TI 2024-028/V
Tinbergen Institute Discussion Paper

Skipping your workout, again? Measuring and understanding time inconsistency in physical activity

Diarmaid Ó Ceallaigh[1]
Kirsten I.M. Rohde[2]
Hans van Kippersluis[3]

1 Economic and Social Research Institute, Dublin

2 Erasmus University Rotterdam, Maastricht University and Tinbergen Institute

3 Erasmus University Rotterdam and Tinbergen Institute

Skipping your workout, again? Measuring and understanding time inconsistency in physical activity

Diarmaid Ó Ceallaigh, Kirsten I.M. Rohde, Hans van Kippersluis*

April 17, 2024

Abstract

Anecdotally, physical activity appears to be a textbook example of time inconsistency, which is the failure to follow through on ex-ante preferences and plans. Interestingly, our longitudinal survey finds that, over a fortnight, exercising more than preferred/planned is actually more prevalent than exercising less. However, over time a majority of our sample exercise less than preferred/planned in at least one of two consecutive fortnights. We find little evidence that time inconsistency is associated with present bias, its most popular explanation in economics. We find instead that it is associated with time-varying affective psychological processes such as willpower and temptations.

1 Introduction

According to the World Health Organization, obesity rates have almost tripled since 1975, with 13% of all adults being obese by 2016, and close to 40% of the

*DÓC (corresponding author - diarmaid.oceallaigh@esri.ie): Economic and Social Research Institute, Dublin. KR (kirsten.rohde@maastrichtuniversity.nl): Erasmus School of Economics (ESE) and Erasmus Research Institute of Management, Erasmus University Rotterdam (EUR); Tinbergen Institute (TI), the Netherlands; Maastricht University. HvK (hvankippersluis@ese.eur.nl): ESE and Erasmus Centre for Health Economics Research, EUR; TI. This project has received funding from the Erasmus Initiative "Smarter Choices for Better Health". The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen). The authors would like to thank Teresa Bago d'Uva, Aurelien Baillon, Han Bleichrodt, Lex Burdorf, John Cawley, Léontine Goldzahl, Dorte Gyrd-Hansen, Martin Hagger, Brendan Kennelly, Lili Kokai, Robin Lumsdaine, Owen O'Donnell, Heather Royer, Benjamin Teireck, Anne Wijtzes as well as attendees at the VERB online seminar, ESA world conference, Birkbeck seminar, Bristol University seminar, Hamburg Center for Health Economics, University of Galway CERIS workshop, and European Decision Sciences Day for helpful suggestions. Further they would like to thank the staff at Lifelines for suggestions and support, and Marthe van der Bijl, Ruby Doeleman, and Amber Stoll for excellent research assistance. This study was approved by the Medical Ethics Review Committee at University Medical Centre Groningen.

world population being overweight.[1] The rapid rise in obesity over the past few decades is one of the main public health concerns of our time. Weight gain often results from an imbalance of calories consumed and calories burned. Whereas there is debate on whether total caloric intake has increased or decreased over time (e.g., Cutler et al., 2003; Griffith et al., 2016; Currie et al., 2010), there is consensus that physical activity levels have decreased over time, as technological developments have led to a more sedentary lifestyle (Finkelstein and Zuckerman, 2008; Lakdawalla and Philipson, 2007). An estimated 28% of the global population now get insufficient levels of physical activity (Guthold et al., 2018), and it has been claimed that around 60% of weight gain can be attributed to declining levels of physical activity (Lakdawalla and Philipson, 2002).

Declining levels of physical activity are not necessarily a policy concern if the new equilibrium levels are the result of a rational and voluntary choice by society. However, there is ample evidence that physical activity is subject to involuntary failures of self-control, since individuals do not fully take into account the long-term costs of physical inactivity – leading to so-called "internalities" (Herrnstein et al., 1993). In economics, self-control failures are commonly inferred from time inconsistency, which is where an individual fails to follow through on their ex-ante preferences or plans (Strotz, 1955; Pollak, 1968). Anecdotally, physical activity would appear to be a textbook example of time inconsistency – just think of the swathes of New Year's resolutions that fail each year. Indeed, the evidence supports this – studies show that many people repeatedly overestimate their future gym attendance (Acland and Levy, 2015; Garon et al., 2015; Carrera et al., 2018, 2022), book exercise classes that they end up not attending (Habla and Muller, 2021), and sign up for gym memberships that they then rarely use (DellaVigna and Malmendier, 2006; Garon et al., 2015).

Given its importance in the obesity epidemic, and the ubiquity of good intentions, physical activity is a popular target for interventions. It seems fair to

[1] https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

say though that most interventions fail to meaningfully improve physical activity habits in the long-run, or even in the short-to-medium term after the intervention is withdrawn (Murray et al., 2017; Milkman et al., 2021; Marcus et al., 2022).[2] In short, it has been notoriously difficult to change physical activity habits sustainably. We took this as motivation to go back to the drawing board to try to get a better understanding of time inconsistency in physical activity, which could in turn help better inform the design of physical activity interventions.

In this paper, we (i) estimate the prevalence of time inconsistency in physical activity in a general population-based cohort; and (ii) perform a theory-based investigation of its empirical drivers. We do so using a longitudinal survey among over 3,000 respondents in *Lifelines*, a large population-based cohort from the north of the Netherlands (Scholtens et al., 2015). We use Choice-matching (Cvitanić et al., 2019) to incentivize honest answering.

We distinguish between two types of time inconsistency: time inconsistent (TI) preferences and TI planning. TI preferences exist when an individual fails to follow through on her ex-ante *preferences* (Strotz, 1955; Halevy, 2015). This can be thought of as failing to follow through on what, ex-ante, she would *like* her future self to do. TI planning occurs where an individual fails to follow through on her ex-ante *plans* (Strotz, 1955; Pollak, 1968), or what she ex-ante *predicts* her future self will do. Sophistication, which is the individual's awareness of her future self's self-control problems or TI preferences, determines the difference between TI preferences and planning. For example, an individual has a preference on Sunday for going to the gym the following morning, but ends up not going, meaning she has TI preferences. If she is sophisticated, she will correctly predict on Sunday that she won't go to the gym and so won't make a plan to go, meaning that she has time consistent planning. If she is not sophisticated (i.e., is naive) she will incorrectly predict on Sunday that she *will* go to the gym

[2] An exception from the economics literature is Royer et al. (2015), which finds treatment effects on gym attendance 2-3 years after an intervention combining financial incentives and a commitment device.

and will make a plan to go, leading to TI planning.

We use self-reported, yet incentivized measures for both of these types of time inconsistency. For TI preferences, we compare an individual's ex-ante *preferences* for physical activity for a two-week period to her actual physical activity for that period. For TI planning, we compare her ex-ante *plans* to her actual. We classify an individual as having time consistent preferences(planning) if her ex-ante preferences(plans) equal her actual, as having *under*-exercise TI preferences(planning) if her ex-ante-preferences(plans) exceeds her actual, and having *over*-exercise TI preferences(planning) if her preferences(plans) are below her actual. This approach to measuring time inconsistency has its roots in the method developed by Ameriks et al. (2007) to measure self-control problems, and has been used to measure time inconsistency in several other contexts (Wong, 2008; Mandel et al., 2017; Hoong, 2021; Cobb-Clark et al., 2021).

By measuring both TI preferences and planning in a general-population sample, we intend to comprehensively estimate the population prevalence of time inconsistency in physical activity. This constitutes our first main aim, and makes a novel contribution to the existing literature, which is limited to measuring TI planning of gym attendance in less general samples of students or gym members (DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Garon et al., 2015; Carrera et al., 2018, 2022; Habla and Muller, 2021). This is important for a number of reasons. Firstly, measuring only TI planning may mask underlying and arguably welfare-decreasing TI preferences if individuals are sophisticated.[3] Secondly, we go beyond exercising in the gym and also incorporate physical activity during regular daily activities (e.g., cycling to work, gardening). This is important given that approximately only 14% of the Dutch adult population are gym members (Centraal Bureau voor de Statistiek, 2022). Thirdly, our approach allows for detecting over-exercising, in contrast to some previous gym studies.

[3] In simple terms, if an individual acts upon her plans (e.g., she attends the two gym classes she signed up for), then existing TI planning-only approaches would classify this individual as time consistent. However, this individual may still experience a welfare loss if she would have *preferred* ex-ante to go to four gym classes, but being sophisticated about her self-control problems, only signed up for two classes.

Finally, we measure time inconsistency in two consecutive two-week periods in order to analyze its within-individual temporal stability, something which has not been examined in the literature to date.

Our second main aim is to estimate the relative importance of various possible explanations for time inconsistency suggested by theories from economics and psychology. The quasi-hyperbolic model is by far the most popular theoretical model of self-control problems in the economics literature (Delaney and Lades, 2017; Ericson and Laibson, 2019). In this model, time inconsistency derives from present bias (Phelps and Pollak, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999), which is generally treated as a "trait" or time invariant variable (Citanna and Siconolfi, 2022). However, the empirical importance of present bias in this regard has been questioned (Read et al., 2012; Delaney and Lades, 2017). Indeed, Halevy (2015) finds evidence to suggest that a large portion of TI preferences is driven by factors other than present bias. Aside from the study of Halevy (2015), little empirical evidence exists to support or reject the theoretical hypothesis that present bias drives time inconsistency, and thus one of our major contributions is in providing much-needed evidence in this regard.

In turn, we go beyond the quasi-hyperbolic model by also estimating the predictive power of a theory-guided list of alternative explanations for time inconsistency. Dual-self models posit that time inconsistency is driven by time varying factors related to automatic or affective psychological processes (i.e., system 1; Kahneman, 2011) such as willpower resources, stress, and temptation intensity (Loewenstein and O'Donoghue, 2004; Benhabib and Bisin, 2005; Brocas and Carrillo, 2008; Loewenstein et al., 2015). Further, changes in risk preferences, and low scores on the psychological concepts trait self-control and self-efficacy, have also been theoretically posited to lead to time inconsistency (Gerber and Rohde, 2010, 2018; Schwarzer, 2001; Hagger, 2014; Pfeffer and Strobach, 2017). While previous studies have provided evidence on the relationship with time inconsistency of projection bias (Acland and Levy, 2015;

Augenblick and Rabin, 2019) and limited attention (Habla and Muller, 2021), we contribute by providing the first empirical evidence on the link between time inconsistency and several additional explanations suggested by theories from the economics and psychology literature.

We find that, over the *study fortnight* (the two-week period which is the focus of our primary analysis), just under a half of our sample had over-exercise TI preferences(planning); that is to say they exercised *more* than their ex-ante preferences(plans). Over a third were under-exercisers (i.e., exercised *less* than their ex-ante preferences(plans)) and the rest were time consistent. The larger prevalence of over-exercisers relative to under-exercisers contrasts to previous studies of time inconsistency in physical activity which find under-exercise TI planning to be predominant (DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Garon et al., 2015; Carrera et al., 2018, 2022; Habla and Muller, 2021). This may reflect our more general sample and broader definition of physical activity.

Additionally analysing the two week period after the study fortnight, which we call the *post-study fortnight*, reveals a lot of within-individual temporal variation in time inconsistency. While time inconsistency patterns at the aggregate level remain broadly the same across both fortnights, a large number of individuals move between categories. This means that over half of our sample are under-exercisers in at least one of the two fortnights analysed. This in turn suggests that a focus by researchers and policymakers on addressing the problem of under-exercise time inconsistency is not misplaced, but that they should take account of this heterogeneity and temporal variation (Heffetz et al., 2022).

Both OLS and quantile regression analyses for the study fortnight show no evidence of a relationship between present bias and TI preferences/planning. When taken in conjunction with the findings of Halevy (2015), our evidence suggests that the almost exclusive focus in the literature to date on present bias as the driver of time inconsistency may have been misguided. Our analysis of other potential drivers provides further evidence to support the case for going beyond

a narrow focus on present bias. In particular, we find evidence to suggest that greater attention needs to be paid to time-varying affective process variables suggested by dual-self theories, namely willpower and temptation. Increased willpower(temptation) is significantly associated with a decrease(increase) in under-exercise time inconsistency in both the study and post-study fortnights. This suggests that the temporal variation in time inconsistency we find may at least in part be explained by the influence of these time-varying affective process variables.[4] In simple terms, time inconsistency is not a fixed personality trait, but temporal variations in willpower strength and temptation makes a good deal of the population susceptible to time-inconsistency from time to time.

Overall, our findings suggest that those wishing to build or estimate theoretical models of time inconsistency should take account of the potential for time inconsistency to vary over time, consider departing from the focus on the quasi-hyperbolic model (e.g., Garon et al., 2015; Acland and Levy, 2015; Carrera et al., 2022), and consider incorporating alternative time-varying drivers, such as temptation and willpower, in their models. Policymakers and researchers who wish to address time inconsistency in physical activity should also widen their focus and consider interventions that target such drivers (e.g., mindfulness exercises: Alem et al., 2021). This last point is particularly pertinent for economists, who, often inspired by the quasi-hyperbolic model, have focused predominantly on interventions which target rational deliberative processes by changing tangible incentives (e.g., through monetary incentives or commitment), rather than on interventions which target affective processes.

The rest of this paper is as follows: Sections 2 and 3 are on the theory and measurement of time inconsistency and its drivers; Section 4 describes the survey design, sample and incentives; Section 5 describes the empirical strategy; Section 6 describes the results; and in Section 7 we give our conclusions.

[4] In a similar vein, Meier (2022) finds changes in risk attitude are related to time-varying emotions.

2 Theory and measurement of time inconsistency

We consider a decision maker (DM) who is choosing between *outcome streams* $(t_1 : x_1, \ldots, t_m : x_m)$, which yield *outcome* x_i in *period* t_i and a neutral outcome 0 in all other periods. In every period τ the preferences of the DM are governed by the preferences of the DM's period τ self, $\succcurlyeq_\tau$, which are assumed to be a weak order. Strict preferences and indifference are defined as usual. The preferences of all selves of the DM are given by $\{\succcurlyeq_\tau\}_{\tau=0}^{\infty}$. Preferences $\{\succcurlyeq_\tau\}_{\tau=0}^{\infty}$ are *time consistent* if for all outcome streams $(s_1 : x_1, \ldots, s_m : x_m)$ and $(t_1 : y_1, \ldots, t_m : y_m)$, for all τ, and for all $\Delta > 0$ with $\tau + \Delta \leq s_1$ and $\tau + \Delta \leq t_1$ we have

$$(s_1 : x_1, \ldots, s_m : x_m) \succcurlyeq_\tau (t_1 : y_1, \ldots, t_m : y_m) \iff$$

$$(s_1 : x_1, \ldots, s_m : x_m) \succcurlyeq_{\tau+\Delta} (t_1 : y_1, \ldots, t_m : y_m)$$

Thus, time consistency in preferences holds when the individual's preference over two outcome streams is independent of when the outcome streams are evaluated (e.g., at τ or $\tau + \Delta$ in the above definition). Preferences are *time inconsistent* if they are not time consistent.

In the vast majority of the literature on intertemporal choice, outcomes are monetary outcomes or consumption. We will think of outcomes as bundles that specify monetary gains or losses, physical activity levels, and health gains or losses. Whenever we describe an outcome, we only describe the relevant component of the bundle and assume that all other components remain as usual. We assume that physical activity can have both costs (usually immediate) and benefits (usually in the future). Physical activity carried out for a given duration x in period s has a net cost in period s represented by $C(x)$, and has a future benefit realized in period $s_f > s$ represented by $B(x)$.[5] The outcome stream $(s : C(x), s_f : B(x))$ represents the bundle of immediate costs and future bene-

[5] $C(x)$ and $B(x)$ are the objective costs (e.g., energy cost) and benefits (e.g., increased life expectancy) of x minutes of physical activity, as perceived by the individual, before they are subjectively evaluated in an individual's utility function. We assume that health benefits accrue only in period s_f, but we have in mind that they may accrue in periods beyond s_f as well. Health benefits $B(x)$ thereby can be interpreted as the present certainty equivalent, at time s_f, of the health benefits accruing in periods s_f and after.

fits of carrying out physical activity for duration x in period s. We can thus say that an individual is time consistent in her physical activity preferences if, for every $x, y, s, s_f, t, t_f, \tau, \Delta$ with $s, t \geq \tau + \Delta > \tau$

$$(s : C(x), s_f : B(x)) \succcurlyeq_\tau (t : C(y), t_f : B(y))$$
$$\text{if and only if}$$
$$(s : C(x), s_f : B(x)) \succcurlyeq_{\tau+\Delta} (t : C(y), t_f : B(y))$$

As a running example throughout, consider Jane, who does not like to engage in physical activity (i.e., faces a short-term utility cost), but does so anyway from time-to-time as she is aware of the future health benefits (i.e., faces a longer-term utility benefit). In period τ, she prefers to do a 30 minute exercise class in period $s > \tau$ rather than not doing any exercise in period s:

$$(s : C(30), s_f : B(30)) \succ_\tau (s : C(0), s_f : B(0)).$$

However, when the time to do the class arrives, she prefers not to do it:

$$(s : C(30), s_f : B(30)) \prec_s (s : C(0), s_f : B(0)).$$

She clearly has time inconsistent preferences: in period s she reversed or didn't follow-through on her ex-ante period τ preferences. Intuitively, we can think of intertemporal decision-making as a game between multiple sequential selves (e.g., an ex-ante self and a future self). Time inconsistent preferences arise when the preferences of a self in a particular period over a given choice differ from the preferences of a self in a later period for that same choice.

Time inconsistent *preferences* may, but need not, result in time inconsistent *planning*. Time inconsistent planning is observed when a person fails to follow through on ex-ante plans. Decision makers who are fully aware of their time inconsistent preferences can still make time consistent plans, i.e., plans that they know they will follow through on in the future. Whether someone exhibits time consistent planning or not thus depends on the degree to which she can accurately predict her future preferences (Strotz, 1955; Pollak, 1968; Gruber and Kőszegi,

2004).[6] Predictions in period τ' about preferences that will prevail in period τ are denoted by $\widehat{\succcurlyeq}_{\tau|\tau'}$. *Predictions* about preferences are *time consistent* if for all τ, τ' we have $\widehat{\succcurlyeq}_{\tau|\tau'} = \succcurlyeq_\tau$. A *time (in)consistent planner* is a person with time (in)consistent predictions about future preferences.

Suppose Jane predicts in period τ that at time s she will be consistent with her time τ preferences and choose to carry out the exercise class such that:

$$(s : C(30), s_f : B(30)) \,\widehat{\succ}_{s|\tau}\, (s : C(0), s_f : B(0)).$$

As we saw above, this prediction is incorrect and she in fact chooses to skip the class in period s, and therefore she exhibits time inconsistent planning. However, if she had correctly predicted the reversal in her preferences then she would exhibit *time consistent planning*, despite having *time inconsistent preferences*.

2.1 Measuring time inconsistency

We measured time inconsistent (TI) preferences and TI planning in physical activity using elicitations of an individual's ex-ante *ideal* and *predicted* physical activity level for a two-week period, which we called the *study fortnight*, along with an ex-post self-report of her *actual* physical activity level for that period. By physical activity, we mean moderate to vigorous physical activity, which are the categories of physical activity recommended by the World Health Organization (WHO) in its physical activity guidelines (Bull et al., 2020). This was clearly explained to participants, and captures physical activity hidden in daily routines such as cycling to work. Ideal and predicted physical activity were measured in a questionnaire completed by participants at the beginning of the study fortnight, *questionnaire 1*, and actual physical activity was measured in a follow-up

[6] When Strotz (1955) and Pollak (1968) refer to "planning", they mean the intertemporal consumption plan or path that an economic agent chooses at a given period t for all consumption from period t onwards. She chooses this path to maximize her overall utility, subject to constraints, one of which is the behavior of her future self. She will not choose a plan that she believes her future self will not implement, as such a plan is infeasible. Thus, the individual's prediction of her own future behavior and her plan are identical.

questionnaire, *questionnaire 2*, at the end of the fortnight.[7] While participants were told in questionnaire 1 that they would be asked to complete questionnaire 2 in two weeks time, they were not told that they would be asked to self-report in questionnaire 2 their actual physical activity for the study fortnight.

To measure ex-ante preferences, we elicited an individual's ideal physical activity level, in hours, for the coming study fortnight, following Ameriks et al. (2007).[8] We measured actual behavior by asking individuals at the end of the fortnight how many hours of physical activity they had completed during the fortnight. TI preferences were then calculated as the proportion of an individual's ideal level of physical activity that she failed to follow through on in her actual behavior:

$$\text{TI preferences} = \frac{\text{Ideal - Actual}}{\text{Ideal}}$$

While this relative measure of TI preferences served as our main measure for analysis, we also analyzed the numerator (Ideal - Actual) for descriptive purposes. Participants with a value of zero for TI preferences were classified as having time consistent preferences. Participants with positive values were classified as having *under-exercise* TI preferences, since they exercised less than ex-ante preferred. A negative value signified *over-exercise* TI preferences.[9]

TI *planning* exists when an individual's actual behavior is not consistent with her ex-ante plan or *prediction*. The predicted physical activity of each participant for the study fortnight was elicited using one of two methods that have been used in previous literature. Two thirds of participants (randomly selected) received the *own prediction* question, which asked the participant to predict, ex-ante, how many hours of physical activity she would do during the study fortnight. This

[7] Each of these measures was elicited using two multiple choice questions – one eliciting total hours of physical activity in the first week of the study fortnight and another eliciting the total hours in the second week. Respondents could choose from the following options for each week: Less than 1 hour, 1 hour, 2 hours,..., 20 hours, More than 20 hours. The full texts of the questionnaires seen by participants are provided in Appendix H.

[8] The question directed participants to state their ideal physical activity if they were free of self-control problems, but still subject to other constraints such as time, physical ability, etc. The rationale for this wording is explained in section 3.1 below.

[9] Note that even though our terminology refers to exercising for linguistic purposes, the physical activity we measured was any kind of moderate to vigorous physical activity.

question was incentivized using Choice-matching (Cvitanić et al., 2019). The other third of participants received the *prediction for a similar other* question. This question asked the participant to predict how many hours of physical activity another participant, similar to herself in terms of ideal and actual physical activity, would do during the study fortnight. This prediction was then used as a proxy for her prediction of her own behavior. This method, developed by Toussaert (2018), has the advantage relative to the *own prediction* method that we could offer monetary rewards to a participant based on the accuracy of her predictions without running the risk of distorting subsequent physical activity behavior.[10] TI planning was calculated as the proportion of an individual's predicted level of physical activity that she failed to follow through on:

$$\text{TI planning} = \frac{\text{Predicted - Actual}}{\text{Predicted}}$$

We also analyzed the numerator for descriptive purposes. As for our measure of TI preferences, we classified participants as having time consistent, under-exercise TI or over-exercise TI planning. For robustness analysis, we also measured TI preferences and planning in the fortnight following the *study fortnight* (i.e., the *post-study fortnight* between questionnaire 2 and a third questionnaire).

3 Drivers of time inconsistency

3.1 Categorization of drivers

The factors that can drive time inconsistent preferences can be categorized as either *exogenous* or *endogenous* drivers. We use the term exogenous drivers to describe those drivers that originate outside the individual's psyche (i.e., exogenous shocks). Examples include unanticipated limitations on the future self's physical resources (e.g., injury, physical illness) and unanticipated time constraints.

[10] The random allocation of participants to one of these two methods of eliciting predictions of future behavior was for the purpose of another experimental study on incentives carried out in parallel to this study and using the same sample, but which is not described in this paper. The results regarding the prevalence and drivers of time inconsistency described in the current paper (described in Sections 6.1 and 6.2) do not differ significantly by incentive and prediction type (see Appendix D). We thus pooled data of those who received the own prediction question and those who received the prediction for a similar other question in our analysis.

On the other hand, we define endogenous drivers as those drivers that do originate within the individual's psyche (e.g., cognitive biases, limited motivation, and impulses affecting the future self's decision-making). These endogenous drivers can be thought of as self-control issues affecting the future self.

Following the framework of time inconsistency set out by Strotz (1955), we assume that an ex-ante preference is what the ex-ante self would like her future self to do, if the future self's decision-making were not restricted by self-control failures (i.e., if anticipated endogenous drivers are ignored). An ex-ante preference therefore ignores unanticipated exogenous or endogenous drivers, but takes into account all anticipated exogenous restrictions. In line with this, we phrased the question eliciting an individual's ideal physical activity so as to elicit an "ideal" that took into account anticipated exogenous constraints, but ignored anticipated *endogenous* constraints. Consequently, an individual will have time inconsistent preferences if there are any endogenous drivers affecting the future self and/or if the ex-ante self incorrectly anticipates exogenous drivers. This is summarized in the first two rows of Table 1.

The ex-ante predictions used to determine time inconsistent planning differ from ex-ante preferences in that they take into account all anticipated drivers, and so additionally take into account anticipated endogenous drivers (Strotz, 1955; Laibson, 1997; O'Donoghue and Rabin, 1999, 2001). The individual makes use of all available information to make as accurate a prediction as possible. Thus, time inconsistent planning can only arise due to unanticipated drivers, be they endogenous or exogenous. See the second two rows of Table 1.

In Table 1 we see that anticipated endogenous drivers are the difference between ex-ante preferences and predictions, and consequently between time inconsistent preferences and time inconsistent planning. Sophistication, which is often defined as the ex-ante self's awareness of her future self's self-control problems (Strotz, 1955; Pollak, 1968; O'Donoghue and Rabin, 1999, 2001), in our context is operationalized as the awareness of future endogenous drivers.

As an example, say that Jane anticipates at the weekend that she won't be able to play tennis the following Monday due to bad weather. She will reduce her ex-ante preferred and predicted physical activity for the coming week to take account of this anticipated exogenous driver, and thus it won't lead to TI preferences or planning. However, if she had not checked the weather this would have been an unanticipated exogenous driver, meaning that she would not have reduced either her ex-ante preference or prediction, and so the bad weather would have led to TI preferences and planning. She also has an ex-ante preference for going for a run on Wednesday morning, but anticipates self-control issues preventing her from getting up in time to do so (anticipated endogenous driver). While she still includes this run in her ex-ante preference for the week, she anticipates the self-control issue and reduces her ex-ante prediction accordingly. Her sophistication prevents the TI preferences from resulting in TI planning.

Table 1: The categories of drivers that can lead to each of TI preferences and TI planning.

	Anticipated Drivers		Unanticipated Drivers	
	Exogenous	Endogenous	Exogenous	Endogenous
TI preferences				
Taken account of in ex-ante preferences?	Yes	No	No	No
Lead to TI preferences?	No	Yes	Yes	Yes
TI planning				
Taken account of in ex-ante prediction?	Yes	Yes	No	No
Lead to TI planning?	No	No	Yes	Yes

3.2 Endogenous drivers: Theory, hypotheses and measures

Present bias: Present bias is the most popular theoretical explanation for time inconsistency among economists (Delaney and Lades, 2017; Ericson and Laibson, 2019). Present bias implies that immediate outcomes are overweighted relative to future outcomes and is captured in the quasi-hyperbolic model (Phelps and Pollak, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999). It is generally

treated as a trait variable (Citanna and Siconolfi, 2022).[11] This present bias can drive TI preferences. Say Jane has an ex-ante preference for doing 30 minutes of physical activity the following Tuesday. If she is present-biased, she may fail to follow through on that preference when Tuesday comes along by doing less than 30 minutes (i.e., she will have under-exercise TI preferences). This is because her present bias leads her to weigh the costs much more heavily when she evaluates the trade-off on Tuesday (when the costs are immediate) compared to when she evaluates it ex-ante (when the costs are in the future). The stronger her present bias, the more her ex-ante preference will exceed her actual, and the larger will be her under-exercise TI preferences.[12]

Present bias can also lead to *TI planning* if the individual is not fully sophisticated about her present bias (i.e., she doesn't accurately predict her future self's present bias). For example, if Jane inaccurately predicts ex-ante that she will not be present biased on Tuesday, and she has under(over)-exercise TI preferences arising from her present bias, then she will also have under(over)-exercise TI planning. The severity of her under(over)-exercise TI planning will be increasing in the strength of her present bias. Given this, we formulated hypotheses H1a and H1b that we could empirically test:

H1a: For under-exercisers, TI preferences and planning are increasing in present bias.

H1b For over-exercisers, TI preferences and planning are decreasing in present bias (i.e., stronger present bias leads to an increase in the severity, or absolute value, of negative (over-exercise) TI preferences/planning.)

We used the Decreasing Impatience (DI) Index method to measure present bias, where an individual completes choice lists from which her DI index, a measure of present bias, can be calculated without making any assumptions on

[11] Exceptions include Ahn et al. (2019), Citanna and Siconolfi (2022) and Duflo et al. (2011).

[12] The opposite will be the case for Jane if she perceives physical activity as having a short-term net benefit but a long-term net cost. For example, she may really enjoy playing tennis all evening, but that would mean having no time for studying for an exam. In this case, present bias may lead to *over*-exercise TI preferences (i.e., TI preferences < 0 due to actual exceeding ideal). The severity, or absolute value, of her over-exercise TI preferences will be increasing in her present bias.

the utility function (Rohde, 2019). An individual's present bias is increasing in the value of her DI index. We measured present bias both in the monetary and physical activity domains at the beginning of the study fortnight.[13]

Affective processes in dual-self models: Several alternative explanations for time inconsistency have been suggested in the economics and psychology literatures. Dual-self models posit that time-varying variables which influence automatic or affective psychological processes, such as willpower resources, stress, and temptation intensity, are all drivers of time inconsistency (Loewenstein and O'Donoghue, 2004; Benhabib and Bisin, 2005; Brocas and Carrillo, 2008; Loewenstein et al., 2015). In such models, decision-making is determined both by these affective processes, which are impulsive and myopic, and deliberative processes, which are rational and far-sighted. Due to the myopia of the affective processes, they only influence decisions with immediate consequences. Deliberative processes influence decisions both when there are immediate and future consequences. Thus, ex-ante preferences for future physical activity are determined by the deliberative processes, but preferences for immediate physical activity are determined by both sets of processes. While deliberative processes will want to stick to the ex-ante preferences, affective processes may want to deviate from this in favor of short-term gratification. Willpower and stress levels determine the deliberative processes' ability to exert self-control and resist a given desire of affective processes to deviate. The level of temptation provided by alternative options determines the size of the utility benefit to the affective processes from deviating, and thus the affective processes' desire to deviate.

Say that Jane has an ex-ante preference for doing 30 minutes of physical activity on Tuesday, as determined by her deliberative processes. When Tuesday comes, the affective processes will want to deviate and not do any physical activ-

[13] The text of the choice lists used to measure present bias, as well as the text of questions used to measure the other endogenous drivers of time inconsistency, can be seen in the full text of the questionnaires provided in Appendix H. The choice lists elicited preferences between a sooner smaller and later larger outcome when the later larger outcome occurred with a delay of between 0 and 52 weeks. For participants who never switched between the outcomes between 0 and 52 weeks, their switch point was assumed to be 53 weeks.

ity, as they only take account of the short-term costs and not the long-term benefits of exercise. The difficulty faced by Jane's deliberative processes in resisting the affective processes' desire to deviate will be decreasing in Jane's willpower reserves, and increasing in her stress and in the temptation she faces to skip physical activity (e.g., to nap or watch TV). The greater this difficulty in resisting, the greater the influence of affective processes on her decision, and consequently the greater will be her under-exercise TI preferences and planning.[14]

We thus can formulate the following hypotheses:

H2a: For under-exercisers, TI preferences and planning are decreasing in willpower and increasing in stress and in temptation to skip physical activity.

H2b: For over-exercisers, TI preferences and planning are increasing in willpower, decreasing in stress and increasing in temptation to skip physical activity.

We measured willpower, stress and temptation to skip physical activity with self-reported Likert scales at the end of the study fortnight (Karvounides et al., 2016).

Change in risk preferences: The quasi-hyperbolic model assumes that the main source of time inconsistency lies in the weights people give to utilities at different points in time, as reflected by the discount function. The dual-self model identifies another source of time inconsistency, which is the utilities people expect to experience at given points in time. When future utility is mispredicted, this can lead to time inconsistency. In the dual-self model, the source of such mispredictions of utility lie in mispredictions of temptation intensity.[15]

Mispredictions of risk preferences are another source of mispredictions of future utility that can drive time inconsistency (Gerber and Rohde, 2018). Such mispredictions can arise when risk preferences change as the delay shortens between the time period in which the risk is evaluated and that in which the risk occurs. Say Jane, in choosing her ex-ante preference and prediction for Tuesday

[14] Conversely, if for Jane physical activity has short-term benefits and long-term costs, then her affective processes will want to do more than 30 minutes of physical activity. In this case, the lower her willpower reserves and the higher her stress on Tuesday, the more severe will be her over-exercise TI preferences and planning.

[15] The projection bias model of Loewenstein et al. (2003) also identifies mispredictions of future utility as a source of time inconsistency.

physical activity, predicts that she will evaluate Tuesday physical activity with the same utility function on Tuesday as she does ex-ante. However, when Tuesday comes, her risk preferences change, and thus her utility function for Tuesday physical activity changes. If it changes sufficiently, she will deviate from her ex-ante preferences and plans, and will have TI preferences and planning.

We measured if the risk preferences of participants were changing in this manner using a hypothetical certainty equivalent (CE) task with a 50/50 gamble for €300 that was due to resolve at the end of the study fortnight. The CE task was completed by participants both at the beginning and end of the study fortnight. Each certainty equivalent was converted into a normalized risk premium (NRP).[16] Change in risk preferences was then calculated by subtracting the first NRP from the second. A positive(negative) value of change in risk preferences signifies that the individual became more risk averse(seeking) as the time delay to the resolution of the risk became shorter. If we assume that such a change in risk preferences is at least partly mispredicted, then we can formulate the following hypothesis that we can empirically test:

H3: A change in risk preferences as the delay shortens between the time period in which the risk is evaluated and the time period in which the risk occurs is a driver of TI preferences and planning.

Trait self-control and self-efficacy: The psychology literature has much to say about the intention-behavior gap, which is where there is a difference between ex-ante intentions and subsequent actual behavior (Sheeran and Webb, 2016), and is therefore closely related to TI preferences and planning. Trait self-control may help the individual to engage the volitional processes necessary to bridge the intention-behavior gap (Hagger and Chatzisarantis, 2014; Hagger, 2014), and may help to resist impulses that prevent an individual from following through on their physical activity intentions (Pfeffer and Strobach, 2017).

Self-efficacy is an individual's "beliefs in their capabilities to produce de-

[16] Normalised risk premium = $\frac{\text{Expected value} - \text{Certainty equivalent}}{\text{Expected value}}$

sired effects by their actions" (Bandura, 1997, page vii), and may also play an important role in the intention-behavior gap. Self-efficacy is necessary in order to develop self-control capabilities (Bandura, 1997) which are crucial to enact the volitional processes needed to turn intentions into action (Schwarzer, 2001).

Trait self-control and self-efficacy are conceptually related to several of the parameters and functions of the economic models we discussed, but the exact relations have not yet been established in the literature. Given that the psychology literature suggests that both trait-self control and self-efficacy reduce the gap between intentions and actual behavior, we assume that a higher value for these variables leads to a decrease in the gap between ex-ante preferences/predictions and actual behavior, and thus a decrease in the absolute value of TI preferences/planning. Thus, we hypothesize that:

H4a: For under-exercisers, TI preferences and planning are decreasing in trait self-control and self-efficacy.

H4b: For over-exercisers, TI preferences and planning are increasing in trait self-control and self-efficacy.

Trait self-control and self-efficacy were measured with self-reported Likert scale measures at the beginning of the study fortnight (Tangney et al., 2004; Morean et al., 2014; Teeuw et al., 1994; Schwarzer and Jerusalem, 1995).

4 Survey design, sample and incentives

Participants in this study were drawn from *Lifelines*, which is a general population cohort study based in the north of the Netherlands.[17] The Lifelines cohort is broadly representative of the general population in that region (Klijs et al., 2015). For this study we used the Lifelines cohort to carry out an additional lon-

[17] Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. See more details on the Lifelines cohort study design at this link: `https://www.lifelines.nl/researcher/data-and-biobank`

gitudinal data collection called *LifeSTYLE*. All eligible cohort members (approximately 85,000 members) were invited to participate.[18] Participants completed three online questionnaires, with a two week gap between each questionnaire.

Table 2: Descriptive statistics for the prevalence sample and the full Lifelines cohort

	Prevalence sample		Lifelines cohort		Difference	
	N	Mean	N	Mean	Mean	P-value
Age	4,333	59.6	156,855	51.8	7.8	0.000
Female	4,333	0.57	150,173	0.59	-0.02	0.011
Ethnicity	4,131	0.01	121,137	0.02	-0.01	0.000
More than highschool	4,254	0.45	146,908	0.32	0.12	0.000
Not in employment	4,305	0.39	148,578	0.25	0.14	0.000
Children in household	4,327	0.34	156,546	0.60	-0.27	0.000
Partner in household	4,274	0.83	132,238	0.80	0.03	0.000
Divorced	4,267	0.16	145,785	0.14	0.02	0.000
Widow(er)	4,057	0.04	116,567	0.03	0.01	0.000
Physical Activity mins/week	4,004	427	129,971	498	-71	0.000

P-values obtained from two-sample t-tests of the equality of means. *Physical Activity mins/week* is average minutes per week of moderate to vigorous physical activity and was measured using the *Short questionnaire to assess health-enhancing physical activity* (Wendel-Vos et al., 2003) in the first wave of data collection carried out by the Lifelines organization between 2007 and 2014 (note that all Lifelines cohort members participated in this first wave). Aside from this variable and age, all other variables are binary.

Table 2 shows descriptive statistics for the *prevalence sample* and the overall Lifelines cohort. The prevalence sample consists of the 4,333 participants who completed the questions in questionnaires 1 and 2 necessary to observe TI preferences and planning for the study period and so are included in our analyses of the prevalence of time inconsistency.[19] The descriptives are largely similar for the *regression sample* of 3,055 participants, which is the sample for whom we also observe the eight potential drivers of time inconsistency analyzed and so are included in our regression analyses (see Appendix Table A1). Both the preva-

[18] The LifeSTYLE data collection was approved by the Medical Ethics Review Committee at University Medical Centre Groningen. Participants had to be at least 18 years old and had to not have been invited to participate in a pilot version of LifeSTYLE carried out in late 2019.

[19] The individuals who opted to participate in the LifeSTYLE data collection were randomized with equal probability to one of two surveys – the survey which formed the basis for our study or another survey which formed the basis for studies being run by researchers at Erasmus Medical Centre. Thus, 4,333 represents approx. 10% of invitees to this study.

lence and regression samples were older, more educated, and less likely to be in employment or to have children in their household than the overall Lifelines cohort. Both samples did less moderate to vigorous physical activity per week prior to this study than the Lifelines cohort. Interestingly, the mean minutes per week in each sample and in the Lifelines cohort exceeded 400 and thereby also exceeded the WHO guideline of at least 150 minutes per week (Bull et al., 2020).

The differences between our samples and the Lifelines cohort are indicative of some selection among respondents to our survey. However, our samples are still arguably more representative than samples of gym members and students used in previous studies measuring time inconsistency in physical activity (DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Garon et al., 2015; Carrera et al., 2018, 2022; Habla and Muller, 2021). Furthermore, to check the impact of this selection on mean time inconsistency levels, we compared the unadjusted means to means adjusted for nonresponse using a logit regression-based response propensity method (Little and Rubin, 2019).

The elicitations of TI preferences and planning, as well as of several of the driver variables, were incentivized monetarily using Choice-matching (Cvitanić et al., 2019), a method for eliciting honest responses to non-verifiable survey questions.[20,21] Choice matching is an extension of the Bayesian Truth Serum (Prelec, 2004), which has been shown to be effective in inducing truth-telling (John et al., 2012; Weaver and Prelec, 2013; Frank et al., 2017) and reducing biases in responses (Weaver and Prelec, 2013; Baillon et al., 2022). To the best of our knowledge, our study is the first empirical application of Choice-matching.

[20] Elicitations where the independent variable being measured and the measurement method are primarily associated with the economics literature (i.e., present bias and risk preferences) were incentivized, but other elicitations where the variable and method are drawn from the psychology literature, and with which incentives are not usually used, were not.

[21] One third of participants (randomly selected) received no incentives so that they could act as a control group against which the responses of the other two thirds of participants, who all received Choice-matching, could be compared. This randomization to different incentive scheme groups was for the purpose of another experimental study on incentives being carried out in parallel to this study and using the same sample, but which is not described in this paper. The results regarding the prevalence and drivers of time inconsistency described in the current paper (described in Sections 6.1 and 6.2) do not differ significantly between incentive types (see Appendix D). We thus pooled data of those who received and didn't receive Choice-matching in our analysis.

5 Analysis strategy

The analysis plan was pre-registered on the Open Science Framework website.[22] To get insight into what endogenous factors drive time inconsistency, we used both OLS and quantile regression to estimate the correlation of each of TI preferences and TI planning with the eight possible endogenous drivers outlined in Section 3. We estimated the following OLS model:

$$y_i = \alpha + \mathbf{X}'_{1,i}\beta_1 + \mathbf{X}'_{2,i}\beta_2 + \epsilon_i$$

y_i was the dependent variable (the relative measure of TI preferences or TI planning) for individual i. $X_{1,i}$ was a vector of our eight possible drivers of time inconsistency: present bias (both in the monetary and physical activity domains), willpower, stress, temptation, change in risk preferences, trait self-control, and self-efficacy. These driver variables were standardized for our analyses. $X_{2,i}$ was a vector of controls.[23] To complement the OLS analysis, we then estimated the following quantile regression model (Koenker and Bassett Jr, 1978):

$$\text{Quant}_\theta(y_i) = \alpha_\theta + \mathbf{X}'_{1,i}\beta_{1,\theta} + \mathbf{X}'_{2,i}\beta_{2,\theta} + \epsilon_{i,\theta}$$

where $\theta \in (0, 100)$ denotes which quantile of the outcome variable is being analyzed. In our analysis, we estimated this model for each $\theta \in \{10, 20, \ldots, 90\}$. Our primary motivation for using quantile regression was that it allowed us to analyze if the relationship between time inconsistency and its potential drivers was heterogeneous across the distribution of time inconsistency. This was particularly important given that we hypothesize the relationship between TI preferences/planning and several of the possible drivers we analyze to be non-monotonic (e.g., we hypothesized that TI preferences/planning would be increasing in stress for under-exercisers, but decreasing in stress for over-exercisers). Such non-

[22] See `https://osf.io/ty9sx`. As recommended by Banerjee et al. (2020), a list of departures from the pre-analysis plan in our final analysis is included in Appendix G.

[23] The control variables were: all variables included in Table 2, except for physical activity minutes per week; self-reported Likert scale measures of whether the person likes PA or not and dispositional optimism (Scheier et al., 1994); self-reported restricted ability to do PA due to medical reasons; dummy variables for responses to the present bias choice lists indicating preferring more physical activity to less, and having no switch point; dummy variables for giving upper or lower bound answers in responses to the risk preference certainty equivalence tasks; dummy variables for the incentive type a participant got. See Appendix I for further details.

monotonicity would be masked in regular OLS.

We used the False Discovery Rate method (Benjamini and Hochberg, 1995; Benjamini et al., 2006) to adjust p-values for multiple hypothesis testing, given that we ran 160 different tests in our primary analysis.[24] We also ran OLS and quantile regressions using the component variables of the time inconsistency measures – ideal, predicted and actual physical activity – as the dependent variables. This was to get additional insight as to the channels through which the driver variables were related to TI preferences/planning.

6 Results

6.1 Prevalence of time inconsistency

Prevalence: On average, individuals in our *prevalence sample* of 4,333 respondents had over-exercise time inconsistency, doing 43(48)%, or 1.8(2.3) hours, more physical activity over the study fortnight than their ex-ante preferences(plans).[25] When we adjust these means for selection bias due to survey non-response (using a logit regression-based response propensity method), we still find mean over-exercise time inconsistency, but a little less severe (37(42)% more than ex-ante preferred(planned)). 45%(48%) of participants had over-exercise TI preferences(planning), while 36(34)% had under-exercise TI preferences(planning). 19(18)% had time *consistent* preferences(planning).[26]

In the two weeks after the study fortnight (the *post-study fortnight* between questionnaires 2 and 3), data for the smaller sample of non-attritors (3,652) shows that the descriptive statistics and spread across categories remains broadly

[24] The 160 tests comprised of 16 OLS tests (8 independent variables for each of TI preferences and planning) and 144 quantile regression tests (8 independent variables for each of TI preferences and planning at 9 deciles).

[25] See Appendix Table A2 for these and other descriptive statistics. The descriptives for the *regression sample* of 3,055 are very similar – see Appendix Table A4. See Appendix Figures A1 to A4 for histograms of the distibutions of TI preferences and planning, as well as of Ideal, Predicted and Actual physical activity. Here we see that a considerable proportion of those categorized as having time consistent preferences(planning) (i.e., with actual = ideal(predicted)) had ideal(predicted) and actual at the upper bound of the response scale (i.e., "more than 20 hours per week"). Some of these could have been inconsistents that had actual and ideal(predicted) that were above 20 hours per week but not equal to each other, and so our estimate of the number of time consistents in our sample represents an upper bound.

[26] The breakdown for the *regression sample* of 3,055 is almost identical: 19(18)% had time consistent preferences(planning), 44(47)% were over-exercisers while 37(35)% were under-exercisers.

similar at the aggregate level.[27] However, at the individual level there is a lot of movement between categories: 51(57)% of those who were under-exercisers in the study fortnight became over-exercisers in the post-study fortnight, while 40(39)% of those who were over-exercisers became under-exercisers. Noteworthy is that 58(56)% of the sample had under-exercise TI preferences(planning) in at least one of the two fortnights. This suggests that within-individual variation over time may be an important feature of time inconsistency.

Comparison with the literature: The substantial fraction of over-exercise time inconsistents is surprising when considering previous studies of gym attendance that find under-exercise TI planning to be the dominant form (DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Garon et al., 2015; Carrera et al., 2018, 2022; Habla and Muller, 2021). A number of factors may explain some of this difference. First, studies using gym membership (DellaVigna and Malmendier, 2006; Garon et al., 2015) or gym class bookings (Habla and Muller, 2021) as proxies for an individual's predictions of their future gym attendance to calculate TI planning can only capture under-exercise TI planning, and not over-exercising. Second, we measure TI preferences and planning in a much broader domain, moderate-to-vigorous physical activity. This incorporates physical activity carried out for practical purposes, such as transport, employment and home maintenance, for which over-exercise TI preferences and planning may be more probable (e.g., working overtime, unplanned trip to the grocery store).

Third, the data collection occurred in the first half of 2021, during the COVID-19 pandemic. As the Netherlands was in lockdown, with shops, restaurants and bars closed, there were few tempting alternatives to engaging in physical activity, which potentially increased the fraction of over-exercisers. We ran a pilot in 2019 before the COVID pandemic, and while the pilot sample also over-exercised on average, our main sample described in this paper had mean over-exercise time inconsistency that was 36% higher in absolute terms than the mean

[27] See Appendix Tables A5 and A6.

for the pre-COVID pilot sample (when compared using age-matched data). The difference was driven by our main sample having higher actual physical activity than the pre-COVID sample. Fourth, our sample differs quite considerably from the samples of gym members and students used in previous studies, where mean age ranged from 22 to 35 (compared to 60 in our study). If older people are more likely to have over-exercise TI preferences/planning, then this could at least partly explain our findings.[28] Finally, participants may have exercised more than usual over the study fortnight due to Hawthorne effects. However, this is unlikely to be an important explanation, as mean actual physical activity for the study fortnight was only 2% higher than what participants reported to be their mean fortnightly physical activity over the previous year.

Characterizing time inconsistent individuals: Table 3 shows how mean outcomes and characteristics differ across the three categories of TI preferences.[29] Those with under-exercise TI preferences failed to follow through on 35(27)%, or 9.9(8.2) hours, of their preferences(plans) on average, whereas the over-exercisers exceeded their preferences(plans) by 125(124)%, or 12.1(11.5) hours, on average. The difference in time inconsistency between the under- and over-exercisers derived from two channels: under-exercisers had both (i) higher ex-ante preferences/plans and (ii) lower actual physical activity. Additionally, individuals with time consistent preferences had higher ex-ante preferences and plans, and higher actual physical activity, than either of the other two groups.

Mean actual physical activity for the study fortnight was 26 hours for the full sample, and 18, 32 and 30 hours for the under-exercisers, time consistents and over-exercisers, respectively. These means are well above the WHO guideline minimum of 2.5 hours/week, and 95% of the full sample exceed this minimum.

[28] In the regression analyses described in the next section (Section 6.2), the coefficient estimates for the association between age (control variable) and TI preferences/planning (dependent variable) are negative in both OLS and quantile regressions. The age-TI preferences relation is significant in OLS, and both the age-TI preferences and age-TI planning relations are significant in quantile regressions for those with severe levels of over-exercise TI preferences/planning.

[29] The patterns we see across the three categories of TI preferences in this table are almost identical to those across the three categories of TI planning, which can be seen in Appendix Table A7.

Table 3: Outcomes and characteristics by TI preferences subcategory

	Mean				P-value		
	Full sample	Under-exercise	Time consist.	Over-exercise	UE= TC	UE= OE	OE= TC
TI preferences							
- Relative	-0.43	0.35	0.00	-1.25	0.00	0.00	0.00
- Absolute	-1.82	9.89	0.00	-12.09	0.00	0.00	0.00
TI planning							
- Relative	-0.48	0.27	-0.13	-1.24	0.00	0.00	0.00
- Absolute	-2.31	8.20	-0.81	-11.49	0.00	0.00	0.00
Ideal	24.21	28.02	32.34	17.68	0.00	0.00	0.00
Predicted	23.71	26.33	31.53	18.29	0.00	0.00	0.00
Actual	26.03	18.13	32.34	29.77	0.00	0.00	0.00
Likes Phys. Activ.	0.82	0.81	0.86	0.82	0.00	0.24	0.02
Age	59.6	58.1	60.9	60.2	0.00	0.00	0.16
Female	0.57	0.59	0.54	0.57	0.02	0.32	0.13
$>$ Highschool	0.45	0.49	0.44	0.42	0.02	0.00	0.38
N	4333	1575	819	1939			

Notes: The number of observations is 4,333, except for $>$*Highschool*, where there are 79 missing values. P-values were obtained from two-sample t-tests of the equality of means. *Likes Phys. Activ.* is a binary indicator equal to 1 if the participant stated in response to a qualitative survey question that they like physical activity.

The broad measure of physical activity and the Dutch sample used may play a role here, especially given that in the Netherlands over a third report cycling or walking to work, over twice the proportion doing so in the UK or the USA (Hallal et al., 2012). Self-report bias may also play a role, though we attempted to mitigate this with the use of Choice-matching. The patterns in Table 3 are largely repeated in the data for the post-study fortnight (see Appendix Table A8), despite the large proportion of participants moving from one time inconsistency category to another between the study and post-study fortnights noted earlier.

It is noteworthy how similar the results are for TI preferences and planning at the aggregate level. Indeed, for 58% of the sample, TI preferences and planning were equal. As discussed in Section 3.1, any difference between TI preferences and planning should be due to anticipated self-control problems, so when there is no difference this means that the individual does not anticipate any self-

control problems. A quarter of the individuals for whom TI preferences and planning were equal were time consistent, and so this anticipation of no self-control problems was correct. For the other three quarters who were not time consistent, there are two possibilities: (1) this anticipation was correct and they have no self-control problems, and so their TI preferences and planning were driven purely by exogenous drivers; or (2) this anticipation was incorrect and thus indicative of a lack of sophistication about their self-control problems. To try and tease out which of these explanations is likely to be more prominent, we analyzed a qualitative question in which we asked participants to rate the importance of a number of possible reasons for their time inconsistency. This analysis showed that, in general, those with under-exercise time inconsistency rate exogenous drivers (e.g., weather) as more important than endogenous drivers (e.g., willpower) (see Appendix Figures A7 and A8). However, the importance of exogenous drivers relative to endogenous drivers was no different between those under-exercisers for whom TI preferences and planning were equal and the rest of the under-exercisers.[30]

6.2 Regression results

Present bias: In the OLS analyses (Table 4), we see that the coefficient estimates for present bias are all negative but not significant.[31] Figures 1 and 2 show the results of the quantile regression analyses. Of the 9 deciles analysed, participants at the lowest deciles from 10-40 were over-exercise time inconsistents, while those at the highest deciles (70-90) were under-exercisers, with time consistents lying in between. The boundaries between these categories are denoted by vertical dotted lines in Figures 1 and 2. We see no significant associations

[30] See Appendix J for further detailed discussion of this concordance between TI planning and preferences. Analysis excluding those with very short response times suggests that the large proportion of our sample with TI preferences equal to TI planning is not driven by insufficient participant attention in responding.

[31] For descriptive statistics of the present bias variables, as well as the other independent variables analysed, see Appendix Table A9.

at any deciles for present bias.[32] We also analyzed the relationship between the components used to calculate the TI preferences/planning variables – ideal, predicted and actual – and the eight drivers. Appendix Table A11 shows that in OLS regressions, a 1 standard deviation (SD) increase in present bias is significantly associated with a reduction in ideal and predicted physical activity of about half an hour per fortnight. Present bias is negatively but not significantly associated with actual physical activity. In Appendix Figures A9 to A11, we see quantile regressions results in line with these OLS results. These results are reassuring that our elicitation of present bias, even if noisy, clearly reflects a genuine signal as it correlates as expected with ideal, predicted and actual.

Willpower: OLS regressions find no significant relation between willpower and time inconsistency, but quantile regressions show that a 1 SD increase in willpower is significantly associated with a 3-5pp (0.02-0.03 standard deviations (SD)) reduction in TI preferences and planning for under-exercisers, in line with our hypothesis for under-exercisers. OLS shows willpower to be significantly associated with increased predicted and actual physical activity, with a 1 SD increase in willpower associated with a 0.6(1.2) hour increase in predicted(actual). Quantile regressions show similar results. This suggests that underexercisers with more willpower may tend to display less time inconsistency because they tend to have higher actual physical activity.

Stress: Neither OLS nor quantile regressions find a significant relation between stress and time inconsistency. In OLS we see no relation between stress and the component variables ideal, predicted and actual, but in quantile regressions we do see some evidence for a positive relation between stress and both ideal and predicted at the lower ends of their respective distributions, and a negative relation with actual at the higher end of its distribution.

Temptation: While we see no significant relation in OLS, we do see that temp-

[32] See Appendix Table A10 and Appendix Figures A5 and A6 for the OLS and quantile regressions run with standardized TI preferences and planning variables.

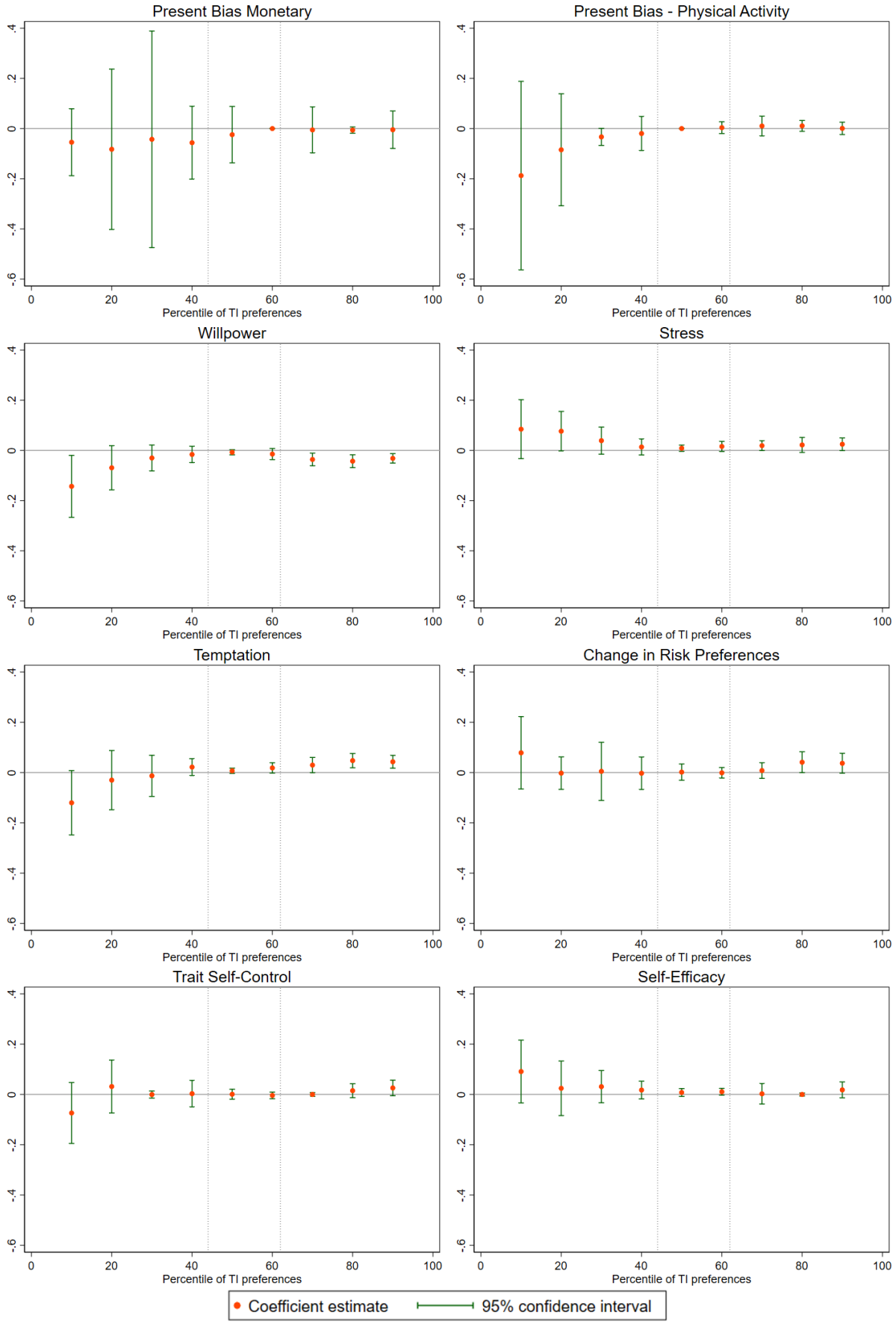
Table 4: OLS regression of TI preferences/planning on eight potential endogenous drivers

	TI preferences	TI planning
Present Bias - Monetary	-0.032	-0.005
	(0.023) [0.368]	(0.019) [0.908]
Present Bias - Physical Activity	-0.023	-0.019
	(0.036) [0.725]	(0.028) [0.686]
Willpower	-0.047	-0.025
	(0.040) [0.447]	(0.047) [0.756]
Stress	0.046	0.090
	(0.033) [0.345]	(0.038) [0.103]
Temptation	0.004	-0.058
	(0.053) [0.946]	(0.052) [0.479]
Change in risk preferences	0.061	0.009
	(0.045) [0.368]	(0.050) [0.935]
Trait self-control	0.031	0.028
	(0.041) [0.672]	(0.036) [0.66]
Self-efficacy	0.000	0.043
	(0.047) [0.946]	(0.062) [0.686]
Controls	✓	✓
N	3055	3055
R^2	0.033	0.027
Dependent variable mean(SD)	-0.406(1.670)	-0.460(1.882)

The *regression sample* of 3,055 participants in this analysis is all participants for whom we observe values for TI preferences, TI planning, and these eight potential drivers of time inconsistency. Robust standard errors in parentheses. False Discovery Rate adjusted p-values (adjusted for 160 tests) in square brackets.

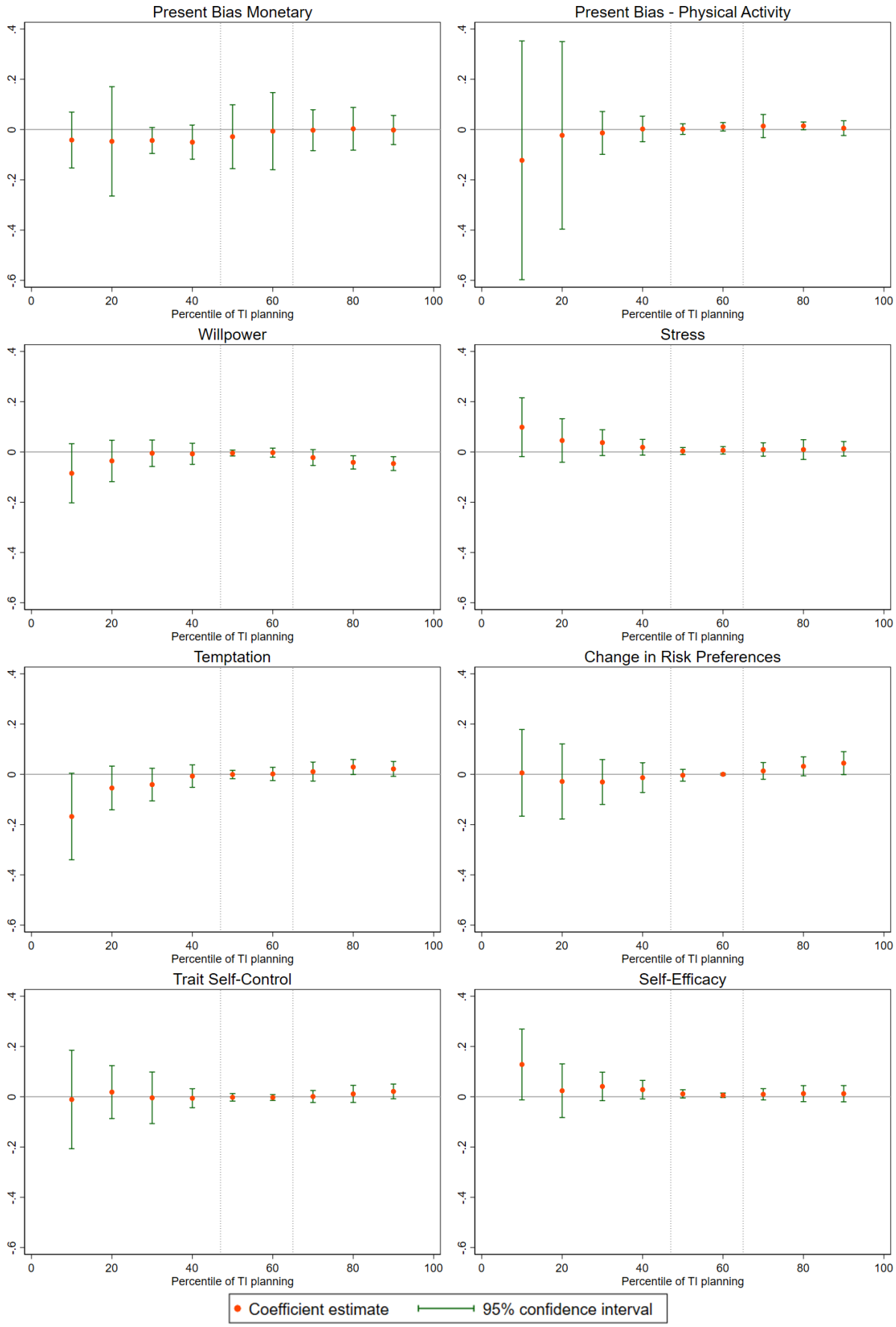
tation is positively related to TI preferences for under-exercisers. Coefficients range from 4 to 5pp (0.02-0.03 SD) per 1 SD increase in temptation, and are in line with our hypothesis for under-exercisers. OLS shows that a 1 SD increase in temptation is significantly associated with a decrease in ideal, predicted and actual of 0.8, 1.1 and 1.2 hours respectively. These results are reflected in quantile regressions. This suggests that temptation leads to under-exercise time inconsistency mostly through lower levels of actual physical activity, given that lower ideal and predicted push under-exercise time inconsistency up rather than down.

Figure 1: Quantile regressions of TI preferences on potential drivers



Notes: Graphs show results of quantile regressions run at each decile $\theta \in \{10, 20, \ldots, 90\}$ of the distribution of TI preferences. Sample is the same as for the OLS regressions (see Table 4). Over-exercisers lie to the left of the first vertical dotted line, time consistents lie between the first and second dotted lines, and under-exercisers lie to the right of the second dotted line. Confidence intervals are calculated to adjust for MHT using FDR-adjusted p-values (adjusted for 160 tests). The dependent variable mean(SD) is -0.41(1.67).

Figure 2: Quantile regressions of TI planning on potential drivers



Notes: Graphs show results of quantile regressions run at each decile $\theta \in \{10, 20 \ldots, 90\}$ of the distribution of TI planning. The dependent variable mean(SD) is -0.46(1.88). Otherwise, notes are as per the notes to Figure 1.

Other variables: We see no significant relations between either change in risk preferences, trait self-control or self-efficacy and TI preferences or planning. OLS and quantile regressions show self-efficacy to be positively related to both ideal and predicted.

As a robustness check, we ran the regressions with data from the post-study fortnight (N=2,097), and the results were broadly the same as those for the study fortnight, notably in terms of willpower and temptation being significantly associated with TI preferences and planning for most under-exerciser deciles (see Appendix C). This finding is particularly interesting given the within-individual variation we find in TI preferences and planning between the two fortnights, and supports an interpretation of time inconsistency varying within-individual due to the influence of time-varying drivers such as temptation and willpower.

As a further robustness check, we checked that our regression results were robust to the incentive type received by participants (Appendix D). We also ran regressions with different combinations of the driver variables (e.g., including the present bias variables in the regression but not the other six drivers – Appendix E). Finally, we ran regressions using different combinations of control variables (Appendix F). In all cases our results turned out to be robust.

We additionally asked under-exercisers to rate the importance, as perceived by themselves, of a number of possible drivers in determining their own TI preferences and planning. The three most important reasons, according to participants, were exogenous drivers: weather, physical constraints and time constraints. Temptation and willpower were the next most important reasons, which aligns with our quantile regression findings.[33]

7 Conclusion

The problem of time inconsistency in physical activity has proven until now to be extremely difficult to solve. This motivated us to go back to the drawing board

[33] See Appendix Figures A7 and A8 for further detail.

to try to gain additional insight into this problem. In a Dutch general population cohort, we first measured the prevalence of both TI preferences and planning. Almost half of the total sample had over-exercise TI preferences(planning) (i.e., they exercised *more* than their ex-ante preferences/plans), compared to just over a third who were under-exercisers (i.e., they exercised *less*). The relatively high prevalence of over-exercisers in our sample contrasts to previous studies and may in part reflect the broader measure of physical activity we used (as opposed to just gym attendance) and our more general sample (as opposed to just gym members or students), and thus highlights the novelty and importance of our contribution to the existing research. Time inconsistency is also shown to vary quite considerably within-individual over time, with large movements between categories from one two-week period to the next. This means that over half of our sample are under-exercisers in at least one of the two fortnights analysed.

Our second main aim was to gain insight into what drives time inconsistency in physical activity. Our regression analyses, both OLS and quantile, give very little evidence for a link between present bias, the most popular explanation for time inconsistency and generally treated as time invariant, and TI preferences/planning. When taken in combination with the findings of Halevy (2015), who finds evidence to suggest that a large portion of TI preferences is driven by factors other than present bias, our results suggest that the narrow focus on present bias in much research to date may have been erroneous.

The case for looking beyond present bias is strengthened by our findings on alternative drivers of present bias. We find reasonably strong evidence for a link between under-exercise TI preferences/planning and time-varying variables that influence affective processes, namely willpower and temptation. Furthermore, we see that the relationship between these variables and time inconsistency holds in a robustness analysis carried out on a subsequent two-week period. This suggests that the temporal variation in time inconsistency we find may at least in part be explained by the influence of these time-varying affective process vari-

ables. Although correlational, our findings make an important contribution to the scant existing empirical evidence on the drivers of time inconsistency.

Our study has some important limitations which should be kept in mind when interpreting our results. First, most of our data is based on self-reports, stated preferences and hypothetical scenarios, although we try to mitigate this limitation by making our measures incentive-compatible using Choice-matching. Second, our data collection occurred when strict COVID-19 restrictions were in place, which may have external validity implications. Third, there is evidence of some selection in our sample, although adjusting mean TI preferences and planning for this selection does not make a meaningful difference.

As for policy implications, our findings suggest that under-exercise time inconsistency measured over a single time period may not be as widespread a problem in the general population as existing gym studies suggest. Having said that, with over a third of our sample experiencing under-exercise time inconsistency in any given two-week period, and over half experiencing it in at least one of the two-week periods examined, the prevalence of *under*-exercise time inconsistency is arguably still high enough to warrant a focus by researchers and policymakers on interventions to tackle such time inconsistency. Our analysis of the drivers of time inconsistency suggests that, in designing such interventions as well as in theory development, a narrow focus on present bias should be avoided by also taking account of time-varying drivers of time inconsistency. In particular, factors which influence affective processes, as suggested by dual-self theories, should be given closer attention.

Our findings also have a number of other important implications for future research. Given the high levels of over-exercise time inconsistency we find, future empirical research should ensure that the measure of time inconsistency used can capture both under- and over-exercisers, not just the former. Secondly, our findings provide motivation for future research to examine whether the correlational patterns we find between time inconsistency and various drivers can

be causally verified. Thirdly, the temporal variation in time inconsistency we find warrants further attention. Fourthly, it would be interesting to analyze if time inconsistency varies over different time horizons. Finally, the relatively high physical activity levels we find using a broad measure of physical activity highlights the importance of being very clear about the domains of physical activity that are relevant in physical activity measurement, target-setting and intervention design.

References

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

Ahn, D. S., Iijima, R., Le Yaouanq, Y., and Sarver, T. (2019). Behavioural characterizations of naiveté for time-inconsistent preferences. *The Review of Economic Studies*, 86(6):2319–2355.

Alem, Y., Behrendt, H., Belot, M., and Bíró, A. (2021). Mind training, stress and behaviour — a randomised experiment. *PloS One*, 16(11):e0258172.

Ameriks, J., Caplin, A., Leahy, J., and Tyler, T. (2007). Measuring self-control problems. *American Economic Review*, 97(3):966–972.

Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86(3):941–975.

Baillon, A., Bleichrodt, H., and Granic, G. D. (2022). Incentives in surveys. *Journal of Economic Psychology*, page 102552.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman, New York.

Banerjee, A., Duflo, E., Finkelstein, A., Katz, L. F., Olken, B. A., and Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics. *National Bureau of Economic Research*.

Benhabib, J. and Bisin, A. (2005). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions. *Games and Economic Behavior*, 52(2):460–492.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Brocas, I. and Carrillo, J. D. (2008). The brain as a hierarchical organization. *American Economic Review*, 98(4):1312–1346.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R., et al. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British Journal of Sports Medicine*, 54(24):1451–1462.

Carrera, M., Royer, H., Stehr, M., and Sydnor, J. (2018). Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. *Journal of Health Economics*, 58:202–214.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2022). Who chooses commitment? Evidence and welfare implications. *Review of Economic Studies*, 89(3):1205–1244.

Centraal Bureau voor de Statistiek (2022). *Trends in the Netherlands 2018*. `https://longreads.cbs.nl/trends18-eng/society/figures/leisure/` [Accessed: 19 October 2022].

Citanna, A. and Siconolfi, P. (2022). Naïve stochastic present bias and credit. *Working paper*.

Cobb-Clark, D. A., Dahmann, S. C., Kamhöfer, D., and Schildberg-Hörisch, H. (2021). Sophistication about self-control. *Life Course Centre Working Paper*.

Currie, J., DellaVigna, S., Moretti, E., and Pathania, V. (2010). The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy*, 2(3):32–63.

Cutler, D. M., Glaeser, E. L., and Shapiro, J. M. (2003). Why have Americans become more obese? *Journal of Economic Perspectives*, 17(3):93–118.

Cvitanić, J., Prelec, D., Riley, B., and Tereick, B. (2019). Honesty via choice-matching. *American Economic Review: Insights*, 1(2):179–92.

De Witt Huberts, J. C., Evers, C., and De Ridder, D. T. (2014). "Because I am worth it": A theoretical framework and empirical review of a justification-based account of self-regulation failure. *Personality and Social Psychology Review*, 18(2):119–138.

Delaney, L. and Lades, L. K. (2017). Present bias and everyday self-control failures: a day reconstruction study. *Journal of Behavioral Decision Making*, 30(5):1157–1167.

DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, 96(3):694–719.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American Economic Review*, 101(6):2350–90.

Ericson, K. M. (2017). On the interaction of memory and procrastination: Im-

plications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association*, 15(3):692–719.

Ericson, K. M. and Laibson, D. (2019). Intertemporal choice. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 1–67. Elsevier.

Finkelstein, E. A. and Zuckerman, L. (2008). *The fattening of America: How the economy makes us fat, if it matters, and what to do about it*. John Wiley & Sons.

Frank, M. R., Cebrian, M., Pickard, G., and Rahwan, I. (2017). Validating Bayesian Truth Serum in large-scale online human experiments. *PloS One*, 12(5):e0177385.

Garon, J.-D., Masse, A., and Michaud, P.-C. (2015). Health club attendance, expectations and self-control. *Journal of Economic Behavior & Organization*, 119:364–374.

Gerber, A. and Rohde, K. I. M. (2010). Risk and preference reversals in intertemporal choice. *Journal of Economic Behavior & Organization*, 76(3):654–668.

Gerber, A. and Rohde, K. I. M. (2018). Weighted temporal utility. *Economic Theory*, 66:187–212.

Griffith, R., Lluberas, R., and Lührmann, M. (2016). Gluttony and sloth? Calories, labor market activity and the rise of obesity. *Journal of the European Economic Association*, 14(6):1253–1286.

Gruber, J. and Kőszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics*, 88(9-10):1959–1987.

Guthold, R., Stevens, G. A., Riley, L. M., and Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global Health*, 6(10):e1077–e1086.

Habla, W. and Muller, P. (2021). Experimental evidence of limited attention at the gym. *Experimental Economics*, pages 1–29.

Hagger, M. S. (2014). The multiple pathways by which trait self-control predicts health behavior. *Annals of Behavioral Medicine*, 48(2):282–283.

Hagger, M. S. and Chatzisarantis, N. L. (2014). An integrated behavior change model for physical activity. *Exercise and Sport Sciences Reviews*, 42(2):62–69.

Halevy, Y. (2015). Time consistency: Stationarity and time invariance. *Econometrica*, 83(1):335–352.

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., Lancet Physical Activity Series Working Group, et al. (2012). Global

physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet*, 380(9838):247–257.

Heffetz, O., O'Donoghue, T., and Schneider, H. S. (2022). Reminders work, but for whom? evidence from new york city parking ticket recipients. *American Economic Journal: Economic Policy*, 14(4):343–370.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., and Vaughan Jr, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, 6(3):149–185.

Hoong, R. (2021). Self control and smartphone use: An experimental study of soft commitment devices. *European Economic Review*, 140:103924.

John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5):524–532.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Karvounides, D., Simpson, P. M., Davies, W. H., Khan, K. A., Weisman, S. J., and Hainsworth, K. R. (2016). Three studies supporting the initial validation of the stress numerical rating scale-11 (stress NRS-11): A single item measure of momentary stress for adolescents and adults. *Pediatric Dimensions*, 1(4):105–109.

Klijs, B., Scholtens, S., Mandemakers, J. J., Snieder, H., Stolk, R. P., and Smidt, N. (2015). Representativeness of the lifelines cohort study. *PloS One*, 10(9):e0137203.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

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

Lakdawalla, D. and Philipson, T. (2002). The growth of obesity and technological change: a theoretical and empirical examination. *National Bureau of Economic Research*, working paper 8946.

Lakdawalla, D. and Philipson, T. (2007). Labor supply and weight. *Journal of Human Resources*, 42(1):85–116.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Loewenstein, G. and O'Donoghue, T. (2004). Animal spirits: Affective and deliberative processes in economic behavior. *Working paper*.

Loewenstein, G., O'Donoghue, T., and Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, 2(2):55.

Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, 118(4):1209–1248.

Mandel, N., Scott, M. L., Kim, S., and Sinha, R. K. (2017). Strategies for improving self-control among naïve, sophisticated, and time-consistent consumers. *Journal of Economic Psychology*, 60:109–125.

Marcus, J., Siedler, T., and Ziebarth, N. R. (2022). The long-run effects of sports club vouchers for primary school children. *American Economic Journal: Economic Policy*, 14(3):128–165.

Meier, A. N. (2022). Emotions and risk attitudes. *American Economic Journal: Applied Economics*, 14(3):527–558.

Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., Beshears, J., et al. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, 600(7889):478–483.

Morean, M. E., DeMartini, K. S., Leeman, R. F., Pearlson, G. D., Anticevic, A., Krishnan-Sarin, S., Krystal, J. H., and O'Malley, S. S. (2014). Psychometrically improved, abbreviated versions of three classic measures of impulsivity and self-control. *Psychological Assessment*, 26(3):1003.

Murray, J. M., Brennan, S. F., French, D. P., Patterson, C. C., Kee, F., and Hunter, R. F. (2017). Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: a systematic review and meta-analysis. *Social Science & Medicine*, 192:125–133.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

O'Donoghue, T. and Rabin, M. (2001). Choice and procrastination. *Quarterly Journal of Economics*, 116(1):121–160.

Pfeffer, I. and Strobach, T. (2017). Executive functions, trait self-control, and the intention–behavior gap in physical activity behavior. *Journal of Sport and Exercise Psychology*, 39(4):277–292.

Phelps, E. S. and Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies*, 35(2):185–199.

Pollak, R. A. (1968). Consistent planning. *Review of Economic Studies*, 35(2):201–208.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466.

Read, D., Frederick, S., and Airoldi, M. (2012). Four days later in Cincinnati: Longitudinal tests of hyperbolic discounting. *Acta Psychologica*, 140(2):177–185.

Rohde, K. I. M. (2019). Measuring decreasing and increasing impatience. *Management Science*, 65(4):1700–1716.

Royer, H., Stehr, M., and Sydnor, J. (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.

Scheier, M. F., Carver, C. S., and Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. *Journal of Personality and Social Psychology*, 67(6):1063.

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., Van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., et al. (2015). Cohort profile: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, 44(4):1172–1180.

Schwarzer, R. (2001). Social-cognitive factors in changing health-related behaviors. *Current Directions in Psychological Science*, 10(2):47–51.

Schwarzer, R. and Jerusalem, M. (1995). Generalized self-efficacy scale. In *Measures in health psychology: A user's portfolio. Causal and control beliefs*, pages 35–37. Windsor, UK: NFER-NELSON.

Sheeran, P. and Webb, T. L. (2016). The intention–behavior gap. *Social and Personality Psychology Compass*, 10(9):503–518.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3):165–180.

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2):271–324.

Teeuw, B., Schwarzer, R., and Jerusalem, M. (1994). *Dutch adaptation of the General Self-Efficacy scale*. `http://userpage.fu-berlin.de/health/dutch.htm` [Accessed: 23 January 2023].

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–889.

Weaver, R. and Prelec, D. (2013). Creating truth-telling incentives with the Bayesian Truth Serum. *Journal of Marketing Research*, 50(3):289–302.

Wendel-Vos, G. W., Schuit, A. J., Saris, W. H., and Kromhout, D. (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *Journal of Clinical Epidemiology*, 56(12):1163–1169.

Wong, W.-K. (2008). How much time-inconsistency is there and does it matter? Evidence on self-awareness, size, and effects. *Journal of Economic Behavior & Organization*, 68(3-4):645–656.

FOR ONLINE PUBLICATION

Online Appendix A Additional tables and figures

Table A1: Descriptive statistics for the regression sample and the full Lifelines cohort

	Regression sample		Lifelines cohort			
	(1) Obs.	(2) Mean	(3) Obs.	(4) Mean	(5) (2)-(4)	(6) p-value
Age	3055	58.75	156855	51.81	6.94	0.000
Gender	3055	0.57	150173	0.59	-0.02	0.029
Ethnicity	2926	0.01	121137	0.02	-0.01	0.000
More than highschool	3004	0.49	146908	0.32	0.16	0.000
Not in employment	3036	0.36	148578	0.25	0.12	0.000
Children in household	3050	0.35	156546	0.60	-0.25	0.000
Partner in household	3010	0.82	132238	0.80	0.02	0.003
Divorced	3014	0.17	145785	0.14	0.03	0.000
Widow(er)	2864	0.05	116567	0.03	0.01	0.002
Phys. Act. mins/week	2839	416.07	129971	498.16	-82.09	0.000

This table shows the same statistics as in Table 2 in the main text, except that they are for the *regression sample* rather than the prevalence sample. The regression sample is all participants who completed the questions in questionnaires 1 and 2 necessary to observe values for their TI preferences and TI planning, and for whom we also observe values for the eight potential drivers of time inconsistency we analyze in our regression analysis.

Table A2: Summary statistics for TI preferences and planning (prevalence sample)

	Mean	Adjusted Mean	P_{25}	Median	P_{75}	σ	N
Relative							
- TI preferences	-0.43	-0.37	-0.50	0.00	0.20	1.65	4,333
- TI planning	-0.48	-0.42	-0.50	0.00	0.17	1.93	4,333
Absolute (hours)							
- TI preferences	-1.82	-1.14	-8.00	0.00	4.00	12.82	4,333
- TI planning	-2.31	-1.76	-8.00	0.00	3.00	12.57	4,333

The Adjusted Mean column shows means adjusted for nonresponse using a logit regression-based response propensity method (Little and Rubin, 2019). P_{25} and P_{75} refer to the 25th and 75th percentiles, respectively. σ refers to the standard deviation. Table A3 shows that the non-response adjustment used to calculate the adjusted means improved the similarity between the prevalence sample and the Lifelines cohort in terms of observable socio-demographics.

Table A3: Effect of nonresponse adjustment on sociodemographic variable means

	Prevalence sample			Lifelines cohort			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	N	Mean	Adj. Mean	N	Mean	(2)-(5)	(3)-(5)
Age	4,333	59.6	53.82	156,855	51.8	7.8	2.02
Female	4,333	0.57	0.61	150,173	0.59	-0.02	0.02
Ethnicity	4,131	0.01	0.02	121,137	0.02	-0.01	0.00
More than highschool	4,254	0.45	0.35	146,908	0.32	0.12	0.03
Not in employment	4,305	0.39	0.24	148,578	0.25	0.14	-0.01
Children in household	4,327	0.34	0.54	156,546	0.60	-0.27	-0.06
Partner in household	4,274	0.83	0.79	132,238	0.80	0.03	-0.01
Divorced	4,267	0.16	0.16	145,785	0.14	0.02	0.02
Widow(er)	4,057	0.04	0.04	116,567	0.03	0.01	0.01
Phys. Act. mins/week	4,004	427	491	129,971	498	-71	-7

Column 3 gives the adjusted means for the prevalence sample after a non-response adjustment (response propensity method using logit regression). Column 6 gives the difference between the Lifelines cohort means and the unadjusted prevalence sample means, while column 7 gives the differences between the Lifelines cohort and the adjusted prevalence sample means.

Table A4: Summary statistics for TI preferences and planning (regression sample)

	Mean	P_{25}	Median	P_{75}	Std. Dev.	Obs.
Relative						
- TI preferences	-0.41	-0.50	0.00	0.21	1.67	3055
- TI planning	-0.46	-0.50	0.00	0.17	1.88	3055
Absolute (hours)						
- TI preferences	-1.55	-8.00	0.00	4.00	12.81	3055
- TI planning	-2.16	-8.00	0.00	4.00	12.49	3055

This table shows the same statistics as Table A2, except that they are for the regression sample rather than the prevalence sample..

Table A5: Summary statistics for TI preferences and planning in the post-study fortnight

	Mean	P_{25}	Median	P_{75}	Std. Dev.	Obs.
Relative						
- TI preferences	-0.20	-0.20	0.00	0.13	1.48	3652
- TI planning	-0.20	-0.25	0.00	0.10	1.25	3652
Absolute (hours)						
- TI preferences	-0.42	-4.00	0.00	3.00	10.00	3652
- TI planning	-0.90	-4.00	0.00	2.00	9.83	3652

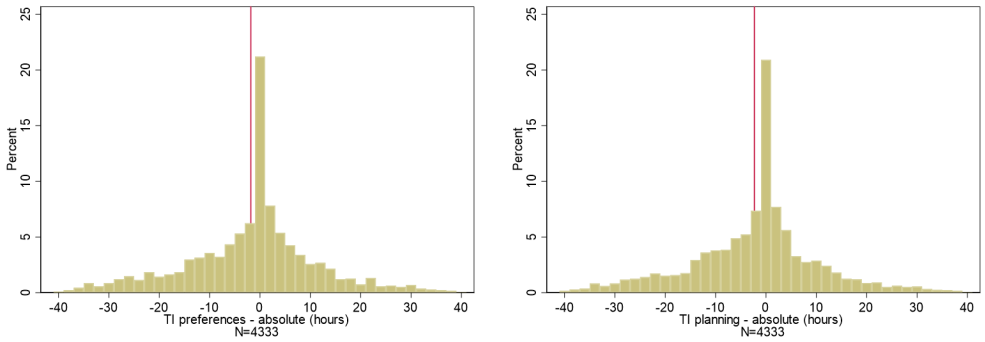
This table shows the same statistics as Table A2 except that they are for the post-study fortnight (i.e., the two weeks subsequent to the study fortnight) and are for the sample of 3,652 participants who had non-missing values for TI preferences and planning in both the study and post-study fortnights.

Table A6: Proportion of participants in each TI preferences and planning category in the study and post-study fortnight

	Post-study fortnight		
	Under-exercise	Time consistent	Over-exercise
TI preferences			
Study fortnight:			
Under-exercise	0.32	0.17	0.51
Time consistent	0.22	0.59	0.18
Over-exercise	0.40	0.26	0.34
All	0.34	0.29	0.37
TI planning			
Study fortnight:			
Under-exercise	0.27	0.16	0.57
Time consistent	0.19	0.60	0.21
Over-exercise	0.39	0.25	0.36
All	0.31	0.29	0.40

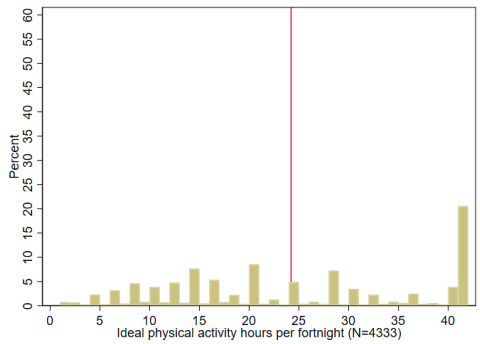
This table shows the proportion of participants in each TI preferences and planning category in the post-study fortnight (one category per column), separated by study fortnight categories (one category per row). Sample is 3652 participants for whom TI preferences and planning data were not missing for either the study or the post-study fortnight.

Figure A1: Distributions of TI preferences and TI planning

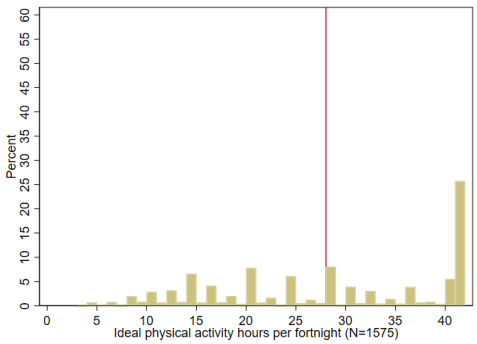


Notes: Red line in each graph shows the sample mean

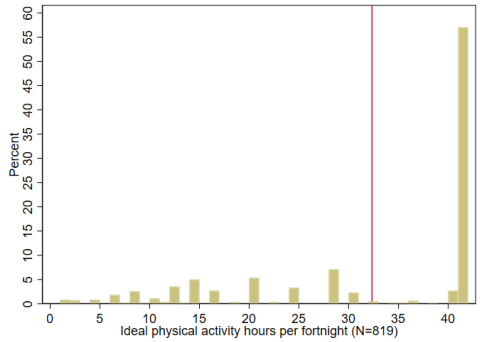
Figure A2: Distributions of Ideal physical activity



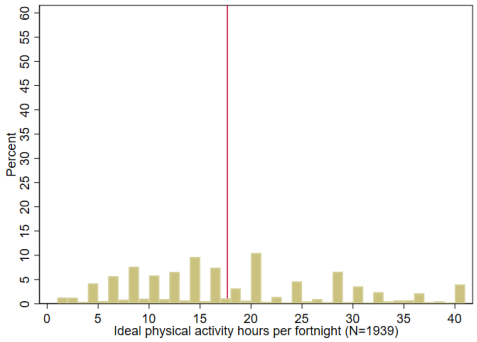
(a) Full sample



(b) Underexercise TI preferences sample



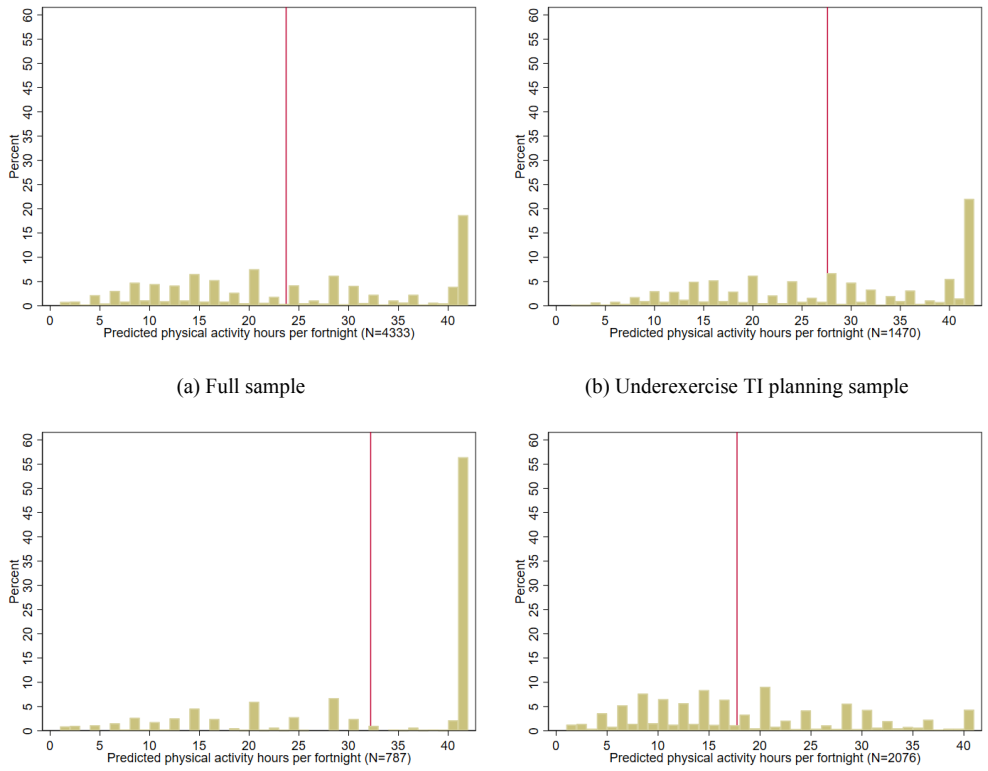
(c) Time consistent preferences sample



(d) Overexercise TI preferences sample

Notes: X-axis shows hours over the study fortnight. Red line shows the mean for each sample. Full sample is the prevalence sample of 4,333. Panel (c) shows that a considerable proportion of those categorized as having time consistent preferences (i.e., with actual = ideal) had ideal and actual at the upper bound of the response scale (i.e., "more than 20 hours per week"). Some of these could have been inconsistents that had actual and ideal that were above 20 hours per week but not equal to each other, in which case the number of time consistents in our sample would be overestimated.

Figure A3: Distributions of Predicted physical activity



(a) Full sample

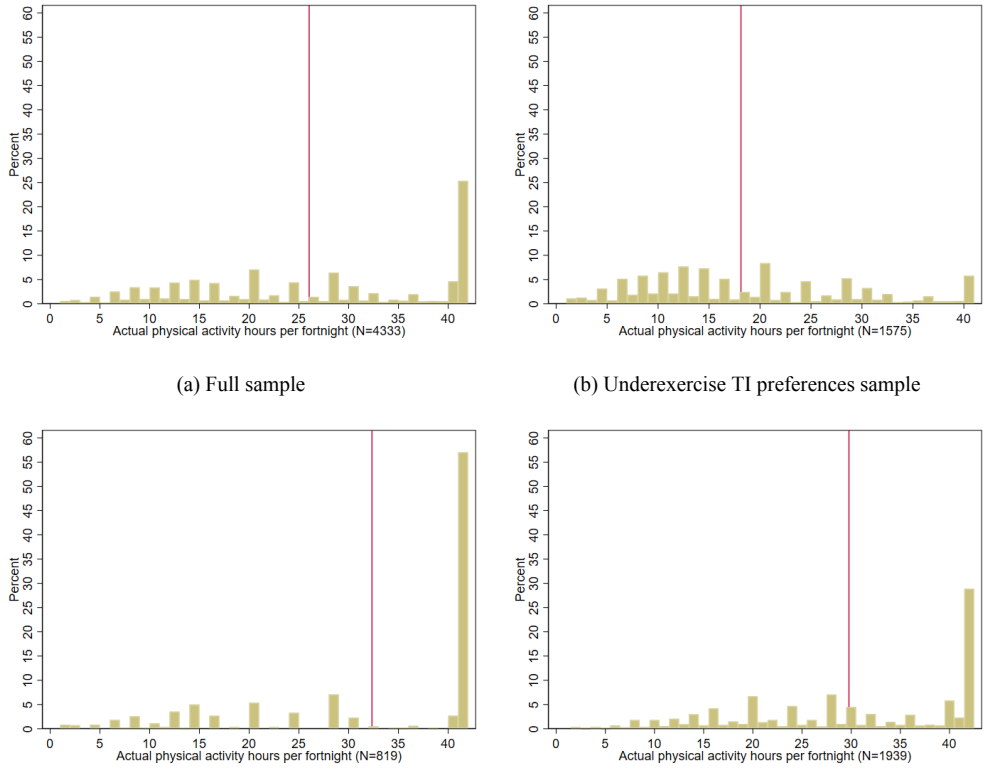
(b) Underexercise TI planning sample

(c) Time consistent planning sample

(d) Overexercise TI planning sample

Notes: X-axis shows hours over the study fortnight. Red line shows the mean for each sample. Full sample is the prevalence sample of 4,333. (C) shows that a considerable proportion of those categorized as having time consistent planning (i.e., with actual = predicted) had predicted and actual at the upper bound of the response scale (i.e., "more than 20 hours per week"). Some of these could have been inconsistents that had actual and predicted that were above 20 hours per week but not equal to each other, in which case the number of time consistents in our sample would be overestimated.

Figure A4: Distributions of Actual physical activity



(a) Full sample

(b) Underexercise TI preferences sample

(c) Time consistent preferences sample

(d) Overexercise TI preferences sample

Notes: X-axis shows hours over the study fortnight. Red line shows the mean for each sample. Full ample is the prevalence sample of 4,333.

Table A7: Outcomes and characteristics by TI planning subcategory

	Mean				P-value		
	(1) Full sample	(2) Under-exercise	(3) Time consistent	(4) Over-exercise	(5) UE= TC	(6) UE= OE	(7) OE= TC
TI preferences							
- Relative	-0.43	0.31	-0.11	-1.08	0.00	0.00	0.00
- Absolute	-1.82	9.61	-0.58	-10.37	0.00	0.00	0.00
TI planning							
- Relative	-0.48	0.35	0.00	-1.25	0.00	0.00	0.00
- Absolute	-2.31	9.65	0.00	-11.66	0.00	0.00	0.00
Ideal	24.21	27.56	31.63	19.03	0.00	0.00	0.00
Predicted	23.71	27.60	32.21	17.74	0.00	0.00	0.00
Actual	26.03	17.95	32.21	29.40	0.00	0.00	0.00
Likes Phys. Act.	0.82	0.81	0.86	0.82	0.01	0.43	0.02
Age	59.6	58.8	61.3	59.50	0.00	0.11	0.00
Gender	0.57	0.57	0.52	0.58	0.02	0.53	0.00
>Highschool	0.45	0.47	0.45	0.43	0.44	0.04	0.41
N	4333	1470	787	2076			

Notes: The statistics in this table are the same as those shown in Table 3 in the main text except that they provide means by TI planning subcategory rather than TI preferences subcategory.

Table A8: Outcomes and characteristics by post-study fortnight TI preferences subcategory

	Mean				P-value		
	Full sample	Under-exercise	Time consist.	Over-exercise	UE= TC	UE= OE	OE= TC
TI preferences							
- Relative	-0.20	0.32	0.00	-0.83	0.00	0.00	0.00
- Absolute	-0.42	8.63	0.00	-9.06	0.00	0.00	0.00
TI planning							
- Relative	-0.20	0.25	-0.04	-0.74	0.00	0.00	0.00
- Absolute	-0.90	7.23	-0.58	-8.62	0.00	0.00	0.00
Ideal	26.42	27.54	32.85	20.24	0.00	0.00	0.00
Predicted	25.94	26.13	32.27	20.69	0.00	0.00	0.00
Actual	26.84	18.91	32.85	29.30	0.00	0.00	0.00
Likes Phys. Activ.	0.82	0.79	0.85	0.83	0.00	0.00	0.20
Age	60.01	59.03	61.21	59.96	0.00	0.06	0.01
Gender	0.57	0.59	0.55	0.56	0.05	0.08	0.79
N	3652	1233	1076	1343			

Notes: The statistics in this table are the same as those shown in Table 3 in the main text except that they show means by *post*-study fortnight TI preferences subcategory, rather than study fortnight TI preferences subcategory.

Table A9: Summary statistics for the eight endogenous drivers

	Mean	Std. Dev.
Present Bias - Monetary	0.123	2.487
Present Bias - Physical Activity	0.765	6.492
Willpower	5.409	1.436
Stress	2.333	2.374
Temptation	2.508	1.576
Change in risk preferences	0.052	0.510
Trait self-control	3.192	0.415
Self-efficacy	3.428	0.550

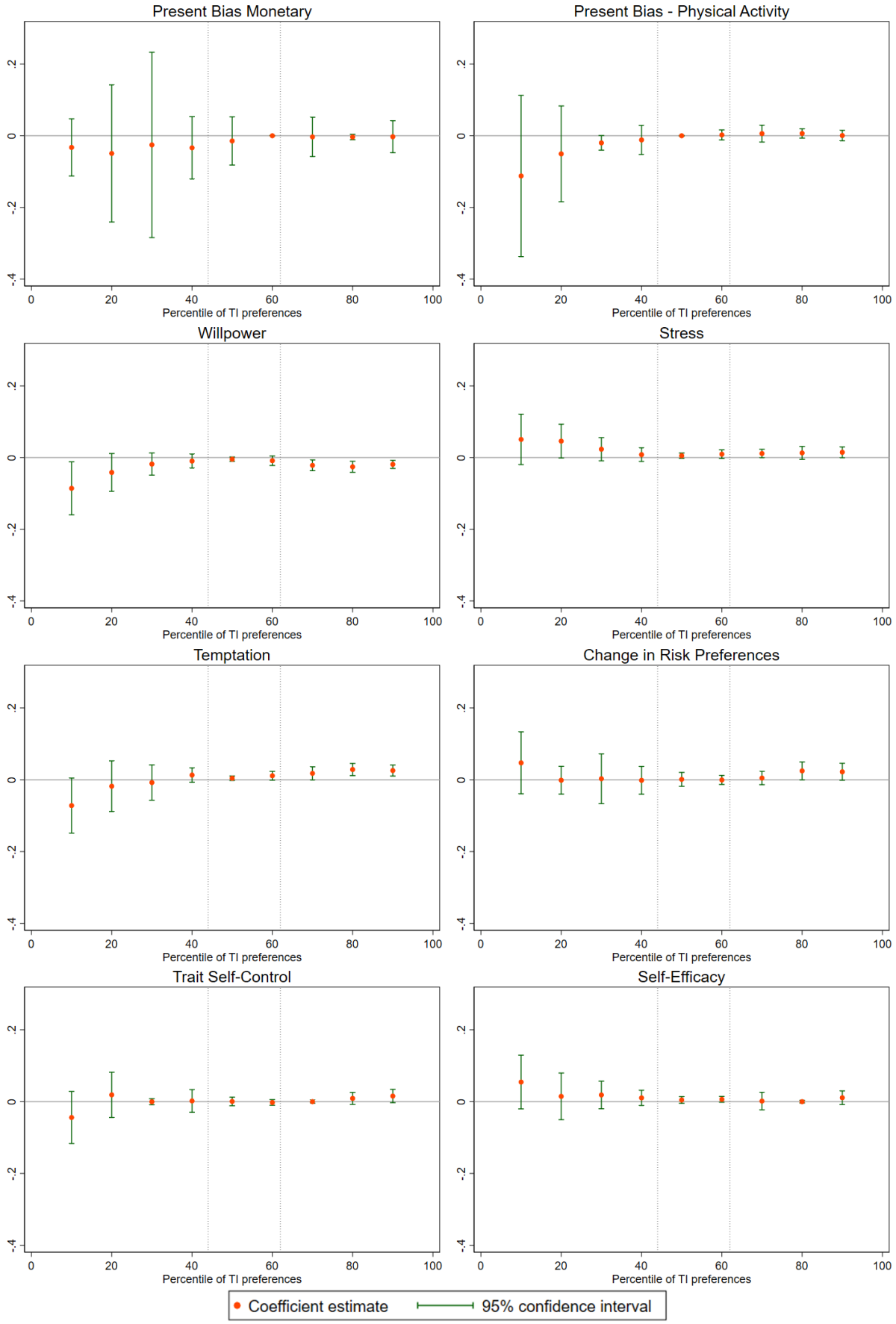
Notes: Statistics in the table are for the regression sample of 3,055 participants. Present bias was measured using the DI index, with present bias increasing in the DI Index. Willpower, stress, temptation are measured with likert-scale measures – willpower and temptation on 7-point scales from 1 to 7, and stress on an 11-point scale from 0-10. Change in risk preferences is measured as the change in normalised risk premium, with a positive value indicating the individual becomes more risk adverse as the time delay to the resolution of the risk shortens. Trait self-control and self-efficacy are likert scale measures, on a 5-point 1 to 5 scale and a 4-point 1 to 4 scale respectively.

Table A10: OLS regression of TI preferences/planning on eight potential endogenous drivers - Standardized dependent variable

	TI preferences	TI planning
Present Bias - Monetary	-0.019	-0.003
	(0.014) [0.368]	(0.010) [0.908]
Present Bias - Physical Activity	-0.014	-0.010
	(0.022) [0.725]	(0.015) [0.686]
Willpower	-0.028	-0.013
	(0.024) [0.447]	(0.025) [0.756]
Stress	0.028	0.048
	(0.020) [0.345]	(0.020) [0.103]
Temptation	0.003	-0.031
	(0.032) [0.946]	(0.028) [0.479]
Change in risk preferences	0.036	0.005
	(0.027) [0.368]	(0.026) [0.935]
Trait self-control	0.018	0.015
	(0.025) [0.672]	(0.019) [0.66]
Self-efficacy	0.000	0.023
	(0.028) [0.946]	(0.033) [0.686]
Controls	✓	✓
N	3055	3055
R^2	0.033	0.027

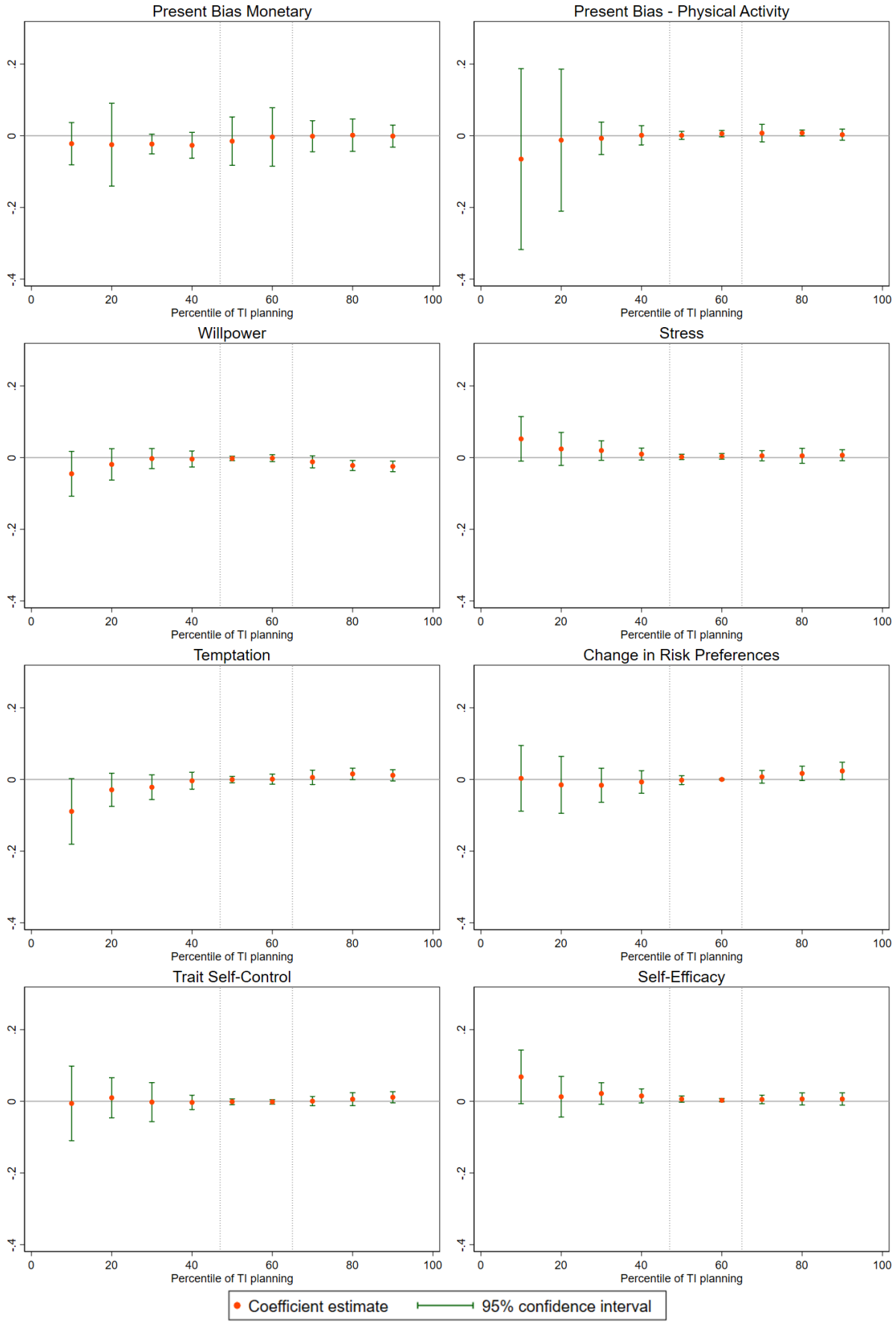
Details of these regressions are the same as for the primary OLS analysis in Table 4, except that the dependent variable is standardized. Robust standard errors in parentheses. False Discovery Rate adjusted p-values (adjusted for 160 tests) in square brackets.

Figure A5: Quantile regressions of TI preferences on potential drivers - standardized dependent variable



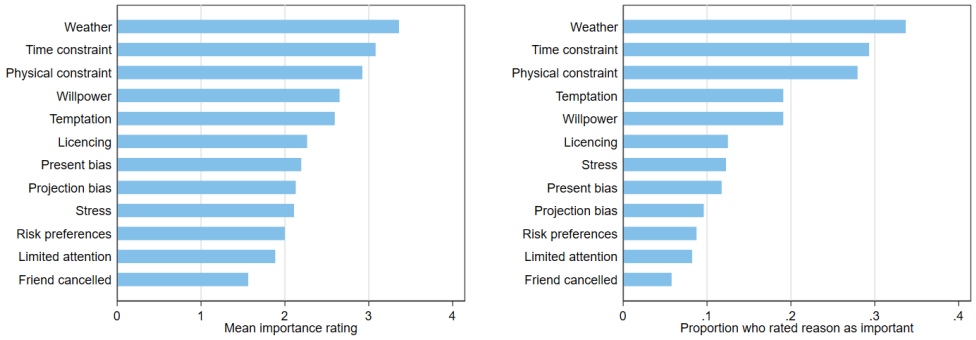
Notes: Details of these regressions are the same as for the primary quantile regression analysis in Figure 1, except that the dependent variable is standardized. Confidence intervals are calculated to adjust for MHT using FDR adjusted p-values method (adjusted for 160 tests).

Figure A6: Quantile regressions of TI planning on potential drivers - standardized dependent variable



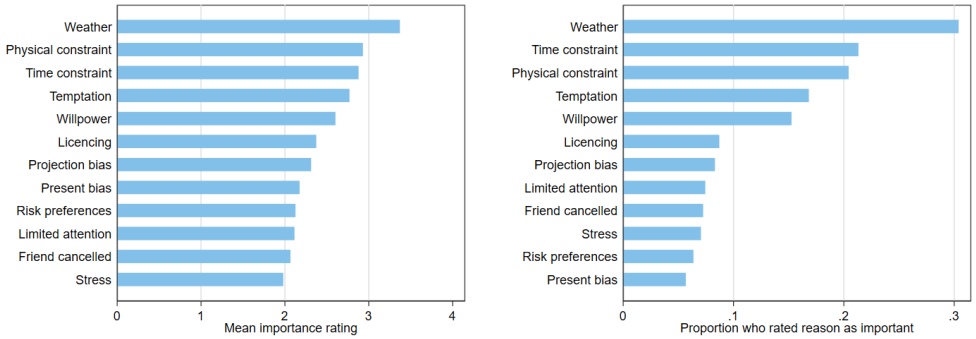
Notes: Details of these regressions are the same as for the primary quantile regression analysis in Figure 2, except that the dependent variable is standardized. Confidence intervals are calculated to adjust for MHT using FDR adjusted p-values (adjusted for 160 tests).

Figure A7: Self-reported reasons for under-exercise TI preferences



Note: Participants who had under-exercise TI preferences in at least one of the two fortnights analyzed were asked at the end of data collection to rate on a scale of 1-7 the importance of each of 12 reasons for their under-exercise TI preferences over the previous four weeks. The list of 12 possible drivers consists of both endogenous and exogenous drivers of time inconsistency, and was drawn from theoretical, empirical and anecdotal evidence. The included drivers reflected: present bias, willpower, stress, temptation, risk preferences, limited attention (i.e., forgetting to exercise: Ericson, 2017; Habla and Muller, 2021), projection bias (i.e., ex-ante overestimating enjoyment from exercise: Loewenstein et al., 2003; Acland and Levy, 2015; Augenblick and Rabin, 2019), licensing (i.e., feeling justified in skipping exercise as have already made "enough" healthy decisions in the recent past: De Witt Huberts et al., 2014), as well as social (e.g., friend cancels), time, physical (e.g., illness, injury) and weather constraints. These were all described to participants in an accessible manner – see the exact wording in the full text of the questionnaires provided in Appendix H. Graphs above show results for the sample of 938 participants who responded to this question, are included in our regression sample, and had under-exercise TI preferences during those four weeks. The proportions shown in the figure on the right-hand side are proportions of this sample just described. In the right-handside figure, the participants were classified as having rated a driver as important if they gave it a rating of 5 or more on the 7-point scale.

Figure A8: Self-reported reasons for under-exercise TI planning in general



Note: All participants were asked at the beginning of the study fortnight to rate on a scale of 1-7 the importance of each of 12 reasons for their under-exercise TI planning. The exact wording was "Please indicate how often the statements below are important reasons for you to postpone physical activity (even if you rarely postpone it)". Graphs above show results for the sample of 1023 participants who responded to this question, are included in our regression sample, and had under-exercise TI planning during the study fortnight. The proportions shown in the figure on the right-hand side are proportions of this sample just described.

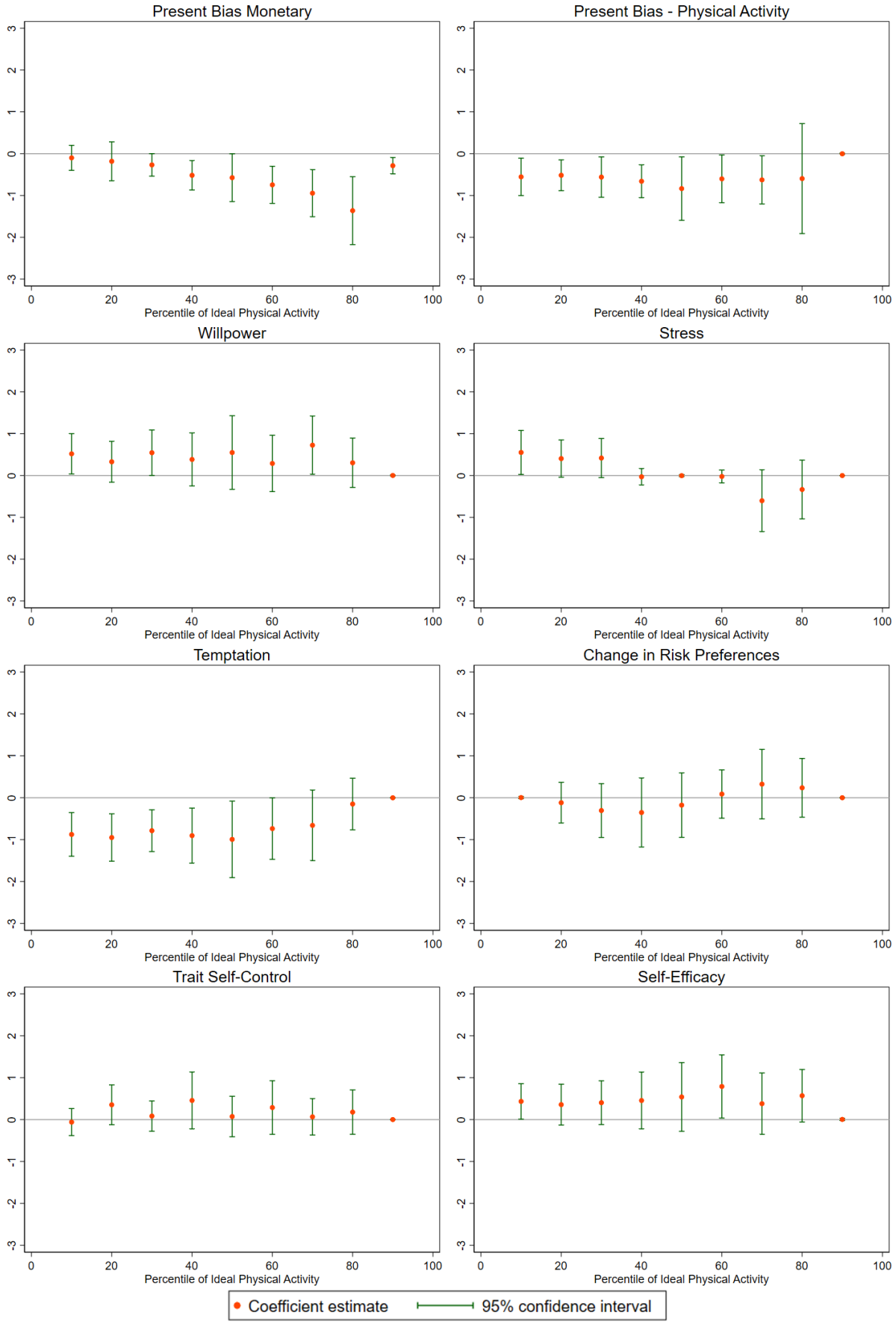
Online Appendix B Ideal, predicted and actual physical activity

Table A11: OLS regression of ideal, predicted and actual physical activity on eight potential endogenous drivers of time inconsistency

	(1) Ideal	(2) Predicted	(3) Actual
Present Bias - Monetary	-0.603	-0.585	-0.243
	(0.181) [0.006]	(0.150) [0.001]	(0.255) [0.354]
Present Bias - Physical Activity	-0.636	-0.407	-0.312
	(0.243) [0.033]	(0.235) [0.093]	(0.212) [0.194]
Willpower	0.438	0.629	1.243
	(0.260) [0.134]	(0.258) [0.034]	(0.261) [0.001]
Stress	0.005	-0.085	-0.481
	(0.247) [0.767]	(0.242) [0.439]	(0.247) [0.105]
Temptation	-0.757	-1.089	-1.247
	(0.263) [0.02]	(0.263) [0.001]	(0.260) [0.001]
Change in risk preferences	-0.194	-0.31	-0.524
	(0.331) [0.512]	(0.325) [0.256]	(0.321) [0.163]
Trait self-control	0.166	0.178	0.092
	(0.250) [0.483]	(0.249) [0.32]	(0.249) [0.578]
Self-efficacy	0.622	0.612	0.277
	(0.254) [0.040]	(0.255) [0.036]	(0.253) [0.316]
Controls	✓	✓	✓
N	3055	3055	3055
R^2	0.112	0.129	0.165
Dependent variable mean	24.284	23.674	25.835
Dependent variable SD	12.661	12.612	12.824

The dependent variables are ideal physical activity (column (1)), predicted physical activity (column (2)), and actual physical activity (column (3)). All are measured in hours per fortnight. Sample and control variables used are as per the primary OLS analysis in Table 4. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT within each outcome variable for 80 tests – 8 OLS tests (8 independent variables) and 72 quantile regression tests (8 independent variables at 9 quantiles).

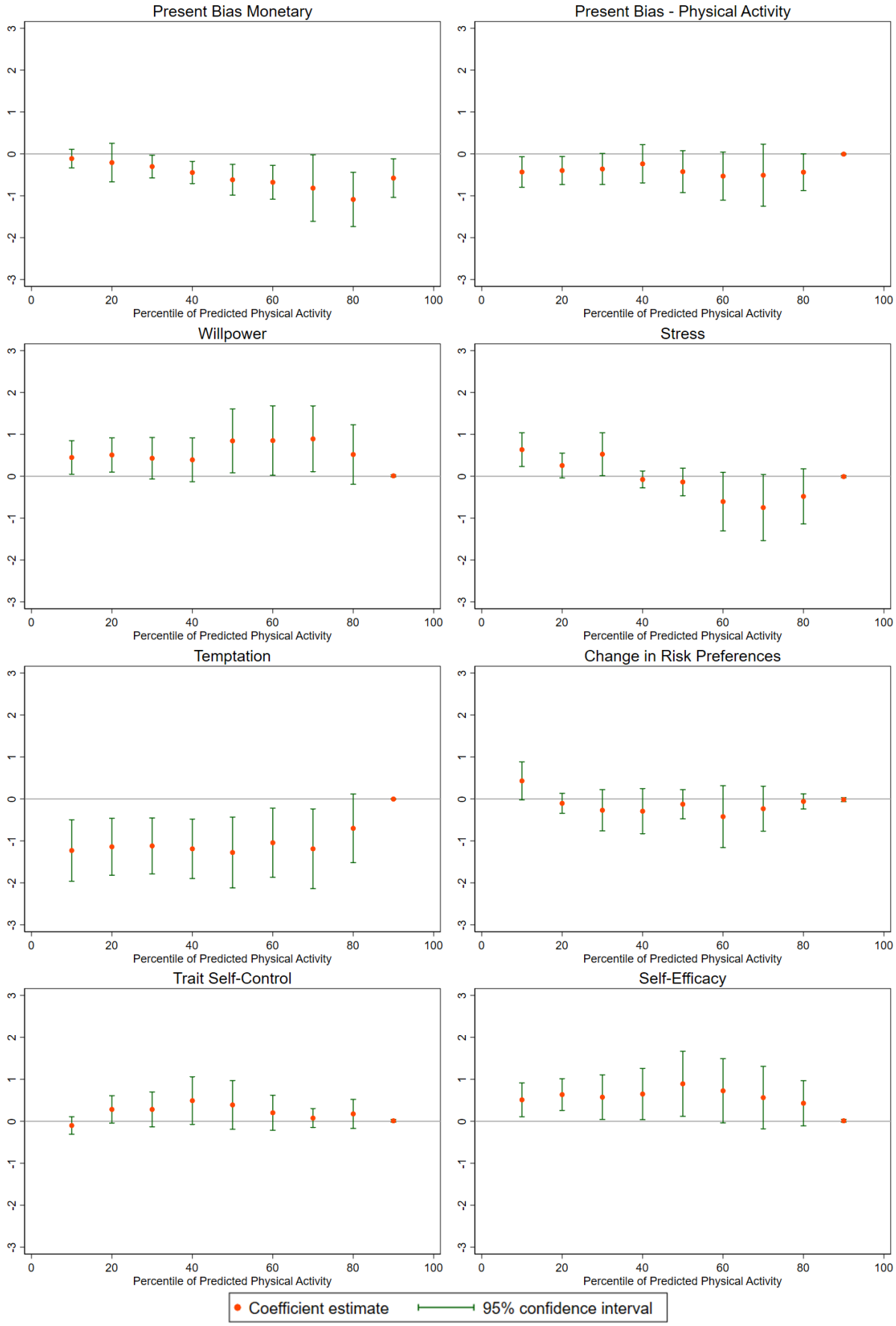
Figure A9: Quantile regressions of Ideal physical activity on potential drivers of time inconsistency



Note: Graphs show results of quantile regressions run at each percentile $\theta \in \{10, 20, \ldots, 90\}$ of the distribution of Ideal physical activity, measured in hours per fortnight. Independent and control variables are the same as for the the main OLS regressions (see Table 4). The dependent variable mean(SD) is 24.28(12.66) hours. Confidence intervals are calculated to adjust for MHT using FDR-adjusted p-values (adjusted for 80 tests - see notes to Table A11)

.

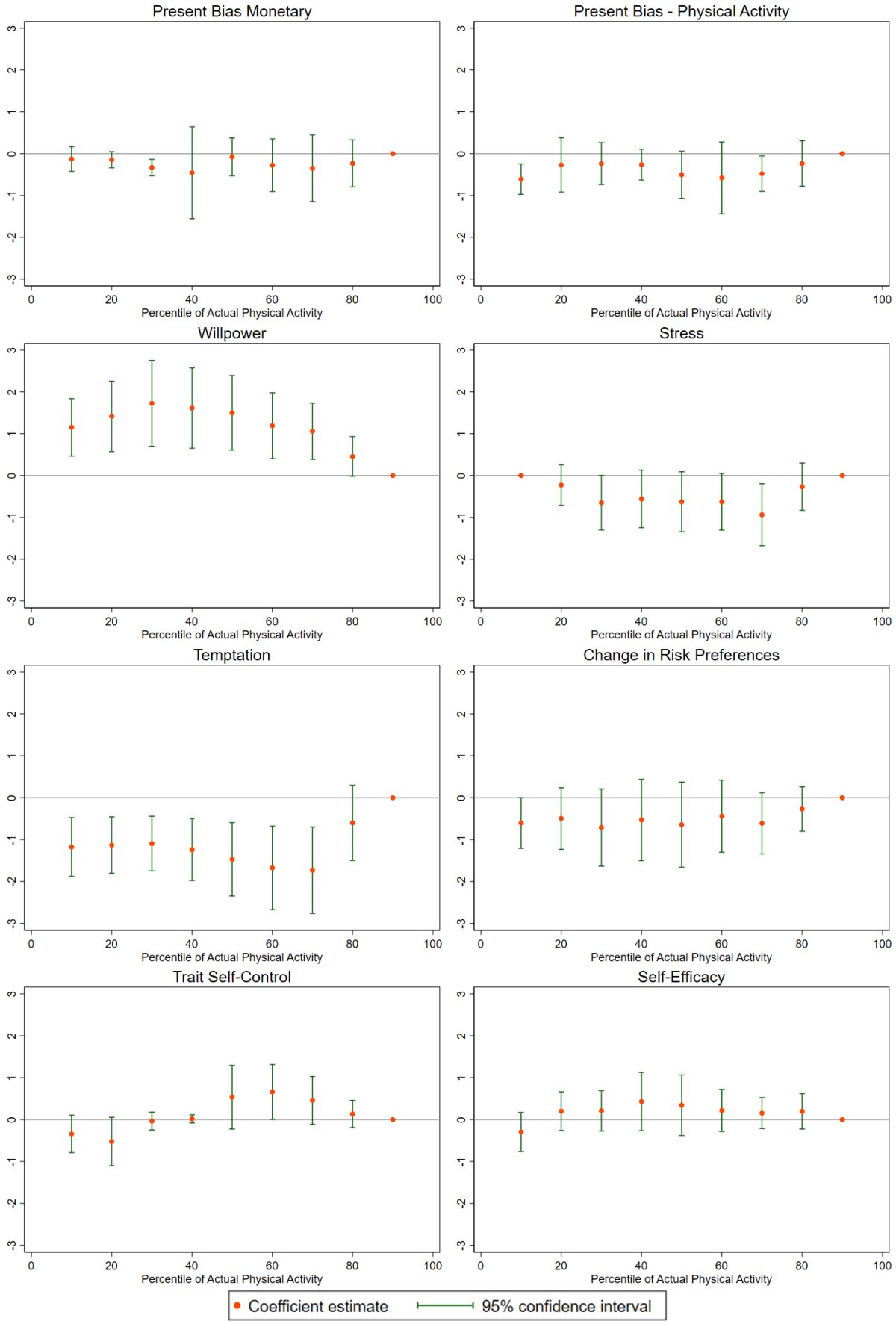
Figure A10: Quantile regressions of predicted physical activity on potential drivers of time inconsistency



Note: Graphs show results of quantile regressions run at each percentile $\theta \in \{10, 20, \ldots, 90\}$ of the distribution of predicted physical activity, measured in hours per fortnight. Independent and control variables are the same as for the main OLS regressions (see Table 4). The dependent variable mean(SD) is 23.67(12.61) hours. Confidence intervals are calculated to adjust for MHT using FDR-adjusted p-values (adjusted for 80 tests - see notes to Table A11)

.

Figure A11: Quantile regressions of actual physical activity on potential drivers of time inconsistency



Note: Graphs show results of quantile regressions run at each percentile $\theta \in \{10, 20, \ldots, 90\}$ of the distribution of actual physical activity, measured in hours per fortnight. Independent and control variables are the same as for the main OLS regressions (see Table 4). The dependent variable mean(SD) is 25.84(12.82) hours. Confidence intervals are calculated to adjust for MHT using FDR-adjusted p-values (adjusted for 80 tests - see notes to Table A11)

.

Online Appendix C Robustness check: Regressions run with post-study fortnight data

Here we ran the same regressions as in the primary analysis in the main text, except we used data for time inconsistency and the endogenous drivers collected in the two weeks after the study fortnight (the post-study fortnight).

Table A12: OLS regression of TI preferences and planning on eight endogenous drivers - using post-study fortnight data

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	0.015 (0.012) [1.000]	0.008 (0.010) [1.000]
Present Bias - Physical Activity	-0.059 (0.064) [1.000]	-0.002 (0.015) [1.000]
Willpower	-0.009 (0.036) [1.000]	-0.005 (0.036) [1.000]
Stress	0.025 (0.025) [1.000]	-0.029 (0.027) [1.000]
Temptation	0.072 (0.032) [0.234]	0.031 (0.027) [1.000]
Change in risk preferences	-0.031 (0.041) [1.000]	-0.042 (0.042) [1.000]
Trait self-control	0.025 (0.053) [1.000]	-0.031 (0.033) [1.000]
Self-efficacy	0.009 (0.026) [1.000]	0.016 (0.026) [1.000]
Controls	✓	✓
N	2097	2097
R^2	0.026	0.025
Dependent variable mean	-0.18	-0.19
Dependent variable SD	1.69	1.42

Details are as per the primary OLS analysis in Table 4, except that the dependent and independent variables are measured two weeks later (post-study fortnight). Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets.

OLS finds no significant relationships (Table A12). Quantile regressions find that each of willpower and temptation are significantly associated with both under-exercise TI preferences and under-exercise TI planning. Willpower is sig-

nificantly negatively associated with TI preferences and planning at two of the three under-exercise percentiles ananlyzed - the 80th and 90th percentiles - with point estimates ranging from -5pp to -8pp. Temptation is significantly positively associated with TI preferences at all under-exercise percentiles (the 70th, 80th and 90th percentiles) and with TI planning at the 80th and 90th perecentiles. Point estimates range from 5pp to 6pp. The quantile regressions find no other significant association.[34]

Online Appendix D Robustness of results to incentives received

As described in Section 2.1 and Footnotes 10 and 21, participants were randomized to various incentive schemes for the purposes of an experiment that will be described in a separate paper. One third of participants (*incentiveless* group) received no incentives, and to elicit their predicted physical activity they received the *own prediction* question (unincentivized). Another third (*Choice-matching Only* group) received Choice-matching incentives, and to elicit their predicted physical activity they also received the *own prediction* question (incentivized with Choice-matching). The final third of participants (*Choice-matching Plus* group) received Choice-matching incentives, and to elicit their predicted physical activity they received the *prediction for a similar other* question (incentivized with monetary rewards for accuracy, rather than with Choice-matching).

As described in Section 6.1, in the full sample we saw overexercise TI preferences and planning on average. In Table A13, we see that analyzing the prevalence of time inconsistency by incentive type subgroup produces the same result, broadly speaking. Using t-tests of equality of means, the only significant difference between incentive types is that predicted PA for the Choice-matching only group is significantly higher than for the Choice-matching plus group, but this

[34] Tables of quantile regression results are available from the first author upon reasonable request.

doesn't result in a significant difference between the incentive groups in TI planning.

In Tables A14 and A15 we see that results from OLS regressions of time inconsistency on endogenous drivers by incentive type subgroup do not differ substantially from those for the full sample described in Table 4, with no significant associations. We also ran the following: quantile regressions by incentive type subgroup, OLS regressions with the full sample where we included interactions between the endogenous driver variables and dummy variables for the incentive type received, and quantile regressions with the same interactions. We find no substantive evidence that results differ across incentive types.[35]

Table A13: Summary statistics by incentive type subgroup

	Mean				P-value		
	(1) Full sample	(2) Choice-Match. Plus	(3) Choice-Match. Only	(4) Incent-iveless	(5) CMP v CMO	(6) CMP v I'less	(7) CMO v I'less
Relative measures							
TI Preferences	-0.43	-0.48	-0.40	-0.41	0.350	0.364	0.708
TI Planning	-0.48	-0.58	-0.45	-0.42	0.350	0.328	0.708
Absolute measures							
TI Preferences	-1.82	-1.82	-1.93	-1.70	0.746	0.708	0.708
TI Planning	-2.31	-2.94	-2.10	-1.96	0.350	0.328	0.708
Components of measures							
Ideal	24.21	24.07	24.43	24.12	0.433	0.708	0.350
Predicted	23.71	22.96	24.26	23.87	0.022	0.350	0.364
Actual	26.03	25.90	26.36	25.83	0.503	0.708	0.433
N	4333	1355	1428	1550			

Columns (5), (6) and (7) show FDR-adjusted p-values (21 tests) from t-tests of equality of means between the Choice-matching Plus group and the Choice-matching Only group (5), the Choice-matching Plus group and the Incentiveless group (6), and the Choice-matching Only group and the Incentiveless group (7).

[35] Tables for OLS regressions including incentive interaction terms, graphs for quantile regressions by incentive type subgroup, and graphs for quantile regressions including incentive interaction terms, are available from the first author upon reasonable request.

Table A14: OLS regression of TI preferences on eight potential endogenous drivers - by incentive subgroup

	(1) Full sample	(2) Choice-Match. Plus	(3) Choice-Match. Only	(4) Incent-iveless
Present Bias - Monetary	-0.032	-0.088	0.03	-0.014
	(0.023) [0.368]	(0.046) [0.531]	(0.030) [1.000]	(0.030) [1.000]
Present Bias - Phys. Act.	-0.023	-0.006	-0.015	-0.013
	(0.036) [0.725]	(0.110) [1.000]	(0.047) [1.000]	(0.033) [1.000]
Willpower	-0.047	-0.137	0.027	-0.04
	(0.040) [0.447]	(0.090) [0.72]	(0.055) [1.000]	(0.060) [1.000]
Stress	0.046	0.028	0.064	0.042
	(0.033) [0.345]	(0.089) [1.000]	(0.041) [0.601]	(0.049) [1.000]
Temptation	0.004	-0.027	0.000	0.002
	(0.053) [0.946]	(0.146) [1.000]	(0.058) [1.000]	(0.044) [1.000]
Change in risk pref.	0.061	0.061	-0.004	0.123
	(0.045) [0.368]	(0.097) [1.000]	(0.061) [1.000]	(0.080) [0.63]
Trait self-control	0.031	0.173	-0.063	0.006
	(0.041) [0.672]	(0.113) [0.72]	(0.048) [0.772]	(0.042) [1.000]
Self-efficacy	0.000	-0.151	0.079	0.073
	(0.047) [0.946]	(0.110) [0.876]	(0.050) [0.601]	(0.074) [0.949]
Controls	✓	✓	✓	✓
N	3055	968	1017	1070
R^2	0.033	0.085	0.082	0.062
Dep. var. mean	-0.406	-0.458	-0.402	-0.362
Dep. var. SD	1.670	2.238	1.369	1.285

The dependent variable in columns (1) to (4) is the relative measure of TI preferences. Sample and control variables used are as per the primary OLS analysis in Table 4. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT within each subgroup for 160 tests – 16 OLS tests (8 independent variables for each of TI preferences and planning) and 144 quantile regression tests (8 independent variables for each of TI preferences and planning at 9 quantiles).

Table A15: OLS regression of TI planning on eight potential endogenous drivers - by incentive subgroup

	(1) Full sample	(2) Choice-Match. Plus	(3) Choice-Match. Only	(4) Incent-iveless
Present Bias - Monetary	-0.005	0.006	0.067	-0.007
	(0.019) [0.908]	(0.031) [1.000]	(0.028) [0.183]	(0.034) [1.000]
Present Bias - Phys. Act.	-0.019	-0.061	0.052	-0.007
	(0.028) [0.686]	(0.072) [1.000]	(0.025) [0.376]	(0.037) [1.000]
Willpower	-0.025	-0.165	0.084	0.006
	(0.047) [0.756]	(0.081) [0.459]	(0.117) [1.000]	(0.057) [1.000]
Stress	0.09	0.09	0.131	0.09
	(0.038) [0.103]	(0.084) [0.98]	(0.063) [0.376]	(0.063) [0.665]
Temptation	-0.058	-0.178	0.027	-0.047
	(0.052) [0.479]	(0.133) [0.879]	(0.093) [1.000]	(0.046) [0.949]
Change in risk pref.	0.009	0.016	-0.01	0.012
	(0.050) [0.935]	(0.098) [1.000]	(0.067) [1.000]	(0.107) [1.000]
Trait self-control	0.028	0.165	-0.055	0.002
	(0.036) [0.66]	(0.095) [0.646]	(0.063) [1.000]	(0.043) [1.000]
Self-efficacy	0.043	-0.12	0.066	0.174
	(0.062) [0.686]	(0.110) [0.98]	(0.073) [1.000]	(0.132) [0.718]
Controls	✓	✓	✓	✓
N	3055	968	1017	1070
R^2	0.027	0.067	0.069	0.063
Dep. var. mean	-0.460	-0.536	-0.462	-0.390
Dep. var. SD	1.882	2.190	1.925	1.499

The dependent variable in columns (1) to (4) is the relative measure of TI planning. Rest of details are as per table A14.

Online Appendix E Robustness of regression results to the endogenous drivers included

We ran the OLS and quantile regressions with different combinations of the eight endogenous drivers we analyze to ensure that our results were robust to the combination chosen. If one of the drivers is a mediator for the effect of another driver on TI preferences/planning, then coefficient estimates may be biased if both drivers are included in a regression together. For instance, present bias may be a mediator for the affective process variables (Loewenstein et al., 2015).

In Table A16 we see the results of OLS regressions where, of the eight drivers we analyse, only the present bias variables are included (columns (1) and (2)), and only the other six drivers are included (columns (3) and (4)). We also ran quantile regressions with these specifications. Results are largely the same as in our primary analysis in the main text. We also ran OLS and quantile regressions using each of the following combinations of drivers: present bias variables and change in risk preferences; willpower, stress and temptation; trait self-control and self-efficacy. The results were broadly similar to those in our primary analysis. Finally, we ran the regressions including each driver on its own and, again, find results that are broadly similar (see the OLS results in Tables A17 and A18).[36]

[36] Tables for OLS and quantile regressions not shown here are available from the first author upon reasonable request.

Table A16: OLS regression of TI preferences and planning on different combinations of endogenous drivers

	(1) TI Pref.	(2) TI Plan.	(3) TI Pref.	(4) TI Plan.
Pres. Bias - Monetary	-0.029	-0.004		
	(0.024) [1.000]	(0.019) [1.000]		
Pres. Bias - Phys. Act.	-0.024	-0.018		
	(0.036) [1.000]	(0.028) [1.000]		
Willpower			-0.046	-0.025
			(0.040) [0.375]	(0.047) [0.684]
Stress			0.047	0.091
			(0.033) [0.313]	(0.038) [0.079]
Temptation			0.007	-0.055
			(0.052) [0.755]	(0.051) [0.417]
Change in risk pref.			0.059	0.006
			(0.044) [0.335]	(0.050) [0.755]
Trait self-control			0.030	0.027
			(0.040) [0.564]	(0.036) [0.564]
Self-efficacy			-0.003	0.040
			(0.047) [0.772]	(0.062) [0.629]
Controls	✓	✓	✓	✓
N	3055	3055	3055	3055
R^2	0.028	0.023	0.031	0.027
Dep. var. mean	-0.406	-0.460	-0.406	-0.460
Dep. var. SD	1.670	1.882	1.670	1.882

The dependent variable in columns (1) and (3) is TI preferences, and is TI planning in columns (2) and (4). Sample and control variables used are as per the primary OLS analysis in Table 4. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for all outcomes estimated for a given combination of endogenous drivers. For columns (1) and (2) this is 40 tests – 4 OLS tests (2 independent variables for each of TI preferences and planning) and 36 quantile regression tests (2 independent variables for each of TI preferences and planning at 9 quantiles). For columns (3) and (4) this is 120 tests – 12 OLS outcomes (6 independent variables for each of TI preferences and planning) and 108 quantile regression tests (6 independent variables for each of TI preferences and planning at 9 quantiles).

Table A17: OLS regression of TI preferences on each endogenous driver separately

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pres. Bias - Mon.	-0.028							
	(0.023) [1.000]							
Pres Bias - PA		-0.023						
		(0.037) [1.000]						
Willpower			-0.046					
			(0.034) [0.149]					
Stress				0.052				
				(0.031) [0.066]				
Temptation					0.022			
					(0.044) [0.674]			
Change in risk pref.						0.06		
						(0.044) [1.000]		
Trait self-control							0.016	
							(0.033) [1.000]	
Self-efficacy								-0.007
								(0.043) [1.000]
Controls	✓	✓	✓	✓	✓	✓	✓	✓
N	3055	3055	3055	3055	3055	3055	3055	3055
R^2	0.028	0.027	0.028	0.028	0.027	0.030	0.027	0.027
Dep. var. mean	-0.406	-0.406	-0.406	-0.406	-0.406	-0.406	-0.406	-0.406
Dep. var. SD	1.670	1.670	1.670	1.670	1.670	1.670	1.670	1.670

The dependent variable in all columns is TI preferences. Only one of the eight potential drivers of time inconsistency is included in each regression. Sample and control variables used are as per the primary OLS analysis in Table 4. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for all tests for a given endogenous drivers. This is 20 tests – 2 OLS tests (one for each of TI preferences and planning) and 18 quantile regression tests (one for each of TI preferences and planning at 9 quantiles).

Table A18: OLS regression of TI planning on each endogenous driver separately

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pres. Bias - Mon.	-0.004							
	(0.019) [1.000]							
Pres Bias - PA		-0.017						
		(0.028) [1.000]						
Willpower			-0.012					
			(0.039) [0.538]					
Stress				0.08				
				(0.036) [0.042]				
Temptation					-0.042			
					(0.044) [0.404]			
Change in risk pref.						0.008		
						(0.049) [1.000]		
Trait self-control							0.036	
							(0.032) [1.000]	
Self-efficacy								0.037
								(0.061) [1.000]
Controls	✓	✓	✓	✓	✓	✓	✓	✓
N	3055	3055	3055	3055	3055	3055	3055	3055
R^2	0.022	0.022	0.022	0.024	0.023	0.024	0.023	0.023
Dep. var. mean	-0.460	-0.460	-0.460	-0.460	-0.460	-0.460	-0.460	-0.460
Dep. var. SD	1.882	1.882	1.882	1.882	1.882	1.882	1.882	1.882

The dependent variable in all columns is TI planning. Only one of the eight potential drivers of time inconsistency is included in each regression. Sample and control variables used are as per the primary OLS analysis in Table 4. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for all tests for a given endogenous drivers. This is 20 tests – 2 OLS outcomes (one for each of TI preferences and planning) and 18 quantile regression tests (one for each of TI preferences and planning at 9 quantiles).

Online Appendix F Robustness of regression results to the control variables included

It is possible that some of the control variables could be outcomes of time inconsistency and some of the endogenous drivers, which may lead to biased coefficient estimates. As a robustness check, we ran the OLS and quantile regressions with no controls except those that cannot be outcomes of other variables in our regression (age, gender, ethnicity, incentive group randomized into). We also ran these regressions with only these control variables and additionally controls for peculiarites in responses to our present bias and risk preferences measure (extreme responses in present bias and/or risk preference measurements, if indicates that prefers more physical activity to less in present bias measurement). The results are in line with those in our primary analyses in the main text. The OLS results can be seen in Tables A19 and A20.[37]

[37] Tables for quantile regressions are available from the first author upon reasonable request.

Table A19: OLS regression of TI preferences and planning on eight endogenous drivers - including only age, gender, ethnicity, incentive group to which randomized as controls

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	-0.022 (0.019) [0.784]	-0.006 (0.017) [1.000]
Present Bias - Physical Activity	-0.021 (0.029) [1.000]	-0.017 (0.027) [1.000]
Willpower	-0.014 (0.039) [1.000]	0.002 (0.051) [1.000]
Stress	0.046 (0.030) [0.426]	0.087 (0.034) [0.079]
Temptation	-0.02 (0.053) [1.000]	-0.079 (0.054) [0.456]
Change in risk preferences	0.009 (0.031) [1.000]	0.006 (0.037) [1.000]
Trait self-control	0.038 (0.040) [0.889]	0.038 (0.034) [0.802]
Self-efficacy	0.015 (0.043) [1.000]	0.044 (0.056) [1.000]
N	3055	3055
R^2	0.007	0.007
Dependent variable mean	-0.406	-0.460
Dependent variable SD	1.670	1.882

Details as per the primary OLS analysis in Table 4, except that the only controls included are age, gender, ethnicity, incentive group to which randomized. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets.

Table A20: OLS regression of TI preferences and planning on eight endogenous drivers - including as controls only age, gender, ethnicity, incentive group to which randomized, present bias measures controls and risk preferences measure controls

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	-0.032 (0.022) [0.357]	-0.007 (0.018) [1.000]
Present Bias - Physical Activity	-0.028 (0.035) [0.743]	-0.027 (0.028) [0.68]
Willpower	-0.02 (0.039) [0.966]	-0.003 (0.050) [1.000]
Stress	0.048 (0.030) [0.303]	0.088 (0.034) [0.055]
Temptation	-0.023 (0.054) [1.000]	-0.081 (0.055) [0.357]
Change in risk preferences	0.057 (0.044) [0.446]	0.007 (0.048) [1.000]
Trait self-control	0.035 (0.041) [0.708]	0.036 (0.035) [0.609]
Self-efficacy	0.017 (0.043) [1.000]	0.047 (0.056) [0.708]
N	3055	3055
R^2	0.011	0.010
Dependent variable mean	-0.406	-0.460
Dependent variable SD	1.670	1.882

Details as per the primary OLS analysis in Table 4, except the only controls included are age, gender, ethnicity, incentive group to which randomized, present bias measures controls and risk preferences measure controls. Robust standard errors in parentheses. False Discovery Rate adjusted p-values in square brackets.

Online Appendix G Departures from pre-analysis plan

A pre-analysis plan (PAP) was pre-registered at `https://osf.io/ty9sx`. In line with the recommendation of Banerjee et al. (2020) that a "populated PAP" should be made available which "can serve as a useful and transparent record of the results of the analysis prespecified in the PAP, or the reasons it was not implemented", see below a list of departures in our final primary regression analyses (the results of which are shown in Table 4 and Figures 1 and 2 in the main text) from the planned primary regression analyses described in the PAP. This is

followed by the results from analyses carried out exactly as specified in the PAP, which do not differ substantively from the results in the final primary regression analyses.

1. In the PAP, we specified that in primary analysis we would run OLS regressions for the under-exerciser subgroup. We did not do this, but instead used quantile regressions in the final primary analysis, which had been specified as secondary analysis in the PAP. We did this as we decided after submitting the PAP that quantile regressions were a better way of analyzing the heterogeneity in relationships between TI preferences/planning and potential drivers across the TI preferences/planning distribution. Quantile regressions gave us more granular information (i.e., we were able to analyze heterogeneity at many different points in the under-exerciser portion of the distribution, rather than just for under-exercisers as a whole) and also allowed us to avoid running regressions with data truncated based on the dependent variable, which may be problematic.

2. Ideal, predicted and actual PA were elicited using an MCQ for each week of the study fortnight with options "less than 1 hour", "1 hour", "2 hours"..."20 hours", "more than 20 hours", as was pre-registered. In the PAP, we specified that we would include as control variables in regressions dummy variables for giving the bottom or top response to the ideal, predicted and actual questions (i.e., the "less than 1 hour" and "greater than 20 hours" responses). In the final analyses, we did not include these controls in the regressions out of a concern that these were mediators between the dependent and independent variables, and thus may bias coefficient estimates.

3. In the PAP we specified that we would include two additional driver variables in our regressions: *risk preferences for the future* and *caring for future self*. We did not include these in the final analysis. We were concerned that there may be a mediation problem between the risk preferences

for the future variable and the changes in risk preferences variable (i.e., that one may mediate the relationship between the other and the dependent variable), as the former variable was an input into the calculation of the latter. We omitted the caring for future self variable for the sake of parsimony.

4. In the final analysis, we included two additional control variables that were not specified in the PAP: ethnicity and employment status.

5. In the PAP, we specified that our present bias variables used in regressions would be an average of measures taken in questionnaire 1 (at the beginning of the study fortnight) and in questionnaire 2 (at the end of the fortnight). In final analysis, we used only the questionnaire 1 measure, as we had many missing values for the questionnaire 2 measure and using the average would have cut our sample by approximately a third, severely impacting statistical power.

6. In the PAP, we specified that we would correct for MHT for the present bias outcomes separately from the other drivers outcomes. In the final analysis, we decided to be more conservative in our corrections by pooling the outcomes for all drivers, including present bias, in adjusting for MHT.

Tables A21 and A22 show the results had we carried out our analysis as per the PAP, except that we apply the more conservative MHT correction used in the final analysis (see point 6 above). Tables A23 and A24 show the same analyses except that only the questionnaire 1 measures of present bias are used in calculating the present bias variables (see point 5 above), with the same being the case for the caring for future self variable. The results for the full sample in Tables A21 and A23 are broadly in line with our findings in Table 4 in the main text. Table A24, which shows results for the under-exerciser subgroup, shows evidence (marginally significant) of TI preferences and planning both

having a negative(positive) relationship with willpower(temptation), which is in line with the quantile regression findings in Figures 1 and 2 in the main text. Table A22, which also shows results for the under-exerciser subgroup, shows similar point estimates as Table A24 for willpower and temptation, but these are not significant, likely due to the reduced statistical power that accompanies the reduced sample size when both questionnaire 1 and 2 measures of present bias and caring for future self are used, rather than just questionnaire 1 measures.

We also stated in the PAP that, following Toussaert (2018), we would test predictions of the quasi-hyperbolic model and models of costly self-control regarding *self-control types*. A self-control type is an individual that may demand commitment even if she anticipates resisting a self-control failure without the commitment. This is because the commitment reduces the self-control costs she incurs in resisting temptation. The quasi-hyperbolic model doesn't allow for such types, but models of costly self-contol, such as dual-self models, predict that such types can exist. Further details of the methodology used to test these predictions can be seen in the PAP. The results of this empirical test are not included in the main text for the sake of parsimony, but we summarize here. We find that 7.1% of our sample are self-control types (n = 3,925; 95% CI: [6.4%, 8.0%]). The proportion of self-control types in our sample is considerably smaller than that found by Toussaert (2018) (23-36%) in her lab experiment with students. However, our estimated proportion likely represents a lower bound on the true proportion in our sample.[38]

[38] There may be other participants who did not reveal self-control type behavior in our data, but may do so if, for instance, some features of the hypothetical commitment device offered to identify self-control types were different (e.g., the time period covered, the duration of physical activity, the size of the monetary deposit required).

Table A21: OLS regression of TI preferences/planning on full sample as per pre-analysis plan (present bias and caring for future self measures are calculated using both questionnaire 1 and 2 measures)

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	-0.017 (0.029) [1.000]	0.011 (0.031) [1.000]
Present Bias - Physical Activity	-0.033 (0.028) [0.652]	-0.026 (0.031) [0.802]
Willpower	-0.069 (0.052) [0.562]	-0.086 (0.051) [0.401]
Stress	0.035 (0.038) [0.765]	0.074 (0.040) [0.354]
Temptation	-0.018 (0.062) [1.000]	-0.073 (0.055) [0.562]
Change in risk preferences	0.034 (0.065) [1.000]	0.001 (0.075) [1.000]
Trait self-control	0.058 (0.035) [0.404]	0.051 (0.033) [0.421]
Self-efficacy	-0.03 (0.056) [1.000]	0.015 (0.061) [1.000]
Preferences for future risk	0.012 (0.059) [1.000]	0.001 (0.079) [1.000]
Caring for future self	-0.034 (0.045) [0.866]	-0.009 (0.049) [1.000]
Controls	✓	✓
N	1969	1969
R^2	0.357	0.36
Dependent variable mean	-0.395	-0.458
Dependent variable SD	1.733	1.913

The dependent variable in columns (1) and (2) are our relative measures of TI preferences and TI planning, respectively. The independent variables are the eight potential endogenous drivers of time inconsistency outlined in Section 3, and additionally *preferences for future risk* and *caring for future self*. The present bias and caring for future self measures are calculated using both questionnaire 1 and 2 measures. All independent variables were standardized for use in regressions. The sample of 1,969 participants used in this analysis consists of all participants for whom we observe values for TI preferences, TI planning, and these ten potential drivers of time inconsistency. Robust standard errors in parentheses. Controls are as per the regressions in Table 4, and additionally we controlled for giving lower or upper bound responses to the ideal, predicted and actual questions. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for 40 tests – 20 OLS tests in this table and 20 in Table A22.

Table A22: OLS regression of TI preferences/planning on under-exerciser subsample as per pre-analysis plan (present bias and caring for future self measures are calculated using both questionnaire 1 and 2 measures)

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	-0.006 (0.006) [0.765]	0.006 (0.007) [0.765]
Present Bias - Physical Activity	0.003 (0.007) [1.000]	0.001 (0.007) [1.000]
Willpower	-0.018 (0.010) [0.354]	-0.021 (0.011) [0.354]
Stress	0.014 (0.009) [0.408]	0.013 (0.010) [0.562]
Temptation	0.021 (0.009) [0.354]	0.021 (0.010) [0.354]
Change in risk preferences	0.033 (0.015) [0.354]	0.048 (0.015) [0.076]
Trait self-control	0.019 (0.009) [0.354]	0.019 (0.010) [0.354]
Self-efficacy	-0.004 (0.010) [1.000]	0.000 (0.011) [1.000]
Preferences for future risk	-0.006 (0.017) [1.000]	0.016 (0.017) [0.765]
Caring for future self	-0.009 (0.009) [0.765]	-0.010 (0.010) [0.765]
Controls	✓	✓
N	744	685
R^2	0.242	0.233
Dependent variable mean	0.341	0.330
Dependent variable SD	0.225	0.227

The dependent variable in columns (1) and (2) are our relative measures of TI preferences and TI planning, respectively. The independent variables are the eight potential endogenous drivers of time inconsistency outlined in Section 3, and additionally *preferences for future risk* and *caring for future self*. Present bias and caring for future self measures are calculated using both questionnaire 1 and 2 measures. All independent variables were standardized for use in regressions. The sample of 744(685) participants used in the analysis of TI preferences(planning) consists of all participants for whom we observe values for TI preferences, TI planning, and these ten potential drivers of time inconsistency, and who have under-exercise TI preferences(planning). Robust standard errors in parentheses. Controls are as per the regressions in Table A21. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for 40 tests – 20 OLS tests in this table and 20 in Table A21.

Table A23: OLS regression of TI preferences/planning on full sample as per pre-analysis plan but using questionnaire 1 measures of present bias and caring for future self

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	0.015	0.054
	(0.041) [1.000]	(0.050) [0.685]
Present Bias - Physical Activity	-0.01	-0.007
	(0.033) [1.000]	(0.027) [1.000]
Willpower	-0.072	-0.074
	(0.036) [0.176]	(0.040) [0.203]
Stress	0.041	0.068
	(0.030) [0.491]	(0.032) [0.142]
Temptation	-0.02	-0.056
	(0.043) [0.978]	(0.044) [0.491]
Change in risk preferences	0.024	0.007
	(0.052) [0.978]	(0.057) [1.000]
Trait self-control	0.038	0.046
	(0.036) [0.685]	(0.035) [0.491]
Self-efficacy	-0.006	0.033
	(0.038) [1.000]	(0.044) [0.888]
Preferences for future risk	0.005	0.013
	(0.056) [1.000]	(0.059) [1.000]
Caring for future self	-0.024	-0.03
	(0.036) [0.939]	(0.038) [0.888]
Controls	✓	✓
N	2974	2974
R^2	0.291	0.331
Dependent variable mean	-0.399	-0.454
Dependent variable SD	1.672	1.891

The dependent variable in columns (1) and (2) are our relative measures of TI preferences and TI planning, respectively. The independent variables are the eight potential endogenous drivers of time inconsistency outlined in Section 3, and additionally *preferences for future risk* and *caring for future self*. Present bias and caring for future self measures are calculated using both questionnaire 1 measures only. All independent variables were standardized for use in regressions. The sample of 2,974 participants used in this analysis consists of all participants for whom we observe values for TI preferences, TI planning, and these ten potential drivers of time inconsistency. Robust standard errors in parentheses. Controls are as per the regressions in Table A21. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for 40 tests – 20 OLS tests in this table and 20 in Table A24.

Table A24: OLS regression of TI preferences/planning on under-exerciser subsample as per pre-analysis plan but using questionnaire 1 measures of present bias and caring for future self

	(1) TI Preferences	(2) TI Planning
Present Bias - Monetary	-0.008 (0.004) [0.203]	0.003 (0.006) [0.978]
Present Bias - Physical Activity	0 (0.006) [1.000]	0.003 (0.006) [0.978]
Willpower	-0.019 (0.008) [0.086]	-0.022 (0.009) [0.084]
Stress	0.011 (0.007) [0.348]	0.010 (0.007) [0.491]
Temptation	0.023 (0.007) [0.066]	0.020 (0.008) [0.084]
Change in risk preferences	0.030 (0.012) [0.084]	0.032 (0.013) [0.084]
Trait self-control	0.019 (0.007) [0.084]	0.020 (0.008) [0.084]
Self-efficacy	0.004 (0.008) [0.978]	0.006 (0.008) [0.888]
Preferences for future risk	0.001 (0.014) [1.000]	0.008 (0.014) [0.978]
Caring for future self	-0.006 (0.007) [0.888]	0.000 (0.007) [1.000]
Controls	✓	✓
N	1122	1044
R^2	0.218	0.21
Dependent variable mean	0.349	0.341
Dependent variable SD	0.227	0.229

The dependent variable in columns (1) and (2) are our relative measures of TI preferences and TI planning, respectively. The independent variables are the eight potential endogenous drivers of time inconsistency outlined in Section 3, and additionally *preferences for future risk* and *caring for future self*. Present bias and caring for future self measures are calculated using both questionnaire 1 measures only. All independent variables were standardized for use in regressions. The sample of 1122(1044) participants used in the analysis of TI preferences(planning) consists of all participants for whom we observe values for TI preferences, TI planning, and these ten potential drivers of time inconsistency, and who have under-exercise TI preferences(planning). Robust standard errors in parentheses. Controls are as per the regressions in Table A21. False Discovery Rate adjusted p-values in square brackets. These adjusted p-values are calculated to adjust for MHT for 40 tests – 20 OLS tests in this table and 20 in Table A23.

Online Appendix H Questionnaire texts

The full text of the questionnaires can be seen at the below links from the pre-registration of the study on the Open Science Framework website:

- Questionnaire 1 - Dutch: click here
- Questionnaire 1 - English translation: click here
- Questionnaire 2 - Dutch: click here
- Questionnaire 2 - English translation: click here
- Questionnaire 3 - Dutch: click here
- Questionnaire 3 - English translation: click here

In order to implement the choice-matching method, each participant was asked in each questionnaire to predict the distribution of responses among all participants for one of the incentivized questions in that questionnaire. The distributional prediction questions can be seen here:

- Choice-matching distributional prediction questions - Dutch: click here
- Choice-matching distributional prediction questions - English translation: click here

Online Appendix I Control variables

In the regressions for our primary analysis included in Table 4 and Figures 1 and 2 in the main text, we include the following control variables:

- All variables included in Table 2, except for physical activity minutes per week. These variables were obtained from the pre-existing Lifelines data which we linked our survey data to. This pre-existing data was collected

from participants by the Lifelines organization in 5 waves of data collection between 2007 and 2020. Where a variable was collected more than once by Lifelines, we used the most recently collected version of that variable in our analysis. When used as controls, a multi-category version of these variables are used where applicable (as opposed to the binary versions used in table 2).

- Self-reported Likert scale measure of whether the person likes PA or not, self-reported restricted ability to do PA due to medical reasons, self-reported Likert scale measure of dispositional optimism (Scheier et al., 1994).
- Present bias measure controls
  - Dummy variable for preferring more physical activity to less, as indicated by the individuals' choices in the physical activity present bias choice list.
  - Dummy variables for participants having no switch point in the present bias choice lists
- Risk preferences measure controls: Dummy variables for giving upper or lower bound answers to the Risk preference measures
- Dummy variables for the incentive type a participant got (i.e., Choice-matching, prediction for a similar other incentive)

Online Appendix J Similarity between results for TI preferences and planning

In what we have seen so far, it is noteworthy how similar the results are for TI preferences and TI planning at the aggregate level. Table A25 provides evidence that this aggregate-level similarity is mirrored at the individual level. 87% of our sample are in the same category for TI preferences as they are for TI planning,

and the correlation between the two measures is positive and very strong (Spearman's $\rho = 0.89$).

As discussed in Section 3.1, both TI preferences and planning are driven by unanticipated endogenous and exogenous drivers, while TI preferences is additionally driven by *anticipated* endogenous drivers (i.e., anticipated self-control problems). Thus, provided that unanticipated exogenous and endogenous drivers of time inconsistency are not trivial, a positive correlation is expected. The very strong correlation we see in this case, however, suggests that something more is at play.

For 58% of the sample, TI preferences and TI planning were equal. This high prevalence of individuals for whom their TI preferences and planning were equal is a major driver of the strong correlation we see. The individuals whose TI preferences and planning were equal anticipate no self-control problems or endogenous drivers. A quarter of these are time consistent, and so this anticipation is correct.[39] For the other three quarters who are not time consistent, this anticipation could also be correct, if their TI preferences and planning arose solely from unanticipated exogenous drivers (e.g., the weather, illness). However, it could also be incorrect, meaning that they lacked sophistication about their self-control problems. The latter seems likely to be the more important explanation. This is because for this group of time inconsistents who have TI preferences equal to TI planning, median levels of the endogenous drivers we measure (present bias, willpower, etc.) are the same as for the rest of the subsample of time inconsistents.[40] If the former explanation was the most important, we would expect this group to score "better" on these measures (i.e., show lower levels of self-control problems) than the rest of the time inconsistents.

[39] As noted in Footnote 25 in the main text, a considerable proportion of those categorized as having time consistent planning (i.e., with actual = predicted) had predicted and actual at the upper bound of the response scale (i.e., "more than 20 hours per week"). Some of these could have been time inconsistents that had actual and predicted that were above 20 hours per week but not equal to each other. Thus the estimate above that a quarter of those with TI planning = TI preferences were time consistent represents an upper bound.

[40] With the exception of stress for which those with TI preferences = TI planning have a lower median level by 0.4 of a standard deviation.

The large proportion of our sample with TI preferences equal to TI planning could also be explained by a lack of attention in answering questions on the part of participants (e.g., skipping through questions quickly and selecting answers without giving them enough thought). When we excluded respondents who were in the bottom or top 5(10)% in terms of response time to any of the questions used to measure TI preferences and planning, we find that the results were almost identical as for the full sample. In this restricted sample, 87% are in the same category for both TI preferences and planning, and 58% have the same value for TI preferences and planning. When we excluded only the bottom 5(10)% in terms of response time, the results are also almost identical (86% are in the same category for both TI preferences and planning, and 57% have the same value for TI preferences and planning). This suggests that the large proportion of our sample with TI preferences equal to TI planning is not driven by insufficient participant attention in responding.

Table A25: Proportion of participants in each of the nine different possible combinations of TI preferences and TI planning subcategories

	TI planning			
	Over-exercise	Time consistent	Under-exercise	Total
TI preferences				
- Over-exercise	0.41	0.02	0.02	0.45
- Time consistent	0.03	0.15	0.01	0.19
- Under-exercise	0.03	0.02	0.31	0.36
Total	0.48	0.18	0.34	1.00

Notes: The descriptive statistics given in this table are for the *prevalence sample* of 4,333 participants, as defined in Section 4.