Journal of Economic Behavior and Organization 166 (2019) 688–708

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Habit formation and activity persistence: Evidence from gym equipment[☆]

Matthew C. Harris*, Lawrence M. Kessler

University of Tennessee, Haslam College of Business, Department of Economics and Boyd Center for Business and Economic Research
722A Stokely Management Center, 916 Volunteer Boulevard, Knoxville, TN 37996-0570 United States

ARTICLE INFO

Article history:
Received 19 November 2018
Revised 5 August 2019
Accepted 11 August 2019
Available online 17 August 2019

JEL classification:
I12
D12
D91

Keywords:
State dependence
Habit
Health behaviors
Survival analysis

ABSTRACT

This paper identifies the effect of habit formation on persistence of an important health-related behavior, namely usage of stationary exercise equipment. Exploiting user-activity data from a leading manufacturer and using rainfall as an instrument for initial exercise behavior, we find that frequent early activity leads to more persistent exercise in the subsequent periods, which is consistent with habit formation. Specifically, individuals who exercise one more time per week over a four-week initial period are more than three times as likely to exercise in each of the next eight weeks without missing a week. However, survival probabilities are quite low, consistent with recent experimental work. The primary implication is that for interventions to lead to persistent behavioral change, they must require more frequent activity or longer intervention periods than recent incentive-based field experiments.

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 for an explanation (Ryder and Heal, 1973; Becker and Murphy, 1988; Constantinides, 1990).[2,3] However, individuals also exhibit persistence in unhealthy choices in some areas for which "addiction" is an unlikely explanation. For

---

[☆] 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 7th World Congress of the International Health Economics Association and Vanderbilt Empirical Applied Microeconomics Festival. All errors are our own. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
E-mail address: mharris@utk.edu (M.C. Harris).

[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.

https://doi.org/10.1016/j.jebo.2019.08.010

example, individuals may choose to abstain from a particular good or activity, such as eating fruits and vegetables or exercising. For these types of decisions, 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 reverting to a default state, rather than because of addiction or switching costs.

Habit-in-utility and inertia (or default bias) are not mutually exclusive explanations for persistence in individuals choices. The combination of these factors can present serious challenges when designing interventions to get individuals to change behaviors. Consider the case of individuals persistently choosing not to exercise and earning the label 'sedentary.' Sedentary lifestyles have become increasingly common among U.S. adults. As a result of this and other factors, obesity and obesity-related diseases such as heart disease, type 2 diabetes, and hypertension have become more prevalent (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. For an intervention to be cost effective, however, it must lead to behavioral changes that persist after the intervention concludes.

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 in the absence (or after the conclusion) of some intervention. If being sedentary is a default option, individuals will tend to revert to that state. If we are designing an intervention, it may be useful to know how greater frequency of exercise during an intervention erodes the tendency to revert to the default state. 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 how individuals' frequency of usage during an initial period affects their exercise decisions in subsequent periods. Equipment users can download a mobile application (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 January 1, 2011 to July 31, 2015, meaning we have multiple years of daily reported workout activity for some users. We investigate the relationship between the frequency with which individuals exercise during an initial period (and the length of that initial period) and the individual's exercise behavior over the ensuing months in two respects. First, we examine whether more frequent workouts during an initial period lead to a higher volume of workouts over the next three to six months. Second, we examine whether the frequency of exercise during the initial period affects the probability that an individual continues to exercise regularly thereafter.

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 three durations of 'initial period': two, four, and six weeks. For each defined period, we estimate a two-stage 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 exercised 16 times over a four week initial period, how many times would he need to exercise during a six week initial period in order to maintain the same predicted level of persistence in exercise behavior going forward? In addition, we estimate a linear 2SLS model to examine how the number of workouts during the initial period affects the number of recorded workouts over the next three and six months.

There are several reasons to suspect there is 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.). For our identification strategy, we merge data on the county in which the individual is located with daily weather data from the National Oceanic and Atmospheric Administration (NOAA). Following Fujiwara et al. (2016), we use the number of days during the initial period in which more than half an inch of precipitation fell as an instrument for the number of workouts.[4] To alleviate concerns about serial correlation in weather patterns, we control for contemporaneous rainfall in the second stage of our main analysis, and also include state and month fixed effects.[5] In our first-stage results, we find that more rainy days during the initial period decreases the number of days that individuals engage in stationary exercise during that period. 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. For example, Connolly (2008) finds that on rainy days, men

---

[4] 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 Chaisemartin (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 (i.e. the compliers and the defiers), the second stage results still approximate the traditional local average treatment effect (LATE). This is further discussed in the empirical methods section.

[5] Regarding the validity of our instrument, we argue that since precipitation 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.

work an additional 30 min on average and salaried workers watch an additional 16 min of television. To the extent that weather induces individuals to allocate their time differently, weather will affect the amount of time available for stationary exercise.[6]

Our identification strategy also helps us address a limiting feature of our data. Specifically, in order to observe an individual exercising, the individual must make all four of the following choices: 1) to exercise, 2) to exercise on stationary exercise equipment, 3) to exercise on our manufacturers brand of stationary equipment, and 4) to 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. However, our use of weather as an instrument eliminates some concerns about how to interpret a non-observation. Specifically, 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. Because our first-stage results imply that precipitation decreases the number of days that individuals engage in exercise, this suggests that weather primarily affects exercise through the first margin (whether or not to exercise at all). However, to the extent that these results are affected by substitution on other margins, they should be interpreted more cautiously (i.e., frequency of stationary exercise during the initial period predicts greater persistence in stationary exercise, rather than physical activity in general).

We find that both the length of the initial period and the frequency with which an individual exercises during the initial period strongly affects both the amount of expected future exercise and the probability of continuing to exercise consistently in subsequent periods. While certain aspects of the data make it infeasible to specifically identify parameters in the individual's utility function, these results are consistent with the hypothesis that increased frequency of exercise during the initial period leads to more rapid habit formation. First, results from the 2SLS model indicate that exercising with greater frequency during an initial period leads to more activity during the ensuing three to six months. If we consider an initial period of four weeks, results indicate that one more workout per week during the initial period (i.e., exercising four more times over the four-week initial period) leads to an additional 5.2 (12.4) workouts during the subsequent three (six) months. The magnitude of the estimated marginal effects is not terribly sensitive to the definition of the length of initial period. If we similarly consider an initial period of six weeks, one more workout per week during the initial period leads to an additional 4.3 (12.9) workouts during the subsequent three (six) months.

In addition to increasing the total number of observed workouts, greater frequency of exercise during the initial period also leads individuals to exercise more consistently. For any given length of the initial period, individuals who exercise four times per week during the initial period are nearly 3.5 times more likely to exercise weekly over the next eight weeks than individuals who exercised three times a week over the same initial period.[7] 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 six weeks were 40.5% more likely to exercise each of the next eight weeks than individuals who exercised four times per week over the first four weeks, and over 80% more likely than those who exercised four times per week over the first two weeks. However, the stylized facts of our data are consistent with a strong default bias towards inactivity. Greater frequency of activity only partially offsets individuals' tendency to stop exercising.

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 (Charness and Gneezy, 2009; Cawley and Price, 2013; DellaVigna and Malmendier, 2006; 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 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.

From a policy perspective, both the response to incentives and the persistence of any incentive-induced behavioral changes are of nearly equal importance. Our results indicate that the lack of persistence in Charness and Gneezy (2009), Acland and Levy (2015), and Royer et al. (2015) could be improved if subjects are required to exercise at 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 require greater exercise frequency may lead to habit formation and affect persistent change. For example, Charness and Gneezy (2009) and

---

[6] More generally, rainy days affect the way individuals allocate their time between work, home production, and leisure. 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. 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.

[7] To put this into perspective using the six-week initial period as an example, those who exercised three times per week over the first six weeks had a 6.7% predicted probability of exercising weekly over the next eight weeks, without missing a week, while those who exercised four times per week had a 23.2% probability of exercising in each of the next eight weeks.

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 participants 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 0.6% probability of exercising every week for the next eight weeks. Similarly, individuals who exercise only 12 times over the first month have a 4.7% probability of exercising each of the next eight weeks. By contrast, an individual who exercises 16 times during the first month has a 16.5% probability of exercising in each of the next eight weeks. While requiring individuals to exercise more frequently may lead to higher attrition rates, our results indicate that interventions requiring greater frequency of activity may generate more persistent behavioral change.

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 stock variable that captures the history of the individual's choices in previous periods, 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 the individual's history of choices, captured by the stock variable, affects the marginal utility of the individual's decision in the current period, $U_{dS} \neq 0$. 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. The second term implies that exercising, d_t, has some positive effect on the stock variable 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 the stock of exercise. Rather, we consider individuals with different values of S_j at the end of the individual's first j weeks (i.e. the initial period) 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 theoretical work indicates that individuals with larger stocks 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'.

Status-quo bias (i.e., inertia) is generally 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 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, S_t, will only partially offset that tendency.

2.2. Intrinsic and extrinsic motivation

We acknowledge that comparing our results to previous studies which randomize individuals 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.[8]

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.

---

[8] 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.

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, but only to those who have some interest in physical activity.

3. Data

The primary data are provided by a large private manufacturer of fitness equipment. Data were collected through a mobile app 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 four-and-a-half-year time period (January 2011 through July 2015), and included information on the day and location of each individual's workout.

In this paper, we undertake two analyses, a linear 2SLS model and a control function survival analysis. The necessary conditions for inclusion in the analytical sample for each are slightly different. In all cases, we restrict our sample to individuals whose locations are known and in the U.S.[9] As noted earlier, we consider three durations of 'initial period:' two, four, and six weeks, and for the 2SLS analysis there are two lengths of post-period (three and six months). For an individual to be included in the 2SLS analytical sample for a combination of initial-period and post-period length, they must have exercised at least once during the initial period and the post-period must be complete before the end of July 2015, when our sample time period ends.[10] Table 1 reports summary statistics for the analytical samples for each initial/post period pair. The average individual was roughly 36 years old and weighed approximately 185 pounds, though more than 40% of the sample did not report their age, and 18 to 24% did not report their weight.

To be included in the analytical sample for our survival analysis, the only condition is that an individual must record a workout at least once during the initial period.[11] Table 2 reports basic summary statistics for the full sample of 24,439 individuals, as well as the three analytical samples where the 'initial period' is defined as two, four, and six weeks. Focusing on the four-week initial period sample, the average observation period was 13.9 weeks, meaning that on average there were 13.9 weeks between the individual's first and final recorded workouts. The average person exercised at least once per week for 6.4 weeks.[12] There was considerable variation in the number of weeks for which we observe individuals exercising. For example, in the four-week initial period sample of 23,490 individuals, over 8200 individuals (35.1%) exercised for one week only, over 7000 individuals recorded workouts for between two and four weeks, and 24 individuals exercised for at least 100 weeks (Fig. 1).

As additional context on the distribution of observed workouts per individual, there were 6951 individuals (29.6% of the sample) who only exercised once during the entire sample period and 2821 individuals who only exercised twice (12.0% of the sample) (Fig. 2). For the entire sample, there is a strong tendency to stop exercising. For example, in the four-week initial-period sample, there were an additional 5758 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 2.0 workouts per week.

When comparing the three initial period samples, we find that the average age, weight, and male-to-female ratio are similar across the three groups. However, the percentage of people who fail to report their weight and gender decrease slightly as the defined length of the initial period increases.

Our identification strategy relies on weather, specifically the number of days with more than half-an-inch of rainfall, to instrument for individuals' exercise choices during the initial period in which they are registered for the fitness mobile app. While the covariates in our primary data are somewhat limited, we are able to identify the zip code in which the individual resides. We use that information to match individuals with the closest WBAN station (Weather-Bureau-Army-Navy) from the NOAA's National Climatic Data Center (NCDC), and use daily weather data from NOAA-NCDC to form weekly measures of

---

[9] 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 24,439. In addition, the individuals who did not exercise again after the initial period are automatically dropped from our survival analysis and as such are statistically irrelevant to our primary estimation.

[10] Roughly 35% of our sample registered for the app during the first half of 2015. Approximately 290 new individuals registered per week over those six months. Therefore, for a given post period length, the difference in sample sizes between the two and four week group is attributable to the intersection of the completeness of the post-period and the end of the data. The much larger differences between the three and six month post periods for a common initial period are for the same reason.

[11] At least a third of our sample (roughly 35 to 56% of the sample, depending on how the initial period is defined) never exercise again after the initial period. To see if this is driving any of our results, we also perfom the survival analysis with a more stringent inclusion restriction of needing to record at least one workout during the initial period *and* at least one after the initial period. The results of which are qualitatively and quantitatively similar and can be found in Appendix D.

[12] This does not mean that the individual worked out for 6.4 consecutive weeks. Rather of the 13.9 weeks that we observe the average individual, they work out at least once during 6.4 of those weeks (on average).

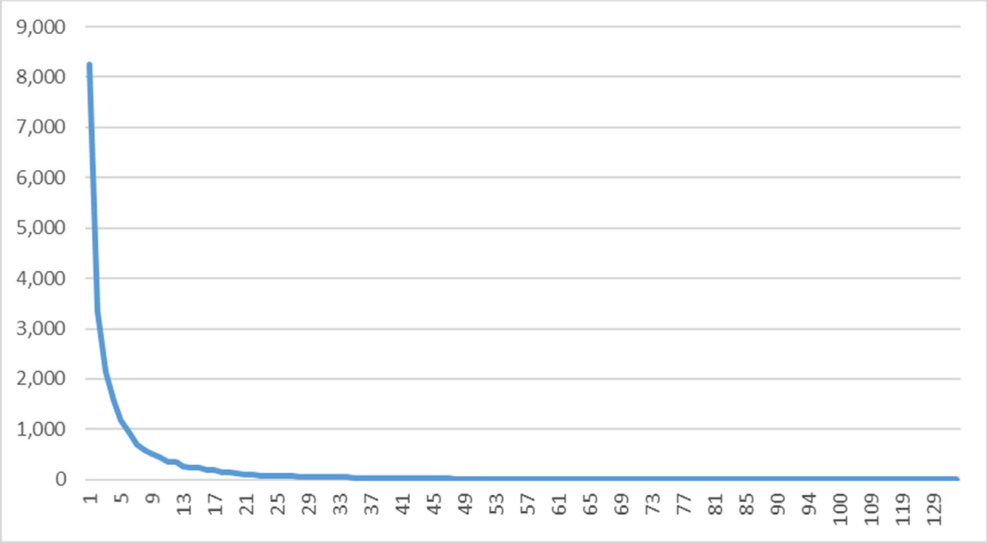

Fig. 1. Distribution of number of total weeks exercised per individual, sample defined by four week initial period. $N = 23,490$ individuals.

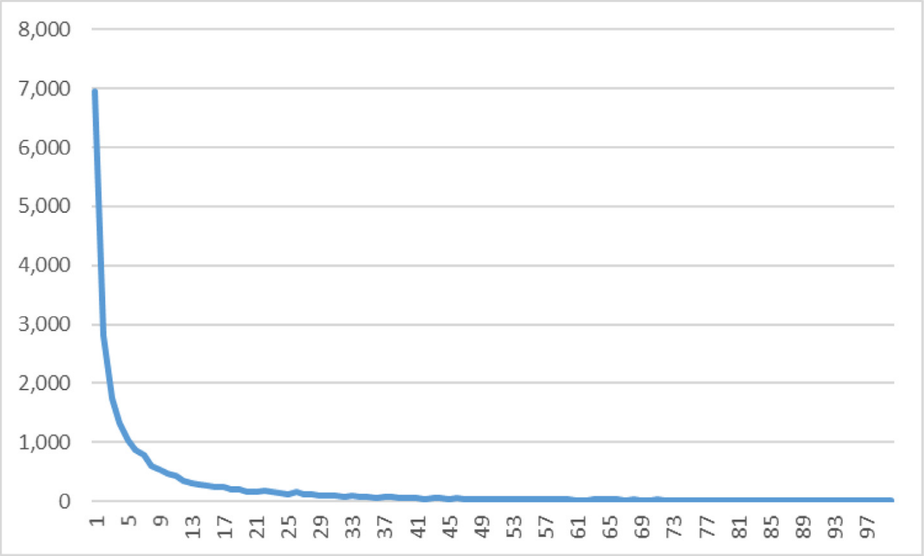

Fig. 2. Distribution of number of total days exercised per individual, sample defined by four week initial period. $N = 23,490$ individuals.

precipitation for each individual in our sample. Fig. 3 shows the counties in the U.S. highlighted in pink where individuals in our sample are present.

One potential challenge to interpreting the descriptive relationship between exercise during the initial period and the probability of continued regular exercise 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 mobile app.

However, using rainfall as an instrument broadens the interpretation of the effect of initial stationary exercise on future stationary exercise of all types. There is little reason to suspect that rainfall induces individuals to switch brands of equipment conditional on exercising, or use the mobile app conditional on using our manufacturer's brand. While rain could lead individuals to substitute between indoor and outdoor exercise, we do not observe outdoor exercise such as running or cycling. Therefore, if rain 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 less direct.[13]

---

[13] 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

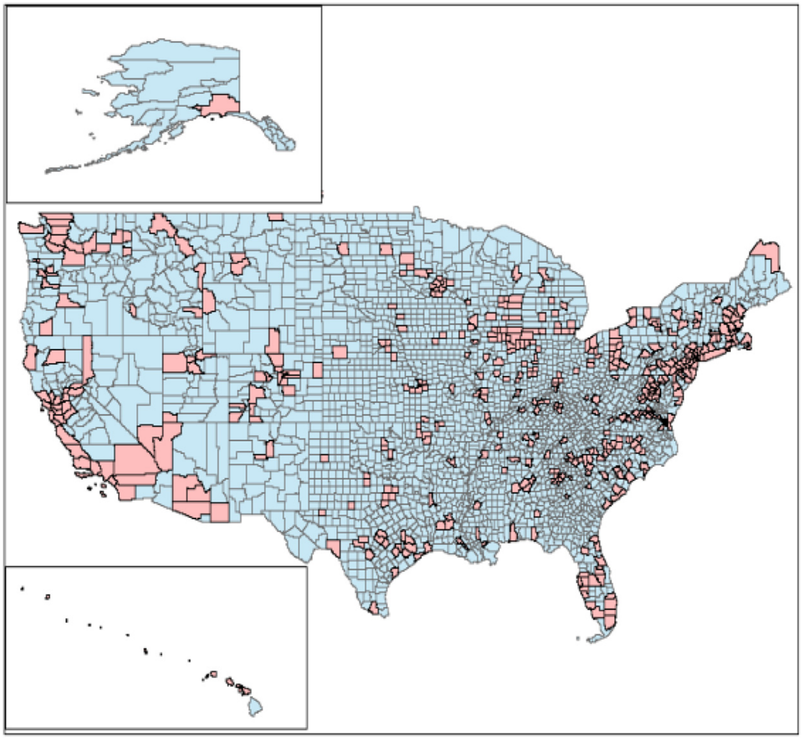

Fig. 3. Locations of individuals in the full sample, $N = 24,439$.

One may also be concerned about substitution between our manufacturer's equipment and other equipment brands not tracked by the mobile 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 rain may encourage people to exercise inside rather than engage in leisure, it is less clear through what mechanism rainfall would cause individuals to use stationary exercise equipment from an alternate manufacturer.

Focusing on the four-week initial period sample, Fig. 4 shows how the number of workout weeks and the number of rainy days (above half-an-inch of rain) vary by month. In total there were 149,265 workout weeks (i.e. weeks with at least one workout day) observed over the full analytic sample time period. The most workout weeks occurred in March (blue line) when there was an average of 0.36 rainy days per week (orange bar). More workouts are recorded in March through July than in any other months. There was also a 51.7% increase in the number of workout weeks from December to January, from 8865 in December to 13,450 in January. Over the full sample time period, the average number of rainy days per week was 0.41 and ranged from 0 to 6 days per week.

In the bottom portion of Table 2 we report summary statistics for the first two, four, and six weeks from when an individual started using the stationary exercise equipment. Focusing on the four-week initial period, the average individual had an initial exercise stock of 3.9 days, meaning they exercised 3.9 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). Fig. 5 shows how the initial exercise stock varies across the sample of 23,490 individuals. There were 8501 individuals who only exercised once during the initial four week period, and a total of 105 people who exercised 20 times or more. In total we observe 49,444 workout weeks that occurred within the first four weeks of when an individual first started using the stationary exercise equipment.[14] Fig. 6 breaks down the number of workout weeks and number of rainy days per week during this four-week initial period, by month. Similar to with the full sample time period, the most workout weeks occurred in March (12.5%). However, a large portion of the initial workout weeks were positioned in January through April, as 43.5% of the sample started exercising in the first four months of the calendar year. Consistent with this finding we also observe an increase from December to January, as the number of initial workout weeks increased by 80.4%.

---

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.

[14] At maximum, the total number of workout weeks over the first four weeks could be 93,960 (23,490 times 4). However, some individuals did not work out every week over those first four weeks. Specifically, there were 44,516 skipped weeks, leaving us with a total of 49,444 workout weeks.

Table 1
Summary statistics - analytical sample for linear 2SLS model.

Initial period (in weeks)	Two		Four		Six	
Post period (in months)	Three	Six	Three	Six	Three	Six
Number of days exercising during initial period (Stock)	2.52	2.52	3.89	3.89	4.91	4.88
	(1.92)	(1.93)	(3.80)	(3.82)	(5.43)	(5.40)
Number of days with rain above 0.5 in. during initial period	0.75	0.77	1.52	1.55	2.28	2.36
	(0.97)	(0.96)	(1.54)	(1.55)	(2.03)	(2.05)
Number of days exercising during post period	5.10	7.73	4.21	6.69	3.65	6.00
	(9.37)	(15.73)	(8.84)	(15.19)	(8.43)	(14.77)
Number of days with rain above 0.5 in. during post period	4.82	9.66	4.88	9.66	4.91	9.58
	(3.56)	(5.89)	(3.57)	(5.89)	(3.57)	(5.91)
Age (in years)	35.94	35.74	35.89	35.74	35.86	35.83
	(13.25)	(13.38)	(13.27)	(13.38)	(13.26)	(13.37)
Percentage with no age reported	40.1	43.1	41.0	43.7	41.3	44.3
Weight (in pounds)	185.10	184.20	185.10	184.16	184.80	183.99
	(50.96)	(51.52)	(51.12)	(51.73)	(51.02)	(51.90)
Percentage with no weight reported	18.5	22.2	18.9	23.1	19.4	24.3
Percentage male (*gender* = 0)	21.1	20.8	21.0	20.7	21.0	20.7
Percentage female (*gender* = 1)	16.9	18.5	17.1	19.0	17.3	19.4
Percentage with no gender reported	62.0	60.7	61.9	60.3	61.7	59.9
Individuals	21,243	17,703	20,747	17,009	20,250	16,123

Table 2
Summary statistics by initial period.

Panel A: full sample time period				
Variable	Full sample*	j = 2 weeks	j = 4 weeks	j = 6 weeks
Weeks observed	13.47	13.69	13.93	14.16
	(20.56)	(20.68)	(20.85)	(21.02)
Number of weeks exercising	6.17	6.26	6.35	6.44
	(10.33)	(10.40)	(10.49)	(10.59)
Number of weeks not exercising	7.30	7.43	7.57	7.72
	(14.51)	(14.61)	(14.73)	(14.87)
Number of days exercising (conditional on exercising that week)	2.04	2.04	2.04	2.04
	(1.30)	(1.30)	(1.30)	(1.30)
Number of days per week with rain above 0.5 in.	0.41	0.41	0.41	0.41
	(0.69)	(0.69)	(0.69)	(0.69)
Age (in years)	36.08	36.06	36.04	36.02
	(13.21)	(13.20)	(13.20)	(13.21)
Percentage with no age reported	39.71	39.14	39.43	39.73
Weight (in pounds)	183.79	183.72	183.60	183.38
	(45.43)	(45.41)	(45.39)	(45.23)
Percentage with no weight reported	21.40	20.50	20.07	19.61
Percentage male (*gender* = 0)	20.69	21.00	21.24	21.25
Percentage female (*gender* = 1)	16.03	16.16	16.13	16.28
Percentage with no gender reported	63.28	62.84	62.62	62.47
Individuals	24,439	24,008	23,490	22,966
Observations	329,203	289,025	263,210	241,665
Panel B: during initial period				
Variable		j = 2 weeks	j = 4 weeks	j = 6 weeks
Number of days exercising over first j weeks		2.52	3.89	4.91
		(1.93)	(3.80)	(5.43)
Number of days with rain above 0.5 in. over first j weeks		0.80	1.60	2.38
		(1.02)	(1.62)	(2.13)

* Full sample includes all individuals with location data in the United States.

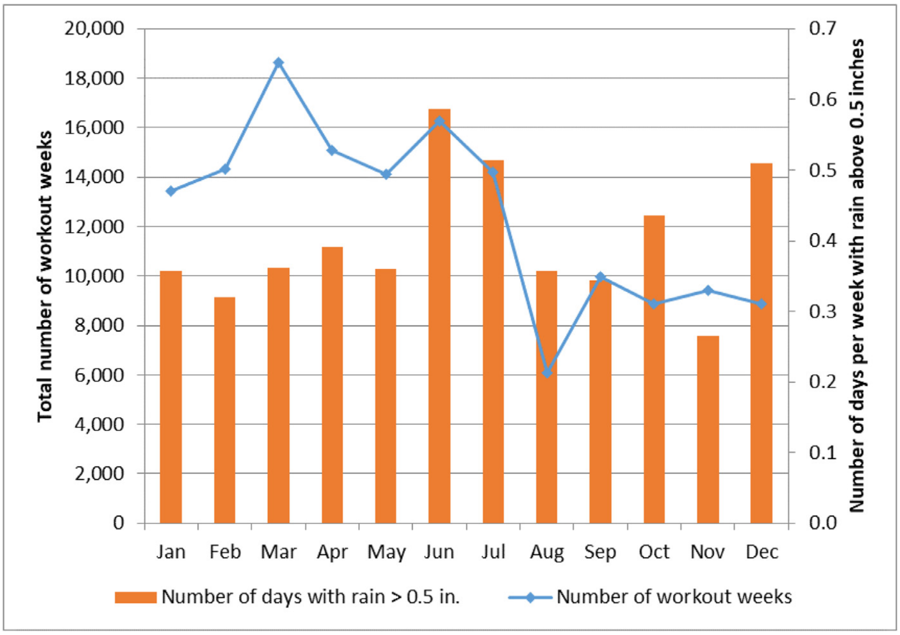

Fig. 4. Workout weeks and number of rainy days per week, by month for the sample defined by a four week initial period - $N = 23,490$ individuals Sample size reflects the analytical sample using the four-week initial period for the survival analysis rather than analytical sample in Table 1.

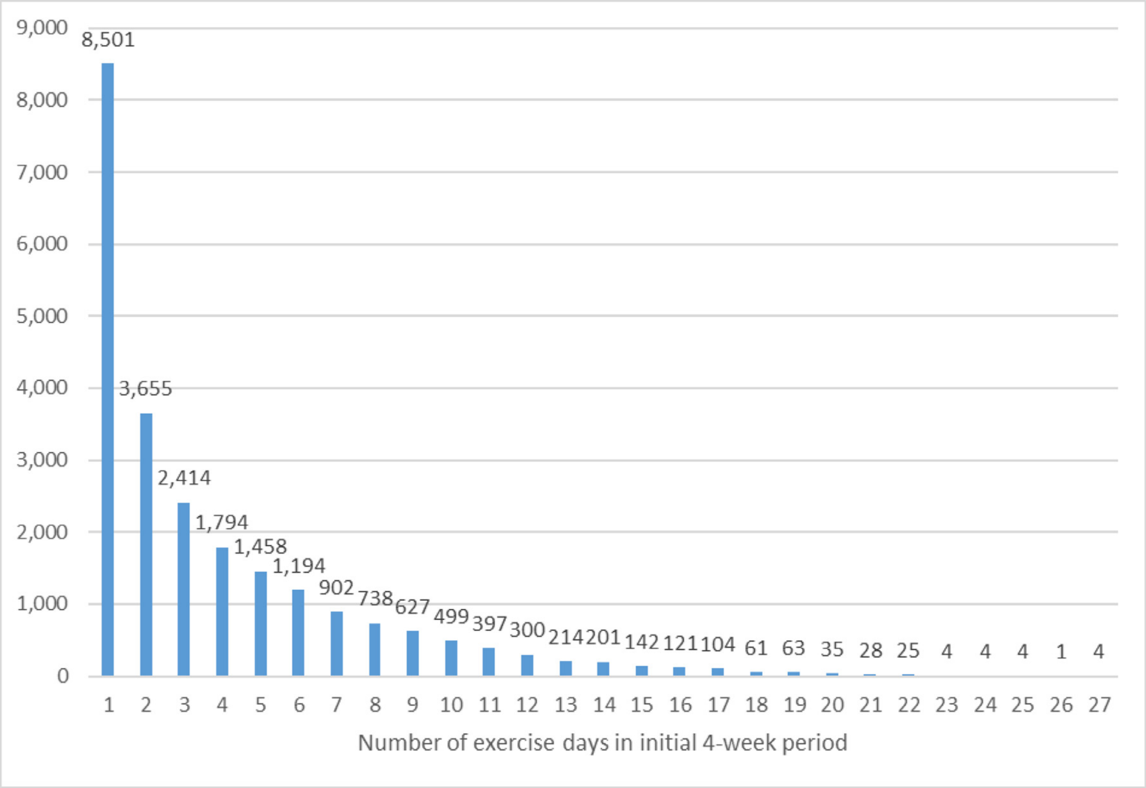

Fig. 5. Distribution of number of workout days over first four weeks of registration, $N = 23,490$ individuals Sample size reflects analytical sample for survival analysis rather than analytical sample in Table 1.

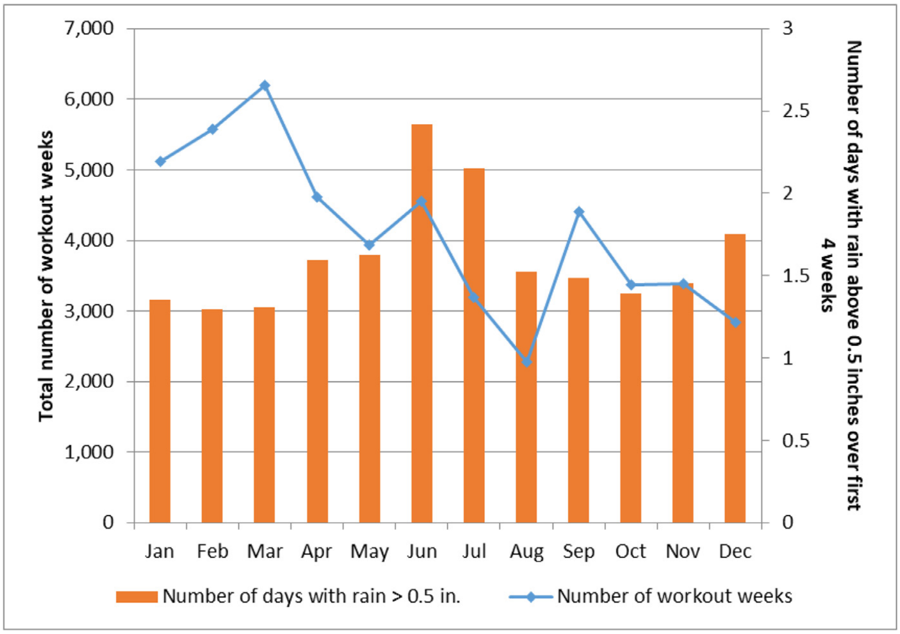

Fig. 6. Workout weeks and number of rainy days per week, by month - initial four weeks. $N = 23,490$ individuals Sample size reflects the analytical sample using the four-week initial period for the survival analysis rather than analytical sample in Table 1.

Table 3
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

4. Econometric methodology

We evaluate how the number of days that an individual exercises during an initial period impacts expected future exercise in two distinct ways. First, using a two-stage least squares approach with rainfall as an exclusion restriction, we examine whether more workouts during an initial period affects the total volume of exercise over the next three or six months. Second, we examine how exercise frequency during the initial period affects the consistency with which individuals exercise going forward, using a Cox proportional hazard model. 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. The outcome of ‘continuing to exercise in each successive week’ or ‘exercising at least every other week’ is particularly important to the notion of ‘persistent’ behavior because missing a week (or two weeks) marks an inflection point in the individual’s expected future frequency of observed workouts. Per Table 3, 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 using rainfall as an exclusion restriction for both the 2SLS and control function specifications.[15] Specifically, we use the number of days during the initial period in which more than half an inch of precipitation fell as our exclusion restrictions for the first stage. We can express

---

[15] The control function approach, also referred as the two-stage residual inclusion model (2SRI), is similar to the common two stage least square instrumental variables approach (2SLS) in that the first stage equation involves regressing instrumental variables onto the endogenous explanatory variable. However, unlike the typical 2SLS model 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). In a linear setting, the 2SLS and control function models produce identical results. However, in a nonlinear model, such as a Cox Proportional Hazard Survival model, the control function method yields consistent estimates whereas the 2SLS model generally does not (Terza et al., 2008).

the first stage as:

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

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. Z_i measures the number of days with more than half-an-inch of rainfall during the initial period and X_i is a vector of individual characteristics including age, gender, and weight at the time of initial registration.[16] While the theoretical ex ante relationship between rainfall and use of stationary exercise equipment is ambiguous and depends on context, recent work in the economics literature has shown that rainy days affect the way individuals allocate their time between work, home production, and leisure (Tucker and Gilliland, 2007; Zivin and Neidell, 2014).[17] Other threats to instrument validity include the possibility that the relationship between weather and exercise is driven by correlations between regional preferences for exercise and weather (i.e., do people in Arizona just exercise more than Idahoans?) or seasonal fluctuations in patterns of physical activity (i.e., are individuals who register in June systematically different from those 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}).[18]

We express the second-stage survival function as[19]:

$$h_{it} = \beta_0 \times exp\left[\beta_1 \sum_{t=1}^{T_j} N_{it} + \beta_2 X_i + \beta_3 \tilde{\epsilon}_{i1} + \beta_4 W_{it} + \alpha_{2s} + \delta_{2m}\right] \quad j = 2, 4, 6 \tag{4}$$

where h is the hazard function for individual i and measures the probability that an individual will continue exercising in week t. The term W_{it} includes a control variable for contemporaneous weekly rainfall (number of days with rain above half-an-inch in week t) to address concerns about serial correlation in weather patterns as a threat to the validity of our exclusions. As is customary in a control function approach, Eq. (4) contains the term $\tilde{\epsilon}_{i1}$, the predicted residuals from our first stage regression.

4.1. Monotonicity and compliers-defiers

An open question for using rainfall as in instrument for stationary exercise is whether the assumption of monotonicity holds. The monotonicity assumption (or no defiers) is necessary for our 2SLS and 2SRI results to be interpreted causally as a Local Average Treatment Effect (LATE) (Angrist et al., 1996). As discussed in Section 3, we only observe exercise of a specific type. When we do not observe individuals exercising, they could be sedentary or could be engaging in outdoor exercise. Therefore, while rainfall (for the majority of our sample) seems to lead individuals to skip the gym and go home, for some individuals, rainfall may lead them to substitute stationary exercise for running or biking. This would violate the monotonicity assumption. However, recent work by de Chaisemartin (2017) shows that monotonicity is not strictly necessary for 2SLS/2SRI results to be interpreted as a LATE, but that a weaker condition, "Compliers–Defiers" (C–D) is sufficient. While we refer the reader to de Chaisemartin (2017) for a comprehensive treatment, we briefly summarize the relevant aspects of the C–D condition for our research question here.

The C–D condition can be summarized as follows: if there are defiers in the population, 2SLS estimates a weighted difference between the effect of the treatment (second stage effect) among compliers and defiers. This is not a fatal concern in and of itself. However, the interpretation of the treatment effect as a LATE can be problematic and depends on: the relative number of compliers/defiers and the second stage treatment effects for the compliers and the defiers. Generally, we can still interpret the second stage estimates as a LATE if a subgroup of compliers accounts for at least the same population as the defiers (that is, the number of compliers is greater than or equal to the number of defiers) and their respective LATEs have certain properties. Per de Chaisemartin (2017), C–D holds if one of the following conditions are true: 1) In each stratum of the population with the same value of treatment effect $|C| > |D|$, that is there are more compliers (C) than defiers (D), or 2) The LATE for the defiers is the same sign as the 2SLS coefficient (which is the LATE for the compliers in a world without defiers), or 3) $\frac{|\beta^D_{LATE}|}{|\beta^C_{LATE}|} < \frac{C}{D}$ even if the compliers' and defiers' LATEs are of different signs.

In the interest of brevity, if either the effect of initial workouts on subsequent exercise behavior is positive among both the compliers and defiers, or if Condition 3 holds, our results are still interpretable as a LATE despite the issues with mono-

---

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

[17] 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.

[18] 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 C.

[19] For results using a linear 2SLS specification, the second stage is just the term inside the brackets, but with predicted values for $\sum_t \widehat{N_{it}}$ and without the first stage residuals $\tilde{\epsilon}_{i1}$. Furthermore, in the 2SLS model the dependent variable h_{it} measures the number of recorded workouts over the three (six) month post period.

Table 4
OLS results.

Initial period (weeks)	Two		Four		Six	
Post period (months)	Three	Six	Three	Six	Three	Six
Number of workouts during initial period	2.604^{***}	3.847^{***}	1.455^{***}	2.252^{***}	1.010^{***}	1.636^{***}
	(0.049)	(0.098)	(0.026)	(0.055)	(0.019)	(0.042)
# of days with rain above 0.5” during post period	-0.047^{**}	-0.038	$-0-.019^{*}$	-0.005	-0.030	0.001
	(0.019)	(0.026)	(0.016)	(0.025)	(0.015)	(0.024)
Age	0.068^{***}	0.127^{***}	0.053^{***}	0.102^{***}	0.047^{***}	0.091^{***}
	(0.006)	(0.013)	(0.006)	(0.013)	(0.006)	(0.013)
Female	-0.981^{***}	-2.300^{***}	-0.818^{***}	-2.022^{***}	-0.623^{***}	-1.708^{***}
	(0.211)	(0.404)	(0.189)	(0.372)	(0.177)	(0.360)
Weight	-0.004^{***}	-0.010^{***}	-0.006^{***}	-0.013^{***}	-0.006^{***}	-0.012^{***}
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
Individuals	21,243	17,703	20,747	17,009	20,250	16,123

Dependent variable: number of recorded workouts during post period. Robust standard errors in parenthesis. All specifications include state and month fixed effects. Additional controls: missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tonicity. First, it seems unlikely that tracked workouts in the initial period negatively affects tracked workouts in subsequent periods among those who substitute between indoor and outdoor exercise (i.e. the defiers). It may be that those who exercise indoors mainly because it is raining outside are less likely to track apps in subsequent periods but that is not the same as finding that more tracked workouts in the initial period have a negative effect on tracked workouts in the future. Second, regarding Condition 3, $\frac{|\beta^{D}_{LATE}|}{|\beta^{C}_{LATE}|} < \frac{C}{D}$, on the right hand side, we likely have substantially more compliers than defiers. On the left, β^{C}_{LATE}, the LATE among the compliers is likely positive and sizable as in our main results. It is also intuitive that if β^{D}_{LATE}, the LATE among the defiers is not positive, it is likely not different from zero. If the defiers, who substitute from outdoor exercise to indoor exercise when it is raining do not habituate to this specific form of behavior then it seems most plausible that exercise during the initial period will have little to no effect on subsequent usage of stationary exercise equipment among this subpopulation.

Unfortunately, statistical tests have not been developed for the above conditions, meaning we cannot provide inference on their validity. While the stylized facts of our data indicate that they likely hold, we acknowledge that if all of the above conditions are false, that our 2SLS and 2SRI results are difficult to interpret without additional information to separate the compliers from the defiers.

5. Results

This section contains results from descriptive, linear 2SLS, and control function survival analyses.[20] All specifications include demographic control variables (unless specified otherwise) and fixed effects for each state and month. As stated above, we evaluate whether exercise frequency during an initial period affects subsequent exercise in two dimensions. First, how do additional days of exercise during an initial period affect the total volume of exercise over the next three to six months? Second, how do additional days of exercise affect the probability that individuals continue to exercise in consecutive weeks after the initial period? If the goal of policy interventions is to affect persistent changes in behavior, both are important measures of how activity during some initial phase affects the individual’s choices in subsequent periods.

5.1. Linear 2SLS results

To examine how exercise frequency during an initial period affects the total volume of exercise in the following months we employ a simple linear cross-sectional model where we regress the number of workouts during the three (or six) months following the initial period on the number of workouts during the initial period (still defined as two, four, or six weeks). In each column of Tables 4–7, the results correspond to a model where the initial period is defined as either two, four, or six weeks, and the dependent variable is the total number of workouts over the ensuing three or six months.

Table 4 presents results from a simple OLS specification, where the variable of interest is the number of workouts during the initial period. For all defined lengths of initial period and post-period, greater frequency of exercise during the initial period is statistically significant and positively correlated with the number of workouts after the initial period. Two interesting patterns emerge from these results. First, if we interpret the marginal effect in per-week terms, an additional workout per week for a two-week initial period is associated with an additional 5.2 workouts over the three months following that

---

[20] In addition, we estimate a dynamic logit model to examine the relationship between past and current exercise behavior. These results are presented in Appendix E.

Table 5
First stage results.

Initial period (weeks)	Two		Four		Six	
Post period (months)	Three	Six	Three	Six	Three	Six
# of days with rain above 0.5” during initial period	-0.043^{***} (0.0141)	-0.065^{***} (0.0158)	-0.036^{***} (0.0097)	-0.041^{***} (0.0108)	-0.028^{***} (0.0077)	-0.032^{***} (0.0087)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (exclusion)	9.26	16.74	16.69	16.73	16.66	17.49
Individuals	21,243	17,703	20,747	17,009	20,250	16,123

Dependent variable: number of recorded workouts during initial period. Robust standard errors in parenthesis. Controls: Age, gender, weight, and missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6
Linear 2SLS Results - No controls.

Initial period (weeks)	Two		Four		Six	
Post period (months)	Three	Six	Three	Six	Three	Six
Number of workouts during initial period	3.989^{***} (1.359)	6.183^{***} (1.773)	0.969^{**} (0.442)	2.088^{**} (0.814)	0.637^{**} (0.268)	1.580^{***} (0.538)
# of days with rain above 0.5” during post period	-0.037 (0.032)	-0.022 (0.034)	-0.044^{*} (0.023)	-0.010 (0.028)	-0.049^{***} (0.018)	0.000 (0.026)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No
F-statistic (exclusion)	9.91	17.89	16.71	15.03	19.12	14.62
Individuals	21,243	17,703	20,747	17,009	20,250	16,123

Dependent variable: number of recorded workouts during post period. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

initial period.[21] An additional workout per week for a four (six)-week initial period is associated with an additional 5.8 (6.1) workouts over the three ensuing months. Therefore, there appears to be a positive interaction effect between the length of the initial period and frequency of exercise during the initial period on the volume of subsequent exercise. Second, all estimates for the six-month post periods are significantly less than two-times the estimates of the three month post period, indicating the association between frequency of initial-period exercise and subsequent exercise diminishes over time. This is consistent with prior findings of non-persistence from the experimental literature.

As stated above, we implement an instrumental variables approach to estimate the causal effects of exercise during an initial period on the number of workouts over the following months. The instrument is the number of days during the initial period in which more than a half-an-inch of rain fell. Table 5 displays the first stage results from the 2SLS model with control variables. For each pair of initial period and post period length, rainy days decrease the number of workouts. Prior work on how rain affects individuals' time allocation decisions provides the most plausible explanation for the negative effect of rain on stationary exercise. Specifically, Connolly (2008) show that on rainy days individuals are more likely to engage in sedentary leisure or supply additional hours of labor. In other words, on rainy days, individuals are more likely to just skip the gym and go home. Aside from the results in Column 1, the F-statistics on our exclusion restriction, the number of rainy days during the initial period, are at least 50% larger than the rule of thumb of 10 established by Staiger and Stock (1997), indicating that precipitation is a strong instrument.

Tables 6 and 7 contain second stage results for 2SLS models with no control variables and control variables respectively. Note that adding controls does not substantively affect first stage F-statistics, nor do the second-stage estimates substantively change. Like the OLS estimates, the 2SLS estimates are positive, indicating that a higher frequency of activity during the initial period leads to more workouts in subsequent months. However, for both the four-week and six-week initial periods the 2SLS estimates are more than twice as large in magnitude in the regressions examining the six month post period than they are for the three month post period. While the OLS models are useful in demonstrating sign and significance, for the causal relationship the difference in the magnitude of the estimate is most likely attributable to the Local Average Treatment Effect from the 2SLS models. Specifically, these estimates capture the effect of initial workout frequency on the number of workouts in subsequent months only insofar as rain affects the number of workouts. Individuals for whom rain makes the

---

[21] For the two-week initial period, one additional workout per week amounts to exercising two more times over the two-week period. Thus the coefficient for 'number of workouts during initial period' from Column 1 of Table 4 of 2.604 is multiplied by 2 to get an additional 5.2 workouts over the ensuing three months. Similarly, for the four-week (six-week) initial period, one additional workout per week amounts to exercising four (six) more times over the respective initial period.

Table 7
Linear 2SLS Results - with controls.

Initial period (weeks)	Two		Four		Six	
Post period (months)	Three	Six	Three	Six	Three	Six
Number of workouts during initial period	3.853^{***}	6.696^{***}	1.289^{***}	3.112^{***}	0.717^{**}	2.148^{***}
	(1.411)	(1.885)	(0.445)	(0.840)	(0.293)	(0.530)
# of days with rain above 0.5” during post period	−0.032	−0.016	−0.022	0.005	-0.035^{**}	0.007
	(0.025)	(0.031)	(0.019)	(0.027)	(0.016)	(0.026)
Age	0.058^{***}	0.107^{***}	0.057^{***}	0.083^{***}	0.057^{***}	0.072^{***}
	(0.013)	(0.019)	(0.011)	(0.022)	(0.012)	(0.023)
Female	-0.973^{***}	-2.160^{***}	-0.845^{***}	-1.804^{***}	-0.732^{***}	-1.413^{***}
	(0.218)	(0.435)	(0.203)	(0.430)	(0.212)	(0.470)
Weight	-0.007^{*}	-0.016^{***}	-0.005^{**}	-0.016^{***}	-0.004^{**}	-0.014^{***}
	(0.003)	(0.005)	(0.002)	(0.005)	(0.002)	(0.004)
F-statistic (exclusion)	9.26	16.74	16.69	16.73	16.66	17.49
Individuals	21,243	17,703	20,747	17,009	20,250	16,123

Dependent variable: number of recorded workouts during post period. Robust standard errors in parenthesis. All specifications include state and month fixed effects. Additional controls: missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8
First-stage control function results.

Dependent variable: Number of recorded workouts during initial period			
Initial period	Two weeks	Four weeks	Six weeks
Number of days with rain above	-0.043^{***}	-0.055^{**}	-0.074^{***}
	(0.013)	(0.017)	(0.019)
State Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
F-statistic (exclusion)	10.87	10.42	14.57
Individuals	24,008	23,490	22,966

Dependent variable: number of recorded workouts during initial period. Robust standard errors are clustered at the individual level. Controls: Age, gender, weight, and missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

difference between going to the gym and not are marginal exercisers. As such, a few more workouts during some initial period may be especially important in affecting workout behavior in subsequent months.[22]

5.2. Survival analysis

Having shown that the frequency with which workouts are observed during an initial period affects the total number of observed workouts in subsequent months, next we examine whether exercise frequency leads individuals to exercise more consistently. Specifically, we evaluate whether exercising more often during the initial period affects the probability that individuals continue to exercise in consecutive weeks after the end of that period. Because we only observe tracked workouts on a particular manufacturer’s equipment, we recommend caution in extrapolating these results as they are limited to habituation of a very specific behavior.

We present the results from the control function survival analysis using the same three lengths of initial period from the 2SLS model and two definitions of ‘failure’ - going one or two consecutive weeks without recording a workout. Table 8 contains the results from our first-stage control function. Regardless of whether the initial period is defined as two, four, or six weeks, more rainy days during the initial period decreases the number of times an individual is observed to exercise over that span, as is the case in the 2SLS model.[23] The F-statistics on ‘rainy days during the initial period’ are all larger than the rule of thumb of 10 established by Staiger and Stock (1997), indicating that days of rain is a strong instrument.

Table 9 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

---

[22] For completeness, we have repeated our analysis examining whether the frequency of workouts during the initial period affects future workouts at the extensive margin - whether individuals exercise at all. Results are available in Appendix A Table XI. Unsurprisingly, the effect of additional workouts is positive in the OLS model, but insignificantly different from zero for the 2SLS model. There is limited room for rain-induced workouts to have an impact at the extensive margin as approximately 75% of those included in the 2SLS analysis exercised at least once after the initial period anyway. We also replicate the analysis from Table 8 and Table 7 but with average daily precipitation rather than ‘rainy days’ as an instrument. These results are also presented in Appendix A Tables XII and XIII. Results are neither qualitatively nor quantitatively different from those in the main text.

[23] Control function approaches usually require a linear first stage. We therefore use a linear first stage even though we are essentially working with count data. We have verified that our first stage estimates do not change in either sign or significance when we estimate the first stage using a Poisson model.

Table 9
Survival analysis results: Failure ≡ Missing one week.

Panel A - baseline survival analysis			
Initial period	Two weeks	Four weeks	Six weeks
Number of workouts-initial period	−0.104***	−0.061***	−0.045***
	(0.004)	(0.002)	(0.001)
Number of days with rain above 0.5 in. in week t	−0.0001	−0.001	−0.001
	(0.002)	(0.002)	(0.002)
State Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	289,025	263,210	241,665
Panel B - control function survival analysis			
Initial period	Two weeks	Four weeks	Six weeks
Number of workouts - initial period	−0.232***	−0.133***	−0.102***
	(0.013)	(0.008)	(0.006)
Fisrt-Stage Residual	0.191***	0.125***	0.104***
	(0.018)	(0.012)	(0.010)
Number of days with rain above 0.5 in. in week t	−0.001	−0.001	−0.001
	(0.002)	(0.002)	(0.002)
State Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
F-statistic (exclusion)	10.87	10.42	14.57
Observations	289,025	263,210	241,665

For the baseline model, robust standard errors are clustered at the individual level. For the control function model, robust standard errors are obtained by bootstrapping, using 500 replications for each individual. Controls: Age, gender, weight, and missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

approach. Regardless of whether or not we instrument for the number of workouts during the initial period, more workouts in the first month leads to a lower probability of failure (i.e. greater persistence). However, under the control function specification, the effect of additional workouts on persistence significantly increases. In the baseline analysis, an additional workout during the four (six)-week initial period decreases the probability of missing a full week of exercise by 6.1 (4.5)%. In the control function models, one additional workout during a four (six)-week initial period decreases the same probability by 13.3% (10.2%). This implies that the baseline estimated effect of initial workouts on persistence is biased downward. While we cannot pinpoint the source of the bias that the control function alleviates, one plausible explanation is an unobservable 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 exercise 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. This sort of variation could create correlations in the unobservable factors that affect both the number of workouts during the initial period and the survival probability in subsequent periods.

The results from the control-function specification indicate that the correction for endogeneity is necessary. In all versions of the model, the first-stage residuals are statistically significant at the 1% 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 10 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 9. Table 10 demonstrates the lack of sensitivity of these results to the definition of failure. Working out with greater frequency during the initial period predicts greater persistence going forward.

5.3. 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. Figs. 7–9 plot the probability that individuals do not go a

Table 10
Survival analysis results: Failure ≡ Missing two weeks.

Panel A - baseline survival analysis			
Initial period	Two weeks	Four weeks	Six weeks
Number of workouts - initial period	−0.128***	−0.078***	−0.058***
	(0.006)	(0.003)	(0.002)
Number of days with rain above 0.5 in. in week *t*	0.004	0.004	0.005
	(0.003)	(0.003)	(0.004)
State Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	289,025	263,210	241,665
Panel B - control function survival analysis			
Initial period	Two weeks	Four weeks	Six weeks
Number of workouts - initial period	−0.291***	−0.170***	−0.135***
	(0.019)	(0.012)	(0.010)
Fisrt-Stage Residual	0.244***	0.161***	0.140***
	(0.027)	(0.020)	(0.015)
Number of days with rain above 0.5 in. in week *t*	0.004	0.004	0.005
	(0.003)	(0.003)	(0.003)
State Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
F-statistic (exclusion)	10.87	10.42	14.57
Observations	289,025	263,210	241,665

For the baseline model, robust standard errors are clustered at the individual level. For the control function model, robust standard errors are obtained by bootstrapping, using 500 replications for each individual. Controls: Age, gender, weight, and missing information dummy variables for age, gender, and weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

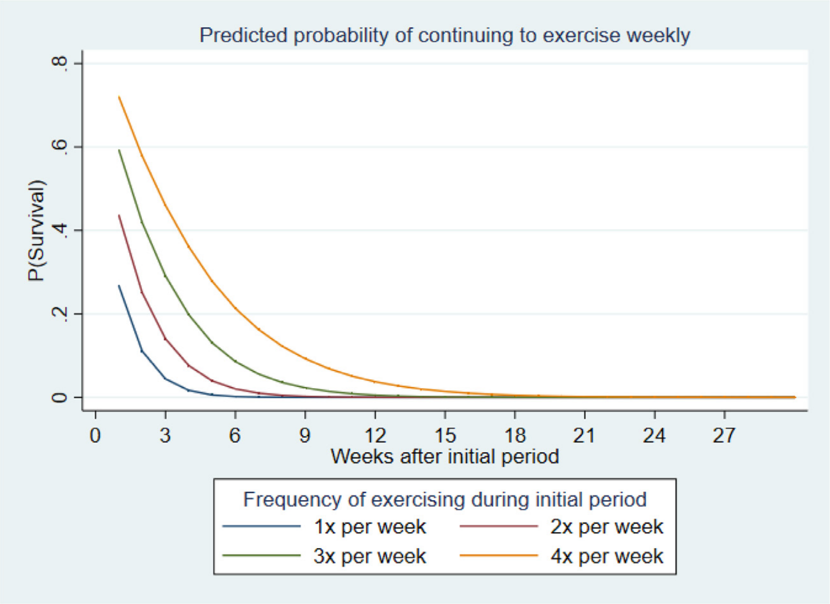

Fig. 7. Survival Function, Two week initial period, Failure ≡ missing one week.

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.

Figs. 7 –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.[24] Comparing the lines within

---

[24] When we restrict the sample so that the 2, 4, and 6 week groups are comprised of the same number of individuals (from the 6-week sample), our results are virtually identical. These results, for the control function models only, are presented in Appendix C Table XV.

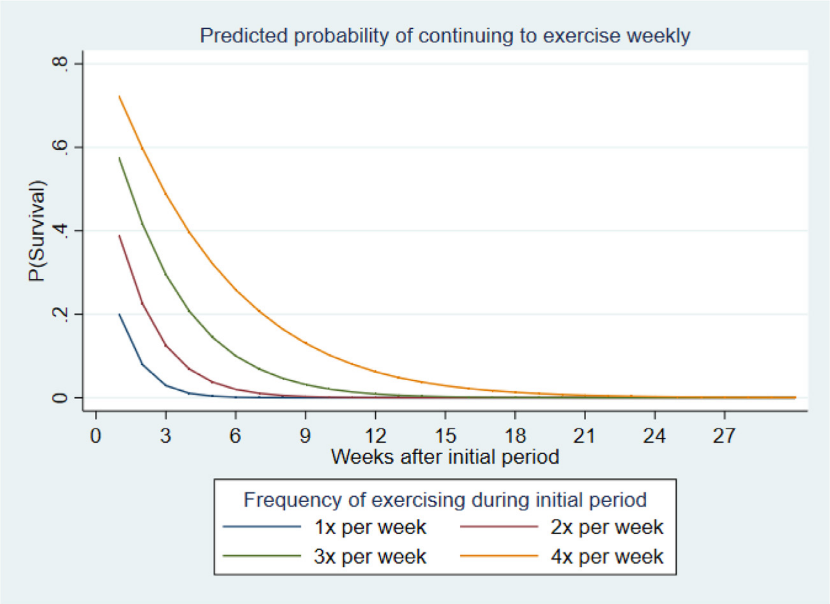

Fig. 8. Survival Function, Four week initial period, Failure ≡ missing one week.

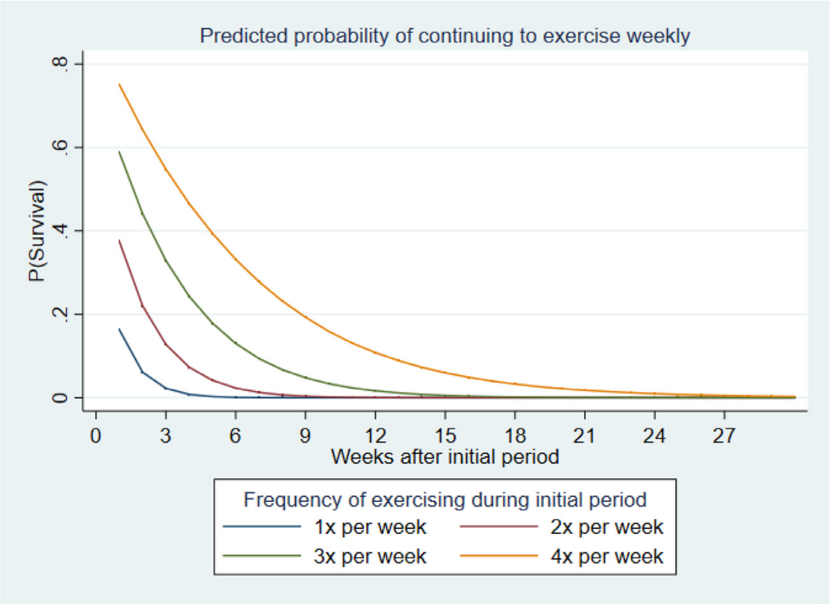

Fig. 9. Survival Function, Six week initial period, Failure ≡ missing one week.

each graph informs about the effect of more frequent activity during an initial period. In Fig. 8, for example, individuals who are observed exercising three times per week over the first four weeks have a 4.7% 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 times per week over the first four weeks have a 16.5% probability of exercising in 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 times per week during the initial period are approximately 3.5 times more likely to exercise in each of the next eight weeks than individuals who exercise three times per week.[25]

---

[25] 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.

Comparing probabilities across Figs. 7–9 illustrates how the length of the initial period affects the probability that individuals continue to exercise, holding the number of workouts per week during the initial period constant. Consider two lengths of initial period, two weeks and six weeks. Individuals who exercise for an average of four times per week over an initial period of six weeks are nearly 50% more likely to continue to exercise in each of the eight successive weeks than when the initial period is defined as two weeks. Indicating that, 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 90.2% 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 69.2% chance of missing two consecutive weeks within the same frame. By contrast, individuals who exercise four times per week for the first month will miss two consecutive weeks during the next two months with a probability of 44.9%. These figures are presented Appendix B.

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 ten 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 4.7% probability of exercising in each of the ensuing eight weeks. The primary takeaway from these results is that the frequency of activity during some initial period is a strong predictor of future persistence. If experimentalists or others are designing interventions with the goal of affecting persistent behavioral change with respect to physical activity, individuals should be encouraged (or incentivized) to participate almost daily. If we extrapolate the estimated relationships to make out of sample predictions, individuals who exercise five times per week over the first four weeks, have a 34.7% chance of exercising every week for the next eight weeks, or more than twice as likely as those who only exercise four times per week. Second, the dominant tendency in the data is for individuals to miss weeks of exercising and eventually attrit.

Although we cannot rule out all other explanations, these results are consistent with the idea that individuals have a strong default bias towards inactivity. During an initial period of any length, greater frequency of activity predicts greater persistence going forward, which is consistent with increased habituation. However, in the absence of continued reinforcement or an intervention of sufficient rigor to alter the individual's default state, these results indicate that individuals exercise patterns will become more intermittent within six months. Eventually, we expect them to revert to inactivity.

5.4. What does it take to maintain persistence?

If we take seriously the above notion that individuals have a default state (here, inactivity) and that part of the goal of an intervention is to form new habits, another way to consider the effects of exercise during an initial period is to ask the following: for a given 'stock', how much continued investment 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, which is a function of the individual's history of exercise choices.[26] We therefore consider the following: suppose an individual is observed exercising four times per week over an initial four week period and therefore has a 16.5% probability of exercising in each of the next eight weeks. How many times on average would he need to exercise during a six week initial period to generate the same survival probability? Formally, we are interested in the value of w that solves:

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

where $S_j(K)$ indicates the stock as a result of exercising an average of K times per week over an initial period of j weeks. Figs. 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.[27] In Fig. 10 for example, tracing a horizontal line at 16.5% will yield the number of workouts over each initial period to generate the same predicted level of persistence eight weeks after the initial period. In general, as the length of the initial period increases, the number of workouts per week needed to generate 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 16.5% after eight weeks, he need only exercise 3.67 times per week for a six week initial period (and approximately 3.4 times per week for an eight week initial period) to generate the same probability of continuing to exercise after that initial period.[28] While we cannot identify the economic primitives, these results are also consistent with habit formation, albeit weak habit formation. Specifically, longer initial periods imply lower average levels of investment are necessary to yield a given stock of habituation.

---

[26] 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 estimates the stock of habituation from persistence. See, for example, Darden (2017).

[27] These graphs also contain predicted survival probabilities from an eight week initial period. Results from the control function regression for an eight week initial period are highly consistent with the two, four, and six week periods reported here, and are available upon request.

[28] This pattern creates a total of 22 workouts during a six week initial period and 27.5 workouts during an eight week initial period.

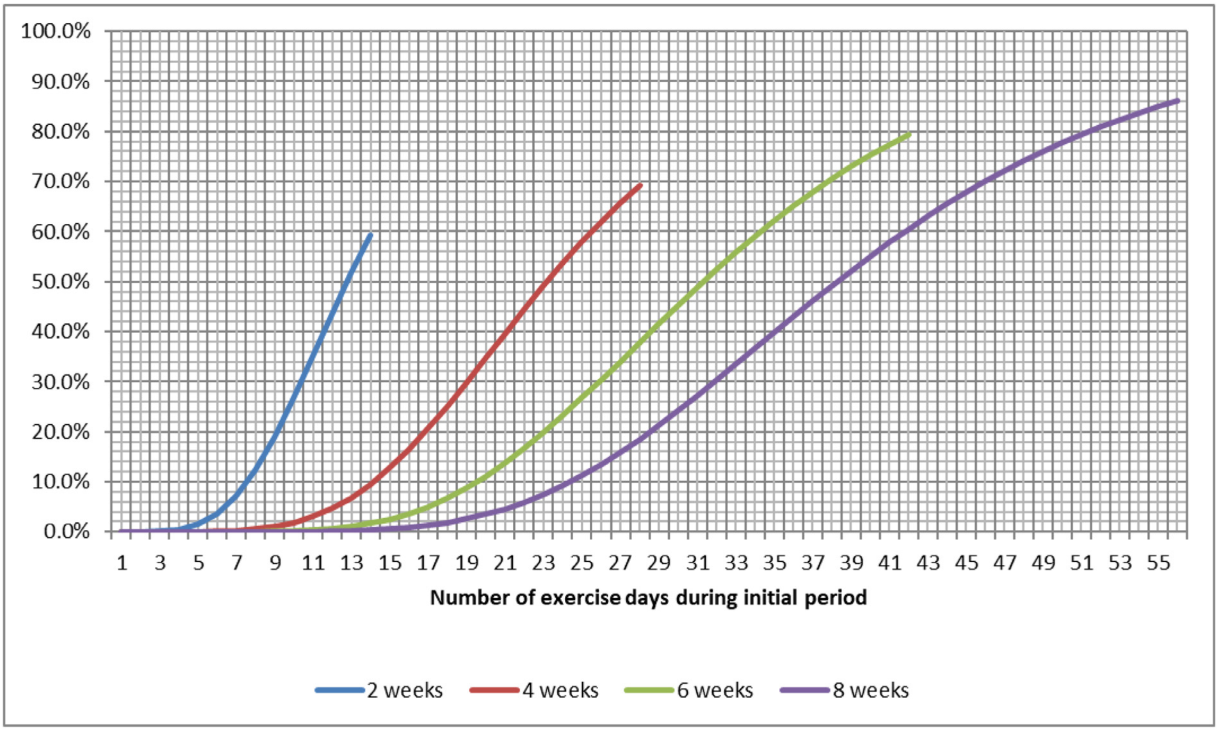

Fig. 10. Persistence and Number of workouts, Eight weeks after initial period.

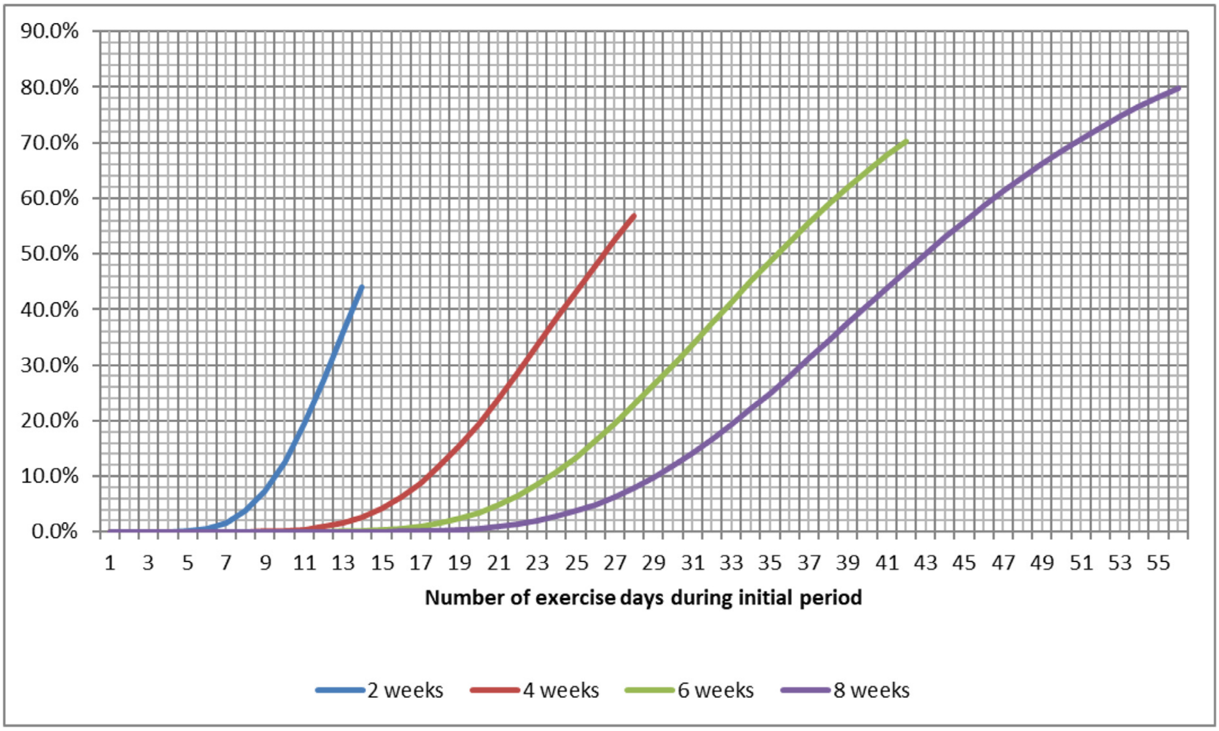

Fig. 11. Persistence and Number of workouts, Twelve weeks after initial period.

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 rainfall as an instrumental variable enables us to identify the causal effect of additional workouts during some initial period on the expected persistence of behavior going forward. 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)% chance of missing whole weeks of exercise in the next two months. Individuals who workout four times per week during the first month are 3.5 times less likely to miss a week than those who exercised three times per week.

The main implication of these results 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 at least 4 times per week are much more likely to lead to persistent behavioral changes than the designs in prior work. One caveat that should be kept in mind when interpreting these results is that we only observe (and therefore study habit formation of) exercise of a very specific type: stationary cardiovascular exercise, done on a particular manufacturer's equipment, and tracked on a specific app. These results, therefore, may not extrapolate to exercise in general, partially because individuals may be exercising in other ways on days that we do not observe them tracking workouts. While we contend these results do inform about habit formation, they should be interpreted as habit formation with respect to a very specific exercise behavior rather than exercise of any type.

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. Additionally, these results indicate that many individuals are likely to revert to inactivity at some point, but that requiring greater frequency of activity during the initial period is likely to postpone that reversion. Whether this trade-off is in fact cost-effective is an area for future work.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2019.08.010.

References

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

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91, 444–472.

Becker, G.S., Murphy, K.M., 1988. A theory of rational addiction. J. Polit. Econ. 96 (4), 675–700.

Benabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Rev. Econ. Stud. 70 (2), 489–520.

Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. J. Polit. Econ. 107 (2), 205–251.

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

de Chaisemartin, C., 2017. Tolerating defiance? local average treatment effects without monotonicity. Quant. Econ. 8 (2), 367–396.

Chaloupka, F., 1991. Rational addictive behavior and cigarette smoking. J. Polit. Econ. 99 (4), 722–742.

Chan, C.A., Ryan, D.A., 2009. Assessing the effects of weather conditions on physical activity participation using objective measures. Int. J. Environ. Res. Public Health 6 (10), 2639–2654.

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

Connolly, M., 2008. Here comes the rain again: weather and the intertemporal substitution of leisure. J. Labor Econ. 26 (1), 73–100.

Constantinides, G.M., 1990. Habit formation: a resolution of the equity premium. J. Polit. Econ. 98 (3), 519–543.

Darden, M.E., 2017. Smoking, expectations, and health: a dynamic stochastic model of lifetime smoking behavior. J. Polit. Econ. Forthcoming.

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

DellaVigna, S., Malmendier, U., 2006. Paying not to go to the gym. Am. Econ. Rev. 96 (3), 694–719.

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

Eisenberg, D., Okeke, E., 2009. Too cold for a jog? weather, exercise, and socioeconomic status. B. E. J. Econ. Anal. Policy 9 (1).

Ferson, W.E., Constantinides, G.M., 1991. Habit persistence and durability in aggregate consumption. J. Financ. Econ. 29, 199–240.

Flegal, K.M., Carroll, M.D., Ogden, C.L., Curtin, L.R., 2010. Prevalence and trends in obesity among us adults, 1999–2008. J. Am. Med. Assoc. 303 (3), 235–241.

Fujiwara, T., Meng, K., Vogl, T., 2016. Habit formation in voting: evidence from rainy elections. Am. Econ. J. 8 (4), 160–188.

Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. J. Econ. Perspect. 25 (4), 191–209.

Handel, B., 2013. Adverse selection and inertia in health insurance markets: when nudging hurts. Am. Econ. Rev. 103 (7), 2643–2682.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46 (6), 1251–1271. doi:10.2307/1913827.

Kahneman, D., Knetsch, J.L., Thaler, R., 1991. The endowment effect, loss aversion, and status quo bias. J. Econ. Perspect. 5 (1), 193–206.

Keane, M.P., 1997. Modeling heterogeneity and state dependence in consumer choice behavior. J. Bus. Econ. Stat. 15 (2), 310–327.

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

Ogden, C.L., Carroll, M., Kit, B., Flegal, K.M., 2014. Prevalence of childhood and aduly obesity in the united states, 2011–2012. J. Am. Med. Assoc. 311 (8), 806–814.

Olekalns, N., Bardsley, P., 1996. Rational addiction to caffeine: an analysis of coffee consumption. J. Polit. Econ. 104 (5), 1100–1104.

Royer, H., Stehr, M.F., Sydnor, J.R., 2015. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. Am. Econ. J. 7 (3), 51–84.
Ryder, H.E., Heal, G.M., 1973. Optimal growth with intertemporally dependent preferences. Rev. Econ. Stud. 40 (1), 1–31.
Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. J. Risk Uncertain. 1 (1), 7–59.
Spence, M.A., 1981. The learning curve and competition. Bell J. Econ. 12 (1), 49–70.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.
Terza, J.V., Basu, A., Rathouz, P.J., 2008. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. J. Health Econ. 27 (3), 531–543.
Thaler, R.H., Sunstein, C.R., 2003. Libertarian paternalism. Am. Econ. Rev. Pap. Proc. 93 (2), 175–179.
Tucker, P., Gilliland, J., 2007. The effect of season and weather on physical activity: a systematic review. Public Health 121, 909–922.
von Weizsacker, C.C., 1984. The costs of substitution. Econometrica 52, 1085–1116.
Wooldridge, J.M., 2015. Control function methods in applied econometrics. J. Hum. Resour. 50 (2), 420–445.
Zivin, J.G., Neidell, M., 2014. Temperature and the allocation of time: implications for climate change. J. Labor Econ. 32 (1), 1–26.