How Do Habits Form? Experimental Evidence from Health Screenings*

Damon Jones
University of Chicago

David Molitor
University of Illinois
and NBER

Julian Reif
University of Illinois
and NBER

February 2026

PRELIMINARY AND INCOMPLETE

Abstract

We develop a framework to identify the mechanisms of habit formation and apply it to experimental data on annual health screenings from a three-year study of 4,799 university employees. We find strong habit formation from initial exposure: completing the first screening raised subsequent screening rates by 32.4–36.0 percentage points (84%–90%) in the second and third years. In contrast, completing a second screening had minimal effect on subsequent screenings. This pattern contradicts reinforcement mechanisms like addiction, where recent consumption drives the habit, and instead supports a threshold model where consumers learn the value of screening through first exposure. Experimental and observational estimates yield markedly different conclusions, underscoring the value of research design for shaping inferences about the sources of habit formation.

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

*This research was supported by the National Institute on Aging of the National Institutes of Health under award number R01AG050701; the National Science Foundation under Grant No. 1730546; the Abdul Latif Jameel Poverty Action Lab (J-PAL) North America U.S. Health Care Delivery Initiative; Evidence for Action (E4A), a program of the Robert Wood Johnson Foundation; and the W.E. Upjohn Institute for Employment Research. This study was preregistered with the American Economics Association RCT Registry (AEARCTR-0001368) and was approved by the Institutional Review Boards of the National Bureau of Economic Research, University of Illinois Urbana-Champaign, and University of Chicago. We gratefully acknowledge co-investigator Laura Payne for vital contributions to the study, Lauren Geary for project management, Michele Guerra for programmatic support, Illinois Human Resources for institutional support, and Nina Kerkebane for research assistance. We are also thankful for comments from Christopher Cronin, Tatyana Deryugina, Riley League, Sarah Miller, Lena Song, and seminar participants at the 114$^{\text{th}}$ Annual Conference on Taxation, the Bozeman Applied Economics Summer Conference, the Chicago Booth Junior Health Economics Summit, the Mid-Midwest Applied Microeconomics Workshop, the Midwest Health Economics Conference, and Toulouse School of Economics. The findings and conclusions expressed are solely those of the authors and do not necessarily represent the views of the National Institutes of Health, any of our funders, or the University of Illinois. An earlier version of this paper was titled "Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study."

1 Introduction

Habits, which emerge when past consumption increases current consumption, shape many social, economic, and health behaviors (Pollak, 1970; Becker, 1992). Distinguishing habit formation from other sources of behavioral persistence, such as fixed preferences or persistent economic conditions, poses a significant empirical challenge. Existing studies often address this challenge by leveraging exogenous variation in past behavior, such as through randomized financial incentives (e.g., Carrera et al., 2020; Hussam et al., 2022). However, these studies are typically not designed to disentangle the underlying mechanisms of habit formation. As a result, the observed habit is often attributed to a specific model, such as addiction or learning, without directly testing competing explanations.

While habit formation can be modeled in many different ways, a central distinction is whether habits are driven more by recent past consumption or by initial consumption. For example, addictive habits like cigarette smoking depend strongly on recent consumption, whereas developing a taste for mangoes might require only a single initial experience to establish a lasting habit. Understanding the specific mechanisms through which habits form and persist is crucial for designing effective taxes and incentives (Becker, Grossman and Murphy, 1994; Hintermann and Lange, 2013).

Our study develops a framework to distinguish between two broad classes of habit formation models: *reinforcement*, in which habits are primarily driven by recent consumption, and *threshold models*, where habits depend mainly on initial consumption. The reinforcement class includes frameworks such as rational addiction (Becker and Murphy, 1988) and transaction costs (Klemperer, 1987), while the threshold class includes experience-good and learning models (Dupas, 2014; Banerjee et al., 2021). In a two-period setting, these classes are observationally equivalent, since initial and most recent consumption are one and the same. We show, however, that in settings with more than two time periods these two classes can be distinguished empirically, since initial and subsequent consumption can be independently identified and their relative effects on current demand directly compared.[1]

[1] A large literature has examined whether specific habit formation mechanisms—most notably addiction—reflect forward-looking behavior, i.e., whether consumers anticipate future consequences when making current decisions. While not the focus of our paper, we note that identifying forward-looking behavior also requires variation across multiple time periods to distinguish anticipatory responses from myopic ones.

We use our framework to study habit formation in annual biometric health screenings, a widely promoted health behavior encouraged by employers and health insurance programs.[2] As part of a large-scale randomized controlled trial, we offered individual-level financial incentives to 3,275 university employees ("Wave 1") each year for three years (2016–2018), paying those who completed a health screening. Incentives were rerandomized and announced at the start of each program year, generating variation across both initial and subsequent participation. We use this design to perform two complementary analyses. First, we conduct a reduced-form analysis that estimates contemporaneous ("direct") effects of incentives on screening completion as well as their persistence effects up to two years later. Second, we use the randomized incentives as instruments to estimate the causal effect of past screening completion on current completion. Because three years of randomization provide variation in both initial and recent screening behaviors, we can directly compare their relative influence on current screening completion. We further validate our findings among a second cohort of 1,524 employees ("Wave 2") who received randomized incentives for two years (2017–2018). The staggered and overlapping design of the two waves enables within-year comparisons of individuals with and without prior exposure, providing direct evidence on the role of initial versus recent screening experience in habit formation.

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

Next, we use our randomized incentives as instruments to estimate the causal effect of past screening completion on current screening completion.[3] We find strong evidence of habit

[2]To our knowledge, no prior study has examined habit formation in the context of biometric health screenings. Related work has documented habit formation in health behaviors such as exercise, bed net usage, and insurance enrollment, where it has been attributed to addiction-like dynamics or experience-good mechanisms (Charness and Gneezy, 2009; Dupas, 2014; Banerjee et al., 2021).

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

formation from initial participation. Completing the first screening raised the likelihood of screening completion one year later by 32.4–36.0 percentage points (84–89%) in Waves 1 and 2, and raised completion two years later by 33.1 percentage points (90%) in Wave 1. The consistency of the one-year estimates across waves underscores the robustness of the first-time screening effect, while its stability over a two-year horizon indicates that the effect does not quickly decay. By contrast, completing the second screening raised third-year completion rates by only 6.3 percentage points, a small and statistically insignificant effect.

These patterns distinguish between the two classes of habit formation models introduced earlier. Our IV estimates show that the more recent second-year screening had a significantly smaller effect on Wave 1's likelihood of completing the third screening than the more distant first-year screening, rejecting a core prediction of the reinforcement class. Instead, the findings align with several predictions of the threshold class of models: conditional on having been screened once, being screened twice does not further increase the probability of future participation; and the impact of completing a first screening is nearly identical whether it occurred one or two years earlier.

Finally, we demonstrate how understanding the mechanisms of habit formation can inform the design of cost-effective interventions. Consider the initial incentive to complete a screening in the first year of Wave 1. Ignoring habit formation, one would conclude that the incentive increased completion rates by 12.4 percentage points for a single year. Accounting for habit formation, however, increases the total effect to 21.3 percentage points over the two-year program period—72% higher than the one-year estimate. Notably, this habit formation effect 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. In other words, habit formation enhanced the cost-effectiveness of the first incentive but not the second. This finding contrasts with predictions from models such as addiction, where reinforcement would amplify the long-run effect of both incentives. More broadly, these results highlight the importance of not only identifying habit formation but also distinguishing among its underlying mechanisms when designing interventions.

Our study builds on a large literature documenting habit formation across a wide range of behaviors, including smoking, voting, energy conservation, and social-media use. A cen-

tral challenge in this work has been distinguishing habit formation from other sources of behavioral persistence (Brandon et al., 2017; Moore and Morris, 2024). Most studies assume rather than identify the underlying mechanism. (Hussam et al., 2022), for instance, interpret persistence in handwashing as an addiction but acknowledge that it may also reflect learning. Other studies have attributed habit formation to reinforcement (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Allcott and Rogers, 2014; Fujiwara, Meng and Vogl, 2016; Allcott, Gentzkow and Song, 2022), inertia (Handel, 2013), environmental cues (Giuntella, Saccardo and Sadoff, 2024), start-up costs (Carrera et al., 2020), switching costs (Polyakova, 2016), and learning through experience (Ackerberg, 2003; Dupas, 2014; Banerjee et al., 2021). A smaller set of studies has attempted to distinguish between different consumer learning mechanisms using structural methods (Ackerberg, 2001; Osborne, 2011).

Our paper advances this literature by introducing a randomized framework to distinguish two broad classes of habit-formation mechanisms, reinforcement and threshold models, which encompass common models such as addiction and experience goods. A key innovation is the use of multiple rounds of rerandomized incentives to isolate the different persistence effects of repeated treatments. This design allows us to identify habit formation—the casual effect of past consumption on current consumption—while ruling out alternatives such as inattention or serially correlation preferences. It also enables a direct comparison of the effects of initial versus subsequent consumption on current demand, thereby distinguishing between the two mechanism classes. Beyond our specific setting, this experimental framework provides a general template for studying other models of behavioral persistence. For example, in industrial organization, switching costs and learning are often cited as important determinants of brand choice (Ackerberg, 2003; Farrell, 2007). In a typical switching cost model, current consumption depends on the immediately preceding period's consumption, whereas in an experience-good model, it depends on earlier consumption when learning is fast. By employing multiple rounds of rerandomized incentives, researchers can empirically separate these mechanisms without relying exclusively on structural assumptions.

We also contribute to the literature on habit formation in health behaviors, which has predominantly focused on high-frequency activities such as exercising.[4] A pioneering study

[4]Even outside the realm of health, few studies have examined habit formation in infrequent behaviors. A

by Charness and Gneezy (2009) finds that financial incentives for attending the gym for one month increased attendance for at least seven weeks after the intervention ended. More recent studies on exercise generally find weaker results: some fail to detect habit formation effects from simple incentives (Royer, Stehr and Sydnor, 2015; Patel et al., 2016; Carrera et al., 2018; Rohde and Verbeke, 2017), while others document short-run gains that fade within months (Volpp et al., 2008; Acland and Levy, 2015; Carrera et al., 2020). However, more sophisticated incentives, such as commitment or time-bundled contracts, have produced durable effects on gym attendance and walking (Royer, Stehr and Sydnor, 2015; Aggarwal, Dizon-Ross and Zucker, 2022). Evidence also extends to other health behaviors: financial incentives have produced lasting effects in children's healthy eating for up to three months (Loewenstein, Price and Volpp, 2016) and in daily handwashing for up to nine months (Hussam et al., 2022), though the effect was diminishing in the latter case. Our study adds to this literature by examining biometric health screenings, a low-frequency (annual) behavior. The results demonstrate that even infrequent health behaviors can become habitual, suggesting that habit formation may play an important role in sustaining preventive practices such as annual vaccinations.

Finally, our study contributes to the extensive literature on using financial incentives to promote healthy behaviors. Similar to our setting, prior work has examined low-frequency activities such as vaccinations (Stone et al., 2002; Campos-Mercade et al., 2021), cancer and cardiovascular screenings (Stone et al., 2002; Alsan, Garrick and Graziani, 2019), and workplace health screenings (Jones, Molitor and Reif, 2019; Song and Baicker, 2019). However, little is known whether incentives produce habit formation for these activities, which typically only occur once per year. One notable exception is Schneider et al. (2023), who find that a $24 financial incentive for the initial COVID-19 vaccine dose in Sweden did not significantly affect uptake of subsequent doses. In contrast, we study much larger incentives ($75–$200) in the context of annual health screenings and show that they produced economically large and statistically significant persistence effects.

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.

prominent exception is Fujiwara, Meng and Vogl (2016), who study habit formation in voting.

Section 3 provides background on our empirical setting and describes our data. Section 4 outlines our empirical strategy. Section 5 describes our results, and Section 6 concludes.

2 Model

We present a flexible framework for habit formation that is capable of nesting a variety of underlying mechanisms. We then focus our attention on two particular classes of habit formation models, which we refer to as "reinforcement models" and "threshold models." Informally, reinforcement refers to cases where yesterday's action consumption has a strictly positive effect on today's stock of habit, and that effect is larger than actions taken in the more distant past, while threshold models capture cases where what matters for the current stock of habit is that a sufficient amount of past activity has taken place, and not *when* that past activity has taken place.

These two classes are widely cited mechanisms—addiction goods, for example, represent a prominent application of the reinforcement model and experience goods are common represntation of threshold-like models. In particular, both are commonly invoked as important drivers of health-related behaviors (Gruber and Köszegi, 2001; Dupas, 2014; Banerjee et al., 2021; Hussam et al., 2022). This framework provides a common structure for analyzing how past consumption influences current choices and yields distinct, testable predictions for each of the two classes of mechanisms of interest. Finally, we show that in our particular setting—a three-period, discrete choice problem—our framework can be used to potentially distinguish between all models that involve non-negative effects of past consumption on current 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. Consumption of these goods, together with other factors z_t, produces flow utility $U(a_t, S_t, c_t, z_t)$. Past consumption of a_t affects current

utility through S_t, the stock of past consumption:

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

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

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

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

where income w_t is allocated between the two goods. The variable z_t represents a potential taste shifter or other instrument that is exogenous. For simplicity, we assume consumers are myopic: they ignore how current consumption affects future utility via the stock S_{t+1}.[5]

For tractability, we treat a_t as a continuous variable and assume utility is concave and quadratic, a common specification in the addiction literature (Becker, Grossman and Murphy, 1994; Gruber and Köszegi, 2001; Reif, 2019).[6] As we show in Appendix A, 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 κ is a constant, π captures price sensitivity, and δ reflects the effect of z_t on demand. The parameter of interest, θ, reflects how the stock of past consumption affects current demand: $\theta > 0$ implies habit formation in behavior—higher past consumption raises current demand—while $\theta < 0$ indicates that past consumption reduces current demand, consistent with satiation or fatigue. Appendix A derives the conditions on preferences under which $\theta > 0$. Identifying θ requires exogenous variation in the stock S_t, which our randomized

[5] In a forward-looking addiction model, current demand would also depend on expected future consumption. In a forward-looking experience-good model, individuals may choose to consume earlier in order to learn their preferences and improve future decisions. However, the qualitative predictions derived below remain unchanged under these forward-looking extensions. Because future incentives were not known to the individuals in our setting in advance, we abstract from forward-looking considerations in both models.

[6] Appendix A presents a discrete-choice version of the model that yields a similar estimating equation, and also solves a model where we relax the quadratic functional form assumption.

incentive design provides. Conceptually, our instruments can be thought of as shifting earlier realizations of z_t, which in turn affect the current stock S_t, or equivalently can be though of shifnt past realizations of p_t.[7]

In what follows, we specify two alternative laws of motion for S_t—one corresponding to reinforcement models, the other to threshold models—and derive their testable implications.

2.2 Reinforcement Models

We generally define reinforcement models as those where the current stock of habit, S_t, is strictly increasing in the action taken in the prior period, a_{t-1}, and the impact of actions in the more distant past are smaller than the most recent action. Formally, we have the following:

Assumption 1 (Habit Reinforcement)

$$\begin{aligned} \frac{\partial S_t}{\partial a_{t-1}} &> 0 \\ \frac{\partial S_t}{\partial a_{t-1}} &\geq \frac{\partial S_t}{\partial a_{t-k}} \quad \textit{for } k > 1 \end{aligned}$$

We operationalize this mechanism using a functional that follows the addiction literature Becker and Murphy (1988) and let past consumption depreciate geometrically:

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

where $d \in (0, 1]$ is the rate of depreciation. Stable demand requires $\theta/d < 1$, ensuring that the effect of past consumption on demand does not grow without bound. Note, that while our results rely heavily on addiction models, other mechanism, such as model of switching costs in consumption, can be thought of as special cases of the reinforcement model.[8]

[7] Equation (4) is a Frisch demand equation, which holds the marginal utility of income constant. Although the randomized incentives are financial in nature, they represent less than 1% of annual income for the median participant. Thus, we assume that any associated income effects are negligible.

[8] As we show below, in a discrete choice case with one good, a switching cost model is equivalent to our reinformcement model with full depreciation, $d = 1$.

In a two-period setting ($T = 2$) with initial stock $S_0 = 0$—analogous to the two-year window in Wave 2 of our study—combining equations (4) and (5) yields:

$$a_2 = \kappa + \theta a_1 + \pi p_2 + \delta x_2. \tag{6}$$

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

$$a_3 = \kappa + \theta a_2 + \theta(1 - d)a_1 + \pi p_3 + \delta x_3 \tag{7}$$

Equation (7) highlights a key empirical prediction: current demand depends on both recent and earlier consumption, but the effect of more recent consumption is stronger, i.e., $\theta > \theta(1 - d)$. In our context, this result implies that current screening participation is more strongly influenced by last year's screening than by screening two years prior.[9]

2.3 Threshold Models

Habit formation can also arise after a sufficient amount of exposure to a good in the past, after which additional experimentation has no further impact on future consumption. A classic example is the so-called experience-good mechanism, where consumers learn their true preferences only after trying the good (Erdem and Keane, 1996; Ackerberg, 2003; Israel, 2005). For example, a consumer who tries mango for the first time may discover they enjoy it—a realization that could lead to a lasting habit. In this model, initial consumption reveals information about taste, and subsequent consumption reflects this updated belief. This mechanism therefore predicts that present consumption depends only on whether the good was ever consumed, not on how recently it was consumed. We formalize this model as the special case where the stock of past consumption captures only the experience of prior use:

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

[9] 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.

When there are only two periods, and the threshold is unknown, a model with initial stock $S_0 = 0$ yields the same qualitative predictions for the demand equation (6) as in the reinforcement model. More consumption in period 1 generally leads to more consumption in period 2. The models diverge, however, when the threshold is known and there are at least three periods: in the threshold model, current demand depends only on whether the individual has ever consumed a sufficient amount of the good, not on the timing of that consumption. Formally:

$$a_3 = \kappa + \theta \mathbf{1}(a_1 + a_2 > \underline{a}) + \pi p_3 + \delta x_3. \tag{8}$$

The threshold model demand specification generates distinct, testable predictions that differentiate it from the reinforcement model discussed in this case. Estimating only θ in equation (8) however, cannot reveal this distinction, since both frameworks imply that past consumption increases current demand ($\theta > 0$). However, when data span at least three periods, two additional implications emerge that uniquely characterize the experience-good model.

First, the effect of additional past consumption is zero once prior consumption of the good exceeds the given threshold. In other words, after an individual's initial exposure, further prior use does not raise current demand. Empirically, those who consumed enough of the good in both previous periods should be no more likely to consume now than those who consumed a sufficient amount in exactly one prior period. This prediction contradicts the reinforcement model, which implies that more frequent past use further reinforces current consumption.

Second, the timing of initial consumption is irrelevant in the threshold model. Because the threshold model mechanism involves no depreciation, the effect of initial consumption is identical whether the experience occurred earlier or more recently. Empirically, among individuals who have consumed enough of the good, those whose experience this exposure in an earlier period should be just as likely to consume now as those whose experience this exposure in a later period. This again contrasts with reinforcement, which predicts that more recent consumption exerts a stronger influence.

2.4 Unified Framework

Both the reinforcement and threshold models impose structural restrictions on how third-period demand depends on past consumption. These restrictions improve statistical power when estimating each model in isolation but require separate specifications to test them individually. This section shows that in a three-period model, a single, less restrictive specification can generate parameter estimates that discriminate between the two mechanisms, allowing both to be evaluated within a unified empirical framework. To better match our empirical setting, we present a discrete version of our model, in which case the threshold amount of consumption is $\underline{a} = 0$.[10]

First, we define the following mutually exclusive indicators of past consumption patterns:

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

and let $D_0 \equiv \mathbf{1}\{a_1 = 0,\ a_2 = 0\}$ denote the omitted reference category. Next, we define three treatment effects of interest:

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

We show in Appendix A.2 that, under our reinforcement model, we predict $\beta_3 = \beta_1 + \beta_2$ and that $\beta_2 > \beta_1$. The former result captures the frequency prediction of the reinforcement model: the impact of having consumed in both periods 1 and 2 on period 3 consumption is greater than consuming in either period 1 or period 2 alone. The latter result captures the timing prediction: the impact of past consumption on period 3 is greater when consumption happens in period 2 than in period 1.[11]

In Appendix A.2, we derive an alternate prediction under the threshold model. Namely,

[10] Appendix A.3 presents a continuous version that yields qualitatively similar results and provides a full derivation of the unified model.

[11] The frequency prediction requires additional assumptions regarding the linearity of the utility gain of consumption, a_t, and a linearity assumption regarding CDF of a random utility shock. See Appendix A.2 for further details.

we have that $\beta_1 = \beta_2 = \beta_3$. This captures both the frequency and timing implications of the model. The effect of past consumption on period 3 consumption is the same whether there was one or two prior periods of consumption. In addition, if there is only one period of prior consumption, it is not relevant when that took place.

The frequency restriction implied by the reinforcement model provides a sharp contrast with the threshold model. Even if the threshold equalities ($\beta_1 = \beta_2 = \beta_3$) are not rejected, the reinforcement model can still be falsified through its sharper prediction that $\beta_3 = \beta_1 + \beta_2$ (for binary a_t). By contrast, the timing implication under reinforcement is an inequality and thus not sharp—it remains compatible with failing to reject the threshold model. Likewise, failing to reject the reinforcement timing restriction does not imply that the threshold equalities hold. However, the frequency restriction under the threshold model implies $\beta_3 < \beta_1 + \beta_2$, while the reinforcement model predicts $\beta_3 = \beta_1 + \beta_2$. We therefore focus on the frequency-based restriction as the central empirical test distinguishing the two mechanisms.

2.5 Discussion

Several studies link the threshold model to health behaviors. Crawford and Shum (2005) and Dickstein (2021) model pharmaceutical demand and drug adherence under experiential learning about drug efficacy. Dupas (2014) subsidizes the purchase of antimalarial bed nets and attributes the subsequent increase in long-run demand to learning about the product's true value. Similarly, Banerjee et al. (2021) find that a one-time health insurance subsidy led to sustained enrollment, which they attribute to learning through experience. In these models, the implied threshold of consumption is typically any non-zero exposure, i.e. $\underline{a} = 0$. Related to the threshold model is the concept of start-up costs, which have been cited as drivers of habit formation in exercise (Charness and Gneezy, 2009; Carrera et al., 2020). In these models, current demand is shaped by initial consumption but does not necessarily depend on recent consumption.

To our knowledge, no prior study has implemented a research design to distinguish reinforcement from a threshold model. Hussam et al. (2022), for instance, interpret persistence in handwashing as a reinforcement—more specifically an addiction model—but acknowledge

that it may also reflect learning, which is more consistent with a threshold model.[12] Similarly, Banerjee et al. (2021) attribute persistence in health insurance enrollment to learning from initial exposure, but it is also possible that persistence is driven instead by more frequent or more recent interaction with the health system, à la reinforcement.

Our analysis demonstrates that reinforcement and threshold models can be empirically distinguished when consumption data span three or more periods, including the initial period in which learning might plausibly occur. Specifically, the two models differ with respect to a frequency prediction and timing prediction.

Other mechanisms, such as inertia and switching costs, can also drive habit formation. Handel (2013) and Polyakova (2016) argue that these mechanisms contribute to persistence in insurance plan enrollment. Like reinforcement, inertia and switching cost models predict that current demand should be most strongly related prior consumption from the most recent period. In our setting, inertia may not be empirically relevant, since our subjects are required to make an active choice in each period. On the other hand, a switching cost model in our setting is equivalent to a reinforcement model with full depreciation, i.e. $d = 1$, and thus our test of the reinforcement model nests this special case.

3 Institutional Setting

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

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

resulting in a final sample size of 4,799.

Employees in the wellness group ($N = 3,275$), referred to as "Wave 1" members, were invited for on-campus biometric health screenings in the fall of 2016, 2017, and 2018. During these annual screenings, clinicians measured each employee's height, weight, waist circumference, and blood pressure, and administered a fingerstick blood test to check for cholesterol, triglyceride, and glucose levels. Employees received their results within minutes and reviewed them with a health coach.

Wave 1 members who completed a health screening in 2016 or 2017 were also invited to complete an online health risk assessment (HRA), a questionnaire designed to assess a person's health habits. Completing the HRA made them eligible to choose from various wellness activities offered throughout the academic year, including a self-paced walking program, weight management classes, and a tobacco cessation program. 34 percent of Wave 1 members completed at least one activity during the two-year program.

Employees assigned to the non-wellness group ($N = 1,524$), referred to as "Wave 2" members, were invited to complete health screenings in 2017 and 2018 so that researchers could compare their biometric health outcomes with those from the Wave 1 group (Reif et al., 2020). Unlike Wave 1 members, they were not eligible to participate in the initial (2016) health screening and were never invited to complete an online HRA or sign up for wellness activities.

Figure 1 presents the study's experimental design. Financial incentives were assigned randomly several weeks before each screening sign-up period. The annual rerandomization ensured that incentives were independent of previous incentives or employee outcomes, including past program participation. In 2016, Wave 1 employees were equally likely to receive an incentive of \$0, \$100, or \$200.[13] In 2017 and 2018, Wave 1 and Wave 2 employees were assigned with equal probability to receive either \$0 or \$125 (in 2017) and either \$0 or \$75 (in 2018) upon successfully completing a screening. The incentives in each year were assigned at the individual level using stratified random sampling, as detailed in Jones, Molitor and Reif (2019). The study's structure creates five distinct experiments—one per wave and year

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

of randomization.

4 Empirical Strategy

Our empirical strategy has two objectives. First, we estimate the causal effects of financial incentives on health screening completion. We define the "direct effect" of incentives as their impact on same-year screening completion and the "persistence effect" as their impact on future completion. The five treatments (incentives) embedded in the study design depicted in Figure 1 all produce measurable direct effects. We can measure persistence effects for three of these treatments: the 2016 incentives' effects on Wave 1's 2017–2018 screenings, and the 2017 incentives' effects on 2018 screenings for both Wave 1 and Wave 2. Second, we estimate habit formation using an IV approach that assumes persistence operates through past completion, drawing on the model in Section 2 to motivate our empirical specifications and tests.[14]

It is uncertain whether financial incentives to complete a screening will increase or decrease future screening completion. If individuals prefer to undergo screenings only intermittently—perhaps because they feel that more frequent screenings do not generate information of sufficient value—then an incentive that boosts current screening could reduce future screenings due to intertemporal substitution. Alternatively, the incentive could raise future screenings if initial participation increases the value of subsequent screenings. For instance, regular screenings allow individuals to monitor changes in biometrics over time, and the logistics of scheduling and completing screenings become easier after the first screening. The parameter θ from Section 2 is therefore of ambiguous sign, both in our model and in light of these more general considerations.

Our experimental framework relies on the assumption that incentives are randomly assigned to participants. To assess this assumption, we test whether assigned incentive amounts are jointly predicted by other screening incentives, baseline demographics, and survey variables. Table 1 reports the averages of these variables (one per row) across treatment arms

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

(one per column). We conduct joint balance tests for each of the five experiments in our study, represented by pairs of adjacent columns (columns (1)–(2), (3)–(4), etc.). The p-values from these tests, all greater than or equal to 0.29, indicate no systematic differences across treatment arms and support the null hypothesis of randomly assigned incentives in each experiment.

4.1 Persistence

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

$$\text{SCREEN}_i^t = \alpha + \sum_{\tau=1}^{T} \beta_\tau \text{INCENTIVE}_i^\tau + \gamma X_i + \epsilon_i. \tag{10}$$

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

The focal parameter, β_τ, represents the average treatment effect (ATE) of monetary incentives assigned in year τ on screening completion. Because individuals receive the assigned incentive only upon completing a health screening in year τ, the regression equation identifies the direct effect of incentives when screening is measured in the same year as the incentive ($t = \tau$) and a persistence effect when screening is measured in a future year ($t > \tau$). Since future incentives were unknown to participants, we expect $\beta_\tau = 0$ for $t < \tau$, providing a falsification test of our identifying assumptions.

Equation (10) includes multiple treatment variables. If the effect of an incentive depends on the receipt of past incentives, β_τ will capture a weighted average of ATEs across the counterfactual scenarios created by different combinations of incentives offered in previous

years (Muralidharan, Romero and Wüthrich, 2023).[15] This weighted average is policy relevant because many real-world wellness programs and other similar interventions commonly offer incentives on an annual basis. In supplemental analyses, we extend equation (10) to include interaction terms.

Our baseline analysis excludes control variables, X_i. Because incentives were randomly assigned, controls are not necessary to remove bias in the focal estimate but may increase precision. In supplemental analyses, we adopt a "post-Lasso" control specification, selecting controls via the Lasso double-selection method of Belloni, Chernozhukov and Hansen (2014). The set of potential control variables includes baseline demographics, health survey responses, health behaviors, and claims-based measures of medical spending and usage, along with all their pairwise interactions.[16] Since randomization was performed at the individual level, we report conventional heteroskedasticity-robust standard errors (Abadie et al., 2023).

4.2 Habit Formation

We define habit formation as the causal effect of past screening completion on current screening completion. We estimate a series of specifications to test our alternative models of habit formation developed in Section 2.

4.2.1 Reinforcement Model

We begin by estimating a single specification that relates current screening behavior to up to two lags of past screening completion:

$$\text{SCREEN}_i^t = \kappa + \theta_1 \text{SCREEN}_i^{t-1} + \theta_2 \text{SCREEN}_i^{t-2} + \gamma X_i + \epsilon_i. \tag{11}$$

The outcome, SCREEN_i^t, is an indicator equal to 1 if individual i completed a screening in year t. The vector X_i denotes individual-specific control variables. Our baseline specifica-

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

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

tion includes the contemporaneous financial incentive, INCENTIVE_i^t, as a control variable, given its strong predictor power for screening completion. In supplemental analyses, we additionally report specifications that include post-Lasso controls.

Equation (11) lets us analyze habit formation in both two- and three-period settings. When only two periods of data are available, the model effectively reduces to a single-lag specification, since SCREEN_i^{t-2} is not observed. In this case, identification is limited to testing whether past screening causally affects current screening, i.e., whether $\theta_1 \neq 0$. Conditional on $\theta_1 > 0$, both the reinforcement and threshold models predict a positive effect of past screening on current screening. As a result, with only one lag, the two mechanisms are observationally equivalent whenever habit formation is present and cannot be empirically distinguished.

The availability of three periods of data introduces mechanism-specific testable restrictions. With two lags observed, initial and more recent past consumption can be separately identified, and their relative effects on current behavior become informative about the underlying habit-formation process. In the reinforcement model, past consumption depreciates over time, so current behavior depends more strongly on recent than on earlier consumption. Under positive persistence, this implies $\theta_1 > \theta_2$; more generally, the model predicts that the effect of the most recent lag dominates that of earlier lags, i.e., $|\theta_1| > |\theta_2|$.

4.2.2 Threshold Model

The threshold model described in Section 2 implies that habit formation operates through initial exposure rather than reinforcement. In particular, we assume that there is a threshold is exceeded with at least one past period of exposure. Because the stock of past consumption records only whether the good has ever been consumed and does not depreciate, current behavior depends on whether an individual has prior experience with the good, but not on the frequency or recency of that experience. This structure yields two testable implications: additional consumption beyond the first has no incremental effect, and conditional on having consumed once, the timing of that initial consumption is irrelevant.

We first test the frequency implication using the following specification:

$$\text{SCREEN}_i^{2018} = \kappa + \theta_3 \text{EVER}_i + \theta_4 \text{TWICE}_i + \gamma X_i + \epsilon_i. \tag{12}$$

Here, $\text{EVER}_i = \mathbf{1}(\text{SCREEN}_i^{2016} + \text{SCREEN}_i^{2017} > 0)$ indicates whether individual i completed at least one prior screening, and $\text{TWICE}_i = \mathbf{1}(\text{SCREEN}_i^{2016} + \text{SCREEN}_i^{2017} = 2)$ indicates whether the individual completed both prior screenings. Under the experience-good model, habit formation operates through initial exposure only, so completing a second screening generates no additional effect on subsequent participation, implying the restriction $\theta_4 = 0$.

We next test the timing implication by distinguishing between first-time screening in 2016 and first-time screening in 2017:

$$\text{SCREEN}_i^{2018} = \kappa + \theta_5 \text{SCREEN}_i^{2016} + \theta_6 \text{ONLY}_i^{2017} + \gamma X_i + \epsilon_i. \tag{13}$$

where $\text{ONLY}_i^{2017} = \mathbf{1}(\text{SCREEN}_i^{2017} > 0, \text{SCREEN}_i^{2016} = 0)$ indicates that the individual completed a screening in 2017 but not in 2016. Because the stock in the threshold model does not depreciate, the model predicts that only first exposure matters for future participation, not its timing, yielding the restriction $\theta_5 = \theta_6$. In other words, conditional on having completed a first screening, subsequent participation is invariant to whether that initial screening occurred in 2016 or 2017.

We estimate equations (12) and (13) separately for Waves 1 and 2. To increase statistical power, we also estimate a pooled specification that combines the two waves. Under the threshold model, the effect of initial exposure is invariant to calendar time and cohort, so pooling across waves is appropriate after controlling for wave fixed effects.[17] These fixed effects absorb level differences in screening participation driven by wave-specific program features (e.g., ancillary wellness activities), while identification comes from within-wave random variation in incentives that induce first-time screening in different years.

[17] For all Wave 2 participants, $\text{SCREEN}_i^{2016} = 0$ by design, since they were not eligible for the 2016 screening. As a result, identification of θ_5 comes entirely from Wave 1, while θ_6 is identified from first-time screening in 2017 in both waves.

4.2.3 Unified framework

Finally, we estimate an unrestricted three-period specification that nests both the reinforcement and threshold models within a single empirical framework. We estimate this specification using the pooled sample of Waves 1 and 2 and include wave fixed effects. Specifically, we relate 2018 screening completion to mutually exclusive indicators for prior screening behavior:

$$\text{SCREEN}_i^{2018} = \kappa + \beta_1 \text{ONLY}_i^{2016} + \beta_2 \text{ONLY}_i^{2017} + \beta_3 \text{TWICE}_i + \gamma X_i + \epsilon_i, \tag{14}$$

For all Wave 2 participants, $\text{ONLY}_i^{2016} = 0$ by design, since they were not eligible for the 2016 screening; as a result, β_1 is identified entirely from Wave 1, while β_2 and β_3 are identified using variation from both waves.

The unified framework allows us to test the core restrictions implied by the two habit-formation mechanisms using a single set of coefficients. In the reinforcement model, past consumption enters current demand through a depreciating stock, so more recent screening has a larger effect than more distant screening, and the effects of multiple prior screenings accumulate. Under positive persistence and the linear structure described in Section 2, this implies the restrictions $\beta_3 = \beta_1 + \beta_2$ and $\beta_2 > \beta_1$.

In contrast, the threshold model implies that habit formation operates entirely through initial exposure. Once an individual has completed a screening at least once, additional screenings do not further affect subsequent participation, and the timing of that initial screening is irrelevant. Accordingly, all prior screening histories involving at least one screening have the same effect on future behavior, yielding the restriction $\beta_1 = \beta_2 = \beta_3$.

4.2.4 Identification

Ordinary least squares (OLS) estimation of equations (11)–(14) will produce biased estimates if past screening completion is correlated with unobserved determinants of current screening completion. This bias is likely to be positive, as many factors that increase screening propensity in one period, such as higher health consciousness or a more proactive attitude toward preventive care, are likely to persist and increase the probability of future screening

completion. However, the bias could be negative if people typically wait one or more years between screenings.

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

The construction of instruments differs across specifications in a way that matches the structure of each habit-formation model. In the reinforcement specifications, the endogenous regressors are lagged screening completion indicators, so we use the corresponding year-specific randomized incentives as excluded instruments. In the threshold model and unified framework specifications, the endogenous regressors are nonlinear functions of past screening histories (e.g., ever, twice), so we use the full set of randomized incentives and their interaction terms as instruments. In pooled specifications, we interact the instruments with wave fixed effects to allow the first-stage relationships to vary across cohorts.

The IV analysis relies on the exclusion restriction that monetary incentives affect future screening completion solely through their influence on concurrent screening completion. We address potential violations of this restriction in Section 5.2. In models with a single endogenous regressor, we can interpret θ as the average causal effect of screening completion in the past among compliers, i.e., those induced to complete a health screening in a prior year by a financial incentive. This interpretation requires the standard monotonicity assumption for local ATE (LATE) interpretations of 2SLS.

When testing linear restrictions across equations with multiple endogenous regressors, we must further assume that the complier populations are comparable across those regressors. To assess this assumption, we proceed in two steps. First, we show that our specification

satisfies the average conditional monotonicity and no cross effects conditions of Bhuller and Sigstad (2024), which ensures that each 2SLS estimand can be interpreted as a properly weighted LATE even in the presence of multiple treatments.[18] Second, we analyze the characteristics of compliers and show that the populations induced by the different instruments are similar across a wide range of observable dimensions.

5 Results

5.1 Persistence

Figure 2 reports annual health screening completion rates by incentive level and experiment wave. These plots separately illustrate the contemporaneous (direct) effects of financial incentives and their persistence over time. The direct effect of an incentive can be assessed by comparing the completion rates of groups that received high and low incentives in the year the incentive was initially introduced. Persistence effects are identified by comparing completion rates in subsequent years among groups that differed only in earlier incentive assignments. The persistence effects of incentives given before 2018 are shown in panels A, B, and D. For the second and third incentives (in panels B, C, and E), completion rates before the incentives were assigned provide a falsification test for whether incentives were successfully rerandomized each year.

Panel A of Figure 2 shows that for Wave 1 subjects, the initial (2016) financial incentive increased completion of that year's health screening from 49.4% to 61.8%. Importantly, this initial incentive also generated sizable persistence effects: those who were offered it were about 4.5 and 4.4 percentage points more likely to complete a health screening in 2017 and 2018, respectively, than those who were not. Error bars confirm the statistical significance of these effects ($p < 0.05$). Panel D shows that the initial incentive offered to individuals in Wave 2 (in 2017) increased screening completion by 27.7 percentage points in that year and by 8.1 percentage points in 2018, demonstrating that the persistence of first incentives

[18] For completeness, the appendix also reports estimates of the coefficients on the two endogenous regressors obtained from separate single-equation IV specifications, which preserve the standard one-instrument-per-treatment LATE interpretation.

replicates in an independent cohort. In contrast, panel B shows that the second incentive offered to Wave 1 subjects had a substantial direct effect but resulted in little to no persistence effect. Finally, the falsification tests in panels B, C, and E all reveal similar screening rates between high- and low-incentive groups in years before the incentive assignment, consistent with successful annual rerandomization.

Table 2 displays regression estimates from equation (10), analyzing the impact of financial incentives on current, future, and past screening completion rates. Each column corresponds to a distinct wave-year outcome, estimated in a separate regression. Rows report the effects of the incentives offered in each of the three years of the program. For clarity, the direct effects on current completion, highlighted in gray, are shown along the diagonals in columns (1)–(3) for Wave 1 and columns (4)–(5) for Wave 2. The off-diagonal entries capture non-contemporaneous effects: the persistence effects on future completion are in bold, while the effects of incentives on past completion (in plain text) serve as falsification tests for the randomization process. Post-Lasso estimates, which yield similar conclusions, are reported in Table A.1.

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

Table A.2 presents estimates from a generalized model incorporating interactions between the incentives assigned in different years.[19] We focus on the 2018 screening outcome,

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

for which the number of previous incentive rounds is greatest. Column (1) of Table A.2 reproduces the baseline estimates from the third column of Table 2, while columns (2)–(5) report different combinations of interactions. Although including these interactions reduces statistical power, the estimates of the main effects remain largely unchanged. We do not detect any significant interaction effects, though the confidence intervals are wide enough that meaningful interaction effects cannot be ruled out.

While all three persistence effects of the first incentives offered to the Wave 1 and Wave 2 groups are large and statistically significant, the effect of the second (2017) incentive on 2018 screening completion in the Wave 1 group is small and insignificant (see panel B of Figure 2 and column (3), row 2 of Table 2). We consider three possible reasons for this discrepancy. First, the 2017 screening experience itself may have differed in ways that attenuated persistence, for example if participants had unusually poor screening experiences. If that were the case, we would expect Wave 2 subjects, who received the same screening treatment in 2017 as Wave 1 subjects, to exhibit similarly weak persistence effects. However, the 2017 incentive produced a positive persistence effect on 2018 screening completion for Wave 2 subjects (see panel D of Figure 2 and column (5), row 2 of Table 2).

A second possibility is that the 2017 incentive had a weaker direct effect on Wave 1 subjects, which could mechanically lead to smaller persistence effects. Comparing the direct effects in columns (2) and (4) of Table 2 provides some limited support for this hypothesis, though the differences are small. Moreover, for Wave 1 subjects, the direct effect of the 2017 incentive exceeds that of the 2016 incentive, which nonetheless generated sizable persistence effects. These patterns suggest that the lack of persistence following the second incentive in Wave 1 was not a consequence of weak direct effects.

A third possibility is that persistence is generated primarily by initial exposure. Unlike Wave 2 subjects, Wave 1 subjects were eligible to participate in the 2016 health screenings. As a result, the 2017 incentive represented a second "dose" for Wave 1 but was the first dose for Wave 2. This interpretation aligns naturally with models in which habit formation operates through initial exposure rather than reinforcement. We consider this possibility in further detail in Section 5.2.

5.2 Habit Formation

This section evaluates the persistence documented in Section 5.1 through the lens of habit formation and tests whether it can be characterized by an reinforcement or a threshold mechanism.

5.2.1 Reinforcement Model

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

Columns (1) and (3) of Panel A report two-period IV estimates from equation (11) for Waves 1 and 2, respectively. In both waves, completing an initial screening substantially increases the likelihood of completing the subsequent screening. For Wave 1, completing the 2016 screening raises the probability of completing the 2017 screening by 36.0 percentage points, an 84 percent increase relative to the mean. For Wave 2, completing the 2017 screening raises the probability of completing the 2018 screening by 32.4 percentage points, an 89 percent increase. Both effects are statistically significant ($p < 0.01$) and similar in magnitude across waves.

Column (2) of Panel A exploits the three-period structure of Wave 1 to test the key prediction of the reinforcement model implied by equation (7): more recent consumption should exert a stronger effect on current demand than more distant consumption. In this context, completing the 2017 screening should have a larger effect on 2018 screening 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 the probability by only 6.3 percentage points and is statistically insignificant. Not only is the more recent effect smaller; it is statistically distinguishable from the earlier effect. In particular, the 95 percent confidence interval for the 2017 effect, $[-8.0, 20.5]$, does not contain the 2016 point estimate, and a Wald test rejects equality of the two coefficients at conventional levels. Together, these results imply

that habit formation is driven primarily by the initial screening: the first screening generates strong and persistent increases in subsequent participation, while the second screening in Wave 1 produces little additional habit formation.

Panel B of Table 3 reports the corresponding OLS estimates. In contrast to the IV results, all OLS estimates are positive and statistically significant. However, these estimates are prone to upward bias because unobserved determinants of screening participation—such as health preferences—are persistent over time. Table 3 therefore demonstrates how failing to account for this bias can lead to incorrect inference regarding habit-formation. Indeed, the OLS estimates spuriously confirm the reinforcement prediction that recent screening has a stronger effect than earlier screening ($\theta_2 > \theta_3$ in column (5)), a conclusion decisively rejected 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, where past screening completion is instrumented with randomly assigned past incentives, while Panel B reports the corresponding OLS estimates.

We begin with the IV estimates in Panel A. Columns (1)–(3) estimate a restricted version of equation (12) that includes only an indicator for having ever completed a screening prior to 2018. This specification imposes the threshold model restriction that additional screenings beyond the first have no incremental effect on current behavior. For Wave 1, Wave 2, and the pooled sample, having ever completed a screening substantially increases the probability of completing the 2018 screening. The IV estimates imply effects ranging from 32.4 to 36.1 percentage points, all statistically significant at conventional levels. These results indicate strong habit formation following initial exposure but, taken alone, do not distinguish between the addiction and experience-good mechanisms. For Wave 2, which has only two screening opportunities, this specification coincides with the two-period habit-formation model estimated in Table 3 and identifies the same habit-formation parameter.

Columns (4) and (5) of Panel A relax the restriction that additional screenings have no effect by separately estimating the incremental effect of having completed a second screening,

corresponding to equation (12) with both terms included. In both the Wave 1 and pooled specifications, the estimated effect of having been screened twice is small and statistically insignificant. Conditional on having completed at least one screening, additional past screenings do not increase the likelihood of screening in 2018. This pattern is consistent with the threshold model's prediction that habit formation operates through initial exposure rather than reinforcement through repeated participation.

Columns (6)–(8) further probe this implication by estimating equation (13), which distinguishes between first-time screening in 2016 and first-time screening in 2017. The experience-good model predicts that the timing of initial exposure should not matter. Consistent with this prediction, the estimated effects of first-time screening in 2016 and first-time screening in 2017 are similar in magnitude in the pooled sample, and we cannot reject equality between them.

Panel B reports the corresponding OLS estimates. In contrast to the IV results, OLS estimates imply that both having been ever screened and having been screened twice have large and statistically significant effects on 2018 screening completion (columns (4) and (5)). Thus, under OLS, the threshold model restriction that additional screenings provide no incremental benefit is rejected. This divergence underscores the importance of accounting for endogeneity in past screening behavior when distinguishing between habit formation mechanisms.

5.2.3 Unified Framework

Table 5 reports estimates from the unified three-period model in equation (14), which relates 2018 screening completion to mutually exclusive prior screening histories. This specification nests both the reinforcement and threshold model mechanisms and allows us to directly test their distinct structural restrictions using a single set of coefficients.

Column (1) reports IV estimates, where prior screening indicators are instrumented with the randomly assigned incentives from earlier years. All three coefficients are positive and statistically significant, indicating substantial habit formation regardless of whether prior screening occurred in 2016, 2017, or both years. Point estimates imply that completing a screening only in 2016 increases the probability of completing the 2018 screening by 44.0

percentage points, completing a screening only in 2017 increases it by 31.6 percentage points, and completing screenings in both years increases it by 40.3 percentage points.

These estimates enable direct tests of the competing model restrictions derived in Section 2. The reinforcement model predicts that the effect of screening twice equals the sum of the effects of screening only in 2016 and only in 2017 ($\beta_3 = \beta_1 + \beta_2$). This restriction is rejected by the IV estimates ($p = 0.017$), 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 restriction is strongly rejected ($p < 0.001$). The estimates also reject the sharp additivity restriction implied by the reinforcement model, but the ordering of the coefficients follows a reinforcement-like pattern ($\beta_3 > \beta_2 > \beta_1$). This pattern suggests that positive serial correlation in unobserved determinants of screening behavior biases OLS estimates upward, producing spurious evidence of reinforcement that does not correspond to any coherent habit-formation mechanism.

Overall, the unified framework consolidates the evidence from the preceding analyses. When past screening completion is instrumented using randomized incentives, the estimates reject reinforcement and instead support a threshold model interpretation in which initial exposure generates persistent participation.

5.2.4 Alternative Explanations

Our exclusion restriction assumes that the effect of financial incentives on future screening completion operates through an increase in prior screening completion. Because incentives were rerandomized each year, serial correlation in assigned incentive amounts is not a threat to validity. However, a potential concern is that subjects may not have fully understood or paid attention to changes in their financial incentives over time. For example, individuals assigned to a high-incentive in 2016 and and a low incentive in 2017 might have mistakenly

thought they would continue to receive a high incentive in 2017. If so, incentives could affect future screening through beliefs about future payments rather than through past screening completion, violating the exclusion restriction.

To investigate that possibility, we estimate the effect of completing the 2016 screening on 2018 completion for the subsample of subjects assigned to the $0 group in 2017. If inattention were driving our estimates, we would expect a smaller persistence effect in this subsample: individuals who attended the 2017 screening expecting a high incentive would have learned they were mistaken, thereby reducing their likelihood of returning in 2018.[20]

We report the results of this investigation in Table A.6. Column (1) shows that the 2016 incentive raised the 2018 completion rate by 4.4 percentage points in the full sample. When we limit the regression to those individuals who were assigned the $0 incentive in 2017, the point estimate rises to 5.4 percentage points, providing no evidence that confusion about incentive payments driving our persistence estimates. Columns (3) and (4) report the corresponding IV estimates. Here, the point estimate from the subsample is slightly smaller than the estimate from the full sample, but the difference is not statistically significant. Including post-Lasso controls yields similar conclusions (Table A.7). Thus, we conclude that inattention does not explain our estimates.

Another potential threat to the exclusion restriction is the possibility that offering a financial incentive prompts individuals to seek information about biometric screenings, thereby increasing the likelihood of future screenings regardless of whether they completed the initial incentivized screening. While we view this channel as unlikely, we cannot rule it out definitively. Importantly, even if the exclusion restriction were to fail, the reduced-form persistence estimates presented in Table 2 would remain valid, as they do not rely on the IV assumptions.

5.3 Discussion

Overlooking habit formation significantly underestimates the effectiveness of financial incentives. For example, consider the initial (2016) incentive offered to Wave 1. Absent habit

[20]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.

formation, one would conclude that this incentive increased completion rates by 12.4 percentage points (or 406 additional screenings) for one year, as shown in column (1) of Table 2. Accounting for persistence, however, nearly doubles the estimated impact: the total effect across all three screenings is an increase of 21.3 percentage points (or 698 people), a 72% increase relative to the one-year estimate.[21] These estimates capture only the habit-formation effect observed over the two-year window of our intervention. Because we find no evidence of decay over this period, the long-run impact of the initial incentive would likely be even larger if the program were extended.

While habit formation substantially increased the effectiveness of our health screening incentives, this amplification is confined to initial exposure. A second dose of incentives generates large contemporaneous responses but had no detectable effect on future screening behavior. As a result, while the effectiveness of initial incentives may be substantially underestimated when habit formation is ignored, the returns to repeated incentives are unlikely to be similarly understated. This patterns contrasts sharply with predictions from standard addiction models, where reinforcement would increase the long-run effects of both initial and subsequent incentives.

Our results also shed light on the role of biased beliefs in shaping health behavior. We find evidence of habit formation, driven by a one-shot mechanism such as an experiential learning model rather than reinforcement through repeated participation. This pattern is consistent with individuals initially holding negatively biased beliefs, for example by overestimating their costs or underestimating their benefits, and updating those beliefs after first exposure. Once these beliefs are revised, additional participation provides little new information, limiting further habit formation. At the same time, different beliefs could generate different dynamics. For instance, if individuals initially believe a preventive behavior will be easy to sustain (e.g., a new exercise routine), early participation may reveal higher-than-expected time or effort costs, leading engagement to fall as expectations adjust.

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

6 Conclusion

We show that even infrequent health behaviors like annual health screenings can become strongly habitual. Completing a health screening for the first time raises the probability of completing a future screening by over 30 percentage points. However, completing a second screening does not further reinforce this habit. This pattern contradicts the a model with reinforcement, but is consistent with a threshold model where consumers learn the value of health screenings.

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

22

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

References

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

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

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

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

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

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

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

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

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

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

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

Becker, Gary S, Michael Grossman, and Kevin M Murphy. 1994. "An empirical analysis of cigarette addiction." *The American Economic Review*, 84(3): 396–418.

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

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

Brandon, Alec, Paul J Ferraro, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer. 2017. "Do the Effects of Nudges Persist? Theory and Evidence from 38 Natural Field Experiments." National Bureau of Economic Research Working Paper 23277.

Campos-Mercade, Pol, Armando N. Meier, Florian H. Schneider, Stephan Meier, Devin Pope, and Erik Wengström. 2021. "Monetary incentives increase COVID-19

vaccinations." *Science*, 374(6569): 879–882.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

Moore, Timothy J., and Todd Morris. 2024. "Shaping the Habits of Teen Drivers." *American Economic Journal: Economic Policy*, 16(3): 367–93.

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.

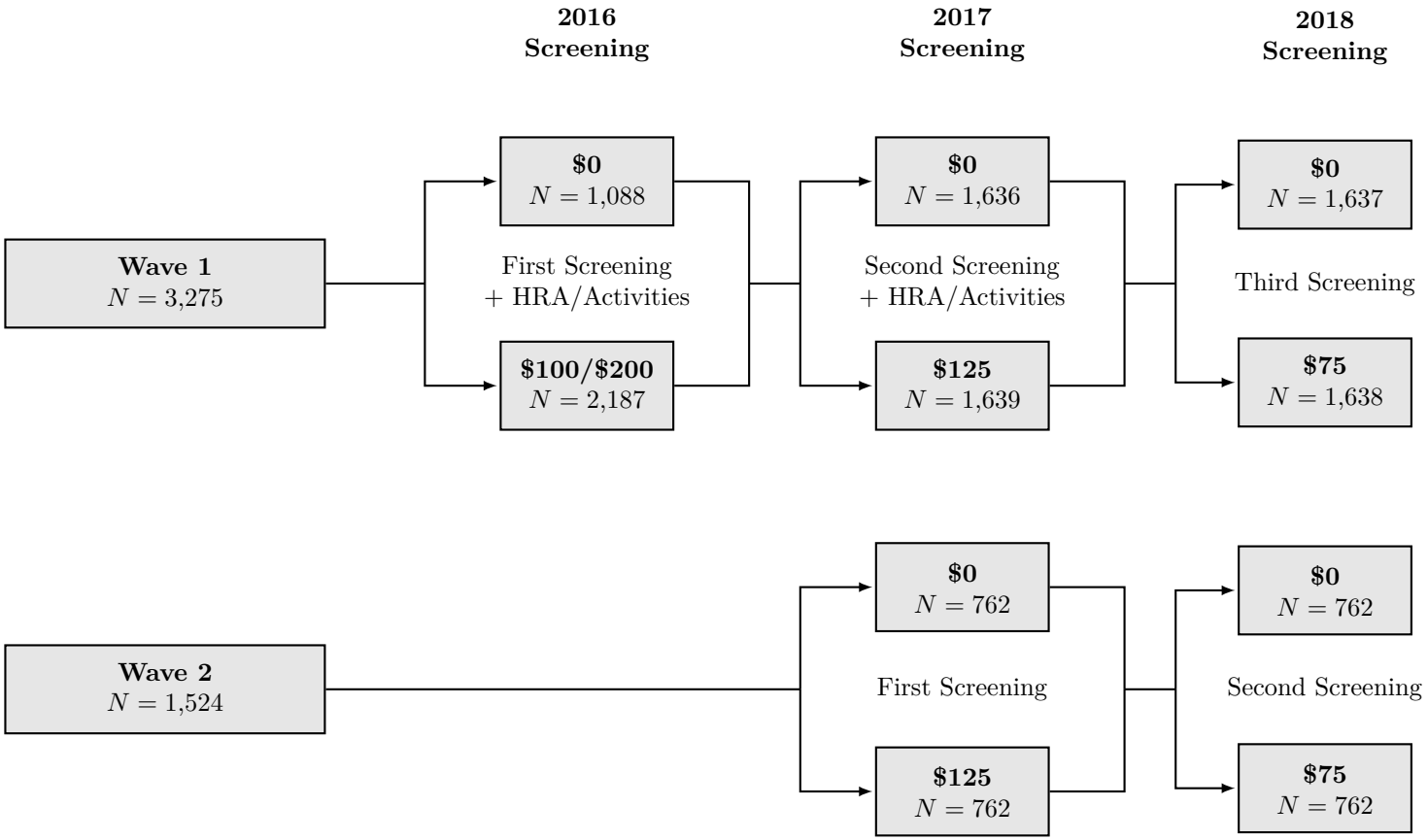
Song, Zirui, and Katherine Baicker. 2019. "Effect of a workplace wellness program on employee health and economic outcomes: A randomized clinical trial." *JAMA*, 321: 1491–1501.

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.

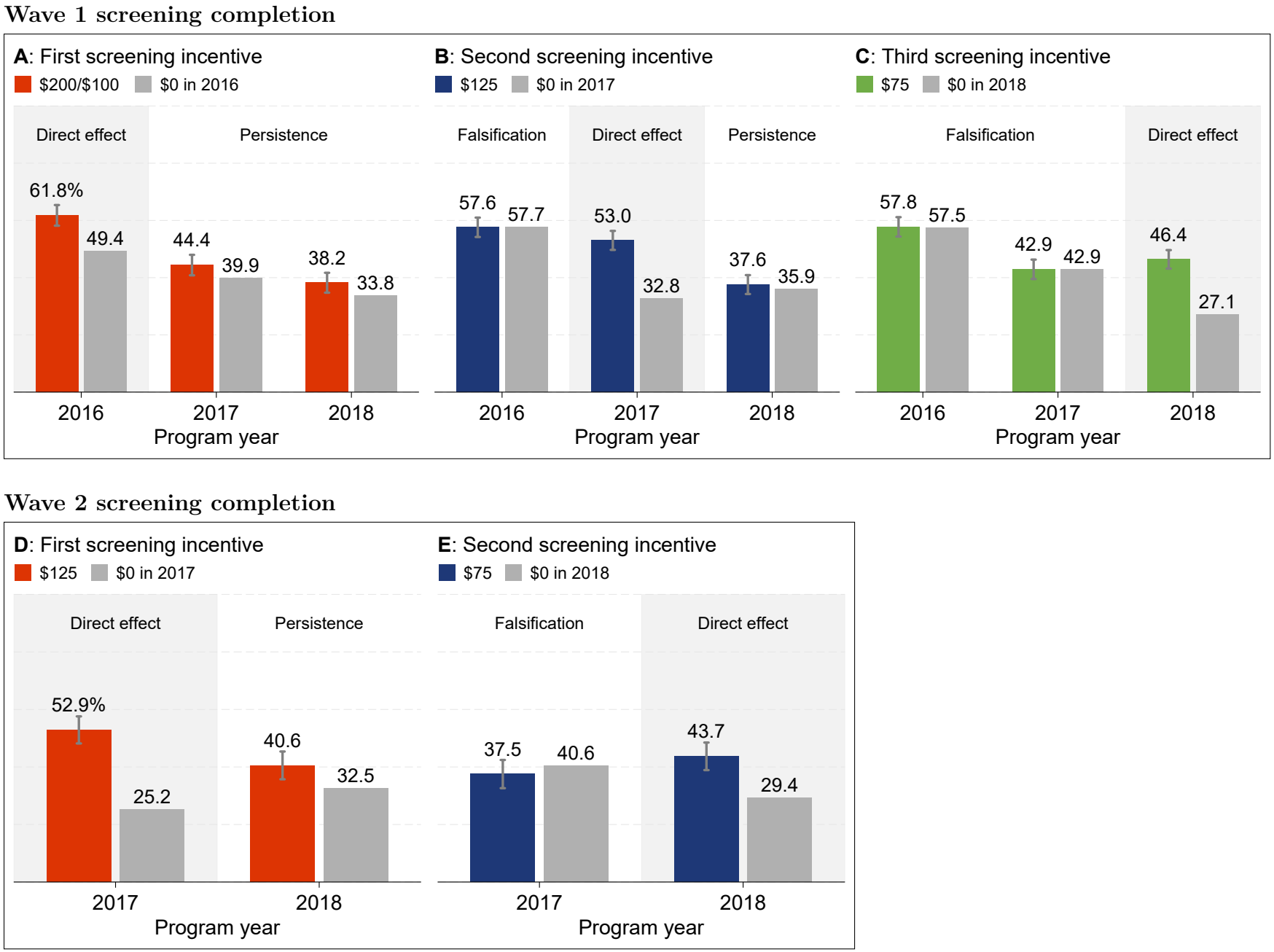
Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008. "Financial incentive-based approaches for weight loss: A randomized trial." *Journal of the American Medical Association*, 300: 2631–2637.

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. Incentive assignments are uncorrelated across years.

Figure 2: Health Screening Completion Rates, by Incentive Groups



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

Table 1: Balance Table

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

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

Table 2: Effect of Financial Incentives on Health Screening Completion

	(1)	(2)	(3)	(4)	(5)
	Wave 1			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 β_τ from Equation (10). The dependent variable is an indicator for whether a screening was completed in the year specified in the column header. Gray-shaded estimates indicate the contemporaneous effect of the incentive on screening completion. Bold-faced estimates indicate the persistence effect of the incentive on future screening completion, and plain-text estimates report a falsification test of an unannounced incentive on past screening completion. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table 3: Addiction Model Estimates

	(1)	(2)	(3)
	Wave 1		Wave 2
	Screened 2017	Screened 2018	Screened 2018
A. IV estimates			
Screened 2016	θ_1: 0.360**	θ_2: 0.331**	
	(0.128)	(0.127)	
Screened 2017		θ_1: 0.063	θ_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	θ_1: 0.467**	θ_2: 0.183**	
	(0.014)	(0.016)	
Screened 2017		θ_1: 0.430**	θ_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 θ_1 and θ_2 from Equation (11). Dependent variable is an indicator for whether the 2017 (or 2018) screening was completed. Reinforcement-class mechanisms like addiction predict $\theta_1 > \theta_2$. All regressions control for contemporary financial incentives. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.

Table 4: Experience Good Model Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wave 1	Wave 2	Waves 1 & 2	Wave 1	Waves 1 & 2	Wave 1	Wave 2	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: All specifications use 2018 screening completion as the outcome. 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: Effects of Mutually Exclusive 2016 and 2017 Screening Choices on 2018 Screening Completion

	(1)	(2)
	IV	OLS
Screened 2016 only	β_1: 0.440**	β_1: 0.188**
	(0.139)	(0.018)
Screened 2017 only	β_2: 0.316**	β_2: 0.503**
	(0.075)	(0.018)
Screened twice	β_3: 0.403**	β_3: 0.628**
	(0.133)	(0.015)
Parameter hypothesis tests (p-values)		
Addiction: $\beta_1 + \beta_2 = \beta_3$	0.017	0.021
Experience: $\beta_1 = \beta_2 = \beta_3$	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 β_1, β_2, and β_3 from equation (14). Dependent variable is an indicator for whether the 2018 screening was completed. All regressions control for contemporary financial incentives, interacted with wave. Robust standard errors are reported in parentheses, and */** indicates significance at the 5%/1% level.