}[p_3] + \delta\mathbb{E}[z_3]$. Taking conditional expectations of the addiction equation and using the invariance of $(p_3, z_3)$ across cells yields:

$$\begin{aligned} \mathbb{E}[a_3 \mid D_0] &= \eta, \\ \mathbb{E}[a_3 \mid D_1] &= \eta + \theta(1-d)\mathbb{E}[a_1 \mid D_1], \\ \mathbb{E}[a_3 \mid D_2] &= \eta + \theta\mathbb{E}[a_2 \mid D_2], \\ \mathbb{E}[a_3 \mid D_3] &= \eta + \theta\mathbb{E}[a_2 \mid D_3] + \theta(1-d)\mathbb{E}[a_1 \mid D_3]. \end{aligned}$$

[3] For identification, it is sufficient that $(p_3, z_3)$ be uncorrelated with the instruments $Z$. The assumption of identical distributions across $\{D_j\}$ cells is imposed only for notational simplicity when deriving the algebraic relationships in this section.

Here, $\mathbb{E}[a_i \mid D_j]$ denotes the mean of consumption conditional on prior use pattern $D_j$. Plugging these into the definition of $\beta_j$ given by (A.7) yields:

$$\beta_1 = \theta(1-d)\mathbb{E}[a_1 \mid D_1], \tag{A.8}$$
$$\beta_2 = \theta\mathbb{E}[a_2 \mid D_2], \tag{A.9}$$
$$\beta_3 = \theta\mathbb{E}[a_2 \mid D_3] + \theta(1-d)\mathbb{E}[a_1 \mid D_3]. \tag{A.10}$$

Solving (A.8) and (A.9) for $\theta$ and substituting into (A.10) yields the continuous analogue of the frequency restriction:

$$\beta_3 = \beta_1 \frac{\mathbb{E}[a_1 \mid D_3]}{\mathbb{E}[a_1 \mid D_1]} + \beta_2 \frac{\mathbb{E}[a_2 \mid D_3]}{\mathbb{E}[a_2 \mid D_2]}. \tag{A.11}$$

A second implication concerns recency. For $d < 1$, from (A.8) and (A.9):

$$\frac{\beta_2/\mathbb{E}[a_2 \mid D_2]}{\beta_1/\mathbb{E}[a_1 \mid D_1]} = \frac{\theta}{\theta(1-d)} = \frac{1}{1-d} > 1,$$

which rearranges to the continuous analogue of the recency restriction:

$$\beta_2 > \beta_1 \frac{\mathbb{E}[a_2 \mid D_2]}{\mathbb{E}[a_1 \mid D_1]}. \tag{A.12}$$

When $d = 1$, we have $\beta_1 = 0$, so the recency ordering $\beta_2 > \beta_1$ holds trivially.

For completeness, we can also derive the continuous analogue of the zero-threshold model. In that model, we have the following expression for consumption in period 3 from equation (8):

$$a_3 = \kappa + \theta\mathbf{1}(a_1 + a_2 > 0) + \pi p_3 + \delta z_3. \tag{A.13}$$

In this three-period setting where any strictly positive prior consumption is sufficient to activate the habit stock, we have:

$$\begin{aligned}
\mathbb{E}[a_3 \mid D_0] &= \eta, \\
\mathbb{E}[a_3 \mid D_1] &= \eta + \theta, \\
\mathbb{E}[a_3 \mid D_2] &= \eta + \theta, \\
\mathbb{E}[a_3 \mid D_3] &= \eta + \theta,
\end{aligned}$$

and therefore:

$$\beta_1 = \beta_2 = \beta_3 = \theta.$$

Thus, the continuous zero-threshold model delivers the same no-recency and no-frequency restrictions as in the binary case.

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

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Wave 1 | | | Wave 2 | |
| | 2016 | 2017 | 2018 | 2017 | 2018 |
| 2016 incentive ($200/$100) | 0.131** | **0.048**** | **0.043*** | | |
| | (0.017) | **(0.017)** | **(0.017)** | | |
| 2017 incentive ($125) | –0.000 | 0.200** | **0.012** | 0.276** | **0.092**** |
| | (0.016) | (0.016) | **(0.016)** | (0.023) | **(0.024)** |
| 2018 incentive ($75) | –0.005 | –0.006 | 0.190** | –0.013 | 0.153** |
| | (0.016) | (0.016) | (0.016) | (0.023) | (0.024) |
| Number of controls | 56 | 55 | 35 | 55 | 35 |
| 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 $\delta_\tau$ from equation (13). 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 (direct) 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. The corresponding baseline estimates are reported in Table 2. 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 $\delta_\tau$ from versions of equation (13) that add interactions between the incentive assignments, allowing each incentive's effect to vary with assignments in other years. The dependent variable is an indicator for whether a screening was completed in 2018, and the sample is Wave 1. Column (1) replicates column (3) of Table 2; columns (2)–(4) each add one pairwise interaction, and column (5) includes all pairwise interactions and the three-way interaction. Estimates from post-Lasso versions of these specifications are reported in Table A.3. 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) | 0.043* | 0.045 | 0.042 | 0.043* | 0.015 |
| | (0.017) | (0.023) | (0.023) | (0.017) | (0.031) |
| 2017 incentive ($125) | 0.012 | 0.014 | 0.012 | 0.032 | –0.005 |
| | (0.016) | (0.027) | (0.016) | (0.022) | (0.036) |
| 2018 incentive ($75) | 0.190** | 0.190** | 0.189** | 0.211** | 0.169** |
| | (0.016) | (0.016) | (0.027) | (0.022) | (0.038) |
| 2016 incentive × 2017 incentive | | –0.004 | | | 0.056 |
| | | (0.034) | | | (0.045) |
| 2016 incentive × 2018 incentive | | | 0.002 | | 0.062 |
| | | | (0.034) | | (0.047) |
| 2017 incentive × 2018 incentive | | | | –0.041 | 0.039 |
| | | | | (0.032) | (0.054) |
| 2016 incentive × 2017 incentive × 2018 incentive | | | | | –0.120 |
| | | | | | (0.067) |
| Number of controls | 35 | 35 | 35 | 35 | 35 |
| 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 $\delta_\tau$ from versions of equation (13) that add interactions between the incentive assignments, allowing each incentive's effect to vary with assignments in other years. The dependent variable is an indicator for whether a screening was completed in 2018, and the sample is Wave 1. Column (1) replicates column (3) of Table A.1; columns (2)–(4) each add one pairwise interaction, and column (5) includes all pairwise interactions and the three-way interaction. 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. The corresponding baseline estimates are reported in Table A.2. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

**Table A.4:** Reinforcement Model Estimates, Post-Lasso Controls

| | (1) | (2) | (3) |
|---|---|---|---|
| | Wave 1 | | Wave 2 |
| | Screened 2017 | Screened 2018 | Screened 2018 |
| **A. IV estimates** | | | |
| Screened 2016 | $\theta_1$: 0.374** | $\theta_2$: 0.322** | |
| | (0.122) | (0.124) | |
| Screened 2017 | | $\theta_1$: 0.060 | $\theta_1$: 0.323** |
| | | (0.071) | (0.071) |
| Number of controls | 55 | 35 | 35 |
| First-stage $F$ (Screened 2016) | 51.1 | 51.4 | |
| First-stage $F$ (Screened 2017) | | 190.6 | 144.1 |
| Mean outcome | 0.429 | 0.368 | 0.365 |
| Sample size | 3,275 | 3,275 | 1,524 |
| **B. OLS estimates** | | | |
| Screened 2016 | $\theta_1$: 0.419** | $\theta_2$: 0.168** | |
| | (0.016) | (0.016) | |
| Screened 2017 | | $\theta_1$: 0.408** | $\theta_1$: 0.522** |
| | | (0.018) | (0.023) |
| Number of controls | 55 | 35 | 35 |
| Mean outcome | 0.429 | 0.368 | 0.365 |
| Sample size | 3,275 | 3,275 | 1,524 |

Notes: This table reports IV and OLS estimates of $\theta_1$ and $\theta_2$ from equation (14). The dependent variable is an indicator for whether the 2017 (or 2018) screening was completed. Reinforcement models like addiction predict $\theta_1 > \theta_2$. 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 the contemporaneous incentive indicator. The corresponding baseline estimates are reported in Table 3. 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: Joint versus Separate Estimation

| | (1) | (2) | (3) |
|---|---|---|---|
| **A. IV estimates** | | | |
| Screened 2016 | 0.331** | 0.354** | |
| | (0.127) | (0.129) | |
| Screened 2017 | 0.063 | | 0.061 |
| | (0.073) | | (0.079) |
| First-stage $F$ (Screened 2016) | 48.1 | 45.6 | |
| First-stage $F$ (Screened 2017) | 182.5 | | 143.1 |
| Mean outcome | 0.368 | 0.368 | 0.368 |
| Sample size | 3,275 | 3,275 | 3,275 |
| **B. OLS estimates** | | | |
| Screened 2016 | 0.183** | 0.383** | |
| | (0.016) | (0.015) | |
| Screened 2017 | 0.430** | | 0.515** |
| | (0.017) | | (0.015) |
| Mean outcome | 0.368 | 0.368 | 0.368 |
| Sample size | 3,275 | 3,275 | 3,275 |

Notes: This table reports IV and OLS estimates of the effects of past screening completion on 2018 screening completion in the Wave 1 sample. Column (1) replicates column (2) of Table 3, which includes both screening lags, instrumented by the 2016 and 2017 incentive indicators. Columns (2) and (3) include one lag at a time, each instrumented by its own year's incentive indicator; the estimate in column (2) therefore includes any effect of 2016 screening operating through 2017 screening. All regressions control for the contemporaneous incentive indicator. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

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

| | (1) | (2) |
|---|---|---|
| | Full sample | Subsample |
| 2016 incentive ($200/$100) | 0.044* | 0.054* |
| | (0.017) | (0.024) |
| Mean outcome | 0.368 | 0.359 |
| Sample size | 3,275 | 1,636 |

Notes: This table reports estimates of $\delta_1$ from equation (13), omitting the 2017 incentive indicator. The dependent variable is an indicator for completing the 2018 health screening. Column (1) includes the full Wave 1 sample. Column (2) restricts to Wave 1 subjects who were assigned the $0 incentive in 2017. All regressions control for the contemporaneous (2018) incentive indicator. Estimates from post-Lasso versions of these specifications are reported in Table A.7. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

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

| | (1) | (2) |
|---|---|---|
| | Full sample | Subsample |
| 2016 incentive ($200/$100) | 0.043* | 0.045 |
| | (0.017) | (0.023) |
| Number of controls | 35 | 35 |
| Mean outcome | 0.368 | 0.359 |
| Sample size | 3,275 | 1,636 |

Notes: This table reports estimates of $\delta_1$ from equation (13), omitting the 2017 incentive indicator. The dependent variable is an indicator for completing the 2018 health screening. Column (1) includes the full Wave 1 sample. Column (2) restricts to Wave 1 subjects who were assigned the $0 incentive in 2017. All regressions control for the contemporaneous (2018) incentive indicator. 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. The corresponding baseline estimates are reported in Table A.6. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

**Table A.8:** Threshold Model Estimates, Post-Lasso Controls

| | (1) Wave 1 | (2) Wave 2 | (3) Waves 1 & 2 | (4) Wave 1 | (5) Waves 1 & 2 | (6) Wave 1 | (7) Wave 2 | (8) Waves 1 & 2 |
|---|---|---|---|---|---|---|---|---|
| **A. IV estimates** | | | | | | | | |
| Screened ever | 0.342** | 0.323** | 0.331** | 0.378 | 0.335** | | | |
| | (0.129) | (0.071) | (0.063) | (0.215) | (0.069) | | | |
| Screened twice | | | | –0.032 | –0.014 | | | |
| | | | | (0.153) | (0.101) | | | |
| Screened 2016 | | | | | | 0.368** | | 0.401** |
| | | | | | | (0.132) | | (0.121) |
| Screened 2017 only | | | | | | 0.162 | 0.323** | 0.310** |
| | | | | | | (0.223) | (0.071) | (0.069) |
| Number of controls | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 |
| First-stage $F$ (endog. 1) | 15.9 | 144.1 | 47.4 | 11.5 | 54.3 | 23.7 | | 24.4 |
| First-stage $F$ (endog. 2) | | | | 16.2 | 62.2 | 29.7 | 144.1 | 66.8 |
| Mean outcome | 0.368 | 0.365 | 0.367 | 0.368 | 0.367 | 0.368 | 0.365 | 0.367 |
| Sample size | 3,275 | 1,524 | 4,799 | 3,275 | 4,799 | 3,275 | 1,524 | 4,799 |
| **B. OLS estimates** | | | | | | | | |
| Screened ever | 0.405** | 0.522** | 0.444** | 0.210** | 0.338** | | | |
| | (0.014) | (0.023) | (0.012) | (0.017) | (0.014) | | | |
| Screened twice | | | | 0.365** | 0.292** | | | |
| | | | | (0.020) | (0.018) | | | |
| Screened 2016 | | | | | | 0.409** | | 0.426** |
| | | | | | | (0.015) | | (0.014) |
| Screened 2017 only | | | | | | 0.370** | 0.522** | 0.469** |
| | | | | | | (0.032) | (0.023) | (0.018) |
| Number of controls | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 |
| Mean outcome | 0.368 | 0.365 | 0.367 | 0.368 | 0.367 | 0.368 | 0.365 | 0.367 |
| Sample size | 3,275 | 1,524 | 4,799 | 3,275 | 4,799 | 3,275 | 1,524 | 4,799 |

Notes: This table reports IV and OLS estimates of equations (15) and (16). The dependent variable is an indicator for whether the 2018 screening was completed. Columns (1)–(3) estimate the effect of having ever been screened prior to 2018. Columns (4)–(5) distinguish between being screened once and being screened twice. Columns (6)–(8) separately estimate the effects of screening in 2016 and screening in 2017 only. Under the zero-threshold model, screening demand increases after an initial screening, but subsequent screenings do not further increase demand. 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 the contemporaneous (2018) incentive indicator; pooled specifications additionally include wave fixed effects and their interaction with the incentive. The corresponding baseline estimates are reported in Table 4. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level. The first-stage F statistics correspond to the first and second endogenous regressors in each specification.

**Table A.9:** Recency and Frequency Estimates, Post-Lasso Controls

| | (1) | (2) |
|---|---|---|
| | IV | OLS |
| Screened 2016 only | $\beta_1$: 0.422** | $\beta_1$: 0.170** |
| | (0.137) | (0.018) |
| Screened 2017 only | $\beta_2$: 0.317** | $\beta_2$: 0.476** |
| | (0.072) | (0.018) |
| Screened twice | $\beta_3$: 0.386** | $\beta_3$: 0.590** |
| | (0.130) | (0.016) |
| Tests of model restrictions ($p$-values) | | |
| Reinforcement: $\beta_1 + \beta_2 = \beta_3$ | 0.014 | 0.039 |
| Threshold: $\beta_1 = \beta_2 = \beta_3$ | 0.747 | 0.000 |
| First-stage $F$ (Screened 2016 only) | 64.5 | |
| First-stage $F$ (Screened 2017 only) | 73.2 | |
| First-stage $F$ (Screened twice) | 27.7 | |
| Number of controls | 35 | 35 |
| Mean outcome | 0.367 | 0.367 |
| Sample size | 4,799 | 4,799 |

Notes: This table reports IV and OLS estimates of $\beta_1$, $\beta_2$, and $\beta_3$ from equation (17). Recency effects arise when $\beta_2 > \beta_1$, and frequency effects arise when $\beta_3 \geq \beta_2$ and $\beta_3 > \beta_1$. The dependent variable is an indicator for whether the 2018 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 the contemporaneous (2018) incentive indicator, interacted with wave. The model restrictions correspond to a reinforcement model with geometric depreciation (e.g., addiction) and a zero-threshold model (e.g., experience goods). The corresponding baseline estimates are reported in Table 5. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.