NBER WORKING PAPER SERIES

# ESTIMATING BIASES IN SMOKING CESSATION: EVIDENCE FROM A FIELD EXPERIMENT

Frank J. Chaloupka IV
Matthew R. Levy
Justin S. White

Working Paper 26522
http://www.nber.org/papers/w26522

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2019

We thank Hunt Allcott, Jing Li, David Powell, Justin Sydnor, Josh Tasoff, Dmitry Taubinsky, and Severine Toussaert for helpful feedback, as well as audience members at the Behavioral Economics & Health Symposium at UPenn (2017), the ASHEcon meeting in Atlanta (2018), the TCORS Economics Working Group (2018), the ESA World Meeting in Vancouver (2019), the SITE Summer Workshop at Stanford (2019), and the Annual Health Economics Conference in San Francisco (2019). This research was funded by the U.S. National Institute on Drug Abuse under grant number P50-DA036128. This trial was registered in the AEA RCT Registry under AEARCTR-0002106. Special thanks go to Elissa Resnick and project managers at Nielsen Healthcare and SSI, notably Deanne Lowery, for research support. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Corresponding authors: Levy and White.

Estimating Biases in Smoking Cessation: Evidence from a Field Experiment
Frank J. Chaloupka IV, Matthew R. Levy, and Justin S. White
NBER Working Paper No. 26522
December 2019
JEL No. C93,D91,I12

## ABSTRACT

We conduct a randomized field experiment to quantify biases that affect consumers of addictive goods: present-biased preferences, naïve beliefs regarding present bias, and projection-biased beliefs over future abstinence. These biases reflect departures from the neoclassical benchmark needed to accommodate intertemporal and state-dependent prediction errors and have important theoretical and policy ramifications. Our experiment employs a new approach for remote monitoring to ensure truthful reporting of behavior and valuations, and a novel identification of subjects' biases based on willingness to pay for future abstinence incentives that serve as partial commitment devices. We find that cigarette smokers overestimate their likelihood of future abstinence by more than 100%, consistent with partially-naïve present-biased preferences. We estimate that on average smokers are present biased and only partially aware of their present bias, with substantial heterogeneity and a positive correlation between the two at the individual level. Smokers mispredict the effects of an abstinence intervention. Ex ante smokers anticipate no effect of the intervention on their future abstinence and ex post fail to recognize the intervention's positive effect. Our estimates highlight that smokers suffer from a constellation of biases: under their own long-run preferences, smokers' choices lead to a private welfare loss of $400 per week.

Frank J. Chaloupka IV
University of Illinois at Chicago
Department of Economics (m/c 144)
College of Liberal Arts and Sciences
601 S. Morgan Street, Room 713
Chicago, IL 60607-7121
and NBER
fjc@uic.edu

Justin S. White
University of California, San Francisco
Philip R. Lee Institute for Health Policy Studies
3333 California Street, Box 0936
San Francisco, CA 94118
Justin.White@ucsf.edu

Matthew R. Levy
Department of Economics
London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
m.r.levy@lse.ac.uk

A randomized controlled trials registry entry is available at AEARCTR-0002106

# 1 Introduction

Tobacco use is a high-stakes behavior with important health and financial consequences. About one in five deaths in the U.S. is attributable to tobacco use, and smokers lose roughly a decade of life expectancy (Fenelon and Preston, 2012; Jha et al., 2013). Premature morbidity and mortality attributable to smoking impose major costs to individual smokers and society. According to one comprehensive assessment, the private cost of smoking is $33 per pack, and the social cost is $40 per pack (Sloan et al., 2006).[1] Scaling up, this is a national social cost over the life cycles of U.S. smokers of more than $200 billion.

Smoking also poses theoretical and empirical challenges for both neoclassical and behavioral economic theory. It is a complex behavior, with meaningful implications for both intertemporal choice (immediate consumption and delayed health consequences) and state-dependent preferences (both long-term addiction and short-term cravings). Becker and Murphy's (1988) model of rational addiction has had a deep and lasting impact on researchers and policy makers. However, Gruber and Kőszegi (2001) among others have shown that observed smoking patterns may be consistent not only with rational addiction but also with an alternative model of agents with time-inconsistent preferences.[2] The multiplicity of interpretations from the same data has consequences for intuitions, welfare calculations, and policy prescriptions. Behavioral model selection becomes especially important—and complicated—for applications such as smoking in which multiple departures from rationality are possible or likely. In this paper, we consider three of the most commonly cited departures from rationality that have been tied to smoking: present-biased preferences, naïve beliefs regarding present bias, and projection-biased beliefs. Present-biased smokers over-weight the immediate pleasure derived from satisfying a craving or avoiding withdrawal symptoms at the expense of longer-term health and financial considerations. Naïve smokers are not fully aware of the extent of their present bias. Projection-biased smokers mispredict how they will feel about smoking after their addiction or craving state changes. We show that all three are important drivers of smoking decisions.

To better understand the behavioral biases in smoking decisions, we undertook an individual-level, randomized field experiment. Over 12 weeks, subjects from 16 large metropolitan areas in the U.S. ($N = 397$) reported their smoking status weekly, and a

[1] The estimate of social costs includes a countervailing fiscal externality of premature mortality to Social Security.

[2] Specifically, Gruber and Kőszegi (2001) focus on an empirical prediction that rational smokers will exhibit forward-looking behavior with respect to future cigarette prices (Becker et al., 1991; Chaloupka, 1991).

random subset verified their smoking status using a saliva test each week. We used a novel lottery-based design for remote monitoring to ensure strict incentive-compatibility at all stages of the experiment, including incentivizing accurate reports of smoking status.[3] We offered abstinence incentives to all subjects during randomly selected weeks of Months 1 and 3, which enables us to estimate the net cost of abstinence under subjects' (biased) preferences. Crucially, we elicited subjects' valuations of these future abstinence incentives at multiple time points and for varying incentive levels. We identify present bias and naïveté from a comparison of subjects' valuations of future abstinence incentives and their subsequent abstinence behavior in response to the abstinence incentives. In Month 2, we offered an abstinence intervention to a random subset of subjects ("treated" subjects), designed to encourage them to exit an addicted state. We identify the treatment effect from the difference-in-differences in *behavior* between treated and control subjects before and after the Month-2 intervention. We identify projection bias from the difference-in-differences in *valuations* between treated and control subjects before and after the Month-2 intervention. We adopt a pre-specified analysis plan to estimate reduced-form and structural estimates.

In the reduced-form estimation, we find that subjects are wildly over-optimistic about future abstinence—by nearly 40 percentage points, overall. They further mispredict the expected effect of the Month-2 abstinence intervention on their subsequent willingness to stay abstinent. Surprisingly, we find that treated subjects' beliefs become more pessimistic following the intervention month despite their actual probability of abstinence increasing.

By combining subjects' valuations of abstinence incentives—which provide imperfect commitment to their future selves—with their subsequent abstinence behavior, our experimental design allows us to structurally estimate subjects' preferences and biases while minimizing the need for arbitrary assumptions.[4] The structural estimates indicate that smokers in our sample have a severe degree of present bias regarding their smoking behavior: an average short-run discount factor $\beta = 0.67$, substantially below the time-consistent benchmark of 1, and with substantial heterogeneity between subjects. While this implies a large departure from time-consistency, it is in line with existing estimates.[5] Moreover, subjects are only partially aware of their present bias, with self-control beliefs $\tilde{\beta} = 0.85$ on average. Although this is below the "fully naive" level of

[3] In other words, subjects were incentivized to accurately report their smoking status.

[4] This approach follows a recent literature which notes that while a discrete choice for commitment is consistent with standard economic models of discrete choice, it is possible to use the willingness-to-pay for incentives to estimate present-biased preferences (Acland and Levy, 2015; Carerra et al., 2019).

[5] For example, Paserman (2008) finds $\beta \in [0.40, 0.48]$ for low- to medium-wage job-seekers; Bai et al. (2017) finds a mean $\beta$ of 0.37 among villagers seeking health care in India; and Laibson (1997) uses an estimate of $\beta = 0.6$.

$\tilde{\beta} = 1$, it is nevertheless above their actual short-run discount parameter for 92% of subjects. Our structural estimates confirm that even this degree of naïveté will induce substantial over-optimism about abstinence behavior, as we observe in the reduced-form results. Moreover, we find a near-linear relationship between individuals' degree of present bias and their beliefs, such that the partial naïveté in our sample of smokers is well-approximated by the relationship $\tilde{\beta} = 0.11 + 1.11\beta$. This relationship is itself a novel finding and is not imposed either by the structural model or theory (if, for example, the cognitive traits relating to a low $\beta$ were associated with the metacognitive traits associated with a high $\tilde{\beta}$).

Regarding the state-dependent prediction errors, we find that a simple model of projection bias does not account for the patterns in the data. Our structural model confirms that the Month-2 abstinence intervention led to a lasting increase in the relative benefit of abstinence, which we estimate to be equivalent in utility terms to 0.56 utils. We find that subjects' end-of-Month-1 valuations of the intervention are consistent with a treatment effect of only 0.13 utils on average, which would correspond to a degree of projection bias of $\alpha = 0.77$ and not significantly different from 1. In other words, we fail to reject that subjects completely project their long-run (pre-treatment) addiction onto how they expect to feel in a future (post-treatment) unaddicted state. However, subjects' end-of-Month-2 valuations—after the treatment has had its effect—are consistent with a (statistically insignificantly) *negative* effect of $-0.08$ utils, or a degree of projection bias of 1.15. Moreover, we can reject that the *ex-ante* and *ex-post* valuations embed the same predicted treatment effect at $p = 0.08$. This suggests that smokers' decisions are subject to a double-hit of pessimism. The intervention is effective at increasing Month-3 abstinence, yet before the intervention they believe it will have no effect, and after the intervention they believe it has made things worse.

Finally, we are able to use the structural estimates to conduct a welfare analysis using subjects' own preferences. We find that, under subjects' own long-run preferences, the private welfare loss of continued smoking is on average approximately \$400 per week. This figure incorporates the effect of abstinence on lowered future addiction, and could not have been estimated without our Month-2 intervention.

We focus on present bias and projection bias for four main reasons. First, they compose the minimal set of departures from the neoclassical benchmark that are needed to accommodate intertemporal and state-dependent prediction errors. Second, these biases are both potentially important for assessing the welfare consequences of addictive goods, and empirical work has not quantified the extent of these biases for welfare calculations. Third, any estimate of projection bias based on predictions would be contaminated by present bias, and mis-specified if researchers failed to account for naïve beliefs about

present bias. This is a notable reason why field-based estimates of projection bias are so rare. We leverage our identification strategy over present bias to estimate the state-dependent prediction error as well, filling an important gap in the literature. Finally, the implications of these models for actual behavior are highly ambiguous, and depending on modeling assumptions can lead to vastly different conclusions about counterfactual predictions, welfare, and optimal policy.[6]

Present bias has long been believed to be a driver of smokers' decisions (Schelling, 1980; Gruber and Kőszegi, 2001). According to this view, addicted smokers will seek to avoid the immediate disutility of quitting by continually revising their plans to quit. Several lines of evidence have been marshaled in support of the link between present bias and tobacco consumption, including smokers' use of commitment devices and high time discounting rates.[7] Present bias may be especially harmful for agents who are not fully aware of their future self-control problems. Such naïve individuals can delay following through on preferred actions today in expectation of performing the action tomorrow (O'Donoghue and Rabin, 1999). Smokers' widespread regret about having started to smoke and low follow-through rates on quit attempts may be indicative that they suffer from some degree of naïveté, or over-optimism (Fong et al., 2004; Hughes et al., 2004). In other health domains such as physical activity, the evidence for naïveté or sophistication has often been mixed (DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Royer et al., 2015).

Smokers may suffer two general forms of projection bias, one involving a long-term transition away from being addicted and one involving short-term fluctuations in craving.[8] In a "hot" addicted state, smokers may fail to predict how their preferences will change once they exit into a "cold" unaddicted state. In particular, they may underestimate the benefits of quitting and their subsequent willingness to stay abstinent. On the other hand, smokers in a "cold" low-craving state may fail to anticipate their behavior in a "hot"

[6] For example, smokers may be pessimistic about future self-control if they are sophisticated, in which case they would not quit in anticipation of failed future attempts. Smokers may be overly optimistic about future self-control if they are naïve, in which case they do not quit in anticipation of successful future attempts. Smokers may be pessimistic about the benefits of quitting if they project addiction. Smokers may be overly optimistic about future consumption if they project satiation.

[7] Several studies have noted smokers' use of commitment devices as an indicator of their degree of present bias. Wertenbroch (1998) shows that some smokers impose self-rationing on themselves by purchasing smaller quantities of cigarettes (e.g., packs instead of cartons) in order to avoid consuming more than prefer *ex ante*. Gruber and Mullainathan (2005) finds that likely smokers have higher subjective ratings of happiness after the passage of tobacco control legislation, which the authors interpret as support for the notion that the laws act as a commitment device for smokers. More direct evidence comes from Giné et al. (2010), in which 10% of smokers agree to take up a commitment contract for smoking cessation. Smoking is also associated with high discount rates for monetary rewards (Chabris et al., 2008; Khwaja et al., 2007; Bickel et al., 1999).

[8] Loewenstein (2005) refers to these general forms as the hot-to-cold empathy gap and cold-to-hot empathy gap.

high-craving state and thereby *overestimate* their future willingness to abstain. Projection bias therefore fails to deliver a clear prediction without further information about the relevant state. Studies have provided evidence on the projection bias of heroin addicts, hungry workers at snack time, and even catalogue shoppers (Badger et al., 2007; Read and Van Leeuwen, 1998; Conlin et al., 2007). For smokers, a cold-to-hot empathy gap has been demonstrated with regard to very short-term craving (Sayette et al., 2008), but it remains unclear whether the short-term craving or longer-term addiction is the relevant state for the cessation decision.

Behavioral economists have shown increasing interest in structural estimation of behavioral parameters as a way to gain traction on otherwise underspecified problems (DellaVigna, 2018). Structural modeling holds the potential to go beyond simple treatment effect estimates to address a number of thorny issues, such as why lasting behavior change (especially around habits) is so difficult to achieve, what are the welfare effects of behavioral interventions, and what mistake is being corrected in behavioral interventions. Answers to these questions are critical for behavioral economists and policymakers to develop and test more effective interventions.

While structural estimation has a number of advantages over reduced-form estimation, it also carries important limitations. Structural methods often rely on weak identification strategies and assumptions that have hidden or complicated implications for interpretation. Chetty (2009) has advocated sufficient statistics as a middle ground between structural and reduced-form methods. DellaVigna (2018) has suggested that an alternate approach involves leveraging experimental variation for structural model identification. While experimental economics has a long tradition of employing these methods, only relatively recently have behavioral economists designed field experiments with structural estimation in mind (some early examples being Chetty et al. (2009) and DellaVigna et al. (2012)).

A key benefit of estimating structural parameters for both utility and behavioral biases is that we can estimate the welfare effects of smokers' biases *under their own preferences*. This avoids any need to impose external judgments about the costs and benefits of smokers' decisions. We estimate that the private welfare loss of continuing to smoke is \$414 per week, under subjects' long-run preferences (absent present bias) in an unaddicted state (absent projection bias). Smokers choose to continue smoking in most weeks, however, because of the two biases. Present bias lowers the discounted value of quitting by \$2,635 (by discounting the long-run benefits of abstinence by $1-\beta$). Projection bias lowers the value of quitting by \$390 (i.e., the difference between anticipated and actual utility from the abstinence intervention). If one divides the welfare loss proportionally over the average number of cigarettes smoked at baseline, it corresponds to a private welfare loss of \$80 per pack.

The remainder of the paper is organized as follows. Section 2 discusses the experimental design and data. Section 3 presents our descriptive and reduced-form results. Section 4 describes our conceptual model incorporating present bias and projection bias. Section 5 presents our structural estimates. Section 6 concludes.

# 2 Experimental Design

To understand beliefs and preferences in smoking decisions, we conducted a longitudinal experiment over 12 weeks (Figure 1). Subjects completed short weekly surveys and three longer online surveys at baseline, at the end of Month 1, and at the end of Month 2. During each of the three longer sessions, we elicited subjects' willingness to pay for abstinence incentives in future weeks. We then implemented a novel incentive-compatible methodology (described below) to remotely verify actual smoking behavior, in which we collected self-reports about subjects' smoking status on a weekly basis and biochemically verified a random subset of subjects each week. We also offered abstinence incentives during certain weeks. By using lottery-based incentives, we ensured that any abstinence incentive was confirmed by biochemical verification. Random variation in the timing and amount of incentives enabled us to compare both the predicted versus actual probability of abstinence (an indication of dynamic inconsistency) and the changes in predictions for subjects who are in different states of craving and addiction (an indication of projection bias).

## 2.1 Recruitment and Eligibility

We recruited 397 smokers from a web-based subject pool maintained by Nielsen Healthcare and its partners. Nielsen invited individuals from 16 large metropolitan areas in the U.S.[9] Eligibility criteria included smoking cigarettes at least 20 of the prior 30 days, being age 21 or older, residing in one of the 16 metropolitan areas, having access to a smartphone or tablet camera, and agreeing to receive in-person visits from a study enumerator.[10] Furthermore, to qualify for the study, each person had to verify that they were a current smoker using a saliva

[9] Metro areas correspond to Nielsen's designated market areas (DMAs) that share a common media market. Metro areas were selected for inclusion in the study to ensure geographic coverage and based on Nielsen data of where likely smokers lived. Included DMAs were: Atlanta, Boston, Chicago, Cleveland, Dallas-Forth Worth, Denver, Detroit, Houston, Los Angeles, Miami, New York, Orlando, Philadelphia, Phoenix, Seattle, and Tampa-St. Petersburg. Study zip codes are mapped in Figure A.1.

[10] We excluded pregnant women, individuals who intended to relocate to another metropolitan area within the next three months, and individuals who were not willing to provide saliva samples for testing cotinine levels.

test of cotinine, a metabolite of nicotine.[11] We mailed two test saliva kits to all individuals who were deemed eligible, one kit for qualification and one for the experiment.[12] Subjects had seven days after enrollment to upload photographic documentation of a positive salivary cotinine test in order to demonstrate their smoking status and fully qualify for the study. This also provided a photograph of each subject's face that we used to verify their identity during each subsequent instance of verification.

Subjects were eligible to receive multiple types of incentives throughout the study. For the purposes of identification, there are four main types of incentive: 1) incentives to verify smoking status, 2) payments offered during elicitation sessions to ascertain valuations of future abstinence incentives, 3) actual abstinence incentives offered during the trial to measure actual willingness to abstain, and 4) abstinence incentives during the Month-2 intervention to encourage a change in addiction state. We describe each incentive type and the associated procedures below.[13]

## 2.2 Verification of Smoking Status

A key challenge for remote data collection is maintaining incentive-compatibility, i.e., an incentive to provide accurate reports of smoking status and valuations. Subjects might be especially likely to misreport their smoking behavior during weeks when they are paid for abstinence. Repeated in-person testing can mitigate this risk, but also imposes substantial burden on subjects—likely leading to a highly selected sample through non-random attrition—and is often prohibitively expensive. We use a novel approach to measure smoking status remotely while maintaining strict incentive-compatibility for all reports and valuations.

In each weekly survey, we asked each subject whether they had smoked during the prior 7 days (a measure known as 7-day point-prevalence abstinence). In order to ensure that truth-telling was strictly incentive-compatible, we implemented a pair of "truth-telling" lotteries in each week. The first lottery awarded $50 to one person per week if the selected person's saliva test matched their self-report, regardless of whether the person had smoked during the prior week. The second lottery, drawn among self-reported abstainers, awarded a $100 payment

[11] We used the Alere™ iScreen® OFD test kit for salivary cotinine. The test involves simple collection and testing procedures, and produces qualitative results within 10 minutes, with reliable remote detection at cotinine levels of 30 ng/mL (Moore et al., 2018).

[12] This ensured that all subjects always had an available saliva test kit in their possession. If a subject used their second test kit later in the experiment, a replacement was immediately sent via two-day mail.

[13] In addition to the study incentives described below, subjects received $5 per week (up to $60 total) for completing a weekly survey of self-reported smoking status, $15 each for completing the surveys at the end of Months 1 and 2, and entry into an end-of-study lottery among those who complied with all aspects of the trial. The lottery consisted of five $100 prizes and one $1,000 grand prize.

to one person per week if the selected person both reported abstinence and tested negative for smoking in a saliva test. Thus, the first lottery gave subjects an incentive to report their smoking status accurately, and the second lottery broke the indifference to make it preferable to be (and to report being) abstinent. Note that this implies that any subject who believed our testing procedures could be gamed had a strict incentive to self-report abstinence during all weeks. However, we find that 98% of subjects reported smoking in at least one weekly survey, suggesting that they believed smoking could be detected. Overall, we observe low self-reported abstinence rates; only 10% of control subjects reported abstaining in at least one week. Moreover, each week we audited 3% of subjects who were not selected for that week's lotteries to provide a saliva test in order to monitor the accuracy of this procedure.

The saliva testing procedure itself followed two stages. In the first stage, all selected subjects were informed at the end of the weekly survey to complete a self-administered saliva test within 12 hours. In the second stage, a random subset of these subjects (one per metro area per week) was asked to complete a second round of saliva testing within 48 hours during an in-person visit from a study field worker. If selected for an in-person test, the subject had to test negative for cotinine during both rounds of testing in order to receive any smokefree-contingent incentives. This second stage was used to validate subjects' self-testing, and, moreover, was made salient to discourage any attempts at deception in the first place. Subjects received a non-contingent $20 payment for each self-administered test and each in-person test completed.[14] A replacement kit was sent by two-day mail to all subjects who were selected for a test.

The testing procedure at both stages involved using a smartphone app to upload a series of three geocoded photographs: one photograph of the person's face while swabbing a cheek with the test sponge, one photograph of the test result window, and one photograph of the window blacked out with a pen to prevent reuse (Figure A.3). A similar approach has been adopted with success in other smoking cessation trials (Ramo et al., 2018). This was the same procedure as the qualification testing, which ensured that all subjects were able to complete the test. In assessing the weekly test results, we further verified the person's identity by comparing their face to the the face shown in the photograph from the initial screening test.

## 2.3 Abstinence Incentives in Months 1 and 3

We offered abstinence incentives to subjects in Months 1 and 3, in order to measure their actual willingness to pay for abstinence incentives. Specifically, each subject was eligible

[14] DellaVigna et al. (2016) estimated that the cost of a home visit from a door-to-door fund-raiser is $10. Our visits are arguably more intrusive, so we increased the compensation accordingly.

for a smokefree-contingent payment during one randomly selected week of Month 1 and one randomly selected week of Month 3. The selected weeks were drawn from among the target weeks about which a valuation had been provided during the baseline elicitation session. The payment amounts, selected at random, varied from \$10 to \$400.[15] Each payment was conditioned on self-reported 7-day abstinence at the end of the target week, verified within 12 hours by saliva test according to the verification procedures outlined above.

Although our main focus will be on the incentives in Month 3, we included the Month-1 incentives in order to familiarize subjects with the study procedures and to establish trust that the incentives would be paid as described in the surveys.[16] Moreover, although the great majority of subjects had reported previous quit attempts during the baseline survey, the Month-1 incentives also ensured that all subjects had experienced trying to abstain from smoking in return for monetary incentives.

## 2.4 Abstinence Incentives During Month-2 Treatment

In Month 2, we offered weekly abstinence incentives to subjects randomly assigned to the treatment group in order to induce them to abstain, thereby creating an exogenous reduction in their level of addiction relative to the remaining subjects.[17] Projection bias is predicated on the existence of a state change, and we use the changes in abstinence valuations across states to draw inferences about biased beliefs regarding abstinence.

As part of the end-of-Month-1 survey, we randomly assigned subjects to the control group or treated group in a 1:2 ratio. Assignment was performed using a rerandomization approach that balanced on several covariates, including baseline cigarette consumption and quit intentions (Morgan and Rubin, 2012, 2015).[18] Control subjects had no intervention during Month 2. Treated subjects were paid for each week they abstained during Month 2. They were entered into a weekly drawing with a 25% chance of winning \$100 and a 75% chance of winning \$20. Thus, they were guaranteed at least \$80 and eligible for up to \$400 in exchange for abstaining during the entire treatment month. Abstinence was self-reported, subsequently verified by saliva test. Alongside the payments, we referred treated subjects to

[15] Possible payment amounts (probabilities) were: \$10 (25%), \$25 (25%), \$50 (33%), \$100 (6%), \$150 (4%), \$200 (2%), \$300 (2%), and \$400 (2%). Each subject was assigned to one "small" payment $\leq$ \$25 and one "large" payment $>$ \$25.

[16] All study payments were deposited in the subjects' existing account with Nielsen.

[17] A duration of four weeks was expected to be sufficiently long to induce a state change among most smokers, based on known relapse patterns over time (Hughes et al., 2004).

[18] Balancing covariates were: metropolitan area, income categories ($<$ \$20,000, \$20,000-29,999, \$30,000-39,999, \$40,000-49,999, \$50,000-59,999, \$60,000-74,999, \$75,000-99,999, $\geq$ \$100,000, decline to answer), average cigarettes per day, number of the last 30 days with electronic cigarette use (0, $\leq$ 10, 10-20, $>$20), and an indicator for plans to quit smoking within 6 months. Covariates were reported during the screening or baseline survey. Among the 1,000 iterations, we implemented the one with the maximum joint $p$-value.

web-based smoking cessation support.[19]

## 2.5 Eliciting Valuations

We elicited valuations of future abstinence incentives during three sessions (baseline, end-of-Month-1, and end-of-Month-2) as part of an incentivized choice experiment. At times, we refer to the valuations as "incentivized predictions" of future abstinence. In each session, subjects made sequential binary choices in a "staircase" or "unfolding brackets" format. Each choice was between a smokefree-contingent payment $p$ or a non-contingent payment $q \in [0.1p, 1.1p]$ during a future target week of the trial. A series of 3-4 choices identified a unique indifference point among the 12 outcome "rows" that represents the person's valuation of abstaining during that target week (Figure 3). The starting point in the staircase was randomized during each iteration to mitigate anchoring effects. In each staircase, there was a different random draw of $p$ for a different randomly selected week. The respondent completed four staircases per elicitation session: at baseline, each subject was randomly assigned two target weeks in Month 1 and two in Month 3; in the other elicitation sessions, target weeks included all four weeks of Month 3.[20] The ordering of staircases was randomized to be either chronological or reverse-chronological in order to account for possible order effects.

We used a random lottery incentive system for incentive-compatibility. In each elicitation, we randomly selected one subject $\times$ week $\times$ row, and we gave the winning subject the (implied) choice from that row during the future target week.

We also elicited unincentivized predictions about future smoking behavior. At the end of each staircase, subjects were asked to predict directly how likely (in 5% increments) they would be to abstain in exchange for a payment $p$ during the future target week. The unincentivized elicitation incorporates the predicted behavioral response to the contingent payment, but not the utility consequences of the payment. These unincentivized predictions can therefore be directly interpreted as predictions about the probability of abstinence conditional on receipt of payment $p$, and moreover we compare them against subjects' valuations to provide reduced-form evidence on the extent to which subjects valued being incentivized to abstain. We will not use them in our structural estimation, however, as truth-telling is not strictly incentive-compatible and indeed it is natural to

[19] Subjects were referred to `Smokefree.gov`, a national service sponsored by the National Cancer Institute that provides cessation support via internet, text message, mobile app, instant message, and phone, and to the BecomeAnEx Program (`becomeanex.org`), a vibrant online social community sponsored by the Truth Initiative and the Mayo Clinic that provides cessation support via text and email. Both resources are free to use. Research shows that Smokefree.gov can promote long-term smoking abstinence (Bricker et al., 2017).

[20] Values of $p \in \{\$10, \$25, \$50, \$100, \$150, \$200, \$300, \$400\}$ were drawn in each elicitation session to assign each subject two "small" values $\leq \$100$ and two "large" values $> \$100$.

suspect that the unincentivized predictions may be subject to a social desirability bias or experimenter demand effects.

# 3 Results

## 3.1 Descriptive Statistics

Table 1 contains baseline characteristics of the sample. Relative to a nationally representative sample of U.S. adult cigarette smokers, our sample contained fewer men (34% in sample vs. 55%), a more compressed age distribution (e.g., 20% vs. 28% age >55), and more highly educated (25% vs. 55% with high school degree or less) and higher income (47% vs. 69% income < $50,000) subjects (Kasza et al., 2017). These differences are not surprising, however, as our sample is conditional on smartphone usage and willingness to participate in an online panel. Nevertheless, some caution must be exercised if applying the study results to the population of smokers as a whole. Demographic and smoking-related characteristics were well balanced across treatment groups. The $p$-value from a global $F$-test of balance by group is 0.74.

We first present an overview of the data on abstinence predictions and abstinence behavior. Figure 4 characterizes subjects' incentivized and unincentivized abstinence predictions by incentive level.[21] The significant positive slope for both types of predictions is reassuring, and confirms that subjects believe they will be more likely to abstain at larger incentive levels. The downward revision in predictions between the baseline and Month-1, and again between Month-1 and Month-2, elicitation sessions implies that subjects became more pessimistic about their ability to abstain as the trial progressed. For example, perfect compliance would require a payment of approximately $1,000 according to the baseline unincentivized predictions, but $1,800 according to the Month-2 predictions. Finally, we note that the abstinence supply curve is significantly steeper for the unincentivized predictions than the incentivized predictions in all three sessions ($p < 0.01$ for each), highlighting that the incentivized predictions incorporate the utility consequences of abstinence, namely the disutility of effort involved in abstaining.

We next turn to comparing abstinence predictions with actual abstinence behavior in the absence of incentives. Figure 5 contrasts smoking behavior to predictions about behavior by

[21] Throughout this section, we will refer to subjects' valuations of Month-3 incentives, divided by the incentive level, as their "incentivized predictions." This is approximately equal to their perceived probability of abstinence, but also includes subjects' perceived utility resulting from the effect of the incentive on their abstinence probability. This second effect will be used in our structural estimation, particularly Equation (10).

week and by treatment group. For each study week, we plot the average of subjects' baseline abstinence predictions about that target week as well as the average abstinence rate observed during that week.[22] We also plot the average end-of-Month-1 and end-of-Month-2 predictions about target weeks during Month 3. Subjects are severely over-optimistic about their ability to abstain, and this naïveté generally persists even after they have had experience with the study procedures during Month 1. Whereas unincentivized predictions of future abstinence start off around 50-60% at baseline, they remain above 35% when elicited at the end of Month 2. In contrast, observed abstinence never exceeds 20% for a given week. The time series for treated and control groups are broadly similar in Month 1. However, the abstinence rate of the treated group starts to climb over the course of Month 2, reflecting the treatment effect, and the treatment effect persists and even grows in Month 3, indicating that there was a shift in addiction state among some treated subjects leading to their continued abstinence. Relative to the control group, the treated group exhibits more pessimistic beliefs during the Month-2 elicitation than the Month-1 elicitation. This double-difference, formally tested below, would provide *prima facie* evidence of projection bias, as it would imply that treated subjects had projected their beliefs about the benefits of quitting from their addicted state onto a future in which they were no longer addicted.

Finally, we examine abstinence behavior versus predictions by incentive level and treatment group in Figure 6. As suggested by the time series, predictions of abstinence are far higher at every incentive level than observed abstinence behavior. Greater pessimism at the end of Month 2 relative to the end of Month 1 is also apparent for treated subjects at all incentive levels. The abstinence supply curve is somewhat flatter for observed behavior, although data on the behavioral response to incentives over $50 are sparse (see footnote 15).

## 3.2 Reduced-Form Estimates

### 3.2.1 Evidence on Addiction

We begin by testing the effect of our Month-2 intervention on the treated group. In Table 2, we test whether the Month-2 intervention leads to a continued increase in smoking abstinence during Month 3 in the treated versus control group. This would indicate that treated subjects are made relatively less addicted by their higher abstinence following the treatment month. We regress an indicator of weekly smoking abstinence on an indicator for treatment status, an indicator for month, and the interaction effect of these two. The treatment effect is

[22] Figure 5 plots subjects' unincentivized predictions for ease of interpretation. Incentivized predictions are shown in Figure A.4 and are qualitatively similar.

the coefficient on the interaction term, analyzed on an intent-to-treat basis. If a subject did not report their smoking status during a weekly survey, we assume that they did not abstain during that week.[23] The model includes week and individual fixed effects, clustering standard errors by individual.

In column 1 of Table 2, we restrict the sample to Months 1 and 3, as the simplest test for a treatment effect is to compare pre-treatment and post-treatment outcomes based solely on experimental randomization. We find that abstinence increases in the treated group by 5.8 percentage points in Month 3 relative to the control group ($p = 0.02$), although in fact abstinence in both groups increased relative to Month 1. In column 2, we include Month-2 data as well. This has no effect on the estimated treatment effect given the inclusion of month indicators, but allows us to examine behavior during the Month-2 intervention. We find that abstinence increased in Month 2 for both groups by 10.5 points. Somewhat surprisingly, the interaction term on $Treated \times Month\ 2$ is not statistically significant, implying that we cannot reject a null effect of the \$100 Month-2 incentives on the treated group's Month-2 behavior. A likely explanation of this comes from examining the time series in Figure 5; it shows a clear upward trend of abstinence for the treated group during Month-2. Although the overall average during the month is not significantly different, this would be consistent with treated subjects taking multiple weeks, and potentially repeated quit attempts, before successfully abstaining. This would also explain the surprising finding that the estimated Month-3 treatment effect is larger than the estimated Month-2 effect, although the difference is not statistically significant.[24]

Finally, in columns 3 and 4, we test whether the treatment effect varies by incentive level, interacting it with the treatment $\times$ month indicators. We find that the response of treated subjects during Month 3 is marginally increased by 3.8 points ($p = 0.09$) for each additional \$100, using all available data. The main effect, however, is a 6.5-point relative increase in the probability of treated subjects abstaining during Month 3 in the absence of any incentives. Taken together, these results show that the intervention exogenously increased the likelihood that treated subjects exit an addicted state relative to controls.

### 3.2.2 Evidence on Overoptimism

We next turn to reduced-form evidence that smokers are naïve, i.e., overly optimistic, regarding their future smoking behavior. We test the hypothesis that the observed behavioral response to incentives (actual willingness to accept to abstain) does not match

[23] The analysis is robust to dropping these week $\times$ subject observations, or to dropping all subjects who fail to answer at least one weekly survey.

[24] Our pre-analysis plan did not cover within-Month-2 behavior or comparing Month-2 and Month-3 behavior, and we therefore leave a fuller investigation of these results to future work.

their stated predictions of willingness to accept. To perform the test, we restructure our data set to stack observed abstinence behavior on top of the abstinence predictions to create a long data set of 7,185 observations. We regress an indicator of weekly smoking abstinence on an indicator equal to 1 for observed abstinence and 0 for incentivized abstinence predictions. The coefficient on the *Behavior* dummy identifies the difference between abstinence predictions and abstinence behavior, and the coefficient on the *Behavior* × *Incentive* interaction identifies how this difference changes with the incentive level.

The results of this regression are presented in Table 3. Overall, subjects are highly over-optimistic about future abstinence. In columns 1 and 2, we restrict attention to the control group to avoid any possible confounds from treated subjects' extra incentives during Month 2. Column 1 estimates the average difference between abstinence predictions and behavior across all weeks and all incentive levels. We find that subjects overestimate the probability of abstinence by 42.9 percentage points relative to their observed abstinence behavior ($p < 0.001$). In column 2, we introduce a covariate for incentive level and interact it with *Behavior* to determine whether the prediction error changes with incentive level. We find that over-optimism grows by 5.0 points for each additional \$100 ($p = 0.001$). These effect sizes are attenuated slightly when we also include treated subjects in the regressions in columns 3 and 4, but predictions remain far above behavior. Across all subjects and all weeks, column 4 indicates an average overestimation of abstinence of 28.6 points without incentives ($p < 0.001$) plus an additional 4.2 points per \$100 incentive ($p < 0.001$).

This systematic pattern of misprediction is highly suggestive that subjects are violating rational expectations about their future smoking behavior. More specifically, subjects are acting consistently with a form of present bias in which subjects have at least partial naïveté about their ability to overcome short-run withdrawal costs, leading them to be overly optimistic about their likelihood of abstaining. Existing evidence of smokers' use of commitment devices and near-universal regret about having started to smoke support the interpretation of partial naiveté (Giné et al., 2010; Wertenbroch, 1998; Gruber and Mullainathan, 2005; Fong et al., 2004). However, it is impossible to separate the degree of present bias from the perceived costs and benefits of abstinence just from these reduced-form results. In Section 5, we will be able to separately identify both the degree of present bias and the extent of naïveté.

### 3.2.3 Evidence on Projection

Next, we assess whether smokers predict the habit effect found in Section 3.2.1. In other words, we ask whether smokers mispredict the effect of the Month-2 abstinence intervention

on their subsequent willingness to stay abstinent. To address this question, we test whether treated subjects revise their abstinence predictions following the end of the treatment month (Month 2), compared with the control group. Doing so would mean that they had failed to fully predict the effect that habit formation has on their smoking behavior. We would expect that, because the treatment raised the actual probability of abstinence, treated subjects may revise their predictions upward if they did not fully predict its effect.

We implement this test using a difference-in-differences estimator of predictions before and after the treatment month for control and treated subjects in Table 4. We regress subjects' abstinence predictions on indicators for the baseline and end-of-Month 2 elicitation sessions, interacted with assignment to the treated group, adjusting for week and individual fixed effects. The omitted category is therefore the end-of-Month-1 elicitation session, and thus the coefficients represent differences relative to this session. Columns 1 and 3 use subjects' unincentivized predictions as the dependent variable, while subjects' incentivized predictions (normalized valuations) are used in columns 2 and 4.

We first confirm the pattern from Figure 4, that subjects' baseline predictions were significantly higher than their later predictions. In column 1 we find that subjects' baseline predictions were on average 10.1 percentage points higher than their end-of-Month-1 predictions ($p < 0.001$). This is not surprising given the lack of familiarity with the abstinence incentives, and was the motivation for including the Month-1 abstinence incentives in the experimental design. When accounting for different incentive levels in column 3, we find that treated subjects were 7.5 points less over-optimistic than control subjects at baseline for weeks with no incentive, but that this is counterbalanced by an increased prediction of their response to incentives of 1.7 points per \$100. Thus, while there is no difference on average between treated subjects and controls at baseline, it is possible that treated subjects believed they would grow more responsive to the incentives—factoring in the removal of their Month-2 incentive.

The key test, however, is the difference-in-differences between treated and control subjects in the end-of-Month-1 and end-of-Month-2 surveys. We find in column 1 that treated subjects revised their predictions downwards by 4.9 percentage points. This is confirmed in column 3, where we moreover find that this downward revision is comparable across all incentive levels. Given that the treatment itself actually *increases* the probability of Month-3 abstinence, this is a somewhat surprising result. One possible interpretation would be that treated subjects are now in a nicotine-deprived state of withdrawal, and are projecting those cravings onto their Month-3 utility. Alternatively, they may have been overly optimistic about their probability of abstinence during Month 2, and are now rationally predicting their behavior conditional on having not abstained during the prior

month. Note that the former interpretation implies that the revised predictions are too pessimistic, and the latter that they are correct (relative to control subjects, given that both groups are present-biased). The reduced-form difference-in-differences cannot separate these alternatives, but we will be able to directly estimate treated subjects' beliefs about the treatment effect in Section 5.

An alternative hypothesis that could explain a downward revision in predictions would be if treated subjects learned about their self-control problems during Month 2 and thus became relatively more sophisticated.[25] Although the existing evidence for such learning is quite thin and, in theory, precluded by the degenerate $\tilde{\beta}$ beliefs of the O'Donoghue and Rabin (2001) model, it is nevertheless worth considering. Indeed, the 10-point drop in predictions for both treated and control groups between the baseline and end-of-Month-1 elicitation reported in Table 4 suggests that some form of learning is quite natural in this context. As we will show in Section 5, however, the difference between subjects' unincentivized predictions and their normalized valuations reveals their demand for commitment. This demand for commitment is increasing in $p$ but at a decreasing rate as $\tilde{\beta}$ increases. We can thus test for learning about $\tilde{\beta}$ by running a triple difference-in-differences between treated and control subjects, unincentivized predictions vs normalized valuations, during the end-of-Month-1 and the end-of-Month-2 surveys. The coefficient on the triple-interaction term is 0.011 with a p-value of 0.696, indicating that we find no evidence of increased sophistication.

### 3.2.4 Evidence on Heterogeneity

Finally, we investigate possible dimensions of heterogeneity in over-optimism focusing on pre-specified baseline characteristics that fit into three categories: demographics (age, gender, household income, and education), addiction-related behavior (cigarettes per day, nicotine dependence, quit plans, and present bias in a hypothetical choice task), and other health behaviors (alcohol use, sunscreen use, and tendency to overeat). We selected the set of covariates to match those used in the structural analysis.[26] We dichotomize the continuous variables at their median and code agreement with the Likert-scale questions as 1. It is plausible to expect that more heavily addicted subjects, as revealed either by a higher nicotine dependence score or no plans to quit smoking, may be less accurate in their predictions. Other research has suggested that demographic and economic covariates may also be related to biases, although we make no causal claims about these relationships. Finally, the unrelated

[25] We thank Justin Sydnor for this hypothesis.

[26] The structural model incorporates the most comprehensive set of covariates for which the model would converge. Two potential covariates were excluded as a result: regret about having started smoking due to a lack of variation (86% of subjects express regret) and dual use of e-cigarettes due to its heavy skew (only 26% used any e-cigarettes).

health behaviors may be associated either with the same behavioral biases as are potentially relevant for smoking, or may also reveal differential health attitudes overall.

As in the regressions in Table 3 for measuring over-optimism, we use a long data set that stacks behavior and predictions in order to test for heterogeneity. We regress a weekly indicator for abstinence on an indicator equal to 1 for *Behavior* and equal to 0 for end-of-Month-1 or end-of-Month-2 predictions, an indicator for the "high" value of the heterogeneity dimension, and the interaction. Thus the *High group* indicator identifies group differences in predictions, and the *Behavior* × *High group* interaction identifies group differences in over-optimism. All models include individual random effects.

In Table 5, we show the results for heterogeneity in over-optimism. We find no differences in either behavior or overoptimism across age, gender, income, or education dimensions. In contrast, we find that more nicotine-dependent are unsurprisingly less likely to abstain, but, more importantly, also over-predict abstinence by 8 percentage points compared with less-dependent smokers ($p = 0.004$). Subjects planning to quit smoking within a year are 18 points more likely to abstain, and in fact are 11 points less over-optimistic about their abstinence. Finally, we see that regular sunscreen users are also more likely to abstain — possibly due to relatively greater health consciousness — but are no more over-optimistic than non-sunscreen-users. We emphasize that these bivariate regressions should not be interpreted as causal, nor do they account for the correlations among the covariates. However, they provide suggestive evidence for heterogeneity among our subjects that we return to in our structural estimation.

# 4 Conceptual Framework

To identify the specific policy parameters of cost and bias, we combine the information from the reduced-form tests in Section 3 with a structural analysis. We adopt a structural model that allows for three departures from a standard neoclassical model: present-biased preferences, naïveté regarding one's degree of present bias, and projection-biased beliefs.

Consider a decision-maker who may choose whether to abstain from smoking during a given week and earn $p$, or continue smoking. We normalize the utility of continued smoking to zero. If the decision-maker abstains, they suffer an immediate disutility (withdrawal) of $c$. They also obtain long-run benefits $b$, which would represent the continuation value of having abstained in a full dynamic problem. A key element of our approach is that we do not need to estimate that dynamic model, as the experimental variation in Month-3 incentives, the Month-2 treatment, and subjects' valuations of the Month-3 incentives provide sufficient identification. We can therefore be agnostic about this continuation value

and instead estimate it directly as the "reduced-form" parameter $b$. We model $b$ as arising in the subsequent period (week), implicitly incorporating any long-run discounting over a stream of benefits. Moreover, $b$ reflects the decision-maker's *perceived* benefits, which may differ from the true benefits (e.g., change in mortality risk $\times$ discounted value of life) either because of information about these benefits or because they mis-perceive their future behavior. When using $b$ in welfare analysis, therefore, we are considering the benefits of abstinence as perceived by our sample of smokers.

We further assume that the agent's utility is quasi-linear in money, at least over the range of incentives we provide. Although in principle it would be possible to allow for curvature in the utility from money, any meaningful departure from linearity over the range used in our estimation would soon run afoul of the calibration arguments in Rabin (2000). We therefore assume a constant marginal utility of $\gamma \in [0, \infty)$ for simplicity. Our experimental design will identify the utility of money off the abstinence incentives, and therefore in theory this utility could be non-parametrically identified over the [\$0, \$400] range of Month-3 incentives. In practice, we did not design the experiment with this goal and therefore lack sufficient variation in our sample to do so.[27] The limits to quasi-linearity in the identification of time preferences are well known in the experimental literature (e.g., Andersen et al., 2008; Andreoni and Sprenger, 2012), though usually misspecification is a concern in elicitation approaches involving money now versus money later. Moreover, the assumption of linear utility over the range of incentives will not lead to falsely rejecting the null of time-consistency ($\beta = 1$) in our setting.[28]

We assume that $c$ is uncertain and distributed according to some differentiable $F(\cdot)$ with mean $\mu_c$. The cost is realized in the period in which the agent decided whether to abstain. As is standard, we assume that the agent knows the distribution $F(\cdot)$. In general, it is impossible to disentangle biased beliefs about the future distribution of costs from biased beliefs about how one will respond to those costs, e.g., naïveté about discounting. Indeed, there is often an equivalence between the models. We therefore make the standard assumption that the departure from rational expectations is through mistaken beliefs about future discounting.

In particular, we consider an agent who may be present-biased, which we capture using the quasi-hyperbolic discounting model of Laibson (1997) and O'Donoghue and Rabin (1999). The agent is characterized by a short-run discount factor $\beta \in [0, 1]$, a long-run discount factor

[27] In particular, an alternative design in which each subject was offered multiple Month-3 incentives and where the upper limit on these incentives was greatly increased would be needed, and would prove prohibitively expensive to achieve this limited generalization. In our robustness checks however, we do estimate an alternative model with CRRA utility over money and obtain qualitatively similar results for the remaining structural parameters.

[28] Indeed, we show in Appendix C that the substantive assumption on utility is that it is additively separable in money and smoking.

$\delta \in [0,1]$, and a degenerate belief $\tilde{\beta} \in [\beta, 1]$ about future selves' impatience.[29] The agent's utility from smoking abstinence at price $p$ is:

$$\beta\delta b + \gamma p - c \tag{1}$$

and the agent believes a future self will obtain utility from abstinence at price $p$ equal to:

$$\tilde{\beta}\delta b + \gamma p - c. \tag{2}$$

We model the effect of the month-long abstinence intervention as an additive shock to the cost of smoking, $\eta$ (or, equivalently and perhaps more intuitively, a positive shock to the utility of abstinence – that is, the habit effect induced by the month-long abstinence intervention). If the effect of the intervention is to lower the subject's degree of addiction, then we expect this shock to be positive. If instead subjects experience increased craving, then the shock will enter negatively. Treated subjects following the intervention will receive utility from abstinence equal to:

$$\eta + \beta\delta b + \gamma p - c. \tag{3}$$

Finally, if subjects experience simple projection bias as in Loewenstein et al. (2003), then they will under-estimate the effect of the abstinence intervention on their future preferences, and therefore on their future behavior. In general, if a projection-biased agent of degree $\alpha$ is currently in state $s_0$, they will mis-perceive utility of being in state $s_1$ as being $\tilde{u}(x, s_1; s_0) = \alpha u(x, s_0) + (1-\alpha)u(x, s_1)$. If the effect of the intervention is additive, then an agent with projection bias of degree $\alpha$ believes that a future (post-intervention) self will only benefit from a $(1-\alpha)\eta$ shock rather than the full $\eta$. Rather than imposing this functional form on our subjects' state-dependent prediction error, we will instead estimate their pre- and post-intervention beliefs over their target-week utility as $\tilde{\eta}_1$ and $\tilde{\eta}_2$, respectively. That is, a subject in the treated group at the end of Month 1 perceives their Month-3 utility from abstinence as:

$$\tilde{\eta}_1 + \tilde{\beta}\delta b + \gamma p - c. \tag{4}$$

and a subject in the treated group at the end of Month 2 perceives their Month-3 utility from abstinence as:

$$\tilde{\eta}_2 + \tilde{\beta}\delta b + \gamma p - c. \tag{5}$$

If subjects exhibit simple projection bias, then the end-of-Month-1 perception identifies

[29] The restriction that $\tilde{\beta}$ lies between the true $\beta$ (full sophistication) and 1 (full naïveté) is the typical constraint imposed in theoretical analyses of present bias. We will not impose this restriction in our structural estimation, but will indeed find that it holds for 80% of our subjects' estimated parameters.

the degree of their bias, as $\tilde{\eta}_1 = (1-\alpha)\eta$. Having been through the treatment, subjects' end-of-Month-2 perceptions should not be distorted due to projection bias and we would predict $\tilde{\eta}_2 = \eta$. Because we estimate $\eta$, $\tilde{\eta}_1$, and $\tilde{\eta}_2$ separately, we will be able to test this prediction in the data.

The above framework allows stochasticity in subjects' behavior, but does not yet allow them to make errors in the valuation task. In order to admit variation in observably-equivalent subjects' valuations, we will assume that they "tremble" around their true valuation (or, equivalently, receive a shock to the utility of receiving the incentive), and that this is drawn from a one-parameter distribution $G(\cdot|\sigma)$.

Finally, as an experimental design choice, we will assume that $\delta$ is known and, in general, will impose $\delta = 1$. We emphasize that our experimental design could be extended to identify $\delta$ by varying the week in which incentives are received,[30] but did not consider doing so to be worthwhile in our particular context. First, as a theoretical standpoint, it is unlikely that there is significant long-run discounting over the course of a week, and thus we would not expect to find $\delta$ to be far from one. Indeed, any noticeable discounting over this horizon would yield implausible predictions once compounded over a reasonable horizon, which was in fact one of the motivations for the quasi-hyperbolic model in the first place.[31] Second, given our limited sample and the high cost of collecting additional data, it was deemed impractical to randomize over an additional dimension. We check the robustness of our results to a range of $\delta < 1$, however, and find qualitatively similar results.

The key parameters to identify in our structural exercise are therefore $\theta = \{c, b, \gamma, \beta, \tilde{\beta}, \eta, \tilde{\eta}_1, \tilde{\eta}_2, \sigma\}$. The limitations of discrete choice models, in particular, in identifying the level of utility as well as the discount factor(s), are well-known. These limitations have been highlighted recently in the literature on present-biased preferences by Carerra et al. (2019), who demonstrate that stochastic choice explains much of the demand for commitment contracts and recommend using demand for linear incentives as in Acland and Levy (2015). Our approach here is similar, using willingness to pay for Month-3 abstinence incentives.

## 4.1 Abstinence Behavior

We make standard parametric assumptions on the error structure of the model in order to identify the utility and discounting parameters. For the discrete choice of whether to abstain in a given week, we assume that the cost shocks for subject $i$ in week $t$ can be decomposed

[30] In particular, rather than paying the Month-3 "non-contingent" payments in the same week as an abstinence choice, one could randomize whether they were paid in that week or the following week.

[31] E.g., a weekly $\delta = 0.98$ would imply an annual discount rate of 186%.

into deterministic and stochastic components: $c_{it} = \mu_c + \varepsilon_{it}$. We further assume that $\varepsilon$ is distributed i.i.d. logistic, which yields the familiar logit formulation of choice probabilities in terms of the deterministic utility parameters. For a control subject, this is:

$$Pr_{control}(\text{abstain}|p, \theta) = \frac{\exp(\beta\delta b + \gamma p - \mu_c)}{1 + \exp(\beta\delta b + \gamma p - \mu_c)} \tag{6}$$

Inspecting equation (6) reveals how $\gamma$ is identified directly from the Month-3 abstinence incentives. Given that these incentives are randomly assigned, it is possible to estimate $\gamma$ by estimating the slope of the probability of abstinence as the incentive increases using a logit model. However, it also reveals why this approach would be insufficient to disentangle the cost and benefit of abstinence from discounting, as the intercept of this regression yields only $\beta\delta b - \mu_c$. This intercept is usually interpreted as the consumption value of continued smoking, and it is clear that if $\beta < 1$ then it will over-state the value under an agent's long-run preferences.

In principle, our assumption that $\varepsilon$ is distributed according to a logit distribution could be relaxed. Existing results provide conditions under which the shock distribution can be non-parametrically identified in binary and discrete choice models (Matzkin, 1993). It is advantageous, however, to have the closed-form solution to equation (6) and, moreover, our statistical power is increased by assuming a functional form. Our results are robust to alternative functional forms, for example, the probit probability model.

Treated subjects' abstinence behavior in Month 3 can be used to identify the treatment effect $\eta$. It is straightforward to extend equation (6) to find that treated subjects' probability of abstinence is given by:

$$Pr_{treated}(\text{abstain}|p, \theta) = \frac{\exp(\beta\delta b + \eta + \gamma p - \mu_c)}{1 + \exp(\beta\delta b + \eta + \gamma p - \mu_c)} \tag{7}$$

As in the reduced-form analysis, the difference between control and treated subjects in Month 3 reveals the effect of the treatment, and therefore equations (6) and (7) identify $\eta$.

## 4.2 Valuations

The key innovation in our identification of $\tilde{\beta}$, $c$ and $b$ is subjects' *valuations* of the Month-3 incentives. From the point of view of a subject during Month 1 or Month 2, the Month-3 incentives serve as partial commitment devices for their Month-3 selves. As the size of the incentive increases, so too does its effect on Month-3 behavior. If a subject perceives their future self as being dynamically inconsistent, then this behavioral response will be perceived as valuable by the earlier self who believes their future self is too likely to continue smoking.

If a subject perceives their future self as being dynamically consistent (e.g., is fully naïve), then this behavioral effect is perceived as an inefficient distortion of a previously optimal policy.

Formally, a subject in the control group perceives that their Month-3 self will receive utility from abstinence according to equation (2), and will therefore abstain for all realizations of $c \leq \tilde{\beta}\delta b + \gamma p$ if holding the incentive, and for $c \leq \tilde{\beta}\delta b$ without the incentive. Their expected utility, from the point of view of the earlier self, will therefore be:

$$U(p|\theta) = \beta\delta \int_{-\infty}^{\tilde{\beta}\delta b + \gamma p - \mu_c} (\delta b + \gamma p - \mu_c - \varepsilon) dF(\varepsilon) \tag{8}$$

with the incentive and

$$U(0|\theta) = \beta\delta \int_{-\infty}^{\tilde{\beta}\delta b - \mu_c} (\delta b - \mu_c - \varepsilon) dF(\varepsilon) \tag{9}$$

without the incentive. Their valuation $V(p|\theta)$ is given by $U(p|\theta) = U(0|\theta) + \beta\delta\gamma V(p|\theta)$, which, by combining with (8) and (9), yields our key equation:[32]

$$V_{it}(p|\theta) = pF(\tilde{\beta}\delta b + \gamma p - \mu_c) + \gamma^{-1} \int_{\tilde{\beta}\delta b - \mu_c}^{\tilde{\beta}\delta b + \gamma p - \mu_c} (\delta b - \mu_c - \varepsilon) dF(\varepsilon) \tag{10}$$

Equation (10) has an intuitive interpretation. From the point of view of the earlier self, the value of the incentive can be decomposed into two terms. The first term represents the expected cash value of the incentive, incorporating the expected behavioral response of the future self. That is simply the incentive amount $p$ times the perceived probability of abstinence when facing that incentive. The second term is the value of the commitment provided by the incentive, as it changes the realizations of $c$ for which the future self is expected to abstain. Note that if $\tilde{\beta} = 1$, i.e., the agent is either time-consistent or naïve, then this term must be (weakly) negative, and the incentives impose a cost on the agent. If $\tilde{\beta} < 1$, however, then this term may turn positive as the incentive is perceived to mitigate the self-control problems of the future self.

By varying $p$, it is possible to identify the perceived costs and benefits of smoking. A noteworthy special case is when $\tilde{\beta} = 1$ (that is, the subject is either time-consistent or naïve). As $p$ grows small, the second term is on the order of $p^2$ and therefore goes to zero faster than the first term, which is linear in $p$, and thus the valuation reveals the subject's perceived probability of abstinence without any incentive, $F(\tilde{\beta}\delta b - \mu_c)$. As $p$ grows large, the subject

[32] Recall that the non-contingent payment in the valuation task is also received during Month 3, and hence must be discounted by $\beta\delta$. It must also be multiplied by $\gamma$ to translate a dollar amount into utils.

will expect to abstain with certainty when facing the incentive, rendering the first term equal to $p$. The second term integrates the commitment value from the known $\tilde{\beta}\delta b - \mu_c$ to infinity. In general, we can write $\partial V/\partial p = pF(\cdot) + (1-\tilde{\beta})f(\cdot)$. Taking the second derivative with respect to $p$ generates an additional moment, which uniquely determines the discounting. We formally establish identification in Appendix C.

Extending equation (10) to treated subjects is quite straightforward, as $\tilde{\eta}_1$ and $\tilde{\eta}_2$ enter additively in all terms with a $\mu_c$. Thus a treated subject's valuations when elicited in the end-of-Month-1 survey are given by:

$$V_{it}(p|\theta) = pF(\tilde{\beta}\delta b + \gamma p - \mu_c + \tilde{\eta}_1) + \gamma^{-1}\int_{\tilde{\beta}\delta b - \mu_c + \hat{\eta}_1}^{\tilde{\beta}\delta b + \gamma p - \mu_c + \tilde{\eta}_1} (\delta b - \mu_c - \varepsilon) dF(\varepsilon) \tag{11}$$

and a treated subject's valuations when elicited in the end-of-Month-2 survey are given by:

$$V_{it}(p|\theta) = pF(\tilde{\beta}\delta b + \gamma p - \mu_c + \tilde{\eta}_2) + \gamma^{-1}\int_{\tilde{\beta}\delta b - \mu_c + \hat{\eta}_2}^{\tilde{\beta}\delta b + \gamma p - \mu_c + \tilde{\eta}_2} (\delta b - \mu_c - \varepsilon) dF(\varepsilon). \tag{12}$$

Finally, we assume that the shocks to subjects' valuations are drawn with mean zero and variance $\sigma^2$. Given (11) and (12), subjects' responses are thus distributed according to:

$$Pr(\textit{Valuation} = v|p,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(V_{it}\left(p|\theta\right) - v\right)^2}{2\sigma^2}\right) \tag{13}$$

The combined likelihood of a subject's abstinence behavior and valuations is then the product of the contributions of the discrete abstinence behaviors and the continuous predictions in equations (6) and (13).

## 4.3 Heterogeneity

As is often the case in maximum likelihood estimation, our structural model is capable of admitting heterogeneity in the parameters based on an individual's observable characteristics $X_i$. In particular, we can allow any parameter $j \in \theta$ to enter the model flexibly as a combination $X_i'\varphi_j$ for some new meta-parameter $\varphi_j$. In our estimation, we include individual characteristics from the baseline survey plausibly related to the structural parameters, namely: age, income, gender, education, baseline nicotine dependence, baseline cigarette consumption, baseline quit intentions, alcohol consumption, regular sunscreen use, and tendency to overeat. As described in Section 3.2.4, the first four are typical demographic and economic covariates in the literature, the next three are plausibly related to the degree of addiction, and the final three represent health behaviors

in other domains. We also include a measure of present bias based on a typical money-now-money-later task conducted at baseline.[33] By incorporating these covariates in this manner, our structural model assumes that they affect smoking choices only through their effect on the parameters of $\theta$.

In practice, the practical constraints of our experimental design mean that we lack sufficient variation conditional on $X_i$, namely within-subject variation, to identify heterogeneity across all parameters. In particular, as each subject received only one incentive level during Month 3 as well as, for treated subjects, the \$100 Month-2 incentives, it is not possible to estimate heterogeneity based on the abstinence behavior discussed in Section 4.1. However, as we were able to quickly obtain a large number of valuations for each subject over a range of *potential* incentive levels, we are very well powered to estimate heterogeneity based on the valuation tasks, namely in the present bias parameters $\beta$ and $\tilde{\beta}$ and in the predicted habit values $\tilde{\eta}_1$ and $\tilde{\eta}_2$. We therefore limit the role of heterogeneity to these four parameters, which are also the parameters most likely to be of general interest. We allow all covariates to explain variation in these parameters, with the exception of the money-task measure of present bias, which we use as an instrument to separate $\beta$ and $\tilde{\beta}$ from the other parts of the utility function.

It is also worth noting that in principle, truly individual-level estimates (i.e., allowing for unobservable variation) are possible within our experimental design and identification strategy. Again the main practical constraint will come from having enough within-subject variation in the Month-3 incentives in order to estimate an individual-level abstinence model. Although that proved impractical in the current smoking experiment, there is no reason in principle that it could not be incorporated into the experimental design in other contexts more practically.

# 5 Structural Estimation

We estimate the structural parameters using a maximum likelihood procedure, as described in Section 4. We include abstinence behavior in Months 2 and 3, as the interpretation of abstinence in Month 1 could be contaminated by subjects' experimentation or lack of familiarity with the incentive structure and verification technology. We include valuations for Month-3 incentives, made both in the end-of-Month-1 survey and end-of-Month-2 survey. For treated subjects during Month 2, we include the \$100 weekly abstinence incentive they

[33] Subjects were asked the following two questions: "Imagine that you have a choice between receiving \$80 guaranteed today, or \$100 guaranteed in 1 month? Which would you prefer?" and "Imagine that you have a choice between receiving \$80 guaranteed in 6 months, or \$100 guaranteed in 7 months? Which would you prefer?"

received.

The results are shown in Table 6, and confirm the pattern of findings discussed in Section 3. Indeed, the purpose of the structural estimation is to be able to map the reduced-form evidence into utility terms in order to make precise statements about subjects' biases and welfare. Panel A of Table 6 presents the main structural parameters; Panel B presents the covariates that predict heterogeneity in the present bias, naïveté, and projection bias parameters; and Panel C presents additional significance tests based on the estimates from Panel A.

## 5.1 Utility Parameters

The first, and perhaps most striking, finding on utility is just how insensitive the smokers in our sample are to financial incentives. We estimate $\gamma = 0.0011$, or that our maximum incentive of \$400 equates to roughly 0.44 utils. Although this confirms the findings in Table 3, it would imply, for instance, that a smoker with the \$400 incentive, who was otherwise indifferent between smoking and abstinence, would have a roughly 61% chance of abstinence. This reflects the limited effects of financial incentives on smoking abstinence.

We next estimate the immediate disutility and long-run benefits of abstinence in util terms, and given $\gamma$, translate them into dollar terms under the assumption of risk neutrality. While this assumption is defensible over the range of \$0 to \$400 incentives, if our subjects are risk averse over larger stakes then our dollar estimates will be lower bounds — and given our low estimated value of $\gamma$, this is likely to be the case. We estimate the immediate disutility of abstinence to be 8.88 utils, equivalent to about \$8,100 per week. Although this is large, it is not outside the scope of previous estimates and is to be expected if large cash incentives do not lead to high abstinence rates. The long-run benefit of abstinence is estimated to be similar, at 8.78 utils or about \$8,000 per week. Both estimates are individually significant, but, importantly, we fail to reject equality of the two parameters at $p = 0.92$.

We can also compute how likely subjects would be to continue to smoke under the counterfactual of no present bias. This is a similar exercise to asking whether they would want to continue to smoke if offered a perfect commitment device, although the perceived value of commitment depends on beliefs about future behavior (for example, if $\tilde{\beta} = 1$ then a smoker would always prefer the option value of being allowed to smoke on some occasions). Under our estimates, the probability of abstinence — in a week in which the subject remained fully addicted — would be 48% if subjects chose under their own long-run preferences.

We note that our estimate of the perceived long-run benefits is in some ways a

reduced-form parameter. It includes both the direct benefits of that weeks' decision and the impact on future behavior of a week of abstinence, and therefore the benefits of any continued abstinence. This multiplier may be more important than the direct effect, and our strategy does not seek to disentangle them. It also incorporates any long-run discounting performed on these direct and indirect effects through $\delta$. Although we have fixed $\delta = 1$ in our estimation because of the weekly frequency of our data, it is reasonable to believe that there is meaningful discounting over the long-term horizon a smoker uses to evaluate $b$. Our estimated $b$ takes this discounting as a given, using only the present value of these benefits, and therefore does not impose $\delta = 1$ when computing this present value over the long run.

## 5.2 Treatment Effect and Predictions

The preceding section considered the preferences of a fully addicted subject. In this section, we turn to the effect of our Month-2 abstinence intervention as well as treated subjects' *ex-ante* and *ex-post* beliefs about its effect. We estimate that on average, a subject in the treated group underwent a shift in their post-treatment utility of abstinence of $\eta = 0.555$, corresponding to \$505 in dollar terms under the maintained assumption of risk neutrality. Note that this corresponds to the average change in the treated group, and is likely to underestimate the impact of a month of successful abstinence. However, as even unsuccessful quit attempts may lower the disutility of future abstinence, we do not attempt to parcel out the effect across successful and unsuccessful Month-2 abstainers.

This estimated treatment effect is large, both in terms of its welfare consequences but also relative to its cost. In a neoclassical model, it would be surprising that an intervention costing at most $4 \times \$100 = \$400$ could be utility-equivalent to \$505. If subjects are present-biased, however, it is natural that the structure of incentives would affect their impact. Importantly, there is a difference between an incentive allocated across four weeks and one allocated as a single lump sum. In particular, we found some reduced-form evidence that the probability of abstinence was increasing for treated subjects over the course of the treatment month. Having sustained incentives could help overcome their dynamic inconsistency during this month, and thus lead to an outsized effect during Month 3.

In contrast to the large actual effect of the treatment, we estimate that treated subjects' *ex-ante* beliefs about the magnitude of the habit effect were statistically indistinguishable from zero: 0.126 utils, or \$115. Moreover, this is also substantially less than the \$400 cost of the incentives, although we cannot reject that it is equal to the actual treatment effect ($p = 0.14$). In principle, we may wish to decompose the difference between the estimates

$\eta = 0.555$ and $\tilde{\eta}_1 = 0.126$ into two components. First, actual and predicted compliance may differ; and second, the actual and predicted effect conditional on compliance may differ. Given that we show in the next section that subjects are naïve about their self-control problems, the first part of this decomposition would lead them to *overestimate* compliance and therefore also overestimate the effect of the treatment. Given that we find subjects underestimate the treatment effect, we instead focus on the latter effect. Viewed through the lens of simple projection bias, we can estimate an *ex-ante* projection parameter $\alpha_{M1}$ by solving $(1-\alpha_{M1})\eta = 0.126$, to obtain $\alpha_{M1} = 0.773$. This implies that subjects are projecting their addicted state onto predictions of future utility following abstinence, deviating from a rational forecast of $\alpha = 0$. Our parameter estimate is comparable to, though somewhat higher than, estimates elsewhere in the literature; for example, Conlin et al. (2007) find $\alpha = 0.499$ for projection over consumption utility, as measured by the probability of returning cold-weather clothes (winter jackets) purchased in mail-order catalogs.

Next, we turn to treated subjects' *ex-post* beliefs about the habit value and find that, whereas the actual treatment effect was significantly positive, subjects' post-treatment predictions are in fact consistent with a *negative* treatment effect of $-0.084$ utils, or $-\$76$. This result is not consistent with most forms of projection bias and most states, which would either require subjects to hold correct beliefs at this point (since the treatment has been realized) or at most under-estimate the magnitude of the treatment but not reverse its sign. Instead we find that subjects' beliefs are more incorrect having been through the treatment than before. Although the point estimate itself is not significantly different from 0, we reject the null hypothesis that mean post-treatment beliefs are equal to mean pre-treatment beliefs at $p = 0.08$. We also reject the null that post-treatment beliefs are correct at $p = 0.035$.

Mapped into the simple projection bias framework, these post-treatment beliefs would imply a value of $\alpha_{M3} = 1.15$. However, this would violate the conditions of Loewenstein et al. (2003), both that post-treatment beliefs should be correct and, even if they are incorrect, that $\alpha \in [0, 1]$. One could attempt to explain this pattern by a model in which subjects project their short-run cravings, given that they are possibly at the maximum level of craving at the end of the treatment. However, this interpretation merely pushes matters back one layer, since for the treatment to have had a positive effect on Month-3 abstinence there must also be a more persistent state for it to operate through. It is possible that the state is multi-dimensional, of course, but such an interpretation would pose significant challenges to a normative analysis.

Figure 7 estimates the joint distribution of $\tilde{\eta}_1$ and $\tilde{\eta}_2$ in our sample, as well as their marginal distributions. Pre- and post-treatment habit beliefs have a strong positive

association at the individual level. The relationship is approximated by $\tilde{\eta}_2 = 0.21 + 1.00\tilde{\eta}_1$, with $R^2 = 0.81$. In other words, there is uniform downward shift of 0.2 utils following the cessation intervention. Some subjects believe the habit effect is negative *ex ante* and learn it is more negative *ex post*, while others believe the habit effect is positive and learn it is less positive.

In Panel B of Table 6, we examine the observable characteristics that are related to $\tilde{\eta}_1$ and $\tilde{\eta}_2$. We find that older smokers have higher (more accurate) predictions of the *ex-ante* and *ex-post* habit value by 0.042 and 0.030 per additional decade, respectively. Sunscreen use is also associated with increases in $\tilde{\eta}_1$ and $\tilde{\eta}_2$.

## 5.3 Discount Parameters

We estimate the average subject's present bias is $\beta = 0.670$.[34] We reject time-consistent preferences of $\beta = 1$ ($p < 0.001$). The distribution across subjects is shown in Figure 8, and reveals substantial hetereogeneity with an interquartile range of 0.15. Our estimated average degree of present bias is severe, but not extreme relative to other estimates in the literature (DellaVigna, 2009). Remarkably, 100% of subjects in our data are present biased, with an estimated $\beta < 1$.

Turning to beliefs, we estimate that our subjects are on average "partially sophisticated," with $\tilde{\beta} = 0.851$ for the average subject. That is, although they appear aware that their future selves will be present-biased, their actual degree of present bias is so great that even $\tilde{\beta}$ is significantly over-optimistic. We estimate that 80% of subjects are somewhat naïve about their present bias, with $\tilde{\beta} < 1$. We reject that $\tilde{\beta} = 1$ ($p < 0.001$) and $\tilde{\beta} = \beta$ ($p = 0.01$), indicating that neither fully sophisticated nor fully naive preferences are a sufficient description of our subjects. Moreover, even a modest degree of over-optimism about $\beta$ translates into a high degree of over-optimism about the probability of abstinence, as in Figure 6. Our estimates imply that an average subject believes they are more than twice as likely to abstain from smoking without additional incentives than they actually are — a perceived probability of 20% vs. an actual probability of 5%. Note that this perceived probability is estimated from the structural model and uses only subjects' valuations of the incentives and their actual behavior. In contrast, the unincentivized predictions obtained from subjects were higher still at roughly 40% probability of abstinence. This suggests that there is a substantial upward bias in the unincentivized predictions, for example due to social desirability bias or experimenter demand effects, which can be avoided in the structural approach.

[34] Given that we estimate a linear model of heterogeneity, this represents both the mean predicted $\beta$ as well as the $\beta$ at the mean value of the covariates.

Given the general interest in quasi-hyperbolic discounting and naïveté and the novelty of our joint identification on these parameters, we utilize our sample to estimate heterogeneity in $\beta$ and $\tilde{\beta}$. We present the joint distribution of $\beta$ and $\tilde{\beta}$ in our sample, as well as their marginal distributions, in Figure 8. It is worth noting that our linear model in no way restricts these parameters to lie on the unit square, and yet this is the case for most subjects. This contrasts with many existing results in the literature that find a large fraction of future-biased subjects and raise concerns about measurement error (see, e.g., Ashraf et al. (2006); Augenblick et al. (2015)).

The most striking pattern in Figure 8 is the near-linear relationship between $\beta$ and $\tilde{\beta}$. Indeed, the two variables are well approximated by a simple regression line of $\tilde{\beta} = 0.11 + 1.11\beta$, indicated by the solid line in the figure, with $R^2 = 0.46$. There is no reason theoretically why this should be the case, as for instance it would be possible to have $\tilde{\beta} = 1$ independently of $\beta$, or for there to be a non-monotonic relationship if the cognitive skills related to a low $\beta$ are related to the metacognitive skills related to a high $\tilde{\beta}$. There is also no reason empirically that the structural estimation would impose a linear relationship, as the two parameters enter the model separately. An important caveat is that our estimator may not give a consistent estimate of the true joint distribution of $\beta$ and $\tilde{\beta}$. That said, we find in Panel B of Table 6 that observable characteristics are related in very similar ways to $\beta$ and $\tilde{\beta}$.

Finally, we note that the covariates in Panel B of Table 6 should not be considered as being causally related to $\beta$ and $\tilde{\beta}$, nor should they be expected to generalize to the general population. On the first point, whereas age is (mostly) exogenous, outcomes such as alcohol consumption and sunscreen use depend on an individual's degree of present bias. Indeed, it is reassuring to see that many of these are related to present bias in the expected way. For example, we estimate that, *ceteris paribus*, someone who regularly uses sunscreen would be expected to have a $\beta$ 0.08 higher than someone who does not. This is a large difference, and suggests a simple "behavioral marker" which can be used in a reduced-form analysis to identify present-biased individuals (similar to the income-tax filing decision in Martinez et al. (2017)). This should be taken with some caution, however, as these relationships are only estimated on our sample of smokers and there may be substantial selection into smoking in the first place. For example, we find that male smokers *ceteris paribus* have a $\beta$ 0.13 higher than female smokers, but given gender differences in smoking rates this pattern may not be representative of the population as whole.

## 5.4 Comparing measures of naïveté

Thus far, we have examined present bias and projection bias separately. However, we can also leverage individual-level variation in the parameters to explore an open question about whether naïve present-biased beliefs and projection-biased beliefs are associated. To do so, we re-parameterize $\tilde{\beta}$ as a linear combination of its smallest and largest possible values, $\hat{\beta} = (1-\omega) \cdot \beta + \omega \cdot 1$. Using this formulation, $\omega = 0$ corresponds to full sophistication, and $\omega = 1$ corresponds to fully naïve beliefs. Thus, $\omega$ is a measure of naïveté independent of the underlying level of time-inconsistency.

In an exploratory analysis, we estimate a bivariate regression of $\omega$ and $\alpha_{M1}$. We find a strong positive association, approximated by $\omega = 0.46 + 0.29\alpha_{M1}$, with $R^2 = 0.15$. This novel finding suggests that naïveté regarding present bias can place a person at increased risk of naïveté regarding projection, and merits further attention.

## 5.5 Welfare calculations

In this section, we apply our structural estimates to welfare calculations regarding the costs and benefits of continuing to smoke. We focus on welfare under three scenarios: 1) present-biased preferences in an addicted state, 2) long-run preferences in an addicted state, and 3) long-run preferences in an unaddicted state. The latter scenario is our best estimate of the welfare effects of smoking, accounting for subjects' biases.

We start by calculating welfare from the perspective of a present-biased smoker in an addicted state, who heavily discounts the future benefits of abstinence. On average, continued smoking for a present-biased smoker generates welfare of $\mu_c - \beta b = 2.83$ utils, equivalent to a welfare gain of $(\mu_c - \beta b)/\gamma = \$2,726$ per week. Under this scenario, it is not surprising that the vast majority of smokers continue to smoke from one week to the next. In fact, this mean utility would predict a 95% probability of smoking in a given week.

Next, we calculate the welfare effects of continued smoking under current smokers' own long-run preferences in an addicted state—that is, under a counterfactual of no present bias. This entails omitting $\beta$ from the welfare calculus. In that case, on average choosing to smoke generates welfare of $\mu_c - b = 0.01$ utils, equivalent to a welfare gain of $(\mu_c - b)/\gamma = \$91$ per week. This suggests that, using the consumption utility of the agent in the addicted state but their long-run time preferences, continued smoking is approximately welfare-neutral.

In contrast to the above results, we find a large private welfare loss when we consider the long-run preferences of a person in an unaddicted state. On average, subjects' choices to continue to smoke generates welfare of $\mu_c - b - \eta = 0.45$ utils, equivalent to a private welfare loss of $(\mu_c - b - \eta)/\gamma = \$414$ per week. This figure is likely a lower bound, as our

experimentally identified $\eta$ is the effect of at most one month of abstinence and therefore is likely to underestimate the long-run adaptation. To interpret this as an "internality" cost per pack, one may further assume that $(\mu_c - b - \eta)$ is linear in the number of cigarettes smoked within a week.[35] In this case, on average subjects incur a private welfare loss from smoking of \$80 per pack of cigarettes consumed, under their own long-run preferences in an unaddicted state. This estimate is plausible when compared with existing estimates in the smoking literature.[36]

We can decompose the private welfare loss from continued smoking into the amounts attributable to present bias and projection bias. Present bias lowers the value of quitting by \$2,635 per week $((1-\beta)b/\gamma)$, because the long-run benefit of abstinence is subject to additional discounting by a relatively low $\beta$. Projection bias lowers the value of quitting by \$390 per week $((\eta - \tilde{\eta}_1)/\gamma)$, because of the wedge between the actual and perceived habit value.

# 6 Discussion

It is widely believed that behavioral biases inhibit the ability of consumers to abstain from addictive goods. Our paper sheds light on this important question, and the results of our field experiment indicate that, on average: (1) smokers are severely present-biased regarding future smoking abstinence; (2) smokers are only partially aware of their problems with present bias; and (3) smokers hold mistaken beliefs that underestimate the value of an abstinence intervention both before and after it has taken place. These three departures from a standard economic model conspire to impose a large private welfare loss on smokers in our sample. Whereas most smokers in our sample would stop smoking in the absence of these departures under their own long-run preferences and in the absence of addiction, present bias and projection bias lead most smokers to continue to smoke in most weeks.

The smoking literature has traditionally focused on the externalities that smokers impose on non-smokers, for example, through secondhand smoke and publicly financed medical care (e.g., Manning et al., 1989). Gruber and Kőszegi (2001) expand this focus to develop

[35] This is a convenient, but implausible assumption. Although medically it is plausible to assume $b$ is linear in consumption, $c$ is likely to exhibit concavity (e.g., a two-pack-a-day smoker may not suffer twice as much withdrawal as a pack-a-day smoker). Our research design allows this to be estimated in principle, but given the limited variation in baseline consumption and limited sample size it is impractical with the current data.

[36] Viscusi and Hersch (2008) estimated that the mortality cost of smoking is \$222 per pack for men and \$94 for women in 2006 dollars. Our estimates should correspond to the mortality cost plus the morbidity cost less the immediate consumption value. If the morbidity cost and immediate consumption value are of similar size, then our per-pack estimate would be highly plausible.

a model that allows for "internalities" that smokers impose on themselves. They further show that observed patterns of smoking are consistent not only with a rational addiction model but also with their calibrated model that introduces a modest degree of present bias (with full sophistication). We advance this literature by quantifying smoking internalities based on present-biased preferences and naïve and projection-biased beliefs. Our study has particular relevance to ongoing debates about the proper way to account for internalities in welfare calculations (Allcott and Sunstein, 2015). Recent cost-benefit analyses in the smoking literature have relied on reduced-form evidence that distinguishes less informed or less addicted smokers compared with a "rational benchmark," in order to incorporate internality costs that offset the lost consumer surplus from tobacco control policies (Cutler et al., 2015; Jin et al., 2015; Levy et al., 2018).

Our experimental identification of preferences and beliefs is aided by a methodological innovation. A common challenge for field studies has been the time, expense, and resources involved in remotely varying behavior. While self-reported behavior may be reliably assessed under certain circumstances (Gorber et al., 2009), it is likely to be distorted when incentives for behavior change are offered. We overcome this challenge by combining remote biochemical verification of smoking status with lotteries that encourage truth-telling and occasional in-person audits from field personnel. This procedure represents a notable advance over existing remote verification technologies. Moreover, our study contributes to a growing literature that leverages experimental designs to estimate structural models of behavioral parameters (DellaVigna, 2018). We provide a novel identification of present bias based on valuations of partial commitment devices, which we hope future work may build upon.[37]

Our findings suggest that policies need to be designed to take into account the confluence of biases that affect consumers of addictive goods. The welfare improvement of a tax is governed by the share of smokers that quit successfully in response. Those smokers who fail to quit might be left worse off by their failed commitment. Given the muted response to even the largest incentives in our study, tobacco taxation may be a second-best policy option for "behavioral" smokers — with the important caveats that our findings speak only to smoking cessation and not to smoking initiation, and taxes may be treated differently than the abstinence incentives we provided. We leave the calculation of the optimal tax rate, which balances the welfare of those who succeed and those who fail, for future work.

An alternative to taxes would be a smoking ban. A ban on the sale of cigarettes has been proposed many times in the public health literature (e.g., Proctor, 2013), but is often criticized for restricting personal freedom (Schmidt, 2016). Typical analyses of the welfare

[37] In particular, settings which permit greater within-subject variation in actual incentives received would expand the extent of heterogeneity that may be estimated.

effects of bans would start by examining the lost consumer surplus as revealed by the demand curve. This overstates the true lost surplus, however, as it includes consumers' present bias and fails to include their adaptation in the event of cessation. Our results indicate that accounting for both biases not only reduces the lost consumer surplus, but in fact leads to a gain in welfare for current smokers of $414 on average. Of course, a ban eliminates the flexibility of consumers to smoke in response to utility shocks, but standard results from discrete choice models indicate that even accounting for the reduced flexibility, a smoking ban would on average increase smokers' welfare by at least $353 per week.[38] Unfortunately, however, our estimates of naiveté and projection bias also imply that some smokers would mis-perceive a ban as reducing welfare.

Our findings also may expand the range of recommended policy options to include libertarian paternalistic interventions. For example, as a way to overcome present bias, one might consider asking smokers to commit to an order of their week's supply of cigarettes in advance, in line with evidence that some smokers use self-rationing to avoid over-indulging (Wertenbroch, 1998). Regardless of intervention, one might expect from our findings that midway through a quit attempt a person may project the negative withdrawal symptoms on future abstinence. One counter-measure would be to use an escalating incentive schedule, as adopted in some smoking cessation trials (e.g., Higgins et al., 2014), to overcome the drag imposed by smokers' pessimistic beliefs about the value of future abstinence. In general, it is clear that a greater understanding of the specific nature of the underlying behavioral biases allows policy responses to be better tailored to the problems they seek to address.

# References

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

**Allcott, Hunt and Cass R. Sunstein**, "Regulating Internalities," *Journal of Policy Analysis and Management*, 2015, *34* (3), 698–705.

**Andersen, Steffen, Glenn W. Harrison, Morten I Lau, and E. Elisabet Rustrom**, "Eliciting Risk and Time Preference," *Econometrica*, May 2008, *76* (3), 583–618.

[38] Note that the change in surplus is at least as large as the change that would occur if smokers were choosing optimally (according to their long-run preferences) in the absence of a ban. Following Rust (1984), the expected utility with i.i.d. extreme value shocks is given by the familiar log-sum formula: $E[\max_i u_i + \varepsilon_i] = \ln(\sum_i \exp(u_i))] = \ln(1+\exp(b-c)) = 0.6445$ utils. Restricting choice but accounting for a single shock yields expected utility $b-c+\eta+E[\varepsilon] = 0.4549+0.5772 = 1.032$ utils. Dividing by $\gamma$ to yield a dollar figure, the difference in utility is $353. An alternative lower bound would be to evaluate that no-ban surplus under unaddicted preferences, i.e., including $\eta$. This alternative yields a $78 weekly increase in surplus from a ban.

**Andreoni, James and Charles Sprenger**, "Risk Preferences are Not Time Preferences," *American Economic Review*, 2012, *102* (7), 3357–3376.

**Ashraf, Nava, Dean Karlan, and Wesley Yin**, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, May 2006, *121* (2), 635–672.

**Augenblick, Ned, Muriel Niederle, and Charles Sprenger**, "Working over time: Dynamic inconsistency in real effort tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

**Badger, Gary J, Warren K Bickel, Louis A Giordano, Eric A Jacobs, George Loewenstein, and Lisa Marsch**, "Altered states: The impact of immediate craving on the valuation of current and future opioids," *Journal of Health Economics*, 2007, *26* (5), 865–876.

**Bai, Liang, Benjamin Handel, Edward Miguel, and Gautam Rao**, "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India," Working Paper 23727, National Bureau of Economic Research August 2017.

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

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

**Bickel, Warren K, Amy L Odum, and Gregory J Madden**, "Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers," *Psychopharmacology*, 1999, *146* (4), 447–454.

**Bricker, Jonathan B, Kristin E Mull, Jennifer B McClure, Noreen L Watson, and Jaimee L Heffner**, "Improving Quit Rates of Web-Delivered Interventions for Smoking Cessation: Full Scale Randomized Trial of WebQuit. org versus Smokefree. gov," *Addiction*, 2017.

**Carerra, Maria, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky**, "How are preferences for commitment revealed?," August 2019. Mimeo.

**Chabris, Christopher F, David Laibson, Carrie L Morris, Jonathon P Schuldt, and Dmitry Taubinsky**, "Individual laboratory-measured discount rates predict field behavior," *Journal of Risk and Uncertainty*, 2008, *37* (2-3), 237.

**Chaloupka, Frank**, "Rational addictive behavior and cigarette smoking," *Journal of political Economy*, 1991, *99* (4), 722–742.

**Chetty, Raj**, "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annual Review of Economics*, 2009, *1* (1), 451–488.

**_ , Adam Looney, and Kory Kroft**, "Salience and taxation: Theory and evidence," *American economic review*, 2009, *99* (4), 1145–77.

**Conlin, Michael, Ted O'Donoghue, and Timothy Vogelsang**, "Projection Bias in Catalog Orders," *The American Economic Review*, September 2007, *97* (4), 1217–1249.

**Cutler, David M., Amber Jessup, Donald Kenkel, and Martha A. Starr**, "Valuing Regulations Affecting Addictive or Habitual Goods," *Journal of Benefit-Cost Analysis*, 6 2015, *6*, 247–280.

**DellaVigna, Stefano**, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, June 2009, *47* (2), 315–372.

_ , "Structural Behavioral Economics," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, Vol. Volume 1, North-Holland, 2018, chapter Chapter 7, pp. 613–723.

_ **and Ulrike Malmendier**, "Paying Not To Go To The Gym," *The American Economic Review*, June 2006, *96* (3), 694–719.

_ **, John A List, Ulrike Malmendier, and Gautam Rao**, "Estimating social preferences and gift exchange at work," Technical Report, National Bureau of Economic Research 2016.

_ **, John List, and Ulrike Malmendier**, "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

**Fenelon, Andrew and Samuel H Preston**, "Estimating smoking-attributable mortality in the United States," *Demography*, 2012, *49* (3), 797–818.

**Fong, Geoffrey T, David Hammond, Fritz L Laux, Mark P Zanna, K Michael Cummings, Ron Borland, and Hana Ross**, "The near-universal experience of regret among smokers in four countries: findings from the International Tobacco Control Policy Evaluation Survey," *Nicotine & Tobacco Research*, 2004, *6* (Suppl_3), S341–S351.

**Giné, Xavier, Dean Karlan, and Jonathan Zinman**, "Put your money where your butt is: a commitment contract for smoking cessation," *American Economic Journal: Applied Economics*, 2010, *2* (4), 213–35.

**Gorber, Sarah Connor, Sean Schofield-Hurwitz, Jill Hardt, Geneviève Levasseur, and Mark Tremblay**, "The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status," *Nicotine & Tobacco Research*, 2009, *11* (1), 12–24.

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

**Gruber, Jonathan H. and Sendhil Mullainathan**, "Do Cigarette Taxes Make Smokers Happier?," *The B.E. Journal of Economic Analysis & Policy*, 2005, *5* (1).

**Higgins, Stephen T, Yukiko Washio, Alexa A Lopez, Sarah H Heil, Laura J Solomon, Mary Ellen Lynch, Jennifer D Hanson, Tara M Higgins, Joan M Skelly, Ryan Redner et al.**, "Examining two different schedules of financial incentives for smoking cessation among pregnant women," *Preventive Medicine*, 2014, *68*, 51–57.

**Hughes, John R., Josue Keely, and Shelly Naud**, "Shape of the relapse curve and long-term abstinence among untreated smokers," *Addiction*, 2004, *99* (1), 29–38.

**Jha, Prabhat, Chinthanie Ramasundarahettige, Victoria Landsman, Brian Rostron, Michael Thun, Robert N Anderson, Tim McAfee, and Richard Peto**, "21st-century hazards of smoking and benefits of cessation in the United States," *New England Journal of Medicine*, 2013, *368* (4), 341–350.

**Jin, Lawrence, Don Kenkel, Feng Liu, and Hua Wang**, "Retrospective and Prospective Benefit-Cost Analyses of US Anti-Smoking Policies 1," *Journal of Benefit-Cost Analysis*, 2015, *6* (1), 154–186.

**Kasza, Karin A, Bridget K Ambrose, Kevin P Conway, Nicolette Borek, Kristie Taylor, Maciej L Goniewicz, K Michael Cummings, Eva Sharma, Jennifer L Pearson, Victoria R Green et al.**, "Tobacco-product use by adults and youths in the United States in 2013 and 2014," *New England Journal of Medicine*, 2017, *376* (4), 342–353.

**Khwaja, Ahmed, Dan Silverman, and Frank Sloan**, "Time preference, time discounting, and smoking decisions," *Journal of Health Economics*, 2007, *26* (5), 927–949.

**Laibson, David**, "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–478.

**Levy, Helen G., Edward C. Norton, and Jeffrey A. Smith**, "Tobacco regulation and cost-benefit analysis: How should we value foregone consumer surplus?," *American Journal of Health Economics*, 2018, *4* (1), 1–25.

**Loewenstein, George**, "Projection bias in medical decision making," *Medical Decision Making*, 2005, *25* (1), 96–105.

**_ , Ted O'Donoghue, and Matthew Rabin**, "Projection bias in predicting future utility," *The Quarterly Journal of Economics*, 2003, *118* (4), 1209–1248.

**Manning, Willard G, Emmett B Keeler, Joseph P Newhouse, Elizabeth M Sloss, and Jeffrey Wasserman**, "The taxes of sin: Do smokers and drinkers pay their way?," *JAMA*, 1989, *261* (11), 1604–1609.

**Martinez, Seung-Keun, Stephan Meier, and Charles Sprenger**, "Procrastination in the Field: Evidence from Tax Filing," October 2017. Unpublished working paper.

**Matzkin, Rosa**, "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 1993, *58* (1), 137–168.

**Moore, Matthew R, Michael J Mason, Aaron R Brown, Claudia M Garcia, Ashlie D Seibers, and Chelsea J Stephens**, "Remote biochemical verification of tobacco use: Reducing costs and improving methodological rigor with mailed oral cotinine swabs," *Addictive Behaviors*, 2018, *87*, 151–154.

**Morgan, Kari Lock and Donald B Rubin**, "Rerandomization to improve covariate balance in experiments," *The Annals of Statistics*, 2012, *40* (2), 1263–1282.

**_ and _** , "Rerandomization to balance tiers of covariates," *Journal of the American Statistical Association*, 2015, *110* (512), 1412–1421.

**O'Donoghue, Ted and Matthew Rabin**, "Doing it now or later," *American Economic Review*, 1999, *89* (1), 103–124.

_ **and** _ , "Choice and Procrastination," *The Quarterly Journal of Economics*, 2001, *116*, 121–160.

**Paserman, M. Daniele**, "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," *The Economic Journal*, August 2008, *118* (531), 1418–1452.

**Proctor, Robert N.**, "Why ban the sale of cigarettes? The case for abolition," *Tobacco Control*, 2013, *22* (Suppl 1), i27–i30.

**Rabin, Matthew**, "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, September 2000, *68* (5), 1281–1292.

**Ramo, Danielle E., Johannes Thrul, Kevin L. Delucchi, Sharon Hall, Pamela M. Ling, Alina Belohlavek, and Judith J. Prochaska**, "A randomized controlled evaluation of the tobacco status project, a Facebook intervention for young adults," *Addiction*, 2018, *113* (9), 1683–1695.

**Read, Daniel and Barbara Van Leeuwen**, "Predicting hunger: The effects of appetite and delay on choice," *Organizational behavior and human decision processes*, 1998, *76* (2), 189–205.

**Royer, Heather, Mark Stehr, and Justin Sydnor**, "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company," *American Economic Journal: Applied Economics*, July 2015, *7* (3), 51–84.

**Sayette, Michael A., George Loewenstein, Kasey M. Griffin, and Jessica J. Black**, "Exploring the Cold-to-Hot Empathy Gap in Smokers," *Pychological Science*, 2008, *19* (9), 926–932.

**Schelling, Thomas C.**, "The intimate contest for self-command," *The Public Interest*, 1980, *60*, 94–118.

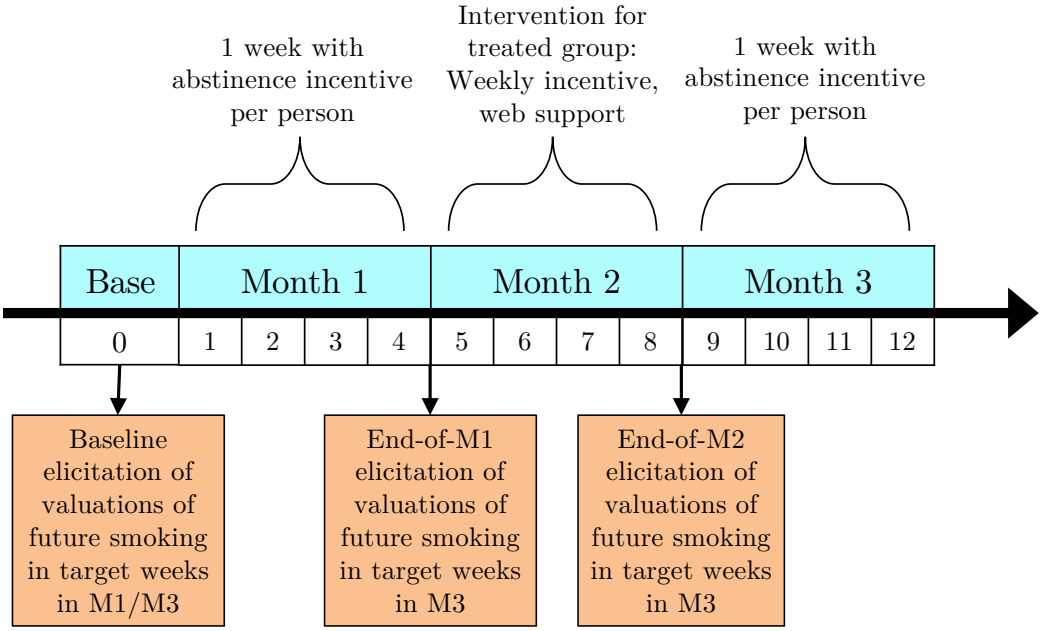
**Schmidt, Andreas T.**, "Withdrawing Versus Withholding Freedoms: Nudging and the Case of Tobacco Control," *The American Journal of Bioethics*, 2016, *16* (7), 3–14. PMID: 27292838.

**Sloan, Frank, Jan Ostermann, Christopher Conover, Donald H Taylor Jr, and Gabriel Picone**, "The Price of Smoking," Technical Report, The MIT Press 2006.

**Viscusi, W. Kip and Joni Hersch**, "The mortality cost to smokers," *Journal of Health Economics*, 2008, *27* (4), 943–958.

**Wertenbroch, Klaus**, "Consumption self-control by rationing purchase quantities of virtue and vice," *Marketing Science*, 1998, *17* (4), 317–337.

# 7 Figures




Note: During one week of Month 1 and one week of Month 3, each subject was eligible for one abstinence incentive payment. Month 2 is the treatment month.

Figure 2: Saliva testing

Photo 1: swabbing


Photo 2: test result

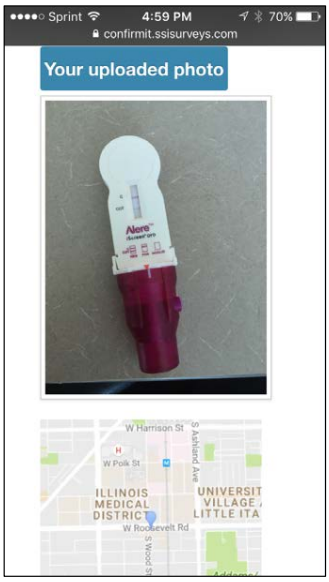
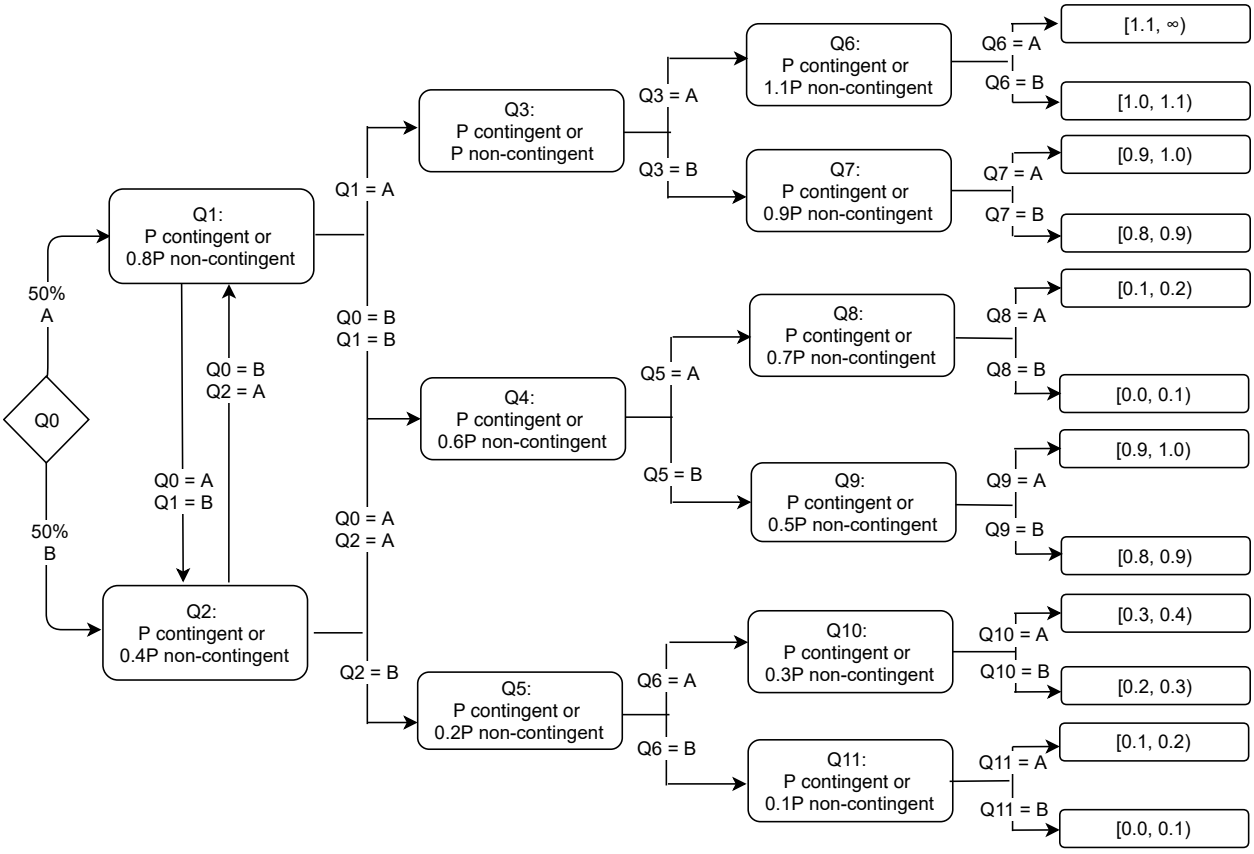



Photo 3: blacked-out window

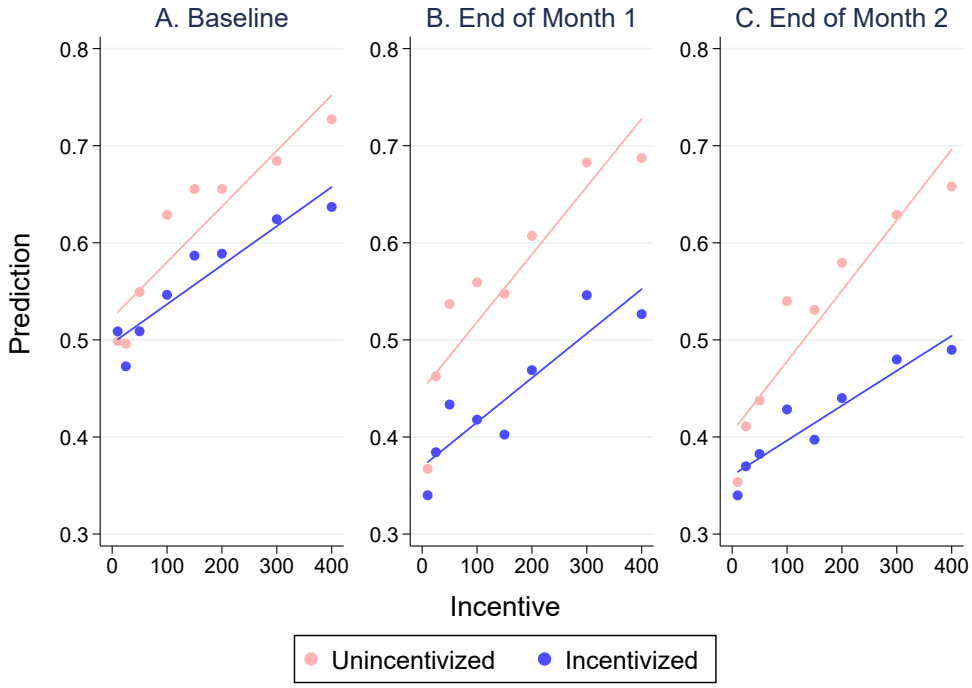

Note: Subjects uploaded a series of three photos as part of the saliva testing procedures.

Figure 3: "Staircase" for Eliciting Valuations



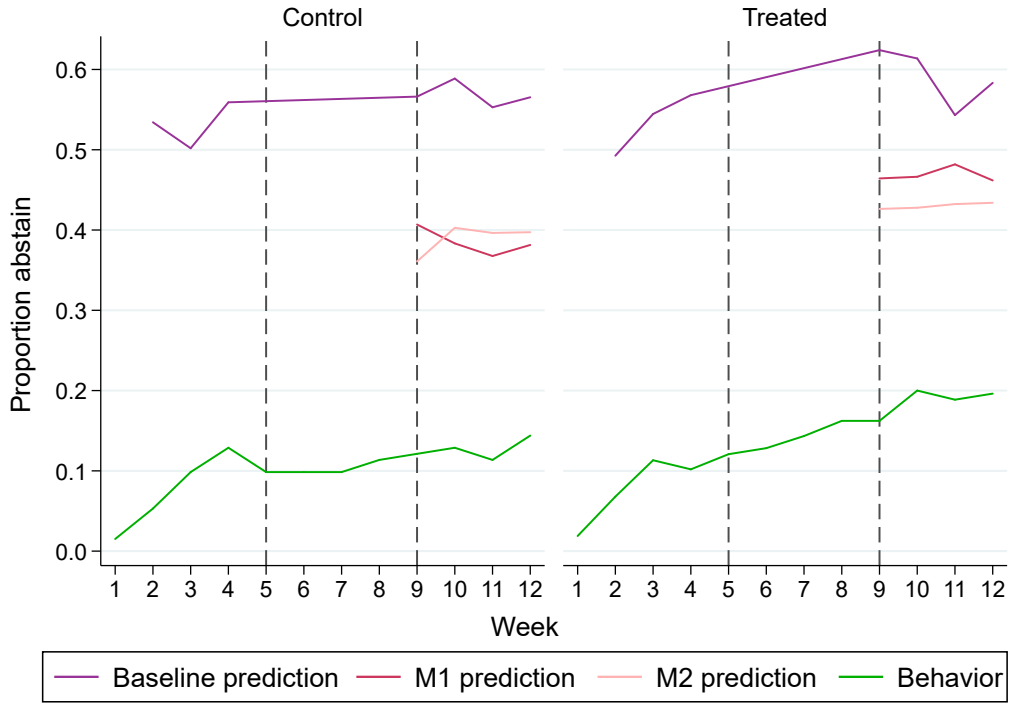

Note: The staircase moves left to right and sorts a person's predicted willingness to accept to abstain into 12 outcome "rows" displayed in far right. The starting point is randomized in question Q0. In questions Q1-Q11, a person chooses between a smokefree-contingent payment worth $p$ (option A) and a non-contingent payment (option B). The target week and payment amount $p$ were randomized for each of the four staircases completed during each elicitation session.

Figure 4: Abstinence Predictions By Incentive Level



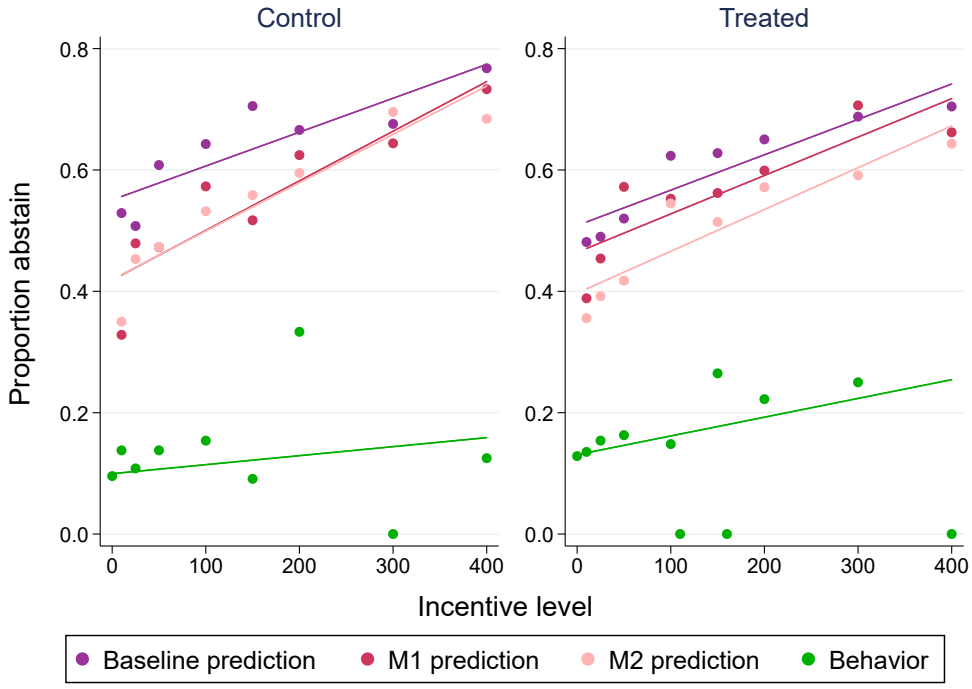

Note: This figure shows the average abstinence prediction (unincentivized) or normalized value (incentivized) by incentive level and elicitation session. The left panel is from the baseline survey; the middle is from the end-of-Month-1 survey; and the right is from the end-of-Month-2 survey. The normalized value for the incentivized task is the incentive valuation divided by the incentive level.

Figure 5: Abstinence Behavior and Predictions by Week and Treatment Group



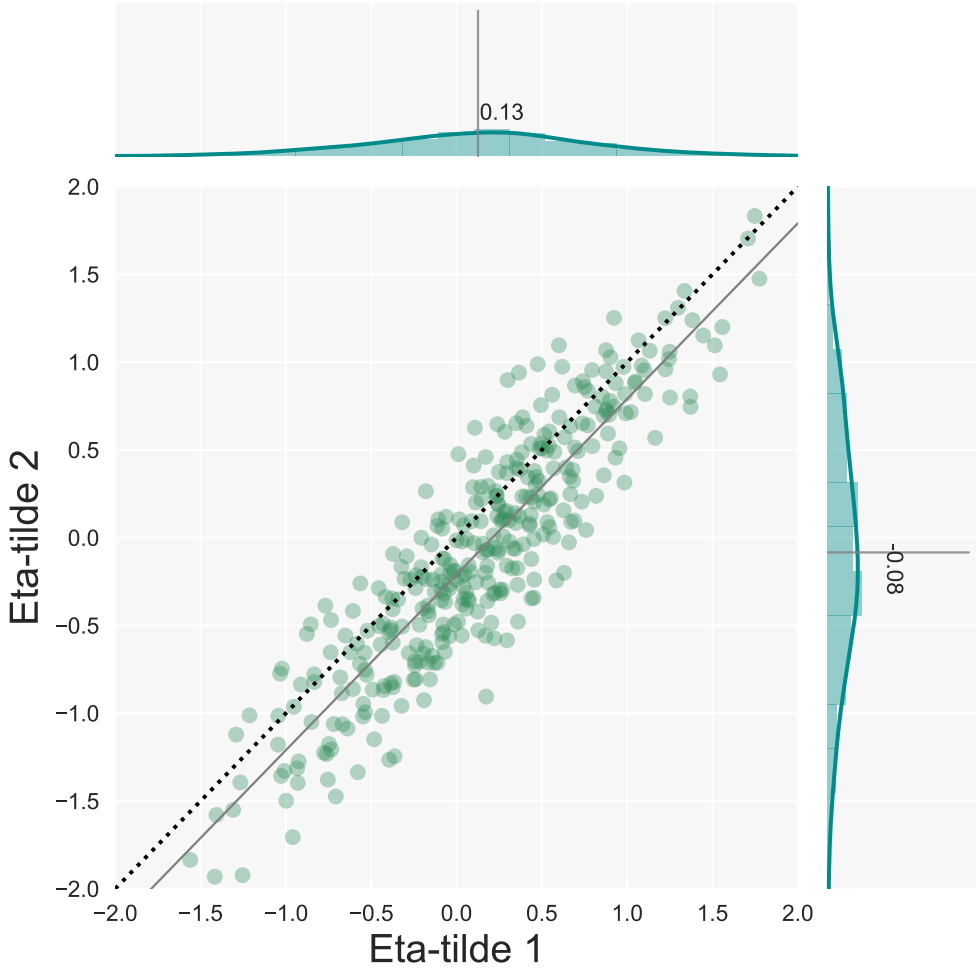

Note: This figure shows the average abstinence behavior or average predictions from each survey (unincentivized), by study week and treatment group. Weeks in which a subject was eligible for a smokefree-contingent payment are included. The left panel is for the control group, and the right is for the treated group.

Figure 6: Abstinence Behavior and Predictions By Incentive Level and Treatment Group



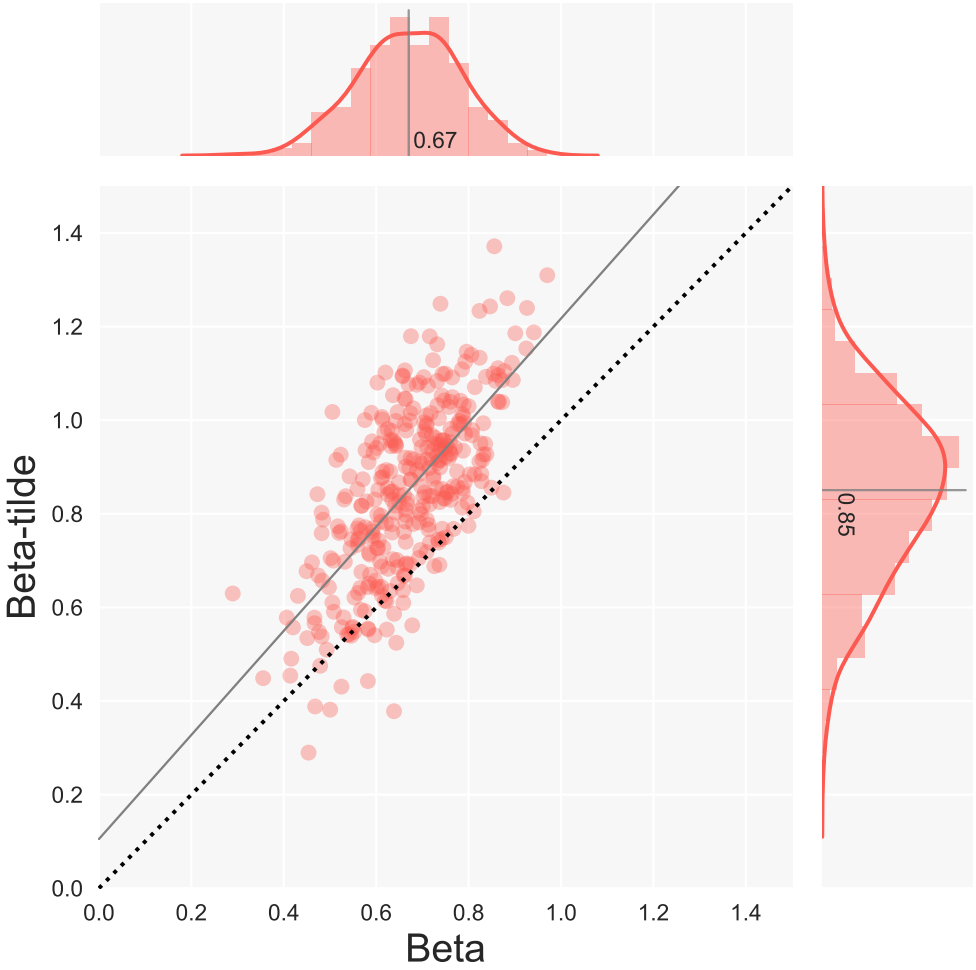

Note: This figure shows the average abstinence behavior or normalized value (incentivized) by incentive level and treatment group. The left panel is for the control group, and the right is for the treated group. The normalized value for the incentivized task is the incentive valuation divided by the incentive level.

Figure 7: Distribution of $\tilde{\eta}_1$ and $\tilde{\eta}_2$




Note: This figure shows the sample distribution of ex-ante predicted treatment effects ($\tilde{\eta}_1$) and ex-post ($\tilde{\eta}_2$) as estimated by the structural model in Section 5. The main figure plots the joint distribution, with the solid line representing a bivariate regression and the dotted line indicating the 45° line. The top and right sub-figures plot the marginal distributions of $\tilde{\eta}_1$ and $\tilde{\eta}_2$, respectively.

Figure 8: Distribution of $\beta$ and $\tilde{\beta}$




Note: This figure shows the sample distribution of discount factors ($\beta$) and beliefs ($\tilde{\beta}$) as estimated by the structural model in Section 5. The main figure plots the joint distribution, with the solid line representing a bivariate regression and the dotted line indicating the 45° line. The top and right sub-figures plot the marginal distributions of $\beta$ and $\tilde{\beta}$, respectively.

# 8 Tables

Table 1: Baseline Characteristics

| | All | Control | Treatment | $p$-value of diff. |
|---|---|---|---|---|
| *Demographics* | | | | |
| Male (%) | 33.8 | 28.8 | 36.2 | 0.16 |
| Age (%) | | | | |
| 21-34 | 23.7 | 28.8 | 21.1 | 0.17 |
| 35-44 | 27.2 | 27.3 | 27.2 | 0.98 |
| 45-54 | 29.5 | 26.5 | 30.9 | 0.34 |
| $\geq$ 55 | 19.6 | 17.4 | 20.8 | 0.55 |
| White (%) | 74.8 | 70.5 | 77.0 | 0.09 |
| Household income (%) | | | | |
| < $30,000 | 24.7 | 28.8 | 22.6 | 0.29 |
| $30,000 - $49,999 | 21.4 | 18.2 | 23.0 | 0.24 |
| $50,000 - $99,999 | 36.3 | 37.9 | 35.5 | 0.60 |
| $\geq$ $100,000 | 17.1 | 15.2 | 18.1 | 0.36 |
| Decline to report | 0.5 | 0.0 | 0.8 | 0.17 |
| Education (%) | | | | |
| High school degree or less | 25.4 | 26.5 | 24.9 | 0.76 |
| Some college or associate's degree | 41.8 | 45.5 | 40.0 | 0.38 |
| College degree | 22.2 | 19.7 | 23.4 | 0.30 |
| Some graduate school | 10.6 | 8.3 | 11.7 | 0.20 |
| | | | | |
| *Smoking characteristics* | | | | |
| Mean cigarettes per day | 14.7 | 14.5 | 14.7 | 0.88 |
| Nicotine dependent (%) | 56.4 | 58.3 | 55.5 | 0.54 |
| Planning to quit < 6 months (%) | 33.2 | 30.3 | 34.7 | 0.31 |
| E-cigarette use (%) | | | | |
| 0 days | 74.1 | 70.5 | 75.8 | 0.28 |
| 1-15 days | 17.9 | 18.2 | 17.7 | 0.90 |
| 16-30 days | 8.1 | 11.4 | 6.4 | 0.15 |
| | | | | |
| $N$ | 397 | 132 | 265 | |
| $p$-value from joint $F$-test | | | | 0.89 |

Notes: $p$-values are reported from Wald tests on the equality of means of treatment and control groups for each variable.

Table 2: Smoking Abstinence by Treatment and Incentives

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Month 3 | 0.122*** | 0.122*** | 0.119*** | 0.120*** |
| | (0.024) | (0.024) | (0.024) | (0.024) |
| Treated × Month 3 | 0.058** | 0.058** | 0.066** | 0.065** |
| | (0.026) | (0.026) | (0.027) | (0.027) |
| Month 2 | | 0.105*** | | 0.046 |
| | | (0.021) | | (0.028) |
| Treated × Month 2 | | 0.035 | | 0.017 |
| | | (0.022) | | (0.043) |
| Incentive | | | 0.046* | -0.042 |
| | | | (0.027) | (0.030) |
| Treated × Incentive | | | 0.023 | -0.043 |
| | | | (0.042) | (0.047) |
| Month 3 × Incentive | | | -0.040 | 0.103*** |
| | | | (0.029) | (0.021) |
| Treated × Month 3 × Incentive | | | -0.046 | 0.038* |
| | | | (0.046) | (0.022) |
| Constant | 0.018 | 0.018 | 0.018 | 0.018 |
| | (0.012) | (0.013) | (0.012) | (0.013) |
| $R^2$ | 0.06 | 0.05 | 0.07 | 0.05 |
| Number of observations | 3,176 | 4,764 | 3,176 | 4,764 |
| Number of clusters | 397 | 397 | 397 | 397 |
| Week FE | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes |

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: The dependent variable is an indicator of weekly smoking abstinence. Incentives are scaled in $100s. Models 1 and 3 are restricted to data from Months 1 and 3 only. Standard errors are clustered at the person level.

Table 3: Predictions vs. Behavior

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Behavior | -0.429*** | -0.314*** | -0.388*** | -0.286*** |
| | (0.026) | (0.028) | (0.015) | (0.015) |
| Incentive | | 0.076*** | | 0.068*** |
| | | (0.009) | | (0.004) |
| Behavior × Incentive | | -0.050*** | | -0.040*** |
| | | (0.014) | | (0.013) |
| Constant | 0.533*** | 0.415*** | 0.520*** | 0.415*** |
| | (0.020) | (0.023) | (0.010) | (0.011) |
| $R^2$ | 0.39 | 0.42 | 0.36 | 0.39 |
| $N$ | 2,024 | 2,024 | 7,185 | 7,185 |
| Individual FE | Yes | Yes | Yes | Yes |

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: The dependent variable is an indicator of weekly smoking abstinence. Incentives are scaled in $100s. The data are stacked with observed behavior on top of predictions. "Behavior" equals 1 for observed abstinence and 0 for incentivized abstinence predictions. Models 1 and 2 restrict to subjects in the Control group, and Models 3 and 4 include all subjects' predictions.

Table 4: Difference-in-Differences of Predictions

| | Unincentivized predictions (1) | Incentivized predictions (2) | Unincentivized predictions (3) | Incentivized predictions (4) |
|---|---|---|---|---|
| Baseline predictions | 0.101*** | 0.177*** | 0.130*** | 0.193*** |
| | (0.022) | (0.030) | (0.027) | (0.037) |
| End-of-M2 predictions | 0.005 | 0.001 | -0.001 | 0.015 |
| | (0.018) | (0.021) | (0.026) | (0.029) |
| Baseline predictions × Treated | -0.044 | -0.065* | -0.075** | -0.088** |
| | (0.028) | (0.036) | (0.034) | (0.043) |
| End-of-M2 predictions × Treated | -0.049** | -0.030 | -0.062* | -0.032 |
| | (0.023) | (0.028) | (0.032) | (0.035) |
| Treated × Incentive | | | -0.016 | -0.011 |
| | | | (0.011) | (0.011) |
| Baseline predictions × Incentive | | | -0.018** | -0.010 |
| | | | (0.009) | (0.012) |
| End-of-M2 predictions × Incentive | | | 0.004 | -0.009 |
| | | | (0.009) | (0.012) |
| Baseline predictions × Treated × Incentive | | | 0.017* | 0.013 |
| | | | (0.010) | (0.014) |
| End-of-M2 predictions × Treated × Incentive | | | 0.007 | 0.001 |
| | | | (0.011) | (0.013) |
| Constant | 0.504*** | 0.359*** | 0.409*** | 0.297*** |
| | (0.015) | (0.019) | (0.016) | (0.020) |
| $R^2$ | 0.03 | 0.07 | 0.20 | 0.11 |
| $N$ | 4,009 | 4,011 | 4,009 | 4,011 |
| Week FE | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes |

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: The dependent variable is an indicator of predicted weekly smoking abstinence. Models 1 and 3 use unincentivized predictions, and Models 2 and 4 use incentivized (implied) predictions. Models 3 and 4 include interactions with the incentive level, scaled in $100s. Standard errors are clustered at the person level.

Table 5: Heterogeneity in Overoptimism

| | Demographics | | | | Addiction-related | | | | Health behaviors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age | Male | Household income | Education | Cigarettes per day | Nicotine dependence | Quit plans | Present bias | Alcohol use | Sunscreen use | Overeats |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| Behavior | -0.401*** | -0.395*** | -0.377*** | -0.383*** | -0.419*** | -0.437*** | -0.329*** | -0.405*** | -0.396*** | -0.391*** | -0.370*** |
| | (0.021) | (0.018) | (0.020) | (0.017) | (0.024) | (0.020) | (0.021) | (0.019) | (0.015) | (0.022) | (0.024) |
| High group | -0.045 | 0.051 | -0.020 | 0.024 | -0.096*** | -0.094*** | 0.176*** | -0.046 | -0.045 | 0.064** | 0.049 |
| | (0.031) | (0.034) | (0.032) | (0.036) | (0.031) | (0.031) | (0.031) | (0.033) | (0.066) | (0.032) | (0.032) |
| Behavior × High group | 0.020 | 0.012 | -0.029 | -0.030 | 0.048 | 0.084*** | -0.111*** | 0.041 | 0.079 | 0.002 | -0.035 |
| | (0.030) | (0.033) | (0.030) | (0.035) | (0.030) | (0.029) | (0.029) | (0.031) | (0.067) | (0.030) | (0.031) |
| Constant | 0.536*** | 0.496*** | 0.523*** | 0.507*** | 0.572*** | 0.565*** | 0.415*** | 0.530*** | 0.516*** | 0.476*** | 0.485*** |
| | (0.021) | (0.019) | (0.021) | (0.018) | (0.024) | (0.022) | (0.024) | (0.020) | (0.016) | (0.024) | (0.026) |
| $R^2$ | 0.25 | 0.26 | 0.25 | 0.25 | 0.26 | 0.26 | 0.28 | 0.25 | 0.25 | 0.26 | 0.25 |
| Number of observations | 7,185 | 7,185 | 7,161 | 7,185 | 7,185 | 7,185 | 7,185 | 7,185 | 7,185 | 7,185 | 7,185 |
| Number of clusters | 397 | 397 | 395 | 397 | 397 | 397 | 397 | 397 | 397 | 397 | 397 |

Notes: The dependent variable is an indicator of weekly smoking abstinence. The data are stacked with observed behavior on top of predictions. "Behavior" equals 1 for observed abstinence and 0 for incentivized abstinence predictions. The "high" value of each dimension of heterogeneity, each measured at baseline, is listed at the top of each column. All models include individual random effects. Age, household income, and alcohol use are dichotomized at the median. Education equals 1 if the person has at least some college or an associate's degree. Cigarettes per day is split at $\geq 10$. Nicotine dependence is split at Fagerström scores $\geq 4$. Quit plans equals 1 if plans to quit within one year. Present bias, based on a hypothetical monetary choice task, equals 1 if prefers larger immediate payment and smaller later payment. Sunscreen use equals 1 if regularly uses sunscreen when outdoors. Overeats equals 1 if regularly eats an amount of food later regretted.

Table 6: Structural Estimates

Panel A: Structural Parameters

| Description | Parameter | Estimate | 95% CI |
|---|---|---|---|
| Weekly disutility of abstinence | $\mu_c$ | $8.883^{***}$ | [8.336, 9.253] |
| Discounted long-run benefit of abstinence | $b$ | $8.783^{***}$ | [8.443, 9.311] |
| Utils per dollar | $\gamma$ | 0.001 | [−0.001, 0.003] |
| Scale parameter | $\sigma$ | $0.105^{***}$ | [0.082, 0.110] |
| Habit value | $\eta$ | $0.555^{***}$ | [0.150, 0.969] |
| Perceived habit value in Month 1 | $\tilde{\eta}_1$ | 0.126 | [−0.353, 0.680] |
| Perceived habit vaue in Month 3 | $\tilde{\eta}_2$ | −0.084 | [−0.545, 0.502] |
| Present bias (mean) | $\bar{\beta}$ | $0.670^{***}$ | [0.542, 0.730] |
| Degree of naïvete (mean) | $\bar{\tilde{\beta}}$ | $0.851^{***}$ | [0.644, 1.013] |

Panel B: Heterogeneity

| Covariate | $\beta$ | $\tilde{\beta}$ | $\tilde{\eta}_1$ | $\tilde{\eta}_2$ |
|---|---|---|---|---|
| Constant | $0.678^{*}$ | $1.432^{***}$ | −1.091 | 0.036 |
| | [−0.076, 1.281] | [0.491, 2.923] | [−7.646, 3.210] | [−6.519, 4.340] |
| Age (decades) | −0.002 | $-0.011^{***}$ | $0.042^{***}$ | $0.030^{**}$ |
| | [−0.006, 0.001] | [−0.019, −0.005] | [0.013, 0.086] | [0.002, 0.076] |
| Ln(Income) | 0.017 | 0.005 | 0.017 | −0.049 |
| | [−0.033, 0.073] | [−0.123, 0.111] | [−0.482, 0.585] | [−0.572, 0.480] |
| Nicotine dependence | −0.000 | −0.002 | −0.127 | −0.137 |
| | [−0.021, 0.025] | [−0.059, 0.064] | [−0.433, 0.152] | [−0.399, 0.127] |
| Cigarettes | $-0.006^{**}$ | −0.006 | 0.011 | 0.008 |
| | [−0.014, −0.000] | [−0.024, 0.009] | [−0.061, 0.093] | [−0.061, 0.093] |
| Male | $0.126^{***}$ | 0.126 | −0.011 | 0.405 |
| | [0.046, 0.200] | [−0.056, 0.298] | [−0.918, 0.802] | [−0.514, 1.178] |
| Education | −0.004 | 0.035 | −0.172 | −0.348 |
| | [−0.047, 0.046] | [−0.094, 0.153] | [−0.651, 0.344] | [−0.777, 0.150] |
| Quit plans | $-0.029^{**}$ | $-0.037^{*}$ | −0.110 | −0.104 |
| | [−0.060, −0.004] | [−0.090, 0.000] | [−0.343, 0.190] | [−0.302, 0.210] |
| Alcohol use | 0.010 | 0.008 | 0.028 | 0.034 |
| | [−0.031, 0.024] | [−0.046, 0.024] | [−0.161, 0.219] | [−0.104, 0.238] |
| Sunscreen use | $0.080^{**}$ | −0.087 | $0.664^{*}$ | $0.956^{**}$ |
| | [0.002, 0.163] | [−0.272, 0.098] | [−0.123, 1.451] | [0.123, 1.723] |
| Overeats | 0.014 | 0.042 | 0.021 | 0.049 |
| | [−0.064, 0.099] | [−0.179, 0.188] | [−0.625, 1.025] | [−0.583, 0.987] |
| Present bias (money) | −0.008 | −0.054 | | |
| | [−0.090, 0.062] | [−0.160, 0.037] | | |

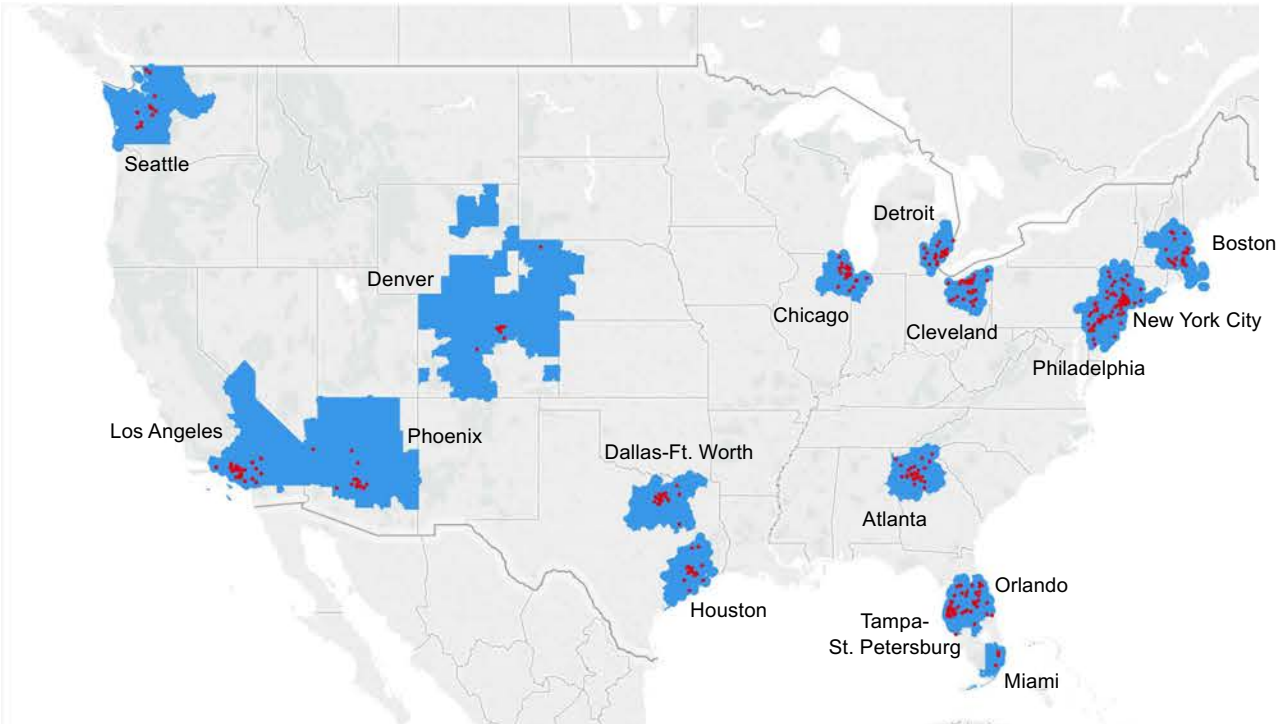
Panel C: Additional Tests

| Test | Value | p-value |
|---|---|---|
| $b - \mu_c$ | -0.100 | 0.916 |
| $b - \mu_c + \eta$ | 0.455 | 0.156 |
| $\bar{\tilde{\beta}} - \bar{\beta}$ | 0.180 | 0.014 |
| $\eta - \tilde{\eta}_1$ | 0.429 | 0.136 |
| $\eta - \tilde{\eta}_2$ | 0.639 | 0.036 |
| $\tilde{\eta}_1 - \tilde{\eta}_2$ | 0.210 | 0.084 |
| $(b - \mu_c + \eta)/\gamma$ | 413.922 | 0.344 |

Notes: Structural estimates are based on a maximum likelihood procedure described in Section 5. Panel A reports the main structural parameters, including the means of $\beta$ and $\tilde{\beta}$, denoted by $\bar{\beta}$ and $\bar{\tilde{\beta}}$. Panel B reports the observed covariates of $\beta$, $\tilde{\beta}$, $\tilde{\eta}_1$, and $\tilde{\eta}_2$. Panel C reports additional tests of linear combinations of parameters in Panel A.

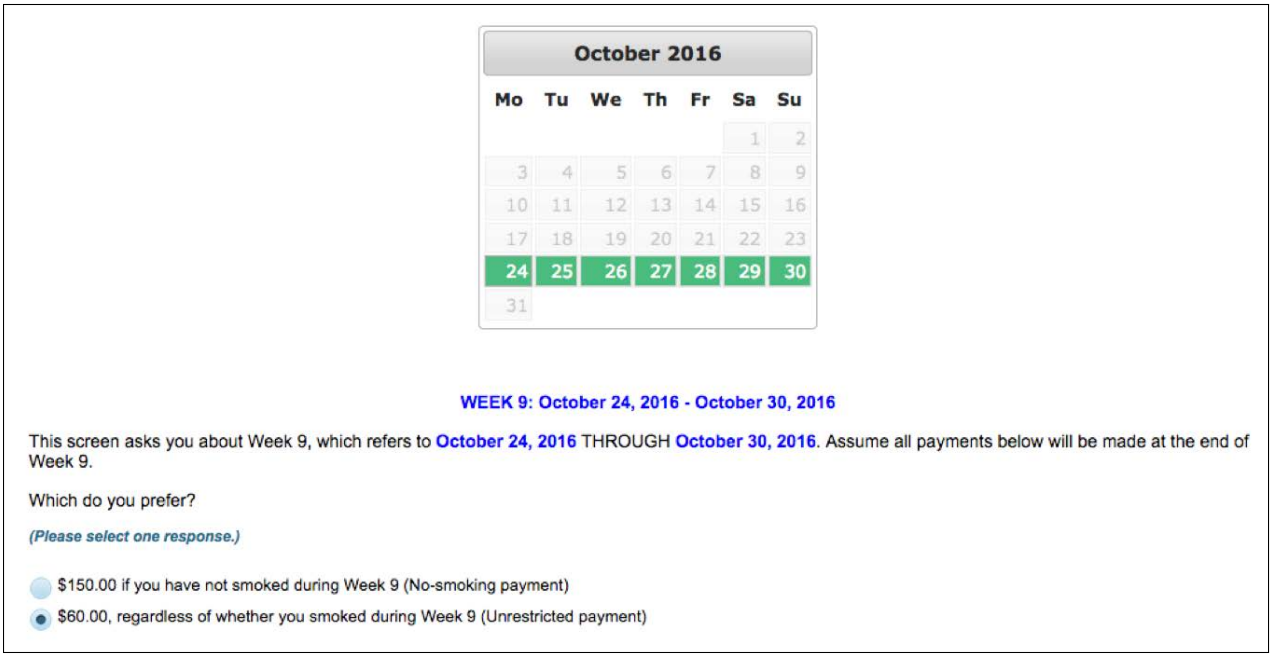
# Appendix A Additional Results

Figure A.1: Zip Codes from Target DMAs




Note: This map shows targeted zip codes in blue ($N = 8{,}820$), along with the name of the targeted DMAs, and subjects' zip codes as red dots.

Figure A.2: Sample Elicitation Choice

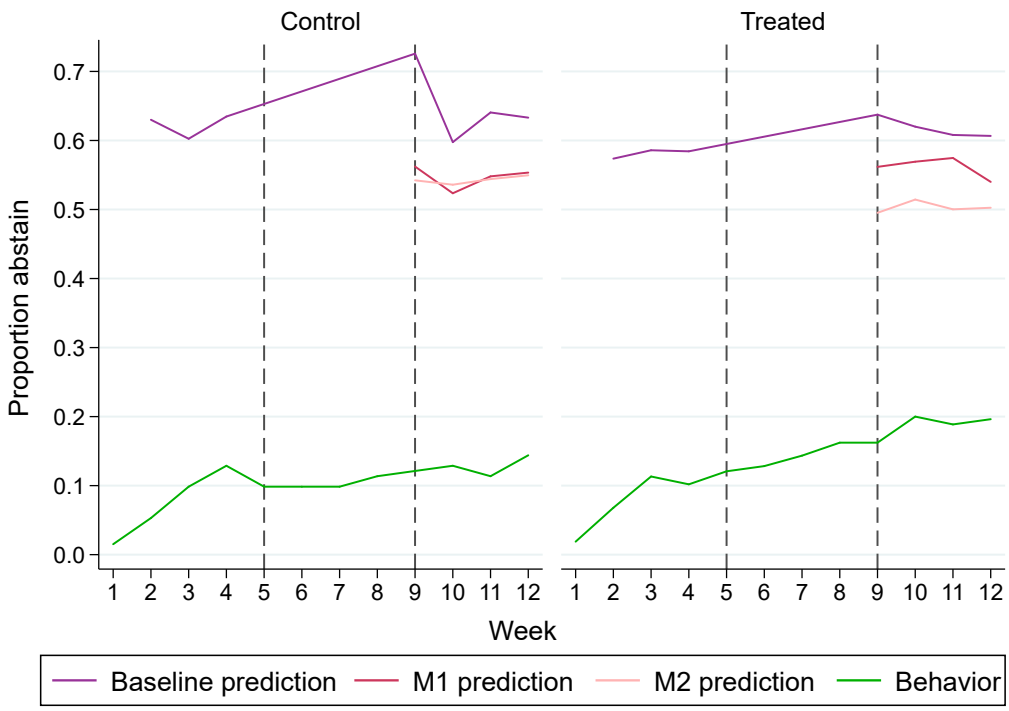



Note: This is an example of one of binary choices presented to subjects during the beliefs elicitation. In this case, $p = \$150$ and the non-contingent payment is $0.4p$.

Figure A.3: Sample Smokefree-Contingent Certificate


Note: In smokefree-contingent weeks, subjects were presented within the survey with a downloadable certificate that detailed the terms of payment.

Figure A.4: Abstinence Behavior and Incentivized Predictions by Week and Treatment Group




Note: This figure shows the average abstinence behavior or average predictions from each survey (incentivized), by study week and treatment group. Weeks in which a subject was eligible for a smokefree-contingent payment are included. The left panel is for the control group, and the right is for the treated group.

# Appendix B Experimental Elicitation Instructions

The instructions for eliciting valuations of abstinence during the baseline session are provided starting on the next page. Programming notes (PNs) are included in caps and brackets. In the instructions, the valuation exercise is referred to as the "smoking choice exercise," the smokefree-contingent payment as the "no-smoking payment," and the non-contingent payment as the "unrestricted payment."

The instructions for the end-of-Month-1 and end-of-Month-2 elicitation sessions followed identical procedures to the ones below.

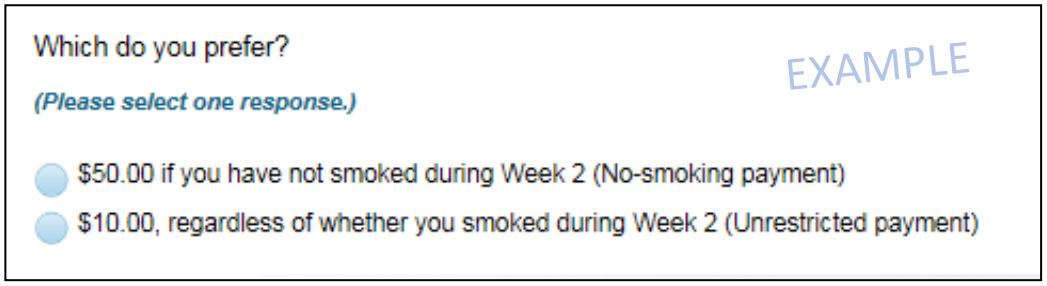
**Baseline Elicitation Instructions**

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q460** On the following screens you will answer 12-16 similar questions about different weeks and different payment amounts. Each question is a decision between a payment for not smoking during a given week (a no-smoking payment) or an unrestricted payment regardless of your smoking that week.

At the end of this survey, we will randomly select just one person. In addition, we will randomly select just one week and one question to determine the no-smoking vs. unrestricted payment preference. The selected person will actually receive his or her chosen payment option during the given week. You should therefore answer these questions truthfully, so that you are awarded your preferred option if you are selected. The no-smoking payment would depend on the results of a saliva test to confirm that you did not smoke that week.

Here is an example question:
[PN: INSERT IMAGE EXAMPLE BELOW]




Which do you prefer?

- If someone chooses the first option, he/she prefers a payment of $50 that depends on not smoking during Week 2 to an unrestricted payment of $10 regardless of whether he/she smoked that week.
- If someone chooses the second option, he/she prefers the unrestricted payment of $10 to the payment of $50 that depends on not smoking.

Both payments would be made at the same time—for example, at the end of Week 2. Please consider one week from this coming Monday to be the start of Week 1.

[PN: LOOP THROUGH Q465-Q501 FOUR TIMES. IN EACH LOOP, RESPONDENT WILL BE ASKED ABOUT A DIFFERENT RANDOMLY SELECTED VALUE FOR $P AND A DIFFERENT RANDOMLY SELECTED WEEK. USE Q465 FOR RANDOMLY SELECTED VALUES. USE Q467 AND Q467a FOR RANDOMLY SELECTED WEEKS. THE ORDERING OF THE FOUR LOOPS SHOULD BE RANDOMZED TO BE EITHER IN CHRONOLOGICAL OR REVERSE-CHRONOLOGICAL ORDER. FOR EXAMPLE, IF THE SELECTED WEEKS IN Q467a ARE 1, 3, 9, AND 11, THEN THE DISPLAY ORDER OF THE LOOPS

WOULD BE WEEKS 1, 3, 9, 11 (CHRONOLOGICAL ORDER) OR WEEKS 11, 9, 3, 1 (REVERSE CHRONOLOGICAL ORDER.)]

[PN: IN EACH LOOP OF Q470-Q494, RESPONDENT ANSWERS 3-4 QUESTIONS BASED ON AN ALGORITHM. THERE ARE TWO STARTING POINTS FOR THE ALGORITHM, 40% OF $P OR 80% OF $P. USE Q468 TO DETERMINE THE STARTING POINT FOR THE LOOP.]

[PN: TRACK THE ORDER OF THE SCREENS SHOWN.]

[PN: FOR ALL THE PERCENT VALUES SHOWN IN Q470-Q501, CALCULATE AND DISPLAY VALUE OF $P.]

[PN: Q465-Q501 ARE MANDATORY.]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE**

**Q465** $P.00 INSERTS FOR Q470-Q501 – DO NOT DISPLAY [ALLOW 2 DECIMAL VALUES FOR P VALUE CALCULATED]

[PN: RANDOMLY SELECT USING LOW READS FOUR AMOUNTS BELOW, TWO OF THOSE LISTED IN 1 THRU 4 AND TWO LISTED IN 5 THRU 8. THESE WILL BE INSERTED INTO Q470-Q501.]

1 $10
2 $25
3 $50
4 $100
5 $150
6 $200
7 $300
8 $400

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q467** WEEK X INSERTS FOR Q470-Q501 – DO NOT DISPLAY

[PN: RANDOMLY SELECT USING LOW READS FOUR WEEKS BELOW, TWO OF THOSE LISTED IN 1 THRU 4 AND TWO LISTED IN 5 THRU 8. THESE WILL BE INSERTED INTO Q470-Q501.]

2 Week 2
3 Week 3
4 Week 4
9 Week 9
10 Week 10
11 Week 11
12 Week 12

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q467a** [DATE INSERTS FOR Q470-Q501 – DO NOT DISPLAY]

[PN: THIS IS A PLACEHOLDER FOR THE DATES TO INSERT INTO Q470-Q501, CURRENTLY LISTED IN THOSE QUESTIONS AS [MONTH DD] – [MONTH DD, YYYY].]

2 When Q467/2 = March 27, 2017 - April 2, 2017
3 When Q467/3 = April 3, 2017 - April 9, 2017
4 When Q467/4 = April 10, 2017 - April 16, 2017
9 When Q467/9 = May 15, 2017 - May 21, 2017
10 When Q467/10 = May 22, 2017 - May 28, 2017
11 When Q467/11 = May 29, 2017 - June 4, 2017
12 When Q467/12 = June 5, 2017 - June 11, 2017

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q468** STARTING POINT FOR ALOGRITHM – DO NOT DISPLAY

[PN: RANDOMLY SELECT A RESPONSE FOR EACH RESPONDENT BASED ON LOW READS.]
[PN: A NEW STARTING POINT SHOULD BE RANDOMLY SELECTED FOR EACH LOOP FOR ANY GIVEN RESPONDENT.]

1 80%
2 40%

[PN: FOR Q470-Q501, IT'D BE IDEAL IF WE COULD INCLUDE A CALENDAR AT THE TOP OF THE SCREEN THAT SHOWED THE MONTH WHEN THE SURVEY IS TAKING PLACE THROUGH THE MONTH IN WHICH WEEK X FALLS IN ORDER FOR RESPONDENTS TO HAVE A CLEAR IDEA OF WHICH DATES ARE BEING REFERRED TO.]

[PN: FOR THE DATES OF WEEK X SHOWN IN Q470-Q501, USE THE MONDAY AFTER THE BASELINE SURVEY DATE AS THE START OF WEEK 1, AND MOVE FORWARD FROM THERE. WEEKS RUN FROM MONDAY TO SUNDAY.]

**BASE: STARTING POINT AT 80% - BASELINE (Q468/1)**

**Q470**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q478]

2 [80% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q476]

**BASE: STARTING POINT AT 40% - BASELINE (Q468/2)**
**Q472**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: GO TO Q474]
2 [40% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q482]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q472/1)**
**Q474**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q478]
2 [80% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q480]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q470/2)**
**Q476**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q480]
2 [40% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q482]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q470/1 OR Q474/1)**
**Q478**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q484]
2 [100% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q486]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q474/2 OR Q476/1)**
**Q480**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q488]
2 [60% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q490]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q472/2 OR Q476/2)**
**Q482**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q492]
2 [20% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q494]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q478/1)**

**Q484**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]

2 [110% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q478/2)**

**Q486**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]

2 [90% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q480/1)**

**Q488**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]

2 [70% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q480/2)**
**Q490**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]
2 [50% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q482/1)**
**Q492**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]
2 [30% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE (Q482/2)**
**Q494**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Which do you prefer?

1 [$P.00] if you have not smoked during Week [X] (No-smoking payment) [PN: SKIP TO Q496]
2 [10% OF $P.00], regardless of whether you smoked during Week [X] (Unrestricted payment) [PN: SKIP TO Q496]

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q496** ASSIGN A PREDICTION RANGE BASED ON Q470-Q494 (DOES NOT APPEAR ON SCREEN)

| | | |
|---|---|---|
| 1 | 0.0 – 0.1 | [PN: IF Q494/2] |
| 2 | 0.1 – 0.2 | [PN: IF Q494/1] |
| 3 | 0.2 – 0.3 | [PN: IF Q492/2] |
| 4 | 0.3 – 0.4 | [PN: IF Q492/1] |
| 5 | 0.4 – 0.5 | [PN: IF Q490/2] |
| 6 | 0.5 – 0.6 | [PN: IF Q490/1] |
| 7 | 0.6 – 0.7 | [PN: IF Q488/2] |
| 8 | 0.7 – 0.8 | [PN: IF Q488/1] |
| 9 | 0.8 – 0.9 | [PN: IF Q486/2] |
| 10 | 0.9 – 1.0 | [PN: IF Q486/1] |
| 11 | 1.0 – 1.1 | [PN: IF Q484/2] |
| 12 | > 1.1 | [PN: IF Q484/1] |

[PN: hPValPipe: Due to suggested change to Q500 below, we can remove calculation for 1%.]

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**

**Q500**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

How much of an unrestricted payment would you need to prefer the unrestricted payment over a no-smoking payment of [$P.00] that depends on not smoking in Week [X]?

[RANGE: 0-900]
$|__|__|__|

**BASE: ALL QUALIFIED RESPONDENTS – BASELINE**

**Q501**

WEEK [X]: [MONTH DD] – [MONTH DD, YYYY]

This screen asks you about Week [X], which refers to **[INSERT: MONTH DD, YYYY]** THROUGH **[INSERT: MONTH DD, YYYY]**. Assume all payments below will be made at the end of Week [X].

Imagine that you have just received:

A payment of [$P.00] if you have not smoked during Week [X].

On a scale between 0% (no chance) and 100% (complete certainty], how likely do you think it is that you would not smoke during Week [X] if you received this payment? Assume that you have to complete a saliva test to confirm that you are smoke-free.

[PN: INSERT SLIDER WITH 5% INCREMENTS AND 0% AND 100% AS END ANCHORS.]

**BASE: ALL QUALIFIED RESPONDENTS**
**Q565** RANDOM USERID SELECTION FOR Q460 DRAWING

[PN**:** SELECT ONE USERID – QUOTA OF 1.]

**BASE: ALL QUALIFIED RESPONDENTS - BASELINE**
**Q568** SUBJECT THE WINNER OF Q460 DRAWING / SMOKING CHOICE EXERCISE – DO NOT DISPLAY

**PROGRAMMING NOTE:**
1. IF THE RESPONDENT USERID IS SELECTED IN Q565, CODE AS 2 (YES).
2. IF THE RESPONDENT USERID IS NOT SELECTED IN Q565, CODE AS 1 (NO).

1 No
2 Yes

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**
**Q570** WINNING WEEK FOR Q460 DRAWING / SMOKING CHOICE EXERCISE – DO NOT DISPLAY

[PN: FILL IN THE RESPONSES BELOW WITH THE FOUR SELECTED WEEKS IN Q467. RANDOMLY CHOOSE ONE WEEK FROM AMONG THE FOUR.]

2 Week 2
3 Week 3
4 Week 4
9 Week 9
10 Week 10
11 Week 11
12 Week 12

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**
**Q572** WINNING "ROW" FOR Q460 DRAWING / SMOKING CHOICE EXERCISE – DO NOT DISPLAY

[PN: RANDOMLY CHOOSE A NUMBER FROM 1-12.]

1 0.1
2 0.2
3 0.3

4 0.4
5 0.5
6 0.6
7 0.7
8 0.8
9 0.9
10 1.0
11 1.1
12 1.2

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**

**Q573** PAYOFF FROM NO-SMOKING PAYMENT IN WINNING WEEK (Q570) – DO NOT DISPLAY

[PN: FOR THE WEEK SELECTED IN Q570, SELECT THE CORRESPONDING $P INSERT (Q465) FROM THE SAME LOOP.]

1 $10.00
2 $25.00
3 $50.00
4 $100.00
5 $150.00
6 $200.00
7 $300.00
8 $400.00

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**

**Q574** PAYOFF FROM UNRESTRICTED PAYMENT IN WINNING WEEK (Q570) – DO NOT DISPLAY

[PN: LET THE RESPONSE BELOW BE THE VALUE IN Q572 MULTIPLIED BY THE VALUE IN Q573. FOR EXAMPLE, LET Q573 BE 100.00. IF Q572/1, THEN ENTER 10.00 = 0.1*100.00. IF Q572/12, THEN ENTER 120.00 = 1.2*100.00. INCLUDE TWO DECIMAL PLACES.]

[RANGE: 1-480]
$|__|__|__|.|__|__|

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**

**Q575** PAYOFF FROM Q460 DRAWING / SMOKING CHOICE EXERCISE – DO NOT DISPLAY

[PN: IF THE RESPONSE CODE IN Q572 ≥ THE RESPONSE CODE IN Q496, SELECT 2 = UNRESTRICTED PAYMENT; IF THE RESPONSE CODE IN Q572 < THE RESPONSE CODE IN Q496, SELECT 1 = NO SMOKING PAYMENT. FOR EXAMPLE, IF Q572/4 AND Q496/3, SELECT 2 BELOW.

1 [PAYOFF FROM **NO-SMOKING PAYMENT** (Q573)]
2 [PAYOFF FROM **UNRESTRICTED PAYMENT** (Q574)]

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**

**Q580** PAYOFF FROM Q460 DRAWING / SMOKING CHOICE EXERCISE – DO NOT DISPLAY

**PROGRAMMING NOTE**: FOR WINNER, Q568/2, STORE THE RESPONDENTS FINAL OPTION SELECTED FROM Q460 DRAWING

1 **PROGRAMMING NOTE**: INSERT PREDICTION RANGE (Q496)
2 **PROGRAMMING NOTE**: INSERT FINAL OPTION SELECTED (Q572)
3 **PROGRAMMING NOTE**: INSERT FINAL OPTION SELECTED, NO-SMOKING PAYMENT VALUE (Q573)
4 **PROGRAMMING NOTE**: INSERT FINAL OPTION SELECTED, UNRESTRICTED PAYMENT VALUE (Q574)

**BASE: WINNER OF Q460 DRAWING - BASELINE (Q568/2)**

**Q585** [IF Q575/1, THEN DISPLAY THE FOLLOWING TEXT AND GRAPHIC:] Congratulations! You were selected to win an electronic gift card worth [INSERT VALUE FROM Q573] if you do NOT smoke during Week [INSERT: NUMBER FROM Q570] of the study. We will remind you of this prize during the prior week.

[PN: FOR CERTIFICATE BELOW, REPLACE $P WITH AMOUNT IN Q573. THE DATES SHOULD MATCH THE START AND END DATE FOR THE WEEK LISTED IN Q570, WHERE THE START AND END DATES ARE DEFINED FOR THAT LOOP OF Q460.]


[IF Q575/2, THEN DISPLAY THE FOLLOWING TEXT AND GRAPHIC:] Congratulations! You were selected to win an electronic gift card worth [INSERT VALUE FROM Q574] at the end of Week [INSERT: NUMBER FROM Q570].

[PN: FOR CERTIFICATE BELOW, REPLACE $P WITH AMOUNT IN Q574. THE DATES SHOULD MATCH THE START AND END DATE FOR THE WEEK LISTED IN Q570, WHERE THE START AND END DATES ARE DEFINED FOR THAT LOOP OF Q460.]

**$P** Reward Certificate **$P**

*This certificate entitles the holder to*

**$P**

*at the end of the week*

*of*

**XXXday, MMM DD** *through* **XXXday, MMM DD**.

**$P** **$P**

[PN: PLEASE SETUP EACH CERTIFICATE TO INCLUDE A PRINT ICON AS A POPUP FOR THE RESPONDENT TO PRINT THE ONLINE CERTIFICATE FOR THEIR RECORDS. INCLUDE ROLL-OVER TEXT ON EACH HYPERLINK THAT STATES “Click here to view, print, and/or save the Certificate”.]

[PN: SETUP PRINT OPTION AS NON-MANDATORY]

# Appendix C Proof of Identification

**Assumption 1 (Utility)** *The utilities from smoking choices are given by:*

$$u(abstain|p, control) = \beta\delta b + v(p) - \mu_c - \varepsilon_{it} \tag{14}$$

$$u(abstain|p, treated) = u(abstain|p, control) + \eta \tag{15}$$

$$u(smoke|p) = -\varepsilon_{s,t} \tag{16}$$

*where* $v(0) = 0$ *and* $v'(\cdot) > 0$.

**Assumption 2 (Discrete shocks)** *The shocks* $\varepsilon_{i,t}$ *are distributed i.i.d. according to a standard type-I extreme value distribution* $F(\cdot)$.

We begin by noting that the difference of two standard type-I extreme value distributions is logistic, as assumed in the text. Under assumption 2, it is possible to write the probability of abstinence as the ratio of exponentiated utilities as in equation 6. The mean abstinence rate in the absence of incentives immediately identifies $(\beta\delta b - \mu_c)$. Moreover, $v(p)$ is identified from the slope of this logit regression, and thus $v^{-1}(\cdot)$. Finally, comparing treated and control subjects in month 3 identifies $\eta$ given the linearity assumption in Assumption 1.

If $\tilde{\beta} = 1$, then as $p \to 0$ the second term of equation (10) goes to zero faster than the first term. Sending $p$ to zero thus identifies $F(\tilde{\beta}\delta b - \mu_c)$, which can be inverted to recover $\tilde{\beta}\delta b - \mu_c$.

If $\tilde{\beta} < 1$, then (10) induces a first-order "commitment value" at all incentive levels. We instead consider how this changes at high values of $p$. Taking the derivative with respect to $p$ yields:

$$\frac{dV}{dp} = F(\chi(p)) + \phi f(\chi(p)) \tag{17}$$

where $\chi(p) = \tilde{\beta}\delta b + v(p) - \mu_c$ and $\phi = (1 - \tilde{\beta})\delta b$. This derivative implies a locus of $(\phi, \chi)$ pairs consistent with the observed response to changes in $p$, in particular:

$$\phi = \left[\frac{\partial V/\partial p - F(\chi)}{f(\chi)}\right] \tag{18}$$

Note that $\phi$ is not a function of $p$. We may therefore evaluate $\partial V/\partial p$ at some other point $p'$. Note that $\chi(p') = \chi(p) + (v(p') - v(p)) \equiv \chi(p) + \Delta$, which is known from the discrete choice. Differencing equation (18) at these two points, and making the above substitution yields:

$$\frac{\partial V/\partial p - F(\chi)}{f(\chi)} - \frac{\partial V/\partial p' - F(\chi + \Delta)}{f(\chi + \Delta)} = 0 \tag{19}$$

Taking the derivative, and substituting in equation (17) yields:

$$\phi \left[ \frac{f'(\chi + \Delta)}{f(\chi + \Delta)} - \frac{f'(\chi)}{f(\chi)} \right] < 0 \tag{20}$$

where the inequality is guaranteed as $\phi > 0$ and $f'/f$ is a strictly decreasing function given assumption 2. Equation (20) is thus a monotone function of $x$, which admits at most one solution. Thus there is a unique value $\chi(p) = \tilde{\beta}\delta b + v(p) - \mu_c$ consistent with the data. Given that $v(p)$ is known from the discrete choice, this identifies $\tilde{\beta}\delta b - \mu_c$.

Finally, evaluate equation (10) as $p \to \infty$. The probability of abstinence goes to one, leaving just the second term. However, this is an integral from the known $\tilde{\beta}\delta b - \mu_c$ to infinity of $\delta b - \mu_c - \varepsilon$. This is an invertible function of $(\delta b - \mu_c)$, which is therefore identified. Equation (17) then identifies a unique value for $\phi = (1 - \tilde{\beta})\delta b$.