Habit Formation and Persistence in the Case of Stationary Exercise

Matthew C. Harris* Lawrence M. Kessler†

January 12, 2018

Abstract

This paper investigates the rate at which habits form - and depreciate - in the context of stationary exercise. Using a novel data set from a manufacturer of stationary fitness equipment, we examine the relationship between the number of days that an individual exercises during their first two, four, six, and eight weeks and the probability that they continue to exercise regularly going forward. Using a two-stage duration model with weather variables as exclusion restrictions, our model predicts that both the length of the initial period and frequency of activity during that initial period lead to more persistent exercise over the next six months, but that any stock of habituation depreciates quickly in the absence of observed workouts. Our findings imply that for interventions to lead to persistent behavioral chance, they must require more frequent activity or longer intervention periods than recent incentive-based field experiments.

Keywords: State Dependence, Habit, Health Behaviors, Survival Analysis

*University of Tennessee, Haslam College of Business, Department of Economics and Boyd Center for Business and Economic Research. Corresponding Author: 722A Stokely Management Center, 916 Volunteer Boulevard, Knoxville, TN 37996-0570, (865) 974-5591 mharris@utk.edu.

†University of Tennessee, Haslam College of Business, Department of Economics and Boyd Center for Business and Economic Research.

‡We are grateful to Blake Abercrombie for his research partnership and Jun Soo Lee for his work as a research assistant. We are grateful to Pravin Trivedi, Rusty Tchernis, Anirban Basu, Marianne Wanamaker, Celeste Carruthers, Seth Gershenson, and Georg Schaur for their helpful comments as well as participants in the 7^{th} World Congress of the International Health Economics Association and Vanderbilt Empirical Applied Microeconomics Festival. All errors are our own.

1 Introduction

Individuals exhibit persistence in many types of decisions including brand of consumer package goods, health insurance, and employment decisions; even when such decisions appear inconsistent with rational forward-looking behavior (Keane, 1997; Dube et al., 2010; Handel, 2013).[1] When individuals persist in making consumption decisions that lead to adverse long-run health outcomes, (e.g., smoking, drinking, etc.) economists frequently appeal to habit formation or rational addiction models (Ryder and Heal, 1973; Becker and Murphy, 1988; Constantinides, 1990).[2,3] However, there are some areas in which individuals exhibit persistence in choices that lead to undesirable long-run outcomes, but "addiction" is an unlikely explanation. For example, individuals may choose to abstain from a particular good or activity, such as eating fruits and vegetables or exercising. While plausible explanations include simple heterogeneity in preferences, the behavioral economics literature posits that status-quo bias or "inertia" may explain why it is difficult to get individuals to change their behavior (Samuelson and Zeckhauser, 1988; Kahneman et al., 1991; Thaler and Sunstein, 2003). The main claim in the inertia literature is that individuals keep repeating the same choices simply by default, rather than because of addiction, switching costs, or even heterogeneous preferences.

Habit-in-utility and inertia (or status-quo bias) are not mutually exclusive explanations for persistence in individuals choices. Consider the case of individuals persistently choosing not to exercise and therefore becoming sedentary. Sedentary lifestyles have become increasingly common among U.S. adults. As a result of this and other factors, obesity and obesity-related diseases have become more prevalent (Flegal et al., 2010). Sedentary lifestyles have been linked to increased likelihood of heart disease, type 2 diabetes, hypertension, and other chronic health conditions (Flegal et al., 2010; Ogden et al., 2014). Employers, researchers, and policy makers continue to seek ways to encourage physical activity and avoid adverse health outcomes. As interventions are costly, affecting not just change, but persistent behavioral change is essential for attaining long-run objectives and maximizing returns on resources invested in these

[1] Several adaptations to forward looking rational optimization models have been proposed to better capture this kind of behavior, including switching costs, learning costs, and hyperbolic discounting (Spence, 1981; vonWeizsacker, 1984; Klemperer, 1987).

[2] See also Chaloupka (1991); Ferson and Constantinides (1991); Olekalns and Bardsley (1996); Campbell and Cochrane (1999); Darden (2017).

[3] The key feature of rational addiction (or habit-in-utility) models is that contemporaneous consumption affects the marginal utility of consuming those goods in future periods, which also implies that current consumption affects the future costs of cessation.

interventions.

Both the concepts of habit-in-utility and inertia are useful in considering how to affect persistent change in individuals patterns of physical activity. First, individuals may have status-quo bias or inertia towards being physically inactive. If being sedentary is a default option, individuals will more likely revert to that alternative. Getting individuals to exercise more frequently during an intervention or more regularly over a long period can erode the inertia of inactivity. Second, for an intervention to affect persistent change, participants will need to build some 'stock of habituation' in physical activity. The notion of 'getting in shape' implies that physical activity today affects the marginal (dis)utility of physical activity in subsequent periods, akin to 'tolerance' in the rational addiction literature. Engaging in physical activity with greater frequency, or engaging in regular physical activity over a longer period of time should increase that stock.

In this paper, we employ a novel data set from a manufacturer of stationary exercise equipment to empirically investigate persistence, which we treat as evidence of habit formation, in the context of equipment usage. Equipment users can download an app that pairs their mobile phones to the company's fitness machines, enabling users to track their workouts over time. Our data set contains all app-tracked workouts on this company's equipment from 2011 to 2015, meaning we have multiple years of daily reported workout activity for some users. We investigate the relationships between the frequency with which individuals exercise during an initial period and whether the length of said initial period affects the probability that an individual continues to exercise regularly over the next six months. By understanding the relationship between intervention duration, frequency of activity, and subsequent persistence; we hope our findings can provide insight on the length of an intervention period and frequency of activity needed to erode inertia and form new habits. We consider four 'initial period' durations: two, four, six, and eight weeks. For each defined period, we estimate a two-stage control-function duration model to analyze how the number of days that an individual exercised in the initial period impacts their likelihood of continuing to exercise regularly. For example, we can use this model to estimate how exercising 10 vs 15 days in a given initial period (e.g., 4 weeks) affects the probability that an individual continues to exercise with a certain regularity. We also use the estimates from our model to make comparisons between initial periods of varying length. For example, if an individual exercises 16 times over a four week initial period, how many times will he need to exercise during the fifth and six weeks to maintain the same predicted level of

persistence in his exercise behavior going forward?

There are two main threats to identification in our data and research design, both of which we address. First, there are reasons to be concerned about correlation in the unobservables that determine both the frequency of exercise within an initial period and the propensity to exercise in subsequent periods (e.g., preferences for physical activity, perseverance, unobserved patterns of exercise prior to registering for the application, etc.). Second, for us to observe an individual exercising, the individual must make all four of the following choices: 1.) to exercise, 2.) exercise on stationary exercise equipment, 3.) exercise on our manufacturers brand of stationary equipment, and 4.) track it using the app. If we do not observe an individual exercising during a given period, we do not know which of these four conditions were not met. Therefore, we cannot discern exercising outside or on a competitor's equipment from inactivity.

Our identification strategy simultaneously addresses both concerns. We merge individuals' location data with weather data from National Oceanic and Atmospheric Administration (NOAA), and use average precipitation and average daily high temperature as instruments for the number of times an individual exercises over the defined initial period. Recall the four choices that an individual must make in order to be observed working out. While weather may affect an individual's propensity to exercise (choice 1), or to use stationary equipment conditional on exercising (choice 2), there is no reason to suggest that conditional on choices 1 and 2, that weather affects users choice of equipment brand (choice 3) or use of the app (choice 4). Therefore, for weather to affect the number of workouts during an initial period, it must be through the extensive margin on whether or not to exercise, or to use stationary equipment. In our first-stage results, we find that higher average precipitation decreases the number of days individuals engage in stationary exercise. While this may initially be counterintuitive, a majority of the individuals in our data are exercising at health clubs. On rainy days, individuals are more inclined to go directly home after work and stay there. While the substitution effect is likely present to some extent, it is dominated by the effect of rain on individuals decision to exercise at the extensive margin.[4] If our instrument was inducing substitution between indoor and

[4]More generally, rainy days affect the way individuals allocate their time between work, home production, and leisure. Connolly (2008) finds that on rainy days, men work an additional average of 30 minutes and salaried workers watch an additional 16 minutes of television. Zivin and Neidell (2014) also find that temperature changes affect the time individuals spend at work. To the extent that weather induces individuals to allocate their time differently, weather will affect the amount of time available for stationary exercise. Eisenberg and Okeke (2009) also analyze individual-level exercise data from the Behavioral Risk Factor Surveillance Survey (BRFSS) and weather data from the NOAA to examine the impact of adverse weather on exercise behavior. They find that outside temperature has a statistically significant impact on physical activity and that in colder temperature ranges, a drop in temperature

outdoor physical activity, we would expect weather to positively affect the number of workouts we observe during an initial period.[5] To further address concerns about correlations between weather and regional preferences or seasonality, we also include state and month fixed effects.

We find that both the length of the initial period and the frequency with which an individual exercises during the initial period strongly affects the probability that they continue to exercise consistently in subsequent periods. Regardless of the length of the initial period, individuals who exercise four times per week during the initial period are 2.4 times more likely to exercise weekly over the next eight weeks than individuals who exercised three times a week over the same initial period. Conditional on exercise frequency, the length of the initial period also matters. Individuals who exercised an average of four times per week over the first eight weeks were 50 percent more likely to exercise each of the next eight weeks than individuals who exercised four times per week over the first four weeks. Finally, if an individual exercises five times per week over the first four weeks, he will only need to exercise four times per week over the subsequent four weeks to maintain the same expected level of persistence after the initial period. This last result indicates that the stock of habituation in our context deteriorates quickly - without constant maintenance, individuals will revert to sedentary behavior.

Our results are consistent with findings from recent studies using field experimental approaches, but offer new insight for future incentive-based interventions. Several recent studies have utilized field experiments to examine whether incentives can motivate physical activity through gym attendance (DellaVigna and Malmendier, 2014; Charness and Gneezy, 2009; Cawley and Price, 2013; Acland and Levy, 2015; Royer et al., 2015). This literature has generally found that financial incentives can successfully induce a strong positive response (i.e. greater gym attendance), but evidence on the long-run persistence of these behavioral changes is mixed (Gneezy et al., 2011). For example, Acland and Levy (2015) found that financial incentives positively impact gym attendance among university students, but these effects decay considerably once the students leave for winter break. While Royer et al. (2015) found some evidence to

leads to a decrease in exercise participation. Weather has also been used as an instrument in the habit literature, as Fujiwara et al. (2016) use weather as an exogenous source of variation in voter turnout in a model about habitual voting. Regarding the validity of our instrument, we argue that since weather is truly an exogenous shock it should have no relationship with other unobserved personal characteristics that impact a person's choice to exercise such as motivation and perseverance.

[5]Because some individuals may respond to rain by substituting from outdoor to indoor exercise (defiers), while others may substitute from stationary exercise to no exercise (compliers), the effect of our instrument on 'initial stock' is likely non-monotonic. However, de Chiasemartin (2017) shows that as long as the second-stage effects (i.e., effects of initial stock on persistence) are of the same sign for both groups, the second stage results still approximate the traditional local average treatment effect (LATE).

suggest that financial incentives combined with commitment contracts can induce some longer-lasting behavioral changes, the overall persistence of any incentive-induced behavioral changes was still relatively weak. However, both the response to incentives and the persistence of any changes are of nearly equal importance from a policy perspective.

Our results indicate that the lack of persistence in Charness and Gneezy (2009), Acland and Levy (2015), and Royer et al. (2015) may not a be a limitation of incentive-based interventions, but rather that the interventions did not require subjects to exercise a sufficiently high frequency (or sufficiently long intervention period) to break the status-quo bias for inactivity or form a new habit. Our model and empirics reproduce this lack of persistence found in earlier work, and suggests that interventions that required greater exercise frequency may lead to habit formation and affect persistent change. For example, Charness and Gneezy (2009) and Acland and Levy (2015) offer incentives for individuals to exercise up to eight times over a one-month period while in Royer et al. (2015) the paticipants are provided with incentives to exercise up to 12 times over the one-month period. In all three cases, the researchers observed increased gym attendance for those who participate in the intervention program, but these effects quickly decay. Our results support this pattern. Our model predicts that individuals who only exercised eight times during the first month have less than a 0.7 percent probability of exercising every week for the next eight weeks. Similarly, individuals who exercise only 12 times over the first month have a 4.3 percent probability of exercising each of the next eight weeks. By contrast, an individual who exercises 20 times during the first month has a 29 percent probability of exercising each of the next eight weeks. While requiring individuals to exercise almost daily would probably lead to higher attrition rates, our results indicate that these types of interventions may generate persistent behavioral change if designed to maximize persistence.

2 Theoretical Background and Other Considerations

2.1 Habit in Utility and Status-quo bias

While the contribution of this paper is entirely empirical, the theoretical underpinnings for our analysis follow from the literature on rational addiction and habit-in-utility (Ryder and Heal, 1973; Becker and Murphy, 1988; Constantinides, 1990; Campbell and Cochrane, 1999).

We present a purely expository model of habit in exercise to motivate our empirical analysis. Similar to the seminal Becker and Murphy (1988) model, suppose the individual has a utility function:

$$U_t = u(d_t, S_t|X_t) \tag{1}$$

where d_t is an indicator variable for whether the individual chooses to exercise in period t, S_t is a measure of the individual's "stock" of habituation, and X_t is a vector of exogenous characteristics. The typical assumption in the literature on rational addiction or other models with habit in utility is that $U_{dS} > 0$. For exposition, we assume that the "stock" evolves according to the following process:

$$S_{t+1} = \delta S_t + f(d_t, S_t) \tag{2}$$

where δ captures the depreciation rate of the stock of habituation. The second term implies that exercising, d_t, has some positive effect on the stock of habituation and that the effect of d_t on S_{t+1} may be a function of S_t.

In this paper, we do not estimate the structural parameters governing the formation or depreciation of habits. Rather, we consider individuals with different values of S_j at the end of the individuals first j weeks of being registered for the manufacturer's application. We estimate the causal relationship between S_j and the probability of exercising in subsequent periods. Prior work from the theoretical literature tells us that individuals with greater stocks of habituation experience more utility (or less disutility) from engaging in a particular activity, and are more likely to continue with that behavior than less-habituated individuals.

For the balance of the paper, we will refer to the number of workouts during the first j weeks as the 'stock of habituation'. While we believe this mechanism to be the most plausible, we acknowledge there could be alternative behavioral mechanisms that can create a causal relationship between the number of workouts in the first j weeks and the probability of exercising in future periods (e.g., tolerance for the physical discomfort of exercising.)[6]

Status-quo bias (i.e., inertia) is usually not modeled as part of the individual's utility function, as it is considered to affect the individual's choices outside of a rational utility maximizing framework. In the context of our model, status quo bias could lead individuals to default into

[6]We rely on our exclusion restrictions and control function approach to address omitted variable bias from factors such as heterogeneous preferences for stationary exercise and unobserved initial habituation.

not exercising, and would therefore enter the mapping between S_t and the probability of exercising, $P(d_t|S_t, X_t, \theta)$, as a large negative constant. In other words, individuals have a strong tendency to stop exercising, and increasing the stock of habituation, S_t, partially offsets that tendency.

2.2 Intrinsic and Extrinsic Motivation

We acknowledge that comparing these results to studies using randomization into incentive programs is imperfect. For example, Deci et al. (1999) and Benabou and Tirole (2003) show that while incentives provided extrinsic motivation to change behaviors, they may partially or completely crowd out intrinsic motivation. Once those incentives are removed, individuals may engage in a targeted behavior less frequently than they would have in the absence of an intervention. These concerns are somewhat mitigated in the case of physical activity, where the central trend of the data is to not exercise.[7]

In our analysis, we use weather as an instrument (discussed further in Section 4). Individuals for whom weather shifts their utility from exercising (relative to leisure) over some margin may have innate differences in extrinsic/intrinsic motivation from individuals for whom incentives provide the necessary nudge. While we motivate our analysis by appealing to results from randomized incentive studies, we acknowledge that weather-induced changes in S_t may differ from incentive-induced changes in S_t.

2.3 Selection Bias and External Validity

For considering factors related to persistence, our data have several positive features which are discussed in the next section. One important feature of our sample is that it is comprised entirely of individuals who have at least shown a small amount of interest in increasing their physical activity by registering for the app. From a policy perspective, this implies that our sample is comprised of individuals who are potential targets for interventions designed to increase physical activity. The self-selected nature of our sample also means that our results do not extrapolate to the full population, only those who have some interest in physical activity.

[7] According to results from CDC's Behavioral Risk Factor Surveillance System, only one-third of adults engage in the recommended amount of physical activity each week.

3 Data

The primary data are provided by a large private manufacturer of fitness equipment. Data were collected through a mobile application which enabled the end-user (exerciser) to link their mobile device to one of the company's cardiovascular fitness machines (e.g. treadmill, recumbent bike, etc.) and track their personal workout results. The original dataset consisted of 36,267 unique users observed over a five year time period (2011 through 2015, and included information on the day and location of each individual's workout). Our analytical samples are derived by excluding those individuals who did not exercise again after the initial j week period and individuals whose locations were unknown or outside the U.S.[8] As a result, our main analytical sample consists of 12,526 individuals, for the four-week initial period.

Table I reports basic summary statistics for the analytical sample where the 'initial period' is defined as four weeks. Within this sample, the average observation period was 26.3 weeks, but one person was observed for 239 weeks. The average person exercised at least once per week for 10.96 weeks.[9] There was considerable variation in the number of weeks for which we observe individuals exercising. In our sample of 12,526 individuals, 3,998 individuals (25.4 percent) exercised for four weeks or fewer, and 24 individuals exercised for at least 100 weeks (Figure 1).

As additional context on the distribution of observed workouts per individual, there were 666 individuals (5.3 percent of the sample) who only exercised twice during the entire sample period (Figure 2), the minimum number of workouts possible based on our sample restrictions. This proportion is fairly large, given that the sample is restricted to those who exercised at least once during their first four weeks and at least once after the initial four week period. For the entire sample, there is a strong tendency to stop exercising. There were an additional 2,984 individuals who exercised between three to seven times over the entire sample period. After seven exercise days, the number of workouts diminishes rapidly. Conditional on exercising at least once in a given week, individuals averaged 2.1 workouts per week.

Our identification strategy relies on weather variables to instrument for individuals' exercise choices during the first four weeks that they are registered for the fitness mobile application.

[8]The restriction on the basis of location is necessary for our identification strategy which relies on the usage of U.S. weather data. Dropping individuals for whom we do not have location information reduces our sample to 16,286. In addition, the approximately 5,000 individuals who do not ever exercise after the first four weeks are automatically dropped from our survival analysis and as such are statistically irrelevant to our estimation sample. A comparison of the two samples is available in Appendix A.

[9]This does not mean that the individual worked out for 10.96 consecutive weeks. Rather of the 24.4 weeks that we observe the average individual, they work out at least once during 10.96 of those weeks (on average).

Table I: Summary Statistics

Panel A: Full Sample				
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Weeks observed	24.36	23.85	5	239
Number of weeks exercising	10.96	12.69	2	206
Number of weeks not exercising	13.40	18.05	0	195
Number of days exercising (conditional on exercising that week)	2.12	1.33	1	7
Average daily maximum temp ^oF	65.05	19.10	8.43	113.71
Average weekly precipitation (in.)	0.62	0.86	0	12.28
Individuals	12,526			
Observations	255,032			

Panel B: First Four Weeks				
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Number of days exercising	5.86	4.30	1	27
Average daily maximum temperature ^oF	63.75	18.84	13.43	107.11
Average weekly precipitation (in.)	0.62	0.53	0.00	5.08

Figure 1: Distribution of number of total weeks exercised per individual

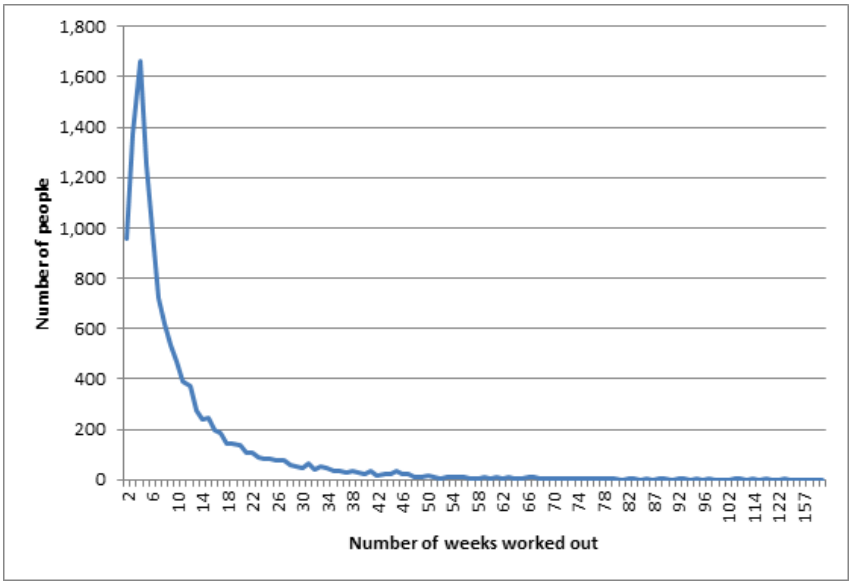
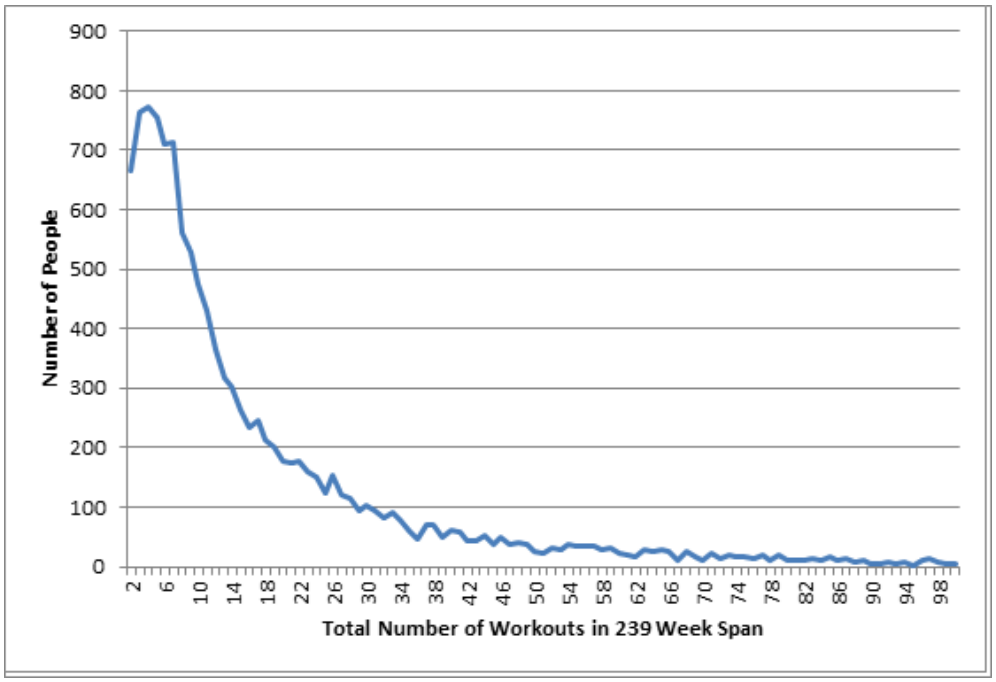


Figure 2: Distribution of number of total days exercised per individual



We incorporate daily weather data from the NOAAs National Climatic Data Center to create weekly measures of temperature and rainfall.

One potential challenge to using weather as an instrument for the number of workouts during the initial period is the fact that the data consist only of workouts recorded through the manufacturers app. For us to observe an individual's workout, they must choose to use stationary exercise equipment, use our manufacturer's brand, and track their workouts on the fitness app. While there is little reason to suspect that weather induces individuals to switch brands of equipment conditional on exercising, or use the app conditional on using our manufacturer's brand, weather could lead individuals to substitute between indoor and outdoor exercise. We do not observe outdoor exercise such as running or cycling. If weather induces individuals to substitute between indoor and outdoor exercise as well as substitute between indoor exercise and leisure, the interpretation of our two-stage estimates becomes more difficult as the instrumental variables assumption of monotonicity is violated.[10]

[10]For example, Eisenberg and Okeke (2009) hypothesize that in colder temperature ranges some people may substitute from outdoor to indoor exercise when the temperature falls. However, they also note that they cannot truly identify within-person substitution between indoor and outdoor exercise because their exercise data is cross-sectional, and the suggestive evidence that does point towards substitution is only apparent among a small subset of their sample.

For additional context, the seminal papers on instrumental variables and local average treatment effects establish a base line case of a binary instrument and treatment variable and categorize the sample as comprised of four groups. The 'always takers' and 'never takers' are individuals that would (or would not) receive the treatment with or without the effect of the instrument. For these individuals, the instrument is irrelevant and the second-stage results do not thusly generalize to these subsamples. The third group are the 'compliers,' or individuals for whom the instrument provides the necessary nudge to receive the treatment. The 'defiers' are individuals for whom the instrument provides a nudge to not receive the treatment. The instrument is said to be monotonic if there are no defiers in the sample, or that the instrument provides a nudge in the same direction for all individuals in the sample. If the instrument is monotonic (as well as relevant and a valid exclusion from the second stage), the second stage result is interpretable as a LATE, which captures the ATE if only for the 'compliers'. If monotonicity does not hold, the second stage is interpretable only as the difference between the average treatment effect of the compliers net of the average treatment of the defiers. The challenge with this interpretation is that the estimated second stage effect may be substantially different from the treatment effect for each group.

However, de Chiasemartin (2017) demonstrates that strict monotonicity does not have to hold for instrumental variables estimates to be interpreted as local average treatment effects. While we refer the reader to that paper for all particulars, we briefly summarize the relevant results here. If a much weaker condition holds (dubbed Compliers-Defiers), the traditional interpretation of IV results is very much intact for a subset of compliers.[11]

Because the dominant effect in our first stage is that rain causes individuals to engage in less stationary exercise, we will refer to this group as the compliers. Results are interpretable as LATEs, even when instruments are non-monotonic under two sufficient conditions: 1.) The second-stage effect must be of the same sign for both compliers and defiers. 2.) The ratio of the LATEs for each group must be smaller than the ratio of the number of members in each group. While neither condition is directly testable, we believe both conditions are met. On the first condition, there is no reason to suspect that exercising more during some initial period would reduce the probability of continuing to exercise, on average, for either group. While one can always make up a story, the overarching associative trend in our data is that higher

[11] Because it is impossible to delineate 'compliers' from 'always takers' and 'never takers' anyway, this is not a big sacrifice.

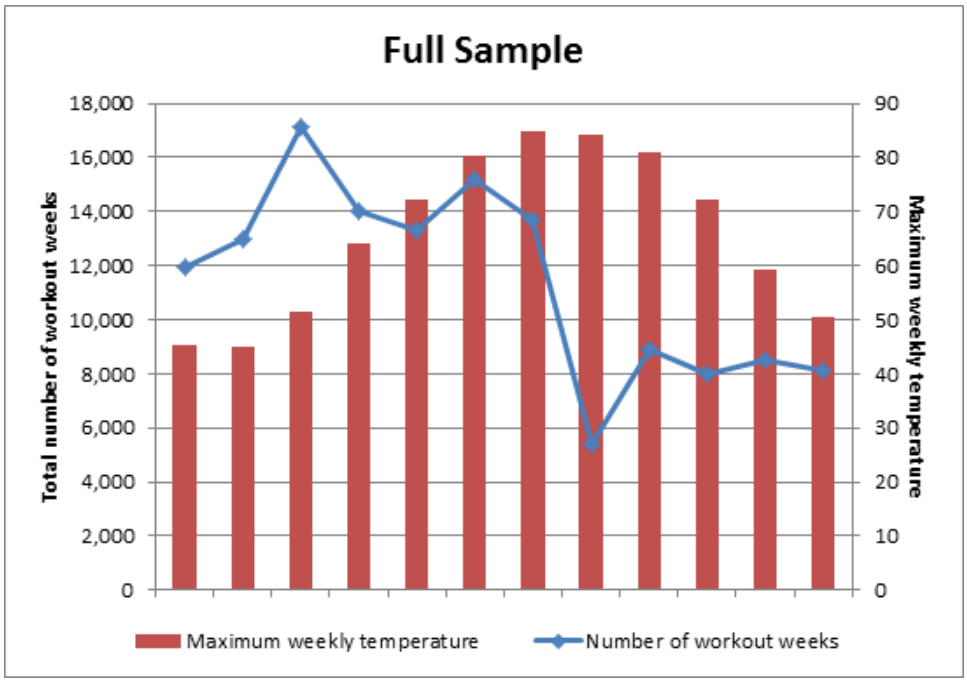
frequency exercise in the initial period is associated with a propensity to continue to exercise in the future. Regarding the second condition, one might characterize the 'defiers' - those who would substitute between indoor and outdoor cardiovascular exercise based on weather as committed exercisers. A brief review of the descriptive statistics indicates that while there are some committed exercises in these data, most users are more likely to be intermittent or non-exercisers. If we informally reframe the ratio of compliers to defiers as the ratio of intermittent or non-exercisers per committed exerciser, the difference in the ATE's for each group would need to be quite large to violate the second condition. We therefore treat the results form our analysis as a LATE for at least a subset of the compliers in the data.

One may also be concerned about substitution between our manufacturer's equipment and other equipment brands not tracked by our app. This is only a concern for individuals who are using these machines in health clubs rather than in their home. However, individuals who take the time to register for the manufacturers app are exhibiting some revealed preference for one manufacturer and therefore are likely to continue to use that manufacturer's equipment at the health club. Finally, while excessively hot or rainy weather may encourage people to exercise inside rather than engage in leisure, it is less clear through what mechanism weather would cause individuals to use stationary exercise equipment from an alternate manufacturer.

Figure 3 shows how the number of workout weeks and the average daily maximum temperature at week t vary throughout the year. In total there were 137,313 workout weeks (i.e. weeks with at least one workout day) observed over the full sample. The most workout weeks occurred in March when the maximum temperature averaged 51.3 degrees Fahrenheit, though a bulk of the workout weeks occurred from March through July. There was also a large spike in workouts in January, as the total number of workout weeks increased from 8,103 in December to 11,968 in January, representing a 47.7 percent increase. The fewest workout weeks occurred in August (4.0 percent), when the maximum temperature averaged 84.1 degrees. Over the full sample, the maximum temperature averaged 65.1 degrees, but varied from 8.4 to 113.7 degrees. The average weekly rainfall was 0.6 inches, and ranged from 0 to 12.3 inches.

In the bottom portion of Table I we report summary statistics for the first four weeks from when an individual started using the stationary exercise equipment. The average individual had an initial exercise stock of 5.9 days, meaning they exercised 5.9 times over the first four weeks. However, this ranged from one to 27 workouts over the initial four week period (no one exercised all 28 days in which they were first observed). Figure 4 shows how the initial exercise stock

Figure 3: Workout weeks and average maximum weekly temperature by month - full sample



varies across the sample of 12,526 individuals. There were 1,609 individuals who only exercised once during the initial four week period, and a total of 113 people who exercised 20 times or more. In total we observe 35,423 workout weeks that occurred within the first four weeks of when an individual first started using the stationary exercise equipment.[12] Similar to with the full sample, the most workout weeks occurred in March (12.6 percent). However, a large portion of the initial workout weeks were positioned in January through April, as 43.5 percent of the sample started exercising in the first four months of the calendar year. Consistent with this finding we also observe a spike in January, as the number of initial workout weeks nearly doubled from December to January, increasing by 78.7 percent. The distribution of initial workouts and maximum temperature in each month is presented in Figure 5.

4 Econometric Methodology

We empirically evaluate the effect of exercise during an initial period on expected future exercise behavior using a Cox proportional hazard model. We examine how the number of days

[12] At maximum, the total number of workout weeks over the first four weeks could be 50,104 (12,526 times 4). However, some individuals did not work out every week over those first four weeks. Specifically, there were 14,681 skipped weeks, leaving us with a total of 35,423 workout weeks.

Figure 4: Distribution of number of workout days over first 4 weeks of registration

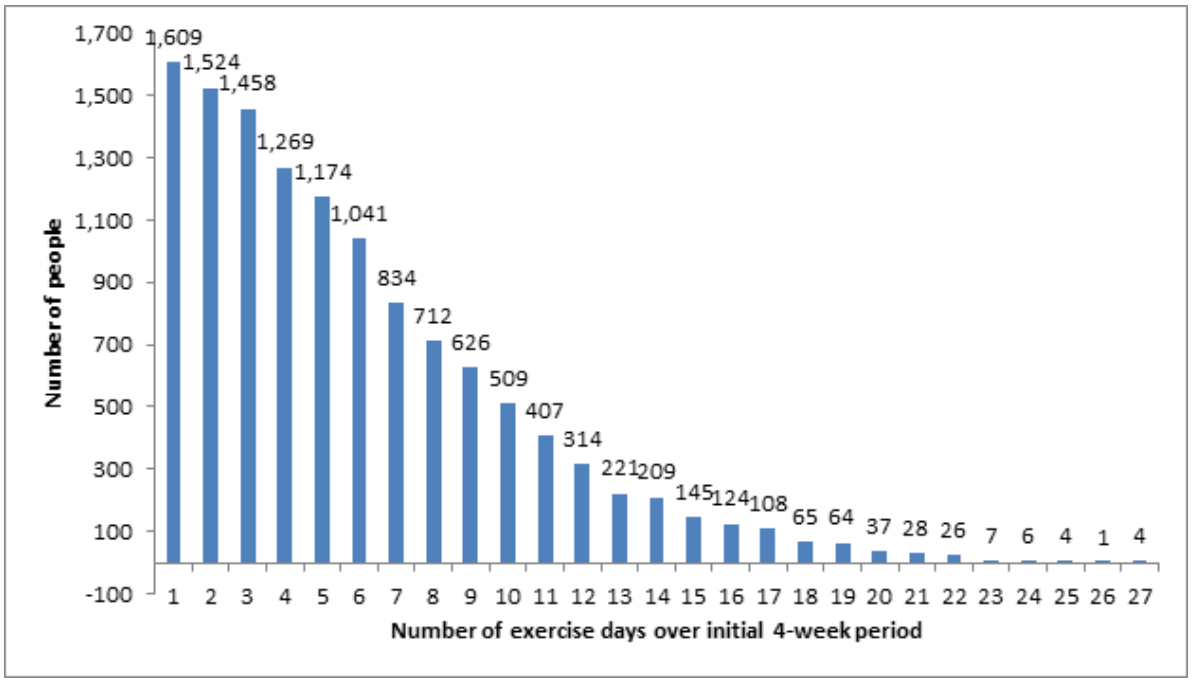


Figure 5: Workout weeks and average maximum weekly temperature by month - first 4 weeks

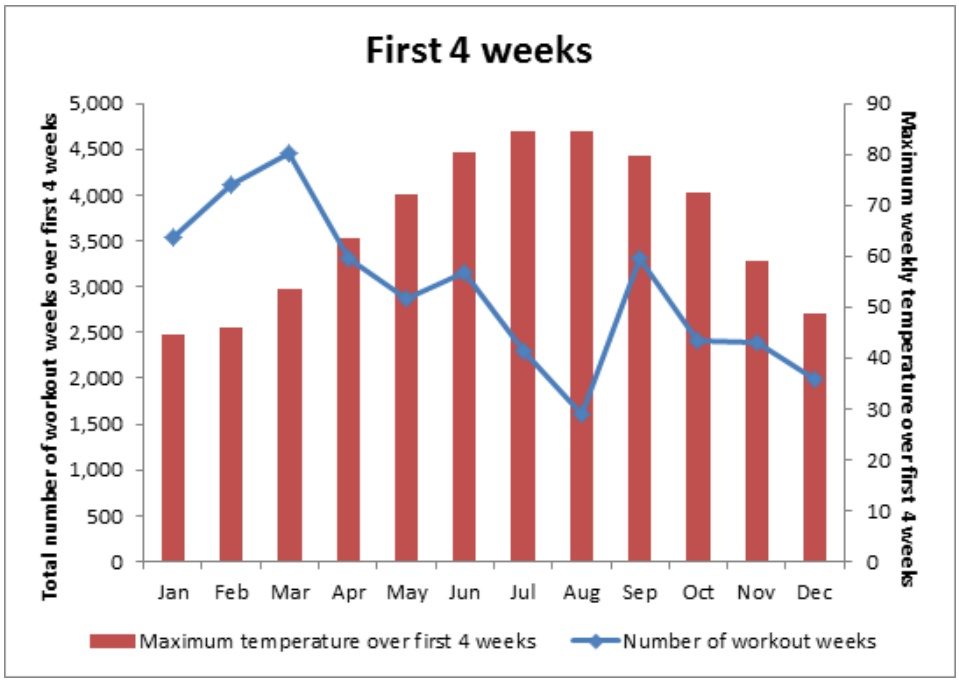


Table II: Consecutive Weeks (not) Exercising and Average Workouts over Next Four Weeks

Consecutive Weeks Worked Out	Average Workouts Next Four Weeks
4	6.79
2	5.40
1	4.32
Consecutive Weeks Missed	Average Workouts Next Four Weeks
1	2.80
2	1.94
4	1.28

that an individual exercises in the first 2, 4, 6, or 8 weeks affects the probability that they continue to exercise for consecutive subsequent weeks. We chose the outcome of 'continuing to exercise in each successive week' or 'exercise at least every other week' because missing a week (or two weeks) marks an inflection point in the individual's expected future frequency of observed workouts. Per Table II, an individual who has not exercised in the last week is expected to workout only half the number of times over the next month as an individual who has exercised at least once in each of the past two weeks.

We identify the causal effect of an individual's exercise behavior over the first K weeks with a control-function approach using key weather variables as exclusion restrictions.[13] Specifically, we use average daily high temperature and average weekly rainfall during the initial period as our exclusion restrictions for the first stage. The control-function approach has been shown to be better suited for controlling for endogeneity in nonlinear models than traditional IV methods (Terza et al., 2008).

We can express the first stage as:

$$\sum_{t=1}^{T_j} N_{it} = \gamma_1 \overline{Z}_{i1} + \gamma_2 \overline{Z}_{i2} + \gamma_3 X_i + \alpha_{1s} + \delta_{1m} + \epsilon_{i1} \tag{3}$$

where i and t index individuals and weeks, respectively. The dependent variable N_{it} captures the number of days that individual i exercised over the first T_j weeks after registering. The $\overline{Z}_{i1}$

[13]The control function approach is similar to the common two stage least square instrumental variables approach (2SLS-IV) in that the first stage equation involves regressing instrumental variables onto the endogenous explanatory variable. However, unlike the typical 2SLS-IV approach where fitted values from the first stage are substituted in for the endogenous explanatory variable in the second stage, the control-function approach uses the residuals (i.e. control functions) from the first stage equation as an additional explanatory variable in the second stage. By adding these control functions, the endogenous explanatory variable becomes exogenous in the second stage structural equation, as the control function in essence controls for the endogeneity (Wooldridge, 2015).

and $\overline{Z}_{i2}$ measure the average weekly rainfall and average daily high temperature over those first four weeks, and X_i is a vector of individual characteristics including age, gender, and weight at the time of initial registration.[14] The theoretical ex ante relationship between weather and use of stationary exercise is ambiguous and depends on context.[15] Additionally, recent work in the economics literature has shown that rainy days affect the way individuals allocate their time between work, home production, and leisure. Connolly (2008) finds that on rainy days, men work an additional average of 30 minutes and salaried workers watch an additional 16 minutes of television. Zivin and Neidell (2014) also find that temperature changes affect the time individuals spend at work. To the extent that weather induces individuals to allocate their time differently, weather will affect the amount of time available for stationary exercise.

Recall that for us to observe an individual exercising, the individual must make four choices jointly, each conditional on the previous: to exercise, to exercise on stationary equipment, to exercise on this particular brand, and to track their workout. The key question for our first stage results is if weather affects the number of times we observe an individual working out, what does that variation really mean? We argue that there is no reason that conditional on deciding to use stationary exercise equipment, weather should affect the individual's choice of brand or whether to track the app. Therefore weather affects only the individuals' decision to exercise and/or their decision to use stationary exercise equipment versus exercising outside. While both effects are likely present, our first stage results imply that precipitation *lowers* the number of workouts during an initial period. If the substitution story dominated (i.e., inducing individuals to exercise outside vs. inside), we would expect precipitation to increase the number of workouts. We therefore posit that bad weather mainly affects the number of observed workouts by encouraging individuals to substitute between stationary exercise and no exercise. However, to the extent that precipitation induces substitution towards using stationary equipment among some subset of users, we refer to the previous discussion of the Compliers-Defiers condition under which our estimates are still interpretable as local average treatment effects for the individuals whom adverse weather provides a disincentive to exercise.

Additionally, one may be concerned that the relationship between weather and exercise is

[14] Unfortunately, over 99 percent of individuals never updated their weight in the app, thereby precluding the possibility of incorporating weight as a state variable in a dynamic model.

[15] A review of studies from the public health and epidemiology literature finds that rainfall is associated with decreased physical activity of many types (Chan and Ryan, 2009). Tucker and Gilliland (2007), review additional literature on the effects of weather on physical activity. While they find that weather generally does affect physical activity, the magnitude of the effect depends on context.

driven by correlations between regional preferences for exercise and weather (i.e., do people in Arizona just exercise more than Idahoans?). Also, we may be concerned about seasonal fluctuations in patterns of physical activity (i.e., are individuals who register in June systematically different from individuals who register in February?). To address both sets of concerns, we include fixed effects for the state in which the individual lives and the month in which they register (α_{1s} and δ_{1m}).[16]

We express the second-stage survival function as:

$$h_{it} = \beta_0 \times exp[\beta_1 \sum_{t=1}^{T_j} N_{it} + \beta_2 X_i + \beta_3 \widetilde{\epsilon}_{i1} + \alpha_{2s} + \delta_{2m}] \quad j = 2, 4, 6, 8 \tag{4}$$

where h is the hazard function for individual i and measures the probability that an individual will continue exercising in week t.[17] As is customary in a control function approach, equation 4 contains the term $\widetilde{\epsilon}_{i1}$, the predicted residuals from our first stage regression.

5 Results

In this section, we present the result from our empirical analysis. In Tables III-VII, all specifications include demographic control variables and fixed effects for each state and month. Table III presents descriptive results from a dynamic logit to motivate our choice of a hazard model. Columns (2) and (3) show that working out in the previous week is a strong predictor of working out in the subsequent week. Individuals who exercise in a given week ($d_t = 1$) are over 4 times more likely to exercise during the next week as compared to an individual who does not exercise that week ($d_t = 0$).

Table IV contains the results from our first-stage control function. When the initial period is defined as two, four, or six weeks, both precipitation and average daily high temperature are at least significant at the 10% level. For all defined lengths of the initial period length, precipitation reduces the expected number of workouts over that span.[18] The signs on our

[16] We have verified that our results are robust to the inclusion of MSA fixed effects. Although our second stage effects are virtually unchanged, our first stage estimates are not as strong, making the state FE our preferred specification. Results are available in Appendix B

[17] We consider two definitions of 'failure' in our survival analysis: not working out in a given week and not working out for two consecutive weeks. Results are qualitatively similar under either definition of failure

[18] Control function approaches usually require a linear first stage. We therefore use a linear first stage even though we are working with, essentially, count data. We have verified that our estimates do not change in either sign nor joint significance when we estimate the first stage using a Poisson model. When we test the joint significance of our two weather instruments for the four-week period, we get $\chi^2_2 = 52.08$. This is well above the critical values for

Table III: Logit analysis: probability of working out.

	(1)	(2)	(3)
Worked Out Last Week		1.960***	
		(0.00)	
Lagged Number of Workouts			0.979***
			(0.004)
Age	0.006***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)
Sex	-0.125***	-0.010***	-0.079***
	(0.010)	(0.011)	(0.012)
Weight	-0.001***	-0.001***	-0.001***
	(0.0001)	(0.0001)	(0.0001)
N	318060	301667	301667

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include state and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

exclusion restrictions also mitigate some concerns about the mechanisms through which our instruments affect behavior. For example, if rain had a positive effect on the number of workouts, one might posit that the rain encouraged exercisers to discover they preferred stationary exercise rather than outdoor exercise. In this case, the second stage results would be more interpretable as learning or developing a preference for a specific sort of physical activity. However, we find the opposite sign, which implies that if rain leads individuals to exercise less on their indoor machines, they are more likely to engage in sedentary leisure or supply additional hours of labor (Connolly, 2008). Because the dominant effect in our first stage appears to be between stationary exercise and no exercise, we interpret our second stage results in the context of habituation to any sort of exercise. The F-statistics of the joint significance of our variables excluded from the second stage in Table IV indicate that our instruments have requisite strength. For the two, four, and six week initial periods, the F-statistic is greater than the rule of thumb of 10 established by Staiger and Stock (1997). Even for the eight week group, an F-statistic of 6.45 is above the critical value for less than 15% bias as detailed in Stock and Yogo (2005) when estimated with LIML.

Before presenting the results from the control-function survival analysis model, we first present results from a linear 2SLS model. In this model, the first-stage is defined as in equation

instrument strength using chi-squared tests as described in Stock and Yogo (2005).

Table IV: First-stage Control Function Results

	(1)	(2)	(3)	(4)
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Precipitation	-0.103***	-0.354***	-0.764***	-0.918***
	(0.027)	(0.085)	(0.163)	(0.257)
Average Daily High	0.007***	0.012*	0.018*	0.003
	(0.002)	(0.005)	(0.009)	(0.014)
Age	0.006***	0.021***	0.058***	0.039***
	(0.002)	(0.003)	(0.008)	(0.006)
Sex	0.012	-0.083	-0.182	-0.210
	(0.049)	(0.113)	(0.255)	(0.183)
Weight1	0.002***	0.004***	0.004	0.004**
	(0.0004)	(0.001)	(0.002)	(0.002)
F-statistic (exclusions)	12.13	10.47	11.95	6.45
N	16286	12526	10503	9196

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include state and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3, but the second stage is the total number of workouts during the three (or six) months following the initial period (still defined as 2, 4, 6, or 8 weeks). These results complement the survival results in two ways. First, this model enables us to evaluate whether exercising with greater frequency during the initial period leads not just to persistent regular activity, but to more exercise in general. Second, tests of instrument validity in overidentified models are developed for linear 2SLS models, but not for control function duration models.

Table V: Linear 2SLS Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Initial Period Length (weeks)	Two	Two	Four	Four	Six	Six	Eight	Eight
Post-initial Length (months)	Three	Six	Three	Six	Three	Six	Three	Six
Stock	2.882***	5.573***	1.333***	2.619***	0.910***	1.624***	1.055***	1.624***
	(0.936)	(0.085)	(0.407)	(0.701)	(0.255)	(0.439)	(0.274)	(0.477)
Age	0.075***	0.122***	0.072***	0.114***	0.067***	0.111***	0.036**	0.076**
	(0.010)	(0.016)	(0.012)	(0.020)	(0.013)	(0.023)	(0.018)	(0.031)
Sex	-0.616***	-1.141***	-0.448*	-0.821*	-0.308	-0.616	-0.463	-0.954*
	(0.239)	(0.417)	(0.250)	(0.183)	(0.264)	(0.469)	(0.284)	(0.504)
Weight	-0.006**	-0.014***	-0.008***	-0.016***	-0.007 ***	-0.012	-0.006 **	-0.010**
	(0.0003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.004)
F-statistic (exclusions)	11.63	11.626	12.302	12.302	12.334	12.334	6.835	6.835
Hansen J-Test P-Value	0.077	0.846	0.974	0.177	0.850	0.101	0.373	0.172
N	16286	16286	12526	12526	10503	10503	9196	9196

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include state and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results from the Linear 2SLS model are displayed in Table V. In each column, the results correspond to a model where the initial period is defined as either two, four, six, or eight weeks, and the dependent variable is the total number of workouts over the ensuing three or six months. The general finding is that more workouts during an initial period has a positive effect on the number of workouts going forward. Additionally, except for the results in column (1), the null that our instruments are valid is not rejected in the overidentification test.

Table VI contains results from our survival analysis in which failure is defined as not working out for one week. The top panel contains baseline results (with no instruments) and the bottom panel contains results from our control-function approach. Regardless of whether or not we "instrument" for stock, (the number of days the individual works out in the initial period) more workouts in the first month leads to a lower probability of failure, or greater persistence. The estimated coefficient on stock is larger when we instrument for the number of workouts in the first month. One plausible explanation is that our control function approach corrects for an initial conditions problem. We do not observe individual's exercise behavior before they register for the app. Consider two types of individuals, those with a pre-established routine and those who are starting a new routine. The individuals with a pre-established workout regimen may exercise a few times a week after registering (i.e, 1-2 time per week) - but be less likely to skip a week. By contrast, individuals who are starting a new regimen may be more likely to exercise frequently at first, but may exhibit more intermittent exercise patterns after the initial period.

The results from the control-function specification indicate the correction for endogeneity is necessary. In all versions of the model, the first-stage residuals are statistically significant at the 1 percent level. One advantage of using the control-function approach is that a test on the first-stage residuals can be used as a Hausman test for exogeneity, robust to heteroskedasticity, on the endogenous explanatory variable (Hausman, 1978; Wooldridge, 2015). Finding that the first-stage residuals are statistically significant provides evidence against the null hypothesis that the number of exercise days is conditionally strictly exogenous.

Table VII contains results from our survival analysis in which failure occurs if the individual does not work out for two consecutive weeks. Qualitatively, the results are similar to the results in Table VI. Table VII demonstrates the lack of sensitivity of these results to the definition of failure. Working out more with greater frequency during the initial period predicts greater persistence going forward.

Table VI: Survival Analysis Results: Failure ≡ Missing One Week

Panel A - Baseline Results, No Control Function				
	(1)	(2)	(3)	(4)
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.099***	-0.060***	-0.044***	-0.035***
	(0.002)	(0.005)	(0.00)	(0.00)
Age	-0.003***	-0.002***	-0.001*	-0.002**
	(0.00)	(0.00)	(0.00)	(0.00)
Sex	0.070***	0.071***	0.073***	0.070***
	(0.02)	(0.01)	(0.01)	(0.01)
Weight	0.001***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
N	285274	255032	210810	231079

Panel B - Control Function Results				
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.214***	-0.114***	-0.085***	-0.072***
	(0.012)	(0.007)	(0.005)	(0.004)
First-Stage Residual	0.172***	0.095***	0.075***	0.068***
	(0.018)	(0.011)	(0.008)	(0.006)
Age	-0.002**	0.000	0.003***	0.0015*
	(0.001)	(0.001)	(0.001)	(0.0006)
Sex	0.078***	0.068***	0.0627***	0.058***
	(0.0150)	(0.014)	(0.014)	(0.014)
Weight	0.001***	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.000)
N	285274	255032	210810	231079

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include state and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VII: Survival Analysis Results: Failure ≡ Missing Two Weeks

Panel A - Baseline Results, No Control Function				
	(1)	(2)	(3)	(4)
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.118***	-0.078***	-0.061***	-0.049***
	(0.006)	(0.003)	(0.002)	(0.002)
Age	-0.003***	-0.002*	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Sex	0.080***	0.080***	0.079***	0.077***
	(0.022)	(0.021)	(0.020)	(0.020)
Weight	0.001***	0.001***	0.001***	0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
N	285274	255032	210810	231079

Panel B - Control Function Results				
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.263***	-0.150***	-0.121***	-0.102***
	(0.019)	(0.010)	(0.007)	(0.006)
First-Stage Residual	0.216***	0.127***	0.110***	0.098***
	(0.027)	(0.016)	(0.011)	(0.009)
Age	-0.001	0.001	0.005***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Sex	0.090***	0.076***	0.065**	0.061**
	(0.022)	(0.021)	(0.020)	(0.020)
Weight	0.001***	0.0014***	0.0011***	0.0013***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
N	285274	255032	210810	231079

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include state and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.1 Survival Function Post-estimation Analysis

We use the estimated parameters from our control-function survival analysis to evaluate how increases in the number of workouts over a given initial period, or increases in the length of the initial period, affect the probability that individuals continue to exercise on at least a weekly basis going forward. Figures 6-9 plot the probability that individuals do not go a week without exercising after an initial period. To facilitate comparisons across durations of initial period, we characterize the frequency of exercise as the average number of days per week over an initial period.[19]

Figures 6-9 illustrate that both the length of the initial period and the frequency of activity over that initial period affect the probability that individuals continue to exercise at least weekly after the initial period. Comparing the lines within each graph informs about the effect of more frequent activity during an initial period. In Figure 7, for example, individuals who are observed exercising three times per week over the first four weeks have a 4.1% predicted probability of exercising in each of the next eight weeks without missing a week. Our analysis predicts that individuals who are observed exercising four (five) times per week over the first four weeks have a 13.3 (27.8) percent probability of exercising each of the next eight weeks. This ratio generally holds at the eight week mark, regardless of the length of the initial period. Individuals who exercise four (five) times per week during the initial period are approximately three (six) times more likely to exercise each of the next eight weeks than individuals who exercise three times per week.[20]

Comparing probabilities between Figures 6-9 illustrates how the length of the initial period affects the probability that individuals continue to exercise, holding the number of workouts per per week during the initial period constant. Consider two lengths of initial period, four weeks and eight weeks. Individuals who exercise for an average of four times per week over an initial period of eight weeks are 67 percent more likely to continue to exercise in each of the eight successive weeks than when the initial period is defined as four weeks. The benefits of a longer initial period are actually largest when individuals exercise less frequently. Further, while there are undoubtedly diminishing returns of the length of the initial period to survival

[19] For example, in Figure 7, the line denoted "3/week" corresponds to an individual working out 12 times over the four week initial period.

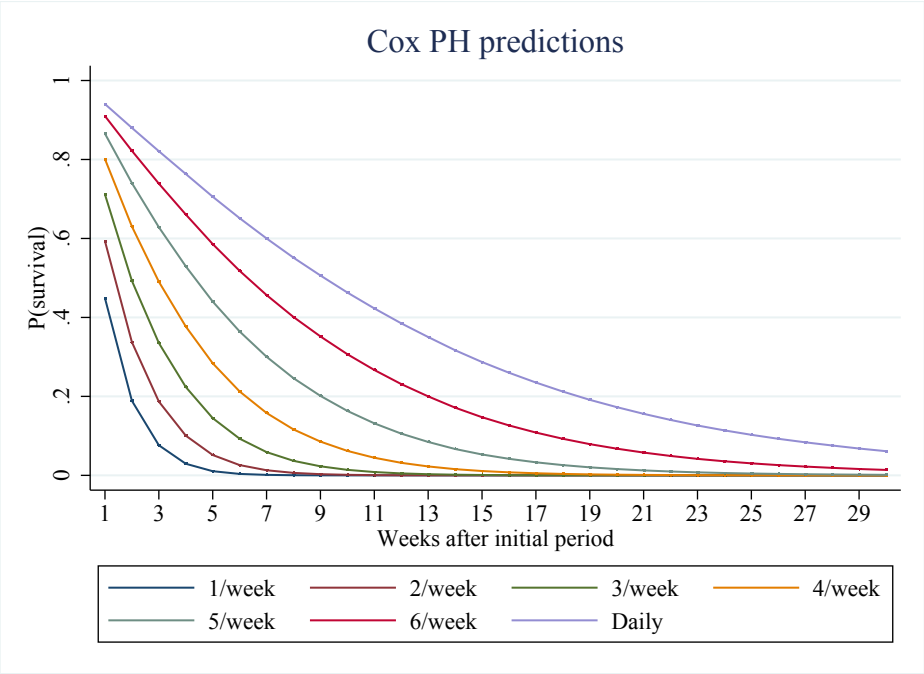
[20] As time increases from the initial period, survival probabilities decrease, and these ratios increase. The eight week mark falls approximately two months after the initial period and therefore makes for a convenient point of comparison, rather than having a special intrinsic meaning.

probabilities, there are increasing returns when the initial period is defined in the two-to-eight week range. Consider an individual who exercises 3-5 times per week during the initial period. Holding the number of workouts per week during the initial period fixed, increasing the initial period from two to four weeks increases the probability an individual is still exercising on a weekly basis eight weeks after the initial period by 14 percent. Increasing the initial period from four to six weeks increases the eight-week post initial period survival probability by an additional 25 percent. Expanding the initial period from six to eight weeks increases the same survival probability by 32 percent. While the marginal effects of increased frequency of activity during a given length of initial period are larger, the length of the initial period also predicts greater persistence in observed exercise.

Our results are qualitatively similar when we define failure as not exercising for two consecutive weeks. Individuals who exercise only twice a week for the first month have a 91.3 percent probability of missing two consecutive weeks within two months (eight weeks) after the initial period. Individuals who exercise three times per week still have a 79 percent chance of missing two consecutive weeks within the same frame. By contrast, individuals who exercise five times per week for the first month will miss two consecutive weeks during the next two months with a probability less than one-third.

These results replicate the lack of persistence found in field experiments by Charness and Gneezy (2009), Acland and Levy (2015), anc Royer et al. (2015). Individuals who exercised 8 times in the first four weeks (i.e., averaging two workouts per week, as Charness and Gneezy (2009) incentivized), have less than a five percent chance of exercising each week over the following month, and less than a one percent chance of exercising in each of the next eight weeks. Similarly, individuals who exercise 3 times per week during the first four weeks have a 4.15 percent probability of exercising in each of the ensuing eight weeks. However, our results indicate that interventions with longer durations or greater frequency of required activity within that period may have more persistent effects. For example, individuals who exercise four (five) times per week over the first four weeks, have a 13.3 (27.8) percent chance of exercising every week for the next eight weeks. Second, the dominant tendency in the data is for individuals to miss weeks of exercising and eventually attrit. Whether this is due to the status-quo bias of sedentarity, or individuals experimenting with the equipment or app and learning about preferences is unknowable from our data. While we show that additional workouts during an initial period has a causal effect on the probability that individuals continue to exercise on a

Figure 6: Survival Function, Two week initial period, Failure ≡ missing one week



weekly basis, the prevailing trend in the data is nevertheless that an individual's frequency of observed workouts decreases over time.

5.2 Habit Formation and Maintenance

We can also use the model to provide insight about the depreciation rate of the stock of habituation, or alternatively, how much continued investment in the stock of habituation is needed to maintain a predicted level of persistence? While we cannot directly measure 'habit', we can observe survival probabilities, which are a function of the stock of habituation.[21] So while we cannot observe the stock of habituation depreciating, we can consider the following: Suppose an individual is observed exercising four times per week over an initial four week period and therefore has a 13.3 percent probability of exercising each of the next eight weeks. How many times would he need to exercise over the next two weeks to maintain a stock of habituation to face the same survival probability after a six week initial period? Formally, we are interested in the value of w that solves:

[21] Most of the empirical literature on habit or rational addiction assumes a functional form on habit formation and the effect of habit on utility, and then and estimates the stock of habituation from persistence. See, for example, Darden (2017).

Figure 7: Survival Function, Four week initial period, Failure ≡ missing one week

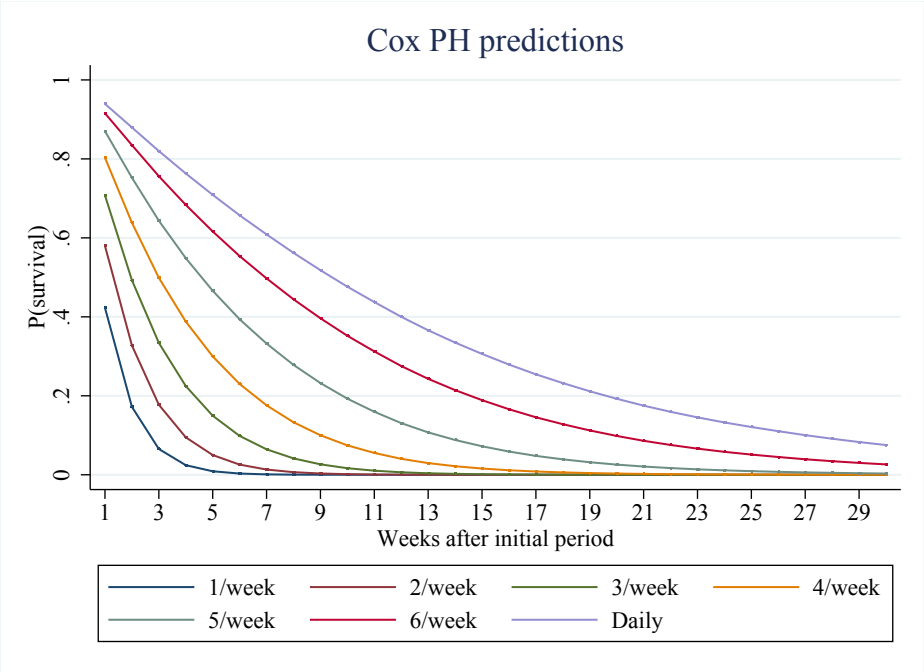


Figure 8: Survival Function, Six week initial period, Failure ≡ missing one week

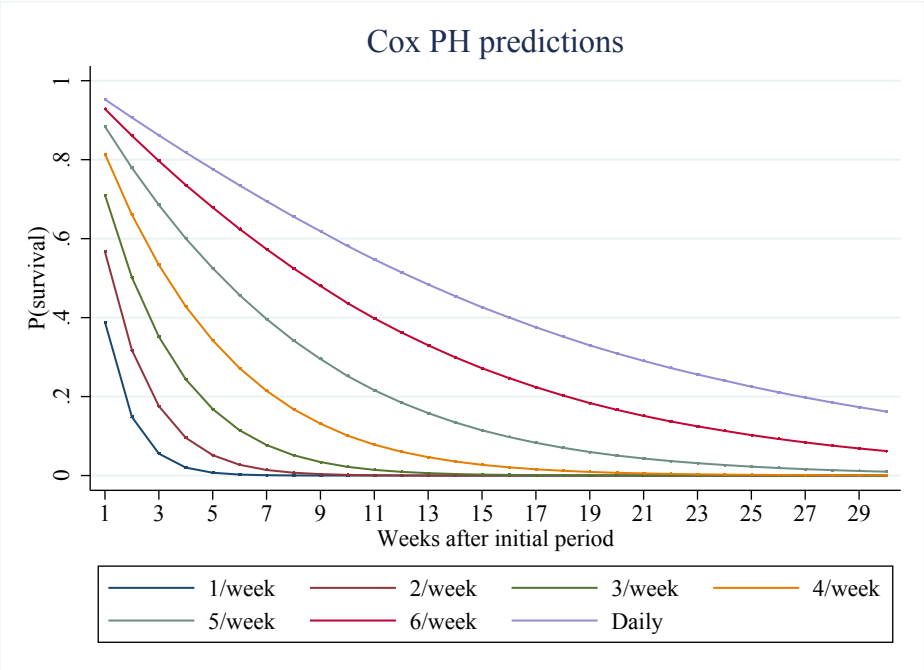
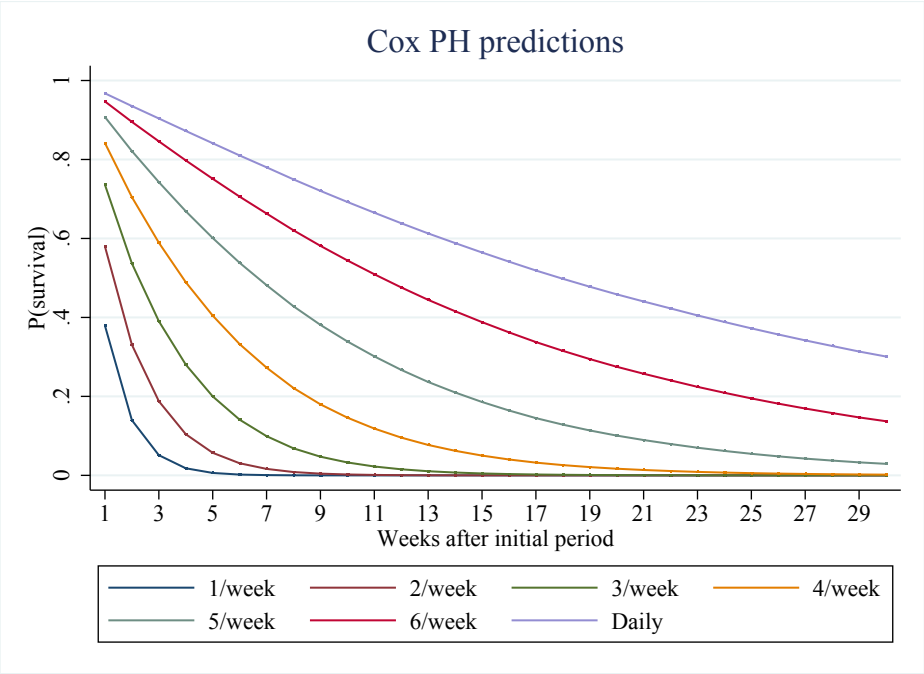


Figure 9: Survival Function, Eight week initial period, Failure ≡ missing one week



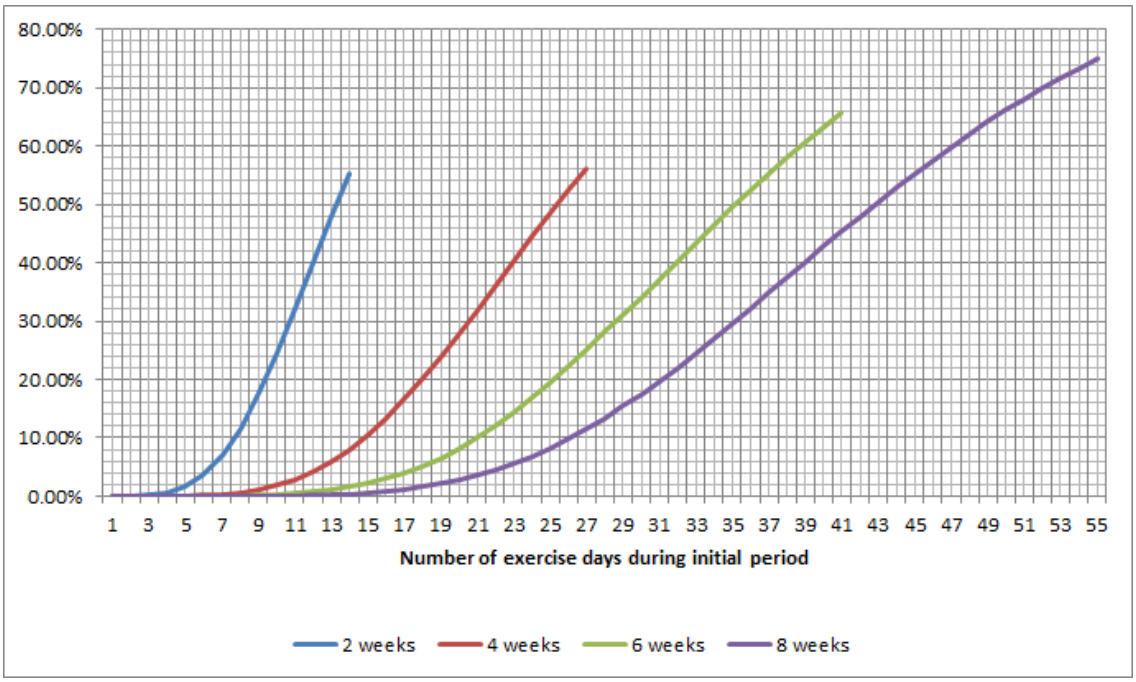
$$\widehat{P}(S_4(4), \mathbf{X}) = \widehat{P}(S_6(w), \mathbf{X}) \tag{5}$$

where $S_j(K)$ indicates the stock of habituation as a result of exercising an average of K times per week over an initial period of j weeks. Figures 10 and 11 show the relationship between the number of exercise days and survival probability after 8 and 12 weeks respectively, for each defined length of initial period.[22] In Figure 10 for example, tracing a horizontal line at 13.3 percent will yield the number of workouts over each initial period to generate the same predicted level of persistence at week 8. In general, as the length of the initial period increases, the number of workouts per week needed to maintain the same survival probability decreases. For example, while an individual must be observed working out 16 times over the first four weeks to generate a survival probability of 13.3 percent after eight weeks, he only needs to exercise 3.3 times in per week in weeks five and six (and 2.7 times per week in weeks 7 and 8) to maintain the same hazard rate.[23] As further evidence of habit formation, longer initial periods imply lower relative levels of investment in habituation to maintain a given stock of habituation. To maintain a 27

[22]The tables that generate these graphs will be available in an online appendix.

[23]This pattern creates a total of 22.6 workouts during a six week initial period and 28 workouts during an eight week initial period.

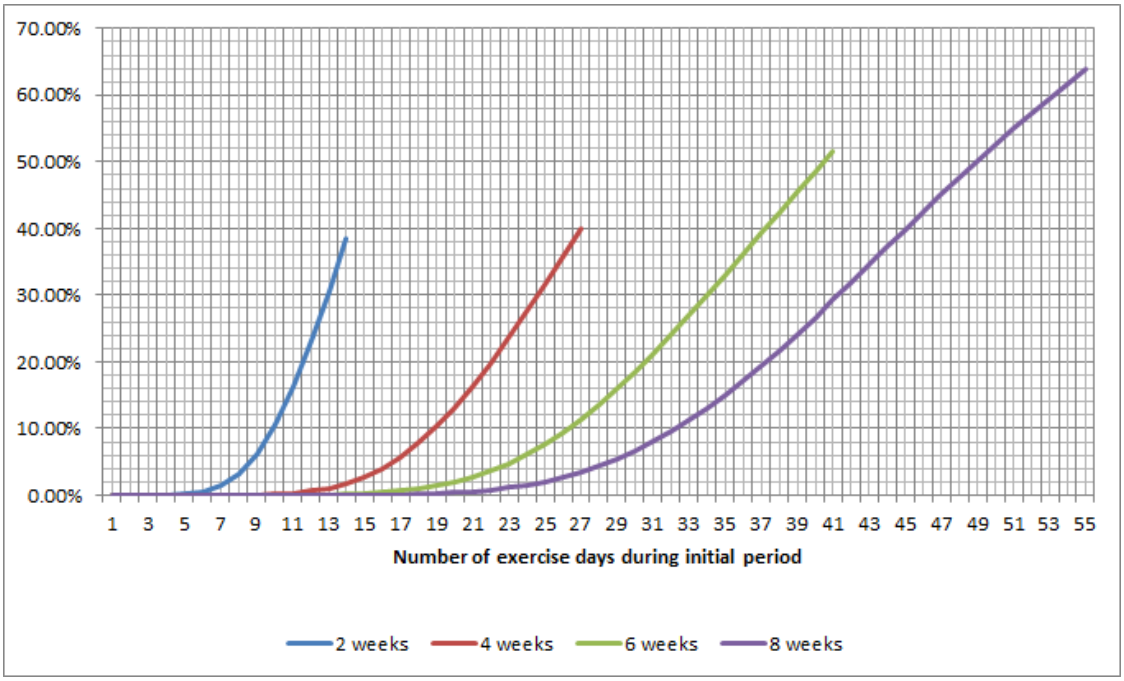
Figure 10: Persistence and Number of workouts, Eight weeks after initial period



percent survival probability eight weeks after some initial period, individuals must be observed working out 5.25 times per week during the first two weeks, but only 3.1 times per week during weeks seven and eight. Longer initial periods therefore not only increase survival probabilities for a constant frequency of activity, but also reduce the amount of incremental activity needed to maintain a given stock of habituation.

These results suggest some evidence of habit formation, but any stock of habituation decreases rapidly if individuals stop exercising for a spell. In a world without habit formation, we would expect that individuals who exercise four times per week for any initial period would generate the same subsequent survival probabilities, regardless of the length of initial period. If, on the other hand, there was no depreciation in the stock of habituation, we should see that an individual who exercises 16 time during the first four weeks would generate the same week 8 survival probability as an individual engaging in 16 workouts over the first eight weeks. Overall, our results indicate strong depreciation in the stock of habituation. Specifically, if an individual has averaged k workouts per week over the first j weeks, he must average approximately $0.8 * j$ workouts in weeks $j+1$ and $j+2$ to maintain the same survival probability.

Figure 11: Persistence and Number of workouts, Twelve weeks after initial period



6 Conclusion

This paper examines the relationship between an individual's observed exercise frequency during an initial period and the individual's probability of continuing to exercise in subsequent weeks. Using weather as an instrumental variable enables us to identify the causal effect of additional workouts in the first approximate month on the expected persistence of behavior. We find that both higher frequency of activity during an initial period and a longer initial period duration have positive causal effects on the probability of continuing to exercise in consecutive subsequent weeks.

Our results provide additional context for previous findings from the field experimental literature on incentivizing physical activity, where behavioral changes induced by incentives are not found to persist. Both Charness and Gneezy (2009) and Royer et al. (2015) incentivized individuals to exercise eight to twelve times per month during the intervention period. Our control-function estimates indicate that individuals who exercise eight (twelve) times in the initial month have a 99 (95) percent chance of missing whole weeks of exercise in the next two months. Individuals who workout 16 (20) times in the first month are 3.2 (6.7) times as likely to not miss a week than those who exercised 12 times. The main implication of these result is

that interventions, perhaps incentive-based interventions, can have persistent effects if they are designed such that individuals are likely to have formed habits by the end of the intervention. Work in the experimental literature typically uses relatively short intervention periods. In these cases, our results indicate that interventions where individuals are incentivized to exercise 4-5 times per week are much more likely to lead to persistent behavioral changes than the designs in prior work. These results suggest that intervention programs aimed at establishing a long-term exercise habit should include incentives for the participants to exercise more often, perhaps even every day, during the intervention period. However, we acknowledge that such a high-frequency threshold for participation will likely increase non-completion and non-participation. A more intense intervention may also require much stronger incentives to get individuals to participate or complete the program. Whether this trade-off is in fact cost-effective is an area for future work.

One drawback of a control-function or IV approach is that our estimates only capture a local average treatment effect (LATE). Specifically, our estimated effect of past exercise on expected future exercise is only valid for those whose propensity to use stationary exercise equipment is affected by weather. These results do not apply to those who have sufficiently strong preferences for exercise, be they positive or negative preferences. Also, from the literature on intrinsic/extrinsic motivation, weather-induced behavioral changes may lead to fundamentally different persistence patterns than incentive-induced behavioral changes.

References

Acland, D. and M. R. Levy (2015). Naivete, projection bias, and habit formation in gym attendance. *Management Science 61(1)*, 146–160.

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

Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70(2)*, 489–520.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107(2)*, 205–251.

Cawley, J. and J. A. Price (2013). A case study of a workplace wellness program that offers financial incentives for weight loss. *Journal of Health Economics 32(5)*, 794–803.

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

Chan, C. A. and D. A. Ryan (2009). Assessing the effects of weather conditions on physical activity participation using objective measures. *International Journal of Environmental Research and Public Health 6(10)*, 2639–2654.

Charness, G. and U. Gneezy (2009). Incentives to exercise. *Econometrica 77(3)*, 909–931.

Connolly, M. (2008). Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics 26(1)*, 73–100.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium. *Journal of Political Economy 98(3)*, 519–543.

Darden, M. E. (2017). Smoking, expectations, and health: A dynamic stochastic model of lifetime smoking behavior. *Journal of Political Economy, Forthcoming*.

de Chiasemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics 8(2)*, 367–396.

Deci, E., R. Koestner, and R. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin 125(6)*, 627.

DellaVigna, S. and U. Malmendier (2014). Paying not to go to the gym. *American Economic Review 96(3)*, 694–719.

Dube, J.-P., G. J. Hirsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics 41(3)*, 417–445.

Eisenberg, D. and E. Okeke (2009). Too cold for a jog? weather, exercise, and socioeconomic status. *B. E. Journal of Economic Analysis and Policy 9(1)*.

Ferson, W. E. and G. M. Constantinides (1991). Habit persistence and durability in aggregate consumption. *Journal of Financial Economics 29*, 199–240.

Flegal, K. M., M. D. Carroll, C. L. Ogden, and L. R. Curtin (2010). Prevalence and trends in obesity among us adults, 1999-2008. *The Journal of the American Medical Association 303(3)*, 235–241.

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

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives 25(4)*, 191–209.

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

Kahneman, D., J. L. Knetsch, and R. Thaler (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives 5(1)*, 193–206.

Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics 15(2)*, 310–327.

Klemperer, P. (1987). The competitiveness of markets with switching costs. *RAND Journal of Economics 18(1)*, 138–150.

Ogden, C. L., M. Carroll, B. Kit, and K. M. Flegal (2014). Prevalence of childhood and aduly obesity in the united states, 2011-2012. *The Journal of the American Medical Association 311(8)*, 806–814.

Olekalns, N. and P. Bardsley (1996). Rational addiction to caffeine: An analysis of coffee consumption. *Journal of Political Economy 104(5)*, 1100–1104.

Royer, H., M. F. Stehr, and J. R. Sydnor (2015). Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics 7(3)*, 51–84.

Ryder, H. E. and G. M. Heal (1973). Optimal growth with intertemporally dependent preferences. *The Review of Economic Studies 40(1)*, 1–31.

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty 1(1)*, 7–59.

Spence, M. A. (1981). The learning curve and competition. *Bell Journal of Economics 12(1)*, 49–70.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65(3)*, 557–586.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, New York, NY. Cambridge University Press.

Terza, J. V., A. Basu, and P. J. Rathouz (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics 27(3)*, 531–543.

Thaler, R. H. and C. R. Sunstein (2003). Libertarian paternalism. *The American Economic Review, Papers and Proceedings 93(2)*, 175–179.

Tucker, P. and J. Gilliland (2007). The effect of season and weather on physical activity: A systematic review. *Public Health 121*, 909–922.

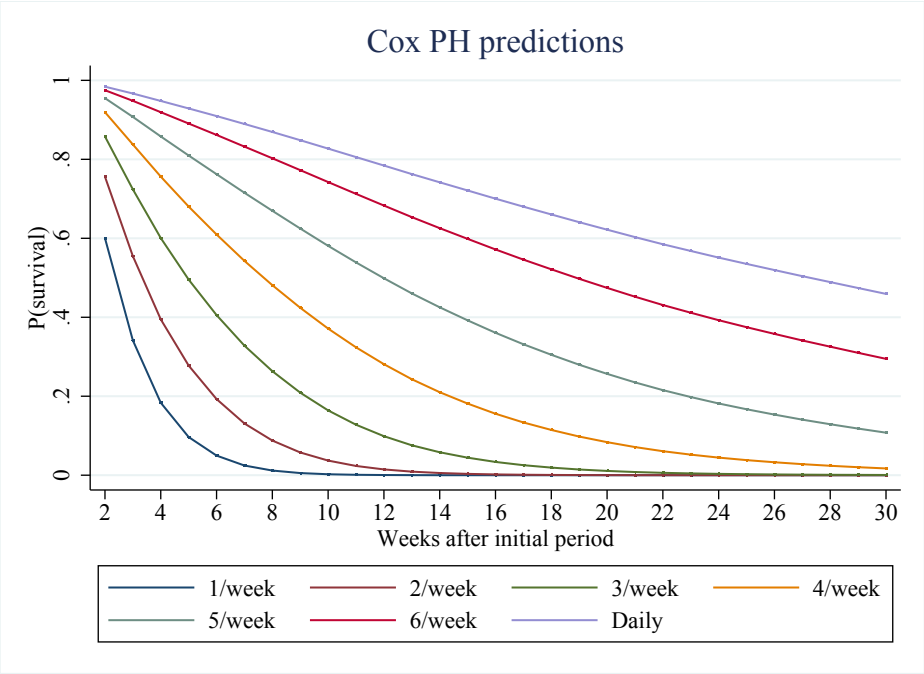
vonWeizsacker, C. C. (1984). The costs of substitution. *Econometrica 52*, 1085–1116.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources 50(2)*, 420–445.

Zivin, J. G. and M. Neidell (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics 32(1)*, 1–26.

A Supplementary Graphs - Two Week Survival Plots

Figure 12: Predicted Survival Functions: Failure ≡ missing two consecutive weeks



B Robustness: Results with MSA Fixed effects

Table VIII: MSA Survival Analysis Results: Failure ≡ Missing One Week

Panel A - Baseline Results, No Control Function				
	(1)	(2)	(3)	(4)
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.098***	-0.059***	-0.044***	-0.035***
	(0.004)	(0.002)	(0.001)	(0.001)
Age	-0.00248***	-0.00175***	-0.00143***	-0.00105*
	(0.00054)	(0.00054)	(0.00054)	(0.00055)
Sex	0.07073***	0.07132***	0.07130***	0.07617***
	(0.01520)	(0.01467)	(0.01455)	(0.01461)
Weight	0.001***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
N	285274	255032	210810	231079

Panel B - Control Function Results				
Initial Period	Two Weeks	Four Weeks	Six Weeks	Eight Weeks
Stock	-0.217***	-0.113***	-0.085***	-0.071***
	(0.012)	(0.007)	(0.005)	(0.004)
First-Stage Residual	0.177***	0.095***	0.076***	0.067***
	(0.018)	(0.011)	(0.008)	(0.006)
Age	-0.001**	0.001	0.002***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Sex	0.071***	0.060***	0.051***	0.060***
	(0.015)	(0.014)	(0.014)	(0.014)
Weight	0.001***	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
F-statistic	6.90	7.44	9.19	6.86
N	285274	255032	210810	231079

Robust standard errors are clustered at the individual level and are obtained by bootstrapping, using 500 replications for each individual. All specifications include MSA and month fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 13: Predicted Survival Functions: Failure ≡ missing two consecutive weeks

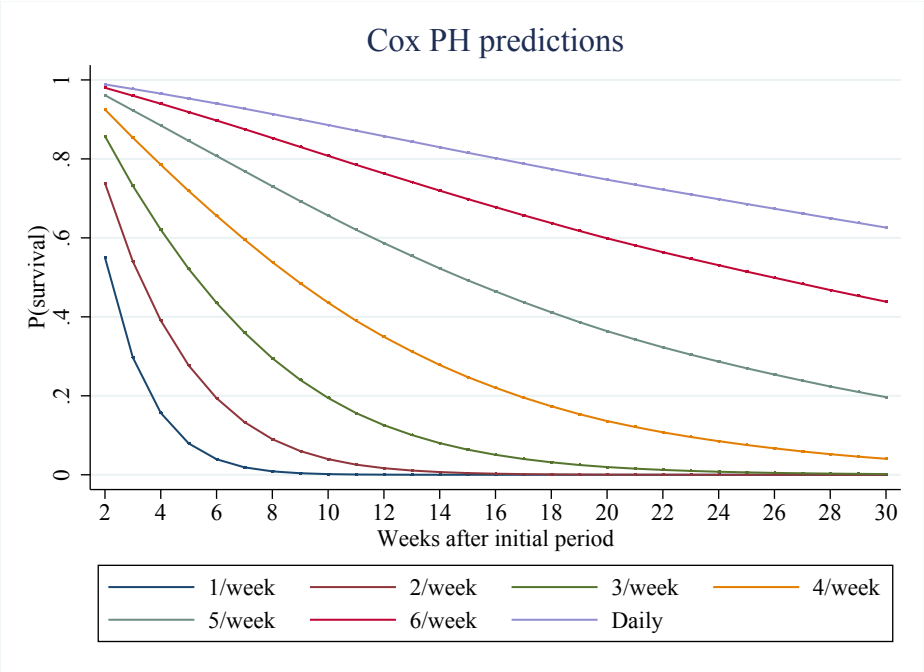


Figure 14: Predicted Survival Functions: Failure ≡ missing two consecutive weeks

