Discussion Papers in Economics

No. 06/2018

HABIT FORMATION, OBESITY, AND CASH REWARDS

Boris Augurzky
RWI – Leibniz-Institut für Wirtschaftsforschung

Thomas K. Bauer
RWI, Ruhr-Universität Bochum (RUB), IZA Bonn

Arndt R. Reichert
The World Bank

Christoph M. Schmidt
RWI, Ruhr-Universität Bochum, IZA Bonn, CEPR London

Harald Tauchmann
University of Erlangen-Nürnberg, RWI, CINCH

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

HABIT FORMATION, OBESITY, AND CASH REWARDS

Boris Augurzky Thomas K. Bauer Arndt R. Reichert
Christoph M. Schmidt Harald Tauchmann*

April 26, 2018

Abstract

This paper examines weight loss and the formation of healthy habits through cash rewards in the context of a multi-phase randomized controlled trial involving 700 obese individuals. We find effects of monetary incentives for weight loss of up to EUR 300 on body weight during all experimental phases, including a period of a year and a half following the exposure to the financial rewards. We also find effects on healthy behavior during this follow-up phase. After the initial incentive period, we additionally provided participants who had lost a substantial amount of weight with monetary rewards of up to EUR 500. These had only short-term effects on body weight and healthy behavior. We argue that our findings are best explained by monetarily incentivized participants having formed healthy habits by the time the experiment ended and that only the speed of the transition to the new (health) equilibrium was affected by the additional rewards. Contrary to the pessimistic perspective presented in earlier empirical research on habit formation, our results suggest that a simple program relying on weight loss rewards can result in long-term health behavioral change.

JEL Codes: I12, I18, D03, C93
Keywords: field experiment, obesity, healthy behavior, habit formation, sustainability, monetary incentives

* Boris Augurzky, RWI – Leibniz-Institut für Wirtschaftsforschung; Thomas K. Bauer, RWI, Ruhr-Universität Bochum (RUB), IZA Bonn; Arndt R. Reichert, the World Bank; Christoph M. Schmidt, RWI, RUB, IZA Bonn, CEPR London; Harald Tauchmann, Friedrich-Alexander-Universität Erlangen-Nürnberg, RWI. We gratefully acknowledge the comments and suggestions of Ronald Bachmann, Alec Brandon, Alfredo Paloyo, John List, Michael Price, Imran Rasul, Johannes Rincke, Daniel Rogger, one anonymous World Bank Research Group staff member and the participants in the Experimental Lunch Seminar at the University of Chicago and the Empirical Economics Research Workshop at the Ludwig Maximilian Universität München, and the Workshop on Risky Health Behaviors at the University of Hamburg. The authors are furthermore grateful to the "Pakt für Forschung und Innovation" for funding and to Rüdiger Budde, Viktoria Frei, Karl-Heinz Herlitschke, Klaus Höhner, Julia Jochem, Hans Klimpel, Wendy Klimpel, Mark Kerßenfischer, Lionita Krepstakies, Claudia Lohkamp, Thomas Michael, Carina Mostert, Stephanie Nobis, Margarita Pivovarova, Gisela Schubert, Marlies Tepaß, and Franziska Valder for research assistance. We explicitly thank the medical rehabilitation clinics of the German Pension Insurance of the federal state of Baden-Württemberg and the Association of Pharmacists of Baden-Württemberg for their support of and dedication to this experiment. In particular, we acknowledge the support of Susan Eube, Michael Falentin, Ina Hofferberth, Marina Humburg, Silke Kohlenberg, Thomas Krohm, Max Lux, Tatjass Meier, Monika Reuss-Borst, Constanze Schaal, and Wolfgang Stiels. All the views expressed in this paper should be considered those of the authors alone and do not necessarily represent those of the World Bank. This paper merges empirical results presented in earlier working papers (Ruhr Economic Papers #368 and #530) and reports results from additional analyses based on the same data. This working paper is also available as Ruhr Economic Paper #750. All correspondence should be directed to Harald Tauchmann, FAU, Findelgasse 7/9, 90402 Nürnberg, Germany. E-mail: harald.tauchmann@fau.de.

1. Introduction

Monetary incentives that seek to align individual behavior with the social optimum have become increasingly popular across a wide range of areas – from contributions to public goods to education and health (Gneezy et al., 2011; De Geest and Dari-Mattiacci, 2013). The emerging literature on financial incentives to motivate behavioral change predominantly finds short-run effects but has not been conclusive as to whether individuals form new habits that would cause the induced changes to persist (habit formation hypothesis). The available evidence seems to point to a decay in the effect after the rewards are removed (e.g., Brandon et al., 2017). Some contributions even argue in favor of adverse effects once the incentives ended because they undermine intrinsic motivation (motivation crowding out hypothesis).[1]

We add to this literature by presenting experimental evidence that a simple program of monetary rewards for healthier behavior can indeed be successful in generating impacts that outlast the incentive scheme — even among populations like the obese, which seem to have characteristics that are hardly conducive to sustainable healthy behavior (Teixeira et al., 2015). We observe a reduction in body weight and favorable weight-related behavioral change in obese individuals throughout the time frame of our experiment, including one year and a half after exposure to the monetary incentives.

Our second contribution is to analyze whether the experimental data are consistent with the presence of habit formation, which would allow us to interpret the observed changes in body weight and weight-related behaviors as long-term effects.[2] The habit formation model explains the persistence of effects by an increase in the future marginal utility of healthy behavior, which is caused by a stock of habit capital that individuals build when engaging in the incentivized behavior (e.g., Becker and Murphy, 1988; Becker et al. 1991; Gneezy et al., 2011).[3] A contextual example of such behavior is physical exercise, which is directly linked to weight loss (the underlying economic model similarly applies to other activities, like healthy eating). If individuals

[1] Some evidence for motivation crowding out is reported in the psychological literature (e.g., Deci 1971; Warneken and Tomasello 2008). Promberger and Marteau (2013) survey the literature and find no convincing evidence for the crowding out in the case of incentivized health behaviors, arguing that those behaviors commonly have non-conducive characteristics, such as low baseline levels and little interpersonal conflict of interest.

[2] While the short run is defined as the period when the incentives are present, we differentiate between post-treatment effects (i.e., after the incentives were removed) and long-run effects. The latter term is used only if, consistent with the habit formation hypothesis, the results indicate that a new steady state is achieved.

[3] Gneezy et al. (2011) and Messinis (1999) directly link it to the seminal theory of rational addiction (Becker and Murphy, 1988; Becker et al., 1991), which itself builds on earlier economic research on habit formation, such as that by Pollak (1970), Ryder and Heal (1973), and Spinnewyn (1981). While the Becker and Murphy model is primarily concerned with detrimental activities, such as the consumption of harmful substances, and is mainly interested in the conditions for individuals to escape from such a "bad" equilibrium, it can easily be transferred to a case in which the addictive behavior is desirable and the focus is on the achievement of a "good" or healthy steady state.

did not exercise in the past, their habit stock is very small and physical exercise has a negative marginal utility that prevents them from exercising despite knowledge about the benefits associated with building fitness capital.[4]

Temporary financial incentives offset the negative marginal utility of physical exercise, causing individuals to start working out. Since the associated gains of fitness capital increase the marginal utility of exercise, individuals may follow a trajectory towards a new equilibrium in which they work out often enough to maintain their fitness capital, i.e., gains equal depreciation.[5] In our experiment the attainment of a new equilibrium characterized by a lower body weight and increased healthy behavior of the participants would be demonstrated if the estimated post-treatment effects remain at a certain level. Unfortunately, the necessary sample size to test statistically for the inter-temporal difference in the estimated effects on weight loss and healthy behavior is arguably difficult to achieve with a constrained budget.

We follow an alternative strategy of testing whether individuals indeed have achieved the healthier equilibrium within the time frame of our experiment. We examine the effects of additional financial incentives provided in the months following the (first) intervention period. According to the habit formation model, these additional incentives will cause an increased stock of habit capital at the end of the experiment if individuals are still on their path towards the healthier equilibrium.[6] Hence, in this scenario the theory predicts significant effects of these additional monetary incentives even after their removal. We do not find such post-treatment effects, which implies that individuals must have achieved the healthier equilibrium without those additional incentives. We argue that the effects of the weight loss rewards are long term in nature. We also assess alternative theoretical explanations for our experimental findings.

Empirically examining habit formation in relation to weight loss and weight-related healthy behavior in the obese is particularly interesting because many obese individuals have already failed in their weight loss attempts, and the majority of those who succeed soon regain weight (Crawford et al., 2000). This is theoretically consistent with little and, by implication, unsustainable initial habit capital accumulation. Our experiment documents that the provision of substantial monetary incentives and sufficiently ambitious weight loss targets overcomes the unsustainability of weight loss attempts. Our results therefore powerfully illustrate the relevance of the mechanism of habit formation.

[4] They operate analogously to a price cut in Becker et al. (1991).

[5] As a genuinely theoretical concept, habit stock is not observable. In the context of this experiment, the most closely related directly observable variable is arguably body weight, its change serving as a key outcome variable in the empirical analysis. We also use (self-reported) weight-related behaviors.

[6] They accelerate or delay the transition to the new equilibrium depending on whether the initial incentives caused a habit capital stock that falls behind or exceeds the steady state, respectively.

From a health policy perspective, the questions of whether cash incentives induce persistent behavioral change in obese individuals and the magnitude of the long-term effects are of great importance. Obesity is widely considered one of the major health problems of developed economies and is associated with considerable economic costs.[7] Relying on cash incentives that exert only short-term or small long-term effects would seem to be a very expensive way to battle the obesity pandemic. Our finding of substantive permanent effects makes temporary financial incentives arguably more appealing to policy makers.

The experiment was administered between spring 2010 and summer 2013 and involved seven hundred participants, who were recruited from four medical rehabilitation clinics. Two randomly assigned treatment groups were first offered EUR 150 (USD 188 in PPP) or EUR 300 (USD 376 in PPP) for achieving an individually assigned contractual target weight loss of between 6 and 8 percent within four months.[8] After the completion of the first intervention phase, those participants who had achieved a reduction in body weight that was presumably high enough to build sufficient habit capital for the achievement of a new health equilibrium[9] were randomly assigned to three experimental groups, again two treatment groups and a control group.[10] The individuals assigned to the treatment groups were promised EUR 250 (USD 313 in PPP) or EUR 500 (USD 627 in PPP) for achieving, for a second time, a body weight below their target weight ten months after enrollment. The body weight information and survey data of the members of the various experimental groups were collected at baseline and four months, ten months, and twenty-two months after the start of the experiment.

Studies examining the effects of interventions aimed to encourage healthier body weight in the obese regularly face the challenge of experiment dropout (e.g., Cawley and Price, 2013). In our experiment the attrition rates range between 25 percent (after four months) and 55 percent (after 22 months). We analyze the sensitivity of our results with respect to attrition. We find that they survive a wide range of assumptions about body weight development experienced by treatment and control dropouts. Only in very unfavorable scenarios concerning group-specific dropout patterns, however, the estimated effects tend to disappear. We further show that the

[7] Obesity increases morbidity, reduces life expectancy and deteriorates life satisfaction (for a comprehensive overview, see Sassi, 2010). Through negative effects on the probability of being employed (Morris, 2006; Reichert, 2015) and wages (Han et al., 2009), as well as positive effects on the risk of early retirement (Houston et al., 2008), the costs of absenteeism (Cawley et al., 2007) and lifetime health care expenditures (Bhattacharya et al., 2011), obesity represents a significant burden for welfare systems.
[8] We use the purchasing power parity exchange rate of 2011 provided by the OECD (2012).
[9] Due to a lack of clear theoretical guidance in relation to the required level of weight loss for the formation of healthy habits, we opt for 50 percent of the contractual weight loss.
[10] We exclude from the estimation of the post-treatment effects the participants who were randomly assigned to the financial incentive in the second phase. To account for likely selection effects of this exclusion rule, we run a weighted regression using the conditional selection probability. This probability is determined exclusively by the experimental design and therefore exogenous to the participants. Importantly, we can derive and use the correct information for the estimation sample weight of each observation (Section 3).

results remain unchanged when controlling for variables that describe the condition at the control weigh-in, capturing possible ways in which the participants may influence their measured body weight other than through weight loss. Hence, potential strategic behavior of the treated participants to achieve their targets is not able to explain our findings.

The remainder of this paper is organized as follows. The subsequent section summarizes the previous work, describes the experimental design and provides some descriptive statistics of the participants. Section 3 discusses the estimation strategy, while Section 4 presents the estimation results. Section 5 concludes.

2. Conceptual Framework, Existing Literature, and Experiment

2.1. Conceptual Framework

Our conceptual framework of habit formation draws on the mechanics of the seminal model of rational addiction (Becker and Murphy, 1988; Becker et al., 1991), which we adapt to weight-related behavior of obese persons. In Figure 1, we display the adapted model and illustrate the conditions required for the formation of weight-related habits as a result of temporary financial incentives.

The concave curve Ψ (*habit capital*) displays the frequency of an "addictive" activity, such as physical exercise (vertical axis), as a function of habit capital (horizontal axis). The more habit capital has been accumulated in the past, the higher is the marginal utility of this activity and, by implication, the desire to engage in it. The straight line[11] through the origin represents possible steady states for which habit capital generation equals habit capital depreciation (δ·*habit capital*). The vertical distance between Ψ and this line represents the net capital increase or decrease from one period to the next period for any realized activity level. The combinations (*habit capital*, activity**) and (*habit capital***, *activity***) are two possible equilibria in which the current habit capital stock induces an activity level that maintains the current stock of habit capital. The former combination represents an unstable equilibrium because even a small negative activity shock would result in a net loss of habit capital and thus ultimately a corner solution. In this state, individuals do not engage in the considered activity. Following the logic of the Becker and Murphy (1988) model, obese individuals are captured in this corner solution (A_0, with 0 denoting the pre-intervention period).

Now we consider a temporary intervention in the first period that incentivizes obese individuals to lose weight by means of a cash reward. Offsetting the negative marginal utility of beneficial weight-related activity (e.g., exercising or healthy eating), the monetary incentive shifts

[11] The properties of the model do not change if one relaxes the assumption of a constant rate of depreciation, that is, a linear steady-state curve, as long as one rules out the possibility that the marginal rate of depreciation exceeds the value of one.

the activity curve from Ψ to Ψ′. This results in a much higher activity level (point A_1). Crucially, this activity level is so large that habit capital generation exceeds its depreciation. Even if the activity is no longer incentivized and the activity curve shifts back to Ψ, individuals still end up gaining habit capital, which yields point A_2. This state implies further net generation (Figure 1a) of habit capital so that the individuals reach A_3. Effectively, the individuals are now on a trajectory towards the stable healthier equilibrium (A_1, A_2, A_3, A_4, …, A_∞), and a healthier habit is formed in the absence of further exogenous shocks. In this equilibrium, the activity level activity** is sustained.[12] If the initial incentive is relatively strong, the levels of habit capital and activity may temporarily exceed their equilibrium values. In this scenario (Figure 1b), the path towards the healthier equilibrium is characterized by endogenous continuous decay of habit capital (C_1, C_2, C_3, …, C_∞).

Next, we consider a second temporary financial incentive. If this incentive is stronger than in the first period, it shifts the activity level further upwards (i.e., from curve Ψ' to curve Ψ”) and changes the trajectory towards the healthier equilibrium (A_1, B_2, B_3, B_4, …, A_∞ versus A_1, A_2, A_3, A_4, …, A_∞ in the case of further net generation of habit capital; Figure 1a).[13] Thus, the model predicts that, if long-term habit formation is achieved during the observation period, we should be able to observe post-treatment effects of the initial incentives but no additional effect of the second incentive. Both incentives exhibiting post-treatment effects at the end of the experiment, on the contrary, would indicate that the adjustment to exogenous shocks is relatively slow and the individuals are still on their path towards the healthier steady state, with the caveat that individuals who receive the second incentive are transitioning at a different speed.[14]

2.2. Previous Empirical Studies

Our experiment is closely related to the experimental study by Charness and Gneezy (2009), who also examine the formation of healthy habits through financial incentives; in this case the rewards are provided for gym attendance. They find that the gym attendance of university students more

[12] The same argument applies if the initial incentive leads to temporary overshooting of the habit capital and the associated activity level (trajectories C_0, C_1, C_2, C_3, C_4, …, C_∞ versus C_0, C_1, C_2, D_3, D_4, …, C_∞; Figure 1b). The model also allows for the case in which the cash rewards are too small to initiate the formation of a health habit, while they may still have significant temporary effects on habit capital and activity. Moreover, the model covers the scenario in which no equilibrium with positive habit capital exists and hence temporary cash rewards will not have sustained effects irrespective of their size. The latter two cases are illustrated in Figures A1 and A2 in the Appendix.

[13] The size of the effect affects the speed at which the new equilibrium is achieved. For instance, a shift from curve Ψ to curve Ψ' will accelerate the process but at a slower pace than a shift from curve Ψ' to curve Ψ” (Figure 1a).

[14] In the logic of the model, the absence of post-treatment effects of the additional incentive is a sufficient yet not necessary condition for equilibrium values being observed by the end of the observation period. One may think, for instance, of a scenario in which the first incentive is too small to initiate the formation of a new habit (Figure A1) and only the additional incentive pushes the individual onto a trajectory towards the healthier equilibrium.

than doubles seven weeks after the provision of financial rewards contingent on exercising several times over the course of about one month. The estimated effects do not decline over the observation period of approximately four months following the intervention. The authors conclude that there is scope for financial intervention in habit formation, although they point out that concerns regarding a strong decline in exercising after removing the incentives cannot be completely rejected. Two subsequent papers examine roughly the same size and length of monetary rewards for gym attendance and find that the effects actually decay within a couple of months after incentive withdrawal (Acland and Levy, 2015; Royer et al., 2015).[15]

Our study differs from this earlier literature in several important ways. First, we focus on obese individuals, who may respond differently to financial incentives than mostly healthy-weight university students. In fact, previous studies find larger effects among participants who did not attend the gym before the experiments began. Second, with the purpose of helping them to succeed with their primary goal of weight loss, we directly incentivize weight loss as opposed to inputs to weight loss. Providing individuals with the option to choose the means of losing weight arguably leads to better (and more sustainable) weight loss results due to the possibility of combining inputs and the use of private information to optimize the mix of inputs. Third, the high (low) weight loss incentive in our experiment is 3.6 times (1.8 times) higher in terms of PPP. Fourth, the incentive period in our study is four times longer. Fifth, we also provide a subsample of the study population with high-powered monetary incentives in the subsequent period, which we exploit to analyze whether individuals reached a new equilibrium within the time frame of the experiment and assess alternative theories that are able to explain the lasting effects of monetary rewards.

Our paper also relates to Loewenstein et al. (2016), who find effects of monetary rewards for the consumption of fruits and vegetables in elementary students two months after the end of the intervention. Additional studies exist that examine monetary rewards for healthy behavior, especially weight loss (e.g., Volpp et al., 2008, 2009; Kullgren et al., 2013; Finkelstein et al., 2017). However, the experimental design choices arguably make the nature of the analyzed incentive schemes different from a genuine price change along the lines of Becker et al. (1991) and hence limit the ability of the studies to explain habit formation. They also abstain from examining the effects of the financial intervention on (weight-related) behavioral change.[16]

---

[15] Royer et al. (2015) additionally analyze the effectiveness of a self-funded commitment contract to improve the lasting effect of monetary incentives for exercise. After the removal of the monetary incentives, the participants are encouraged to deposit money that is refundable contingent on the continuation of regular exercise. The authors detect the effects of the deposit contract even one year after the incentive ends, concluding that a commitment option may allow users to lengthen their incentive period endogenously to reach critical thresholds in habit formation.

[16] We briefly discuss these experiments in Appendix 1.

2.3. Experimental Design and Implementation

To determine whether monetary incentives are effective means for obese people to lose weight sustainably and change their weight-related behavior, we conducted an experiment in cooperation with the Association of Pharmacists of Baden-Württemberg and four medical rehabilitation clinics operated by the German Pension Insurance of the federal state of Baden-Württemberg. The project was funded by the joint initiative for research and innovation (*Pakt für Forschung und Innovation*), which is part of the excellence initiative of the German government. Obese patients of the four rehabilitation clinics were invited to participate in the experiment in the final week of their rehabilitation stay, which included a weight loss program that varied from clinic to clinic. Only patients with a BMI above 30 at admission, aged between 18 and 75 years, and registered as a resident in the German federal state of Baden-Württemberg were eligible to participate in the experiment. Patients with considerable language barriers, psychological and eating disorders, a tumor disease within the last five years, a history of alcohol and drug abuse, or serious general diseases, as well as patients who were pregnant were excluded from the experiment. All participants were informed about the procedures of the experiment through handouts, and clinic personnel gave personal instructions. The study protocol was approved by the ethics commission of the Chamber of Medical Doctors of Baden-Württemberg.

Conditional on the agreement to participate in the experiment, the staff of the rehabilitation clinics conducted baseline measurements of several medical variables of the patients, such as the body mass index (BMI), blood glucose level, and cholesterol level. The participants further answered a detailed questionnaire related to their socioeconomic background, additional health outcomes, and healthy behavior. In relation to the latter, the participants were asked a battery of questions on (i) food consumption (vegetables, fruits, whole grains, and meat) and (ii) attention to eating behaviors (food composition, calorie content, and eating speed), as well as (iii) eating habits (frequency of cooking at home, eating at fixed times together with household members, and snacking between meals). The questionnaire also covered the frequency and intensity of exercise.[17] Moreover, the physician in charge assigned an individual weight loss target to the participants that they were supposed to realize within four months after leaving the clinic. The target was chosen to lie between 6 and 8 percent of the participants' current body weight, which moderately exceeds the critical threshold associated with beneficial health effects (Vidal, 2002).

The experiment consisted of four phases (see Figure 2).[18] Phase zero was the rehabilitation stay. After being discharged from the clinic, the participants entered phase one of

---

[17] The questionnaire was designed in close collaboration with the medical staff of the rehabilitation clinic. The questions reflect the learning content of the behavioral change training given to the participants at the clinic.

[18] We present a flow chart in Figure A3 in the Appendix.

the experiment, which was the first intervention period of four months. It was followed by the second intervention period of six months (phase two) and a twelve-month follow-up/post-intervention period (phase three). Two randomizations took place after the discharge from the clinic (for which medical staff members were blinded): at the start of phase one and at the start of phase two. The former was carried out stratified by the clinics and without replacement within blocks of 51 participants. Based on this randomization procedure, the participants were assigned to one of three groups with equal probability: either the control group or one of two treatment groups. While the members of the control group were not promised any reward for achieving their weight loss target, the members of the treatment groups were promised up to EUR 150 (henceforth called *Group 150*) or EUR 300 (henceforth called *Group 300*).

In the second randomization, all participants who substantially (≥ 50 percent of the targeted weight loss) had lost weight by the end of the first phase (irrespective of group assignment) were eligible. Consequently the analysis is restricted to the participants from the phase one premium groups who met this criterion. Randomization (without replacement and without stratification by the clinics) was used to produce three additional experimental groups with equal shares of participants. In this phase of the experiment, two premium groups were promised up to EUR 250 (henceforth called *Group 250*) or EUR 500 (henceforth called *Group 500*) for achieving the same target weight. These additional incentives had not been announced at the beginning of the experiment. The participants assigned to the control group were not informed that a second randomization had taken place. All the participants were told to ensure that their weight did not exceed the individually assigned target weight during phase two and phase three.

The members of the premium groups were paid the full bonus if they reached or even exceeded their weight loss target at the end of the respective phase. Once the achieved weight loss exceeded 50 percent, they were rewarded proportionally to the maximum reward. As an example, consider a participant with an initial body weight of 120 kg (264.5 lbs) and a target weight loss of 8.4 kg (18.5 lbs) who loses 6 kg (13.2 lbs) within four months and is able to maintain her reduced body weight during phase two. As a member of the control group in both phases, she receives no premium. As a member of the treatment group in phase one, she obtains EUR 107 (USD 134 in PPP) if she is in *Group 150* or EUR 214 (USD 268 in PPP) if she is in *Group 300*. At the end of phase two, she receives another EUR 179 (USD 224 in PPP) or EUR 357 (USD 477 in PPP), depending on whether she is a member of *Group 250* or *Group 500*. In contrast, if she loses only 4.1 kg (9 lbs), she receives no reward regardless of her group assignment and phase. For a weight loss of 8.4 kg (and maintaining the target weight later on), she receives the entire group-specific premiums.

The participants were informed by regular mail about their maximum possible premium (this does not apply to the members of the control groups) and about the week during which they had to attend the weigh-in at a pharmacy.[19] Since the participants spent the intervention periods outside the medical rehabilitation clinic, interactions between the participants were very unlikely. Thus, we do not expect a perception of unfairness that may be associated with randomization. Most importantly, the control group participants should not be affected by the treatment status of other participants.[20]

We asked the participants with any health complaints throughout the experiment to consult their general practitioner or the rehabilitation clinic. Two weeks prior to the end of each experimental phase, a reminder for the control measurement of their body weight was sent to the participants. The letter contained a questionnaire with the same set of questions on time-varying variables as the one collected at the initiation of the experiment. To avoid relying on self-reported weight, the reminder indicated that the participants should attend a nearby pharmacy for the control measurement. The pharmacies had been called by the project staff beforehand to ask for their participation. By assigning participants to specified pharmacies, we ruled out the possibility of the members of the treatment group switching from one pharmacy to another to take advantage of probable measurement errors of the scales, i.e., strategic behavior to achieve their targets.

Dropping out of the experiment occurred in two ways. First, some participants left the experiment by actively canceling their participation. Second, a larger number did not return the required documents at the end of an experimental phase. To reduce the dropout rate in the experiment, all participants whose documents were still pending three working days after the specified week were contacted by phone. We encouraged them to make up for the weigh-in and to send in the documents. All participants received EUR 25 (US$31 in PPP) if they returned the documents, regardless of their weight loss success and group assignment. The premiums were still paid if the date of measurement indicated by the pharmacist was within 14 days after the end of the supposed weigh-in week. We did not follow up dropouts in subsequent periods. For instance, a participant who discontinued experiment participation at the end of phase one was not contacted at the end of phase two and phase three.

2.4. Participants

The recruitment of a total number of 700 participants took place between March 2010 and August 2011. Of these individuals, five had to be excluded from the trial because of a missing consent form, becoming pregnant, developing cancer, or internal documentation problems.[21] The last participant finished phase three (follow-up) at the end of July 2013.

[19] Participants could postpone the date of measurement or move it forwards by means of an early phone call.
[20] See Angrist and Lavy (2009) for a similar argument in the context of a within-school randomized trial.
[21] The results are robust with respect to treating these individuals as dropouts in the sensitivity checks described in Section 3.

Table 1 provides descriptive statistics of the study population (see Section 3 for a definition of the variables). The average body weight at the start of the experiment (at the end of the rehabilitation stay) was 113.0 kg (249.4 lbs) or a BMI score[22] of 37.6. The level of self-reported healthy behavior was rather low. Only 13.5 percent and 12.9 percent of the sample reported eating healthily and exhibited good eating habits, respectively, whereas only about 4 percent used unhealthy ingredients. Fewer than a third of the participants engaged in exercise with moderate intensity.

About 68 percent of the participants were men, and 21 percent had a migration background.[23] For most patients of the co-operating clinics, medical rehabilitation was paid for by the German pension fund, the predominant goal of which is to avoid work disability and early retirement. Since there are many obese people in the overall population who have already retired, our study population oversamples persons who are available for the labor market.

The four rehabilitation clinics involved in the trial are located in different towns. About 42 percent of the participants were recruited by the clinic in Bad Mergentheim, 33 percent in Bad Kissingen, 18 percent in Isny, and roughly 7 percent in Glottertal. The clinics in Bad Mergentheim and Isny primarily focus on orthopedic interventions, while the clinics in Bad Kissingen and Glottertal specialize in gastroenterology as well as endocrinology and psychosomatic disorders, respectively. Many participants attended the clinics because of diagnoses other than adiposity, although their symptoms were related to their body weight. All the participants were medically indicated to lose weight.

3. Hypotheses and Methods

In our empirical analysis, we aim to analyze weight loss and the formation of healthy habits through cash rewards. Our objective is to examine whether individuals achieved a new equilibrium characterized by a lower body weight and increased healthy behavior within the finite time frame of our experiment. Informed by our conceptual framework, presented in Section 2, we formulate four testable null hypotheses to examine the presence of habit formation in our data: the financial incentives for weight loss had no effect on body weight and healthy behavior while they were in place (i) and during the observation period following their removal (ii); even higher additional monetary rewards likewise did not exert an effect on body weight and healthy behavior during the intervention period (iii) and after the intervention had ended (iv). Rejection of the second hypothesis is not sufficient to conclude that the point estimates represent long-term

[22] Figure A4 shows the distribution of the BMI over time.

[23] These shares are substantially lower than the corresponding averages for obese individuals in Germany, which we obtain using the German Socioeconomic Panel (SOEP) – a representative panel of German households (Table 1). The mean age of the study population (48 years) lies about ten years below the average age of obese individuals in Germany, while the share of employed participants (82 percent) is almost twice as large. Only the share of married participants does not deviate considerably from the respective share of obese people in Germany.

effects of the financial incentives, since individuals might still be following the path towards the new equilibrium. The examination of the latter two hypotheses will enable a better interpretation of the main findings. In particular, rejecting the fourth hypothesis would reveal that the study participants are still on their path towards a new equilibrium at the end of the observation period of the experiment, whereas failure to reject it would indicate that they have already achieved the steady state.

Comparisons

Concerning the first two hypotheses, we examine whether individuals who were exposed to financial incentives during phase one (*Group 150* and *Group 300*) lost more weight than the control group four, ten and 22 months after the start of the experiment. If we find significant differences in weight change across the experimental groups between the start of the intervention period and six as well as 18 months after the intervention, respectively, the effects of monetary rewards for weight loss will not appear to be only short term.

To examine the third and fourth hypotheses, we compare the mean weight loss over phases two and three between the individuals who were promised the additional rewards (*Group 250* and *Group 500*) and the control group. We investigate weight loss between the fifth month and the tenth month and between the fifth month and the twenty-second month. The latter time window allows us to test the fourth hypothesis, which addresses the question of whether the estimated short-run effects of (the initial) weight loss rewards can theoretically be considered as long-term effects, i.e., incentivized individuals achieved a new steady state during the observation period. Only members of the premium groups of the previous period who successfully reduced their body weight are taken into account. This means that we exclude (successful) individuals who belonged to the control group in phase one of the experiment and received monetary incentives for the first time in phase two.

In addition to the percentage weight loss of the participants, we use the first difference of four healthy behavior dummy variables as outcome variables. The first concerns unhealthy ingredients, taking the value one if the participants reported infrequent consumption of vegetables, fruits or whole grains (less than once a week) and frequent consumption of meat (every day). The second approximates healthy eating behavior. We assign the value one if the participants reported a tendency (more than three on a scale from one to six) to pay attention to food composition and calorie content as well as eating speed. The third binary variable indicates good eating habits, which are defined as frequently cooking at home, eating at fixed times together with household members and not snacking between meals (values above three on a scale from one to six for all three dimensions). The fourth variable addresses engaging in exercise and takes the value one if the participants reported exercising more than once per week with moderate intensity (less than four on a scale from one to six).

Inverse probability weighting

To isolate the post-treatment effects of the first intervention (not the short-term effects) and to avoid the estimation of a combined effect of the first and the second intervention, we exclude individuals from the analysis who were promised monetary rewards for realizing their target weight ten months into the experiment. However, a comparison of weight development across the experimental groups of the first incentive may give a disproportionately large weight to individuals who failed to reach the target weight. A simple inverse probability weighting estimator is able to restore the original distribution (Wooldridge, 2002). Importantly, due to the experimental design, the probability that an individual will be excluded from the sample is exogenous, i.e., the ignorability assumption is fulfilled. Moreover, we know the exact conditional selection probability or, in other words, the correct estimation sample weight for each observation. As opposed to standard observational studies, there is no need to estimate it. We refer to Section A.1 in the Appendix for a more technical description of the estimation method.

Multivariate regression (part one of the sensitivity analysis)

Since the present analysis rests on data generated in the course of a randomized trial, in principle, simply comparing means across treatment and control groups yields an unbiased estimate of the causal effects because randomization ensures that the different experimental groups differ only in terms of receiving the treatment. To address random imbalance of individual characteristics and potential strategic behavior of the participants to achieve their target, we rely on multivariate regressions in a robustness check. As covariates, we include age, gender, month of recruitment to the experiment, and variables that relate to the weigh-in at the pharmacies. We asked the pharmacists to indicate whether the participants' last food intake was more than half an hour ago, or more than two hours ago, whether they were wearing shoes (and if so whether these were heavy), a pullover, or long trousers, and whether they attended the control weigh-in within the specified time. An additional set of dummy variables captures the timing of the control weigh-in. We specifically use indicators for whether the participants were weighed prior to the specified date of measurement, within the right week (reference category), or afterwards (one week, two weeks, or more than two weeks later). The variables that describe the condition at the control weigh-in allow us to capture possible ways in which the participants may have influenced their measured body weight other than through weight loss. This may be particularly relevant to the analysis of monetary rewards at the end of both intervention periods because members of the treatment groups may have behaved strategically to achieve their target and increase their bonus.[24]

[24] This does not equivalently apply to the effects after the rewards have been removed, i.e., members of the treatment groups no longer had incentives to behave strategically.

Except for the variables related to the weigh-in, all the variables enter the analysis as pre-treatment values. Following a standard approach (e.g., Morris, 2006; Spenkuch, 2012), we deal with missing values in covariates by replacing them with zero and including additional dummy variables indicating missing values. Only the gender of the participants is imputed using predictions from a probit regression of the variable on relevant individual characteristics. Imputation is preferred to excluding observations with missing information because the latter would reduce the sample size substantially, despite the fact that the share of missing values is rather low for most covariates.

Accounting for experiment attrition (part two of the sensitivity analysis)

Despite substantial efforts to keep the attrition rates low, among the initial 695 participants, 177 dropped out during phase one, another 106 did not attend the weigh-in at the end of phase two, and an additional 96 dropped out during phase three. If the dropping out from the experiment was random, our estimates of the effects of the financial incentives would be unbiased. This seems to be the case when analyzing the second hypothesis because the members of the groups receiving the additional incentive did not have significantly lower dropout rates than the members of the control group (see Table 5).[25] We are concerned, however, that the termination of experiment participation is endogenous in the case of the first incentive. The reason is that we observe lower cumulative dropout rates four, ten, and 22 months after the start of the experiment for members of the weight loss premium groups in the first period relative to the members of the respective control group. This is explained by the fact that the treated participants had financial motives to comply with the weight measurement after four months because they would receive a premium if they could prove weight loss success.[26] Indeed, *Group 300* members continued with the experiment significantly more often than the members of the control group.

Importantly, the difference in attrition rates did not increase over time (see Table 2). This means that the marginal attrition rates in phase two and phase three did not significantly differ between *Group 300*, *Group 150*, and the control group. This seems plausible considering that, compared with the control group, the period-one weight loss reward groups no longer had greater financial motives for experiment continuation after the incentives had been removed. We attribute the perpetuation of attrition to the experimental protocol that involved no longer following up participants once they had missed a weigh-in. While our experimental results may be biased in the presence of non-random dropout, the concern is primarily attrition during phase one of the experiment and much less during subsequent periods. We use several estimation methods in the sensitivity analysis in relation to the problem of sample attrition that occurs during

[25] To strengthen our argument for the type of dropout being indifferent between the control and the treatment groups, we show balance of the pretreatment covariates across all three experimental groups in the selected sample after dropout in Table A1 in the Appendix.

[26] Please note that we treat participants who did not show up in the pharmacy for weigh-ins as dropouts.

the first four months of the experiment (treating attrition in post-intervention periods as random).

The first approach aims to address the selection problem by using self-reported information on body weight. Individuals with pending documents were called by phone and asked to make up for the weigh-in at the pharmacy. During the course of the phone call, these participants were also asked about their current body weight. We argue that the participants had no financial incentive to misreport their body weight, especially after the first four months, as monetary rewards for weight loss were no longer promised (and the participants who belonged to the premium groups in phase two are excluded from the analysis). Using the self-reported weight for individuals without measured weight information substantially increases the number of observations and, by implication, reduces the dropout rate in the estimation of the incentive effects. This does not apply to the healthy behavior variables, which the phone calls did not cover.

Our second and third approaches consist of methods that estimate the treatment effects under extreme assumptions about the distorting effect induced by non-random dropout from the experiment. On the one hand, we follow the baseline carried forward approach, which is the most commonly used method in the medical literature and is also pursued more recently in the economic literature (e.g., Cawley and Price, 2013).[27] It assumes that the outcome of the participants who dropped out of the experiment remained at the baseline level. However, this may not be considered a very conservative assumption in the case of the first outcome variable because we observe individuals deviating from the baseline value in both directions.[28] In addition, assuming the same level of weight reduction for dropouts from the treatment and the control groups is unlikely to be appropriate since experiment continuation may be a function of weight loss during the intervention period. We therefore additionally vary the assumption about changes in body weight during the first intervention period over the entire distribution of the outcome variable between the treatment and the control group.[29] In doing so, we aim to understand how extreme the assumptions about heterogeneity in dropping out need to be for the incentive effects on weight loss to disappear.[30] Here the statistical inference is based on the variance of observed, not hypothesized, weight loss.

[27] In the medical literature, it is frequently referred to as the intention-to-treat approach (see, e.g., Hollis and Campbell 1999), although, in the present context, rather than selection into treatment, selection into the estimation sample is the relevant problem.

[28] A similar problem applies to the bounds proposed by Horowitz and Manski (2000).

[29] We cannot rule out the possibility that the weight change among the dropouts exceeds the range of changes in body weight observed in the data. However, the purpose of this exercise is to simulate how extreme the assumptions about the process that drives sample attrition need to be to neutralize any effect found in a naïve analysis that assumes that sample attrition is purely random. We expect that assuming maximum observed weight loss for control group dropouts and maximum observed weight gain for treatment group dropouts would drastically reduce the treatment effect.

[30] Assumed rather than actual levels of weight loss enter the analysis in the case of dropouts. To account for this artificial inflation of the size of the estimation sample, we calculate the standard errors on the basis of observed changes in body weight.

On the other hand, we rely on the trimming procedure proposed by Lee (2009) to obtain bounds for the estimated treatment effects. This procedure trims the distribution of the outcome variable for the experimental group (treatment or control) that suffers less from sample attrition during the intervention period (which has relatively more participants with information on the outcome variables, i.e., "excess observations") in the quantile that corresponds to the share of excess observations in this group. Then the difference in means for the trimmed sample of one group and the untrimmed sample of the other group yields the estimated treatment effect bound. This method assumes that "excess observations" are those with either the most favorable or the least favorable development of the outcome variables (e.g., those with the largest and smallest percentages of lost body weight). This yields a lower and an upper bound of the treatment effect, respectively, depending on whether the trimming is performed from below or from above. According to Lee (2009), this procedure provides bounds for the average treatment effect among always compliers.

4. Estimation Results

This section first presents the results regarding the short-run effects of monetary incentives to lose weight and their effects after the incentives were removed before discussing the results of the additional monetary rewards.

4.1. Effects During and After Exposure to Monetary Rewards

Before investigating the effects of the weight loss premiums, we provide reassurance that the randomization procedure that assigned the participants to the experimental groups worked properly and that the inverse probability weighting estimator restores the original distribution of observations. Table 2 displays the descriptive statistics of the individual characteristics of the study population for each arm of the experiment before the first intervention was implemented. Odd column numbers refer to the population that is used for estimating the short-term incentive effects, while even column numbers indicate the inversely weighted population that is relevant to estimating the effects after the incentives' removal (see Section 3).

Almost all the variables appear to be balanced between the experimental groups, including body weight and healthy behaviors. This holds for both populations, which indicates that inverse probability weighting adequately approximates the original distribution of the experiment population. The only individual characteristic that does not seem to be balanced across the groups after inverse probability weighting is age, which is significantly smaller in *Group 300* than in the control group. A weakly significant deviation between the two incentive groups is found with respect to the use of unhealthy ingredients. Overall, these few significant deviations are in line with a conventional type one error probability. The lower panel of Table 2 displays figures concerning the previously discussed attrition issue (i.e., statistically significant heterogeneity in cumulative dropout rates across experimental groups, with *Group 300* exhibiting the smallest and the control group the largest rates).

Effects on Body Weight

The upper panel of Table 3 displays the results for weight loss, measured as percentages, over the first intervention period. Members of all the groups, on average, reduced their body weight, and the weight loss was significantly larger in the treatment groups.[31] There is no significant difference between the two treatment groups, which suggests that mere exposure to a monetary incentive matters more than the size of the reward. Comparing the entire weight loss distribution across groups corroborates those findings (Figure 3). According to Kolmogorov–Smirnov tests (e.g., Heathcote et al., 2010), the weight loss distributions of the treatment groups stochastically dominate the distribution of the control group. This is a strong result, as it means that receiving financial incentives increases the likelihood of realizing any level of weight reduction. Comparing the distributions of the two treatment groups with each other, the Kolmogorov–Smirnov test becomes statistically insignificant.[32]

The results for the effects after the withdrawal of the weight loss rewards are displayed in the middle and lower panels of Table 3. Six and 18 months after the weight loss incentives ended, all experimental groups, on average, continued to weigh less than at the start of the experiment. The weight loss of *Group 300* (but not of *Group 150*) is significantly larger than the weight loss of the control group in both post-intervention periods. The effect even appears to increase from phase one to phase three of the experiment, although we lack the statistical power to detect this difference in point estimates as argued in the introduction (this result is not shown in the table).

We do not observe any statistically significant difference in weight loss across the two treatment groups in the two phases after the incentives were removed. Pooling both treatment groups together yields a statistically significant difference from the control group at the end of phase two (p-value of 0.07) and a statistically insignificant difference at the end of phase three (p-value of 0.11). Figures 4 and 5 show the distribution of weight loss by experimental group at the end of both post-intervention periods, indicating that the effects of the monetary incentives are not primarily due to a small number of participants with very large changes in body weight. The analysis hence suggests that weight loss incentives exert not only short-term effects but also effects beyond the end of the intervention period.

[31] In quantitative terms, the weight loss amounts to roughly 5 percent in both treatment groups and hence just meets the threshold that is frequently mentioned in the literature as being required for inducing health improvements (Vidal, 2002; Rakel and Rakel, 2015). The weight loss of the control group may be attributable to the lasting effects of the clinic weight loss program, the effect of receiving a specified weight loss target from a physician or self-selection into the experiment of individuals motivated to lose weight.

[32] Paloyo et al. (2015) build on this result and examine heterogeneity in the short-term effects of financial incentives for weight loss, using data from the same experiment. Reichert (2015) uses this short-term variation in weight loss caused by the monetary rewards for the estimation of the causal effect of BMI growth on employment.

The estimation results appear to be robust when individual characteristics and variables related to the weigh-in are taken into account in a multivariate regression (Table A2 in the Appendix). The coefficients of *Group 300* are statistically significant (one-sided test) in the intervention period and the first post-intervention period. It is significant at the 7 percent level at the end of the observation period. Moreover, the point estimates all have the expected positive sign and are of a similar magnitude to the difference in means.

Regarding the sensitivity of our results to accounting for non-random sample attrition, we observe that including observations with self-reported body weight information in the analysis does not remarkably alter the results (Table 4). The point estimates for the short-term effects on weight loss are somewhat smaller but still highly significant for both incentive groups. For *Group 300*, a significant deviation from the average weight loss in the control group after 10 and 22 months is also confirmed (p-values of 0.06 and 0.02, respectively). Although there is a weight loss differential between the two premium groups at the end of phase three that is significant at the 8 percent level when including self-reported information, the results for the pooled treatment are roughly the same as before (not displayed in the table).

The pre-intervention weight carried forward estimates (the circle on the 45-degree line in Figures 6a–f; Table 4) also yield significantly positive effects of monetary weight loss rewards, except for the post-treatment effects of the EUR 150 rewards. Our sensitivity analysis further shows that positive incentive effects on weight loss are even consistent with strong hypothetical heterogeneity in attrition between the treatment groups and the control group (see for instance Figures 6b and 6d). We still find statistically significant effects if we assume that dropouts from the control group, on average, lost more weight than their counterparts from the intervention group (the shaded area above the 45-degree line in the graphs). In fact, the short-term effects of both the reward sizes and the post-treatment effects of the EUR 300 reward survive scenarios with pronounced group-specific heterogeneity in assumed weight loss of the dropouts. For instance, we observe a positive and statistically significant difference between *Group 300* and the control group at the end of the intervention period if we assume that dropouts from the control group, on average, lost as much weight as the 70th percentile of the observed weight loss distribution while treatment group dropouts lost as little as the 25th percentile.

This general picture is confirmed by the results obtained from the trimming procedure proposed by Lee (2009). As displayed in Table 4, the estimated conservative bounds for the effects at different points in time are, with a single exception (*Group 150*, months 0–22), positive, indicating that even extremely selective dropout will not generate the observed pattern in mean weight loss.[33] However, this does not hold for the corresponding estimated confidence intervals for the effects after incentive removal, which overlap the value of zero.

Overall, our results are in line with those obtained by Charness and Gneezy (2009), who show that financial incentives to exercise have lasting positive effects. Our findings clearly argue against motivation crowding out playing a major role in the post-intervention effects of cash rewards on body weight.[34] Importantly, they tend to support the habit formation hypothesis. In fact, our results are consistent with individuals having achieved either the healthier equilibrium or temporary overshooting, i.e., pushing individuals beyond the steady state.[35] The analysis of the additional financial incentive in the second phase will indicate which of the scenarios applies.

Effects on Healthy Behaviors

To shed light on the behavioral change that is necessary for weight reduction as well, we present the estimated incentive effects on the improvement of "using unhealthy ingredients," "healthy eating," "good eating habits" and "frequency of exercise." Analogous to weight loss, a numerically positive change in healthy behavior indicates an "improvement,", i.e., behavioral change that works towards the reduction of body weight. We find for all the experimental groups significant and considerable behavioral change during the intervention period in terms of improved eating habits and particularly healthier eating but not in terms of using fewer unhealthy ingredients or engaging in more exercise (Table 3). However, the improvements are rather similar across the experimental groups, which points to the absence of any significant short-term effects of weight reduction incentives on healthy behavior variables. Adjusting for covariates (Table A2) and taking selective dropout into account, through the pre-intervention outcome carried forward analyses and estimation of conservative treatment effect bounds, do not alter this finding (Table 4).[36]

The estimated effects after incentive removal exhibit a rather different pattern. For *Group 300*, we find significant deviations from the control group with respect to using fewer unhealthy

[33] Only the lower limit is reported in Table 5. The unreported upper limits are all positive and large. The estimated optimistic (i.e., in absolute terms upper) bounds indicate that non-random dropout might mask very strong treatment effects. For instance, the upper bound for the deviation of mean weight loss between *Group 300* and the control group after 22 months is as large as 6.1 percentage points. We abstain from reporting these results because (i) assuming that self-selection causes a bias away from zero seems to be less likely than selection bias towards zero and (ii) the focus of the analysis is on documenting habit formation rather than bounding the size of the effect estimate from above.

[34] This is remarkable evidence against the motivation crowding-out theory because our experimental set-up rather tended to favor the odds of finding results that are consistent with the crowding out of intrinsic motivation. The participants were recruited during a medical rehabilitation stay involving obesity treatment (Section 2). Medical rehabilitation may build intrinsic motivation for weight loss that can be destroyed through extrinsic rewards.

[35] Weight regain in both reward groups indicates a scenario in which the monetary incentives may have yielded a habit capital stock at the end of the intervention period that exceeded the equilibrium.

[36] Aiming to avoid reporting too many results, we abstain from trying various different assumptions of the group-specific outcome of dropout but confine ourselves to reporting only the results of the pre-intervention outcome carried forward method.

ingredients and exercising more at the end of both post-treatment periods. In addition, we observe a positive treatment effect on healthier eating at 22 months.[37] Compared with the control group, *Group 150* exhibits improvements in eating habits (p-value below 0.1) and the frequency of exercise (p-value below 0.05) six months after the end of the intervention period. Both effects are no longer statistically significant 12 months thereafter.

Regression analyses that take covariates into account qualitatively yield the same results (Table A2). As displayed in Table 4, the pre-intervention outcome carried forward estimates are also similar to the results from the comparison of raw means. However, for the EUR 300 reward, the effects on physical exercise become somewhat smaller and the associated p-values exceed the 10 percent threshold of statistical significance. Though very conservative, Lee's bound estimates confirm the positive post-treatment effects of the larger incentive on fewer unhealthy ingredients and physical exercise six months after its removal. For physical exercise, this also holds after 22 months. The confidence intervals include the value of zero. The post-intervention effects of the EUR 150 reward appear to be sensitive with respect to sample attrition.

4.2. Equilibrium Analysis

The short-term and post-intervention effects of the additional incentives to which successful premium group members of the previous period were exposed between month five and month ten of the experiment (phase two) allow us to determine whether individuals achieved the healthier equilibrium within the observation period and thus whether our post-intervention results above can be interpreted as long-term effects.

The upper panel of Table 5 presents descriptive statistics of the estimation sample and each experimental group (Columns 2–4). Except for the share of female participants, all background characteristics appear to be balanced between the different experimental groups. As for body weight, the majority of the participants achieved their individually assigned target weight after four months (see the middle panel of Table 5). The remaining participants achieved more than 50 percent of the weight loss target during the intervention period. Importantly, the other outcome variables at both the start and the end of the first intervention are also uncorrelated with treatment. The only exception is the use of unhealthy ingredients, which was significantly more

[37] This difference between the short-term effects and the effects after incentive removal can partly be explained by the low granularity of the available (qualitative) information about healthy behavior. All the participants – irrespective of their assignment to an experimental group – accepted a weight loss target. It is not surprising that most of them made an effort to realize it, with the consequence that in qualitative terms we observe largely uniform behavioral change across the groups during phase one. In quantitative terms, however, the members of the incentive groups may still have struggled much harder, which may not be captured in the data. This possibly translated into a higher likelihood of continuing with healthy behavior at the two later stages of the experiment. The reason is that, according to the habit formation theory, maintained behavioral change requires the creation of sufficient habit capital, which, in turn, is a function of the intensity with which an activity was carried out in the past.

prevalent in *Group 250* than in the control group (*p*-value of 0.037 percent) and *Group 500* (*p*-value of 0.056 percent).

In the lower panel of Table 5, we show the average attrition rates in our sample. The attrition rate among the participants in the second intervention is lower than that of the average participant in the first intervention (see Table 2). We do not observe any structural attrition pattern for the additional incentives, i.e., the treatment groups are not significantly more likely to continue the experiment than the control group. For this reason, we abstain from extensively discussing of the sensitivity of the results for the second intervention with respect to sample attrition.[38]

The results of the short-term effects of the additional monetary incentives are presented in the upper panel of Table 6. Both treatment groups have a significantly larger weight loss than the control group.[39] There is no statistically significant difference between the two treatment groups. The difference in weight loss between the two treatment groups pooled together and the control group is significant. Figure 7 displays the distributions of weight loss achieved during the intervention period by the experimental groups. Concerning health-related behavior at the end of the intervention period of the additional incentives, we also find significant differences in healthier eating and improved eating habits between the treatment groups and the control group. We do not observe any effects on the use of unhealthy ingredients and the frequency of physical exercise. Multivariate regressions that take into account individual characteristics and variables related to the weigh-in confirm the previous results for the additional monetary rewards promised during the second phase (Table A3 in the Appendix). As a single exception, the covariate-adjusted results indicate that the second incentives had a beneficial short-term effect on good eating habits.

After 22 months, we no longer observe statistically significant differences across the different experimental groups of phase two.[40] Figure 8 shows the distribution of weight loss at the end of the experiment for individuals exposed to the additional incentives and individuals not promised the additional monetary incentives. These findings are in line with the results regarding healthy behavior. The considered behaviors no longer exhibit any significant change among the experimental groups. Furthermore, we do not see post-treatment effects of the additional incentives on weight-related behavior (with a weakly significant effect for the smaller incentive on the use of unhealthy ingredients as the sole exception).

[38] All the point estimates are robust with respect to the various approaches to taking sample attrition into account, corroborating that sample selection is not a major issue here (the results are available on request).
[39] The members of the control group significantly regained body weight during phase two of the experiment. In contrast, the incentivized participants did not significantly regain weight. *Group 250* even lost slightly more weight.
[40] The average weight loss over the entire observation period among the successful premium group members of the second phase is statistically significant for all experimental groups, although it tends to be smaller than in the first four months of the experiment (compare Table 5 with Table 6).

These results suggest that there were no effects of the second incentives after the intervention had ended. This is consistent with individuals who were exposed to the weight loss rewards having reached the healthier equilibrium during the observation period of this experiment and being temporarily pushed beyond the steady state due to the standard price effect of the additional incentive. Albeit less pronounced, the finding regarding weight-related behaviors supports this interpretation. On the basis of the pattern found for the post-intervention effects of the two differently timed cash incentives, we can interpret the effects of the initial rewards as long term due to habit formation.

4.3. Assessing Alternative Interpretations of the Experimental Findings

The goal of our theoretical framework was to provide a parsimonious economic model that is able to explain (i) why obese individuals struggle with a persistent reduction of body weight and (ii) how monetary incentives can induce the formation of healthy habits as well as initiating persistent weight change. The empirical results prove to be consistent with the predictions of this theoretical model, which suggests the interpretation of the results as evidence for habit formation. However, competing theories can also explain lasting effects of monetary rewards. If these alternative explanations are consistent with our empirical results, interpreting them as evidence for the habit formation hypothesis would be questionable. In this subsection, we discuss whether our results are equally aligned with other theories, specifically the reference-dependent preference theory and the theory of motivation crowding-in.

First, one can argue that individuals who have reference-dependent preferences might exhibit sustained behavioral change if exposed to monetary incentives. Assuming that, at any given time, the reference point is the body weight in the previous period, individuals who were successful in the first phase of the experiment will aim to maintain the new reference level. This would imply that the effects of the monetary incentives for weight loss persist. As the reference point moves, we expect this pattern to apply analogously to the second financial incentive. However, even though we find a short-term effect of these additional incentives on body weight, they tend to disappear in our experiment. Hence, our findings are not consistent with a model in which the utility depends on the lagged outcome level.[41]

Alternatively, individual utility might depend on an expectation-based reference point. One obvious candidate for the latter is the weight loss goal that was assigned to the study participants at the start of the experiment.[42] Under this preference structure, the study participants who were exposed to the weight loss incentive and fully achieved the weight loss goal

[41] Reference-dependence utility models are generally difficult to reconcile with the common observation that the majority of obese individuals with a successful weight loss attempt soon regain body weight (Crawford et al. 2000).

[42] It is worth noting that the average weight loss of the individuals exposed to financial incentives remained below the weight loss target throughout the duration of the experiment. This argues against the weight loss target being an important reference point for the study participants.

at the end of the first experimental phase had an incentive to maintain their target weight in subsequent periods. For them, we expect neither a short-term nor a lasting effect of the second financial incentive. On the contrary, for partially successful participants, this particular variant of the reference-dependent preference model would predict lasting effects of the second incentive. The reason is that the standard price effect is likely to help these participants to achieve their weight loss target at the end of the second period, which makes them want to confirm it in subsequent periods to avoid missing out on the bonus utility payoff. Hence, we expect to observe heterogeneity in the effects of the second-period financial incentive between participants who, in the first phase, were fully and partially successful, yet our results do not reveal such effect heterogeneity (Table A4).[43]

Second, the financial incentives may have conveyed new information about the cost of obesity or the benefits and prospects of weight loss. Alternatively, it may have given more weight to information that was already available (e.g., information provided to the participants by the medical staff at the outset of the experiment). This might imply that the financial rewards reinforced the message that the medical doctor sent by assigning a weight loss target.[44] After the removal of the incentive, in theory, this information would remain with the participants and, as a result, sustainable behavioral change would occur.[45] This is consistent with our findings of lasting effects of the weight loss rewards. However, this pattern should apply equally to the second-period financial incentive conditional on it conveying further information to the participants. We expect the latter to be the case since the reward size increased substantially from the first to the second phase (for some participants from EUR 150 to EUR 500), arguably suggesting that sustainable behavioral change is even more important than temporary improvements. The absence of persistent effects of the additional incentive seems to be at odds with the motivation crowding-in theory.

Eventually, a recent contribution by Brandon et al. (2017) raises the question whether our results are attributable to investments in new technology rather than to the formation of new

[43] While we observe an effect at the end of the second period among the participants who achieved their target weight in period one (note that, for the second-period control group, the weight loss target does not seem to represent an important reference point because many individuals failed to achieve it for a second time), there is no statistically significant difference in body weight between the experimental groups twelve months later. Among the partially successful participants, the second financial incentive significantly affected the probability of achieving the target weight at the end of the intervention period but not at the end of the third phase (the results for binary target weight achievement outcomes are available on request).
[44] We refer to Bénabou and Tirole (2003), who model the effects of extrinsic incentives as a principal–agent problem in which the agent has complementary private information about the task or the agent's prospects from it.
[45] We acknowledge John List's comment regarding the possibility of a motivation crowding-in effect of monetary incentives. It can be regarded as a one-to-one analogy to motivation crowding out (Frey and Jegen, 2001; Festré and Garrouste, 2015), albeit with the difference that monetary incentives shift the supply of the incentivized behavior in the opposite direction. The thin existing literature on motivation crowding in primarily examines non-monetary incentives, such as democratic participation possibility in the context of tax-paying behavior (Frey and Jegen, 2001).

habits. Although this argument seems to be less relevant to the context of health behavior than to other domains such as energy consumption, bariatric surgery or medical treatments may explain the persistence of effects in addition to genuine behavioral change. Our data provide some indication for such technologies playing a minor role for observed weight loss. First, very few participants in our experiment report having used obesity surgeries or diet drugs. Second, the utilization of such technologies does not seem to vary across the experimental groups (results available upon request).

5. Conclusion

This paper presents the results of a large randomized experiment aiming to answer the question of whether obese individuals exhibit weight-related patterns consistent with the habit formation hypothesis when exposed to monetary weight loss incentives.

The study was motivated by the increased popularity of the use of monetary incentives for healthy behavior in the design of policy interventions. While there is a general need for more knowledge about effective interventions to improve public health, this seems to be particularly evident for obesity, the public health challenge of our time. Finding effective means to fight the obesity pandemic represents an urgent need for many obese individuals who fail in their weight loss attempts and for welfare systems around the globe that are overloaded with costs attributable to obesity. We test whether temporary cash rewards help the obese to achieve persistent weight loss to shed light on the promise of this popular policy instrument.

Our results suggest that monetary rewards for weight loss have not only short-term but also long-term effects on the body weight of obese individuals. We observe similar results for a comprehensive set of health behavior variables (except in the short run). We comprehensively assess the sensitivity of our results with respect to accounting for sample attrition and find that they survive a wide range of assumptions about the change in body weight among treatment and control dropouts.

Our experimental data are consistent with the presence of habit formation. i.e., financial weight loss rewards cause obese people to achieve a healthier equilibrium. This interpretation is based on remarkable asymmetry found in the post-intervention effects of the two differently timed cash incentives of the experiment. While habit formation provides a rationale for our findings, it is difficult to reconcile this effect asymmetry with alternative theories. Evidently, the motivation crowding-out hypothesis is inconsistent with our results as it predicts adverse rather than desirable post-intervention effects of cash rewards.

Given the substantial yearly costs of obesity, a one-time investment of EUR 300 that generates a small but permanent improvement in body weight is likely to pay off soon. Rough back-of-the-envelope calculations based on lower-bound estimates of the direct as well as the indirect costs of morbidity and mortality attributable to obesity suggest that it would require approximately 30 months in total to yield net savings.[46]

Our finding of evidence that cash rewards are likely to cause sustainable behavioral change in obese individuals is remarkable since the obese arguably have characteristics that are hardly conducive to (weight-related) habit formation. Hence, our findings can be regarded as a powerful illustration of the relevance of the mechanism of habit formation. Previous relevant studies report more ambiguous results with respect to habit formation. One possible explanation is that the intervention in our program in many ways differs from the interventions previously studied. For instance, we focused on obese individuals who may respond differently to financial incentives than mostly healthy-weight university students. We also directly incentivized a desired health outcome, used up to 3.6 times larger rewards, and examined a longer incentive period. Nevertheless, our experiment did not attempt to attribute the differences in results to any of these design variations.

[46] Konnopka et al. (2010) estimate that obesity and being overweight caused costs of EUR 9,873 million in Germany in 2002, which is in line with the results for other countries (Dee et al. 2014) but is likely to be an underestimate (Cawley and Meyerhoefer 2012). Under the lower-bound assumption that three-fourths of these costs are attributable to obesity as opposed to being overweight (cf., Tsai et al. 2011) and using 10.6 million obese people as the basis of calculation, the costs per obese individual amount to about EUR 928. In our study, the average participant needs to lose roughly 7.6 BMI points to fall below the obesity threshold (Table 2), which implies a yearly cost reduction of EUR 122 per BMI unit if we assume a linear relationship between BMI and costs in our sample. The estimated effect of the EUR 300 reward ranges between 2.8 and 3.3 percentage points at four, ten, and 22 months. This corresponds to a weight reduction of about 1.1 to 1.3 BMI points compared with a baseline of 38.2 points.

References

ACLAND, D. and LEVY, M., 2015. Naiveté, projection bias, and habit formation in gym attendance. *Management Science*. vol. 61, no. 1, pp. 146–160.

ANGRIST, J. and LAVY, V., 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. *American Economic Review*. vol. 99, no. 4, pp. 1384–1414.

BHATTACHARYA, J., BUNDORF, M. K., PACE, N. and SOOD, N., 2011. Does health insurance make you fat? In: M. GROSSMANN and N. H. MOCAN, eds. *Economic aspects of obesity*. University of Chicago Press, chapter 2, pp. 35–64.

BECKER, G. S., GROSSMAN, M. and MURPHY, K. M., 1991. Rational addiction and the effect of price on consumption. *American Economic Review*. vol. 81, no. 2, pp. 237–241.

BECKER, G. S. and MURPHY, K. M., 1988. A theory of rational addiction. *Journal of Political Economy*. vol. 96, no. 4, pp. 675–700.

BÉNABOU, R. and TIROLE, J., 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies*. vol. 70, no. 3, pp. 489–520.

BRANDON, A., FERRARO, P., LIST, J. A., METCALFE, R. D., PRICE, M., and RUNDHAMMER, F. (2017). Do the Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments. NBER Working Paper w23277.

CAWLEY, J. and MEYERHOEFER, C., 2012. The medical care costs of obesity: an instrumental variables approach. *Journal of Health Economics*. vol. 31, no. 1, pp. 219–230.

CAWLEY, J. and PRICE, J. A., 2013. A case study of a workplace wellness program that offers financial incentives for weight loss. *Journal of Health Economics*. vol. 32, no. 5, pp. 794–803.

CAWLEY, J., RIZZO, J. A. and HAAS, K., 2007. Occupation-specific absenteeism costs associated with obesity and morbid obesity. *Journal of Occupational and Environmental Medicine*. vol. 49, no. 12, pp. 1317–1324.

CHARNESS, G. and GNEEZY, U., 2009. Incentives to exercise. *Econometrica*. vol. 77, no. 3, pp. 909–931.

CRAWFORD, D., JEFFERY, R. W. and FRENCH, S. A., 2000. Can anyone successfully control their weight? Findings of a three year community-based study of men and women. *International Journal of Obesity*. vol. 24, no. 9, pp. 1107–1110.

DECI, E. L. (1971): Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology.* vol. 18, no. 1, pp. 105–115.

DEE, A., KEARNS, K., O'NEILL, C., SHARP, L., STAINES, A., O'DWYER, V., FITZGERALD, S. and PERRY, I. J., 2014. The direct and indirect costs of both overweight and obesity: a systematic review. *BMC Research Notes*. vol. 7, no. 242, pp. 1–9.

DE GEEST, G. and DARI-MATTIACCI, G., 2013. The rise of carrots and the decline of sticks. *University of Chicago Law Review*. vol. 80, no. 1, pp. 341–392.

FESTRÉ, A. and GARROUSTE, P., 2015. Theory and evidence in psychology and economics about motivation crowding out: a possible convergence? *Journal of Economic Surveys*. vol. 29, no. 2, pp. 339–356.

FINKELSTEIN, E., THAM, K., HAALAND, B. and SAHASRANAMAN, A., 2017. Applying economic incentives to increase effectiveness of an outpatient weight loss program (TRIO) – a randomized controlled trial. *Social Science & Medicine*. vol. 185, pp. 63–70.

FREY, B. S. and JEGEN, R., 2001. Motivation crowding theory, *Journal of Economic Surveys.* vol. 15, no. 5, pp. 589–611.GNEEZY, U., MEIER, S. and REY-BIEL, P., 2011. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*. vol. 25, no. 4, pp. 191–210.

HAN, E., NORTON, E. C. and STEARNS, S. C., 2009. Weight and wages: fat versus lean paychecks. *Health Economics*. vol. 18, no. 5, pp. 535–548.

HEATHCOTE, A., BROWN, S., WAGENMAKERS, E.-J. and EIDELS, A., 2010. Distribution-free tests of stochastic dominance for small samples. *Journal of Mathematical Psychology*. vol. 54, no. 5, pp. 454–463.

HOLLIS, S. and CAMPBELL, F., 1999. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*. vol. 319, no. 7211, pp. 670–674.

HOROWITZ, J. L. and MANSKI, C. F., 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*. vol. 95, no. 449, pp. 77–84.

HOUSTON, D. K., CAI, J. and STEVENS, J., 2008. Overweight and obesity in young and middle age and early retirement: the ARIC study. *Obesity*. vol. 17, no. 1, pp. 143–149.

KONNOPKA, A., BÖDEMANN, M. and KÖNIG, H.-H., 2010. Health burden and costs of obesity and overweight in Germany. *European Journal of Health Economics*. vol. 12, no. 4, pp. 345–352.

KULLGREN, J. T., TROXEL, A. B., LOEWENSTEIN, G., ASCH, D. A., NORTON, L. A., WESBY, L., TAO, Y., ZHU, J. and VOLPP, K. G., 2013. Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. *Annals of Internal Medicine*. vol. 158, no. 7, pp. 505–514.

LEE, D. S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *Review of Economic Studies*. vol. 76, no. 3, pp. 1071–1102.

LOEWENSTEIN, G., PRICE J., VOLPP, K., 2016. Habit formation in children: Evidence from incentives for healthy eating. *Journal of Health Economics*. vol. 45, pp. 47-54.

MESSINIS, G., 1999. Habit formation and the theory of addiction. *Journal of Economic Surveys*. vol. 13, no. 4, pp. 417–442.

MORRIS, S., 2006. Body mass index and occupational attainment. *Journal of Health Economics*. vol. 25, no. 2, pp. 347–364.

OECD., 2012. *Statistics, Organisation for Economic Co-operation and Development* [online] [viewed 8 June 2012]. Available from: http://stats.oecd.org.

PALOYO, A. R., REICHERT, A. R., REINERMANN, H. and TAUCHMANN, H., 2014. The causal link between financial incentives and weight loss: an evidence-based survey of the literature. *Journal of Economic Surveys*. vol. 28, no. 3, pp. 401–420.

PALOYO, A. R., REICHERT, A. R., REUSS-BORST, M. and TAUCHMANN, H., 2015. Who responds to financial incentives for weight loss? Evidence from a randomized controlled trial. *Social Science & Medicine*. vol. 145, pp. 44–52.

POLLAK, R., 1970. Habit formation and dynamic demand functions*. Journal of Political Economy*. vol. 78, no. 4, pp. 745–763.

PROMBERGER, M. and MARTEAU, T. M., 2013. When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. *Health Psychology*. vol. 32, no. 9, pp. 950–957.

RAKEL, D. and RAKEL, R. E., 2015. *Textbook of family medicine*. 9th ed. Philadelphia: Elsevier Saunders.

REICHERT, A., 2015. Obesity, weight loss, and employment prospects – evidence from a randomized trial. *Journal of Human Resources*. vol. 50, no. 3, pp. 759–810.

ROYER, H., STEHR, M. and SYNDOR, 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*. vol. 7, no. 3, pp. 51–84.

RYDER, H. E. and HEAL, G., 1973. Optimal growth with intertemporally dependent preferences. *Review of Economic Studies*. vol. 40, no. 1, pp. 1–31.

SASSI, F., 2010. *Obesity and the economics of prevention: fit not fat*. Paris: OECD Publishing.

SPENKUCH, J., 2012. Moral hazard and selection among the poor: evidence from a randomized experiment. *Journal of Health Economics*. vol. 31, no. 1, pp. 72–85.

SPINNEWYN, F., 1981. Rational habit formation. *European Economic Review*. vol. 15, pp. 91–109.

STEPHENS, D. J., SAUNDERS, J. K., BELSLEY, S., TRIVEDI, A., EWING, D. R., IANNACE, V., CAPELLA, R. F., WASIELEWSKI, A., MORAN, S., SCHMIDT, H. J. and BALLANTYNE, G. H., 2008. Short-term outcomes for super-super obese (BMI > or = 60 kg/m2) patients undergoing weight loss surgery at a high-volume bariatric surgery center: laparoscopic adjustable gastric banding, laparoscopic gastric bypass, and open tubular gastric bypass. *Surgery for Obesity and Related Diseases*. vol. 4, no. 3, pp. 408–415.

SYKES-MUSKETT, B. J., PRESTWICH, A., LAWTON, R. J. and ARMITAGE, C. J., 2015. The utility of monetary contingency contracts for weight loss: a systematic review and meta-analysis. *Health Psychology Review*. vol. 9, no. 4, pp. 434–451.

TEIXEIRA, P. J., CARRAÇA, E. V., MARQUES, M. M., RUTTER, H., OPPERT, J.-M., DE BOURDEAUDHUIJ, I., LAKERVELD, J. and BRUG, J., 2015. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. *BMC Medicine*. vol. 13, no. 84, pp. 1–16.

TSAI, A. G., WILLIAMSON, D. F. and GLICK, H. A., 2011. Direct medical cost of overweight and obesity in the USA: a quantitative systematic review. *Obesity Reviews*. vol. 12, no. 1, pp. 50–61.

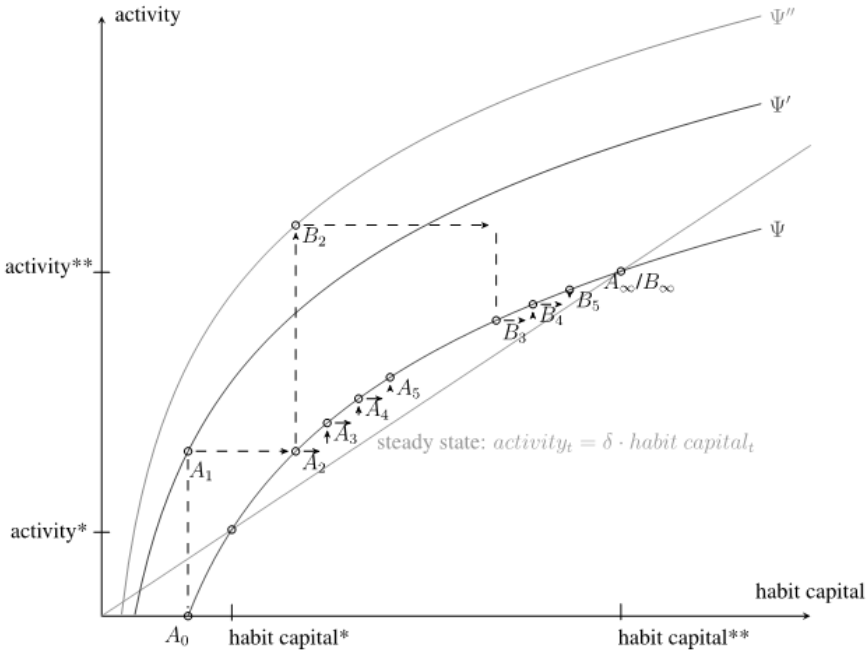
VIDAL, J., 2002. Updated review on the benefits of weight loss. *International Journal of Obesity Related Metabolic Disorders*. vol. 26, suppl. 4, pp. 25–28.

VOLPP, K., JOHN, L., TROXEL, A., NORTON, L., FASSBENDER, J. and LOEWENSTEIN, G., 2008. Financial incentive-based approaches for weight loss. *Journal of the American Medical Association*. vol. 300, no. 22, pp. 2631–2637.

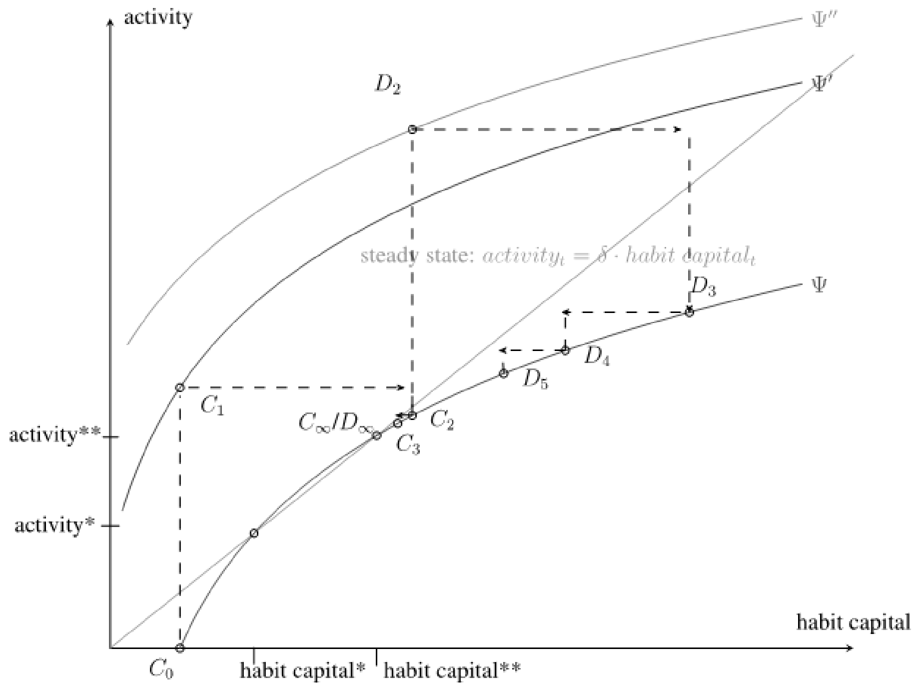
VOLPP, K. G., TROXEL, A. B., PAULY, M. V., GLICK, H. A., PUIG, A., ASCH, D. A., GALVIN, R., ZHU, J., WAN, F., DEGUZMAN, J., CORBETT, E., WEINER, J. and AUDRAIN-MCGOVERN, J., 2009. A randomized controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*. vol. 360, no. 7, pp. 699–709.

WARNEKEN, F. and TOMASELLO, M., 2008. Rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology*. vol. 44, no. 6, pp. 1785–1788.

WOOLDRIDGE, J. M., 2002. Inverse probability weighted M-estimators for sample selection, attrition and stratification. *Portuguese Economic Journal*. vol. 1, no. 2, pp. 117–139.



1a. Acceleration of Transition due to Additional Financial Incentives



1b. Transition exhibits Decay of Habit Capital

Figure 1: Trajectory towards Healthy Equilibrium

Note: Ψ' and Ψ'' denote the shift of the activity curve following the temporary cash incentives of periods one and two.

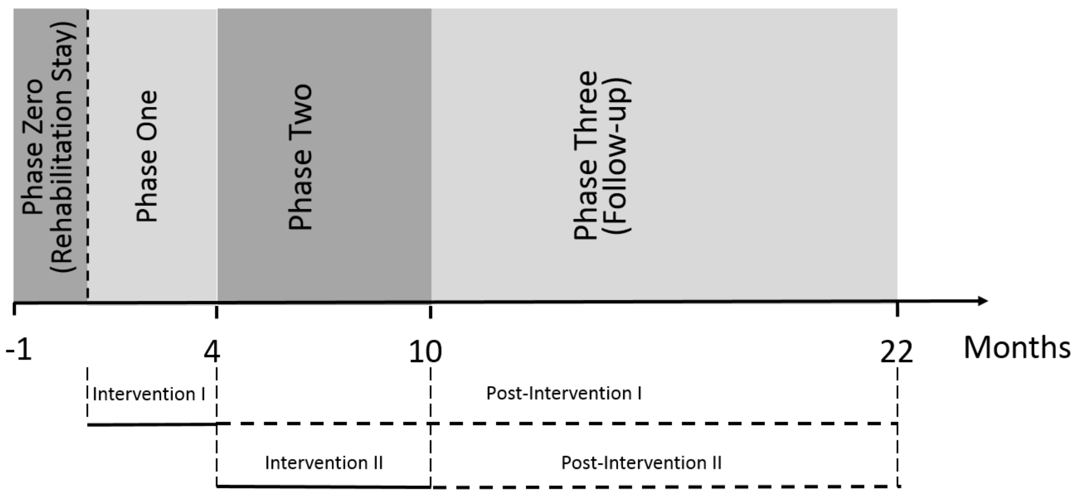


Figure 2: Experimental Design

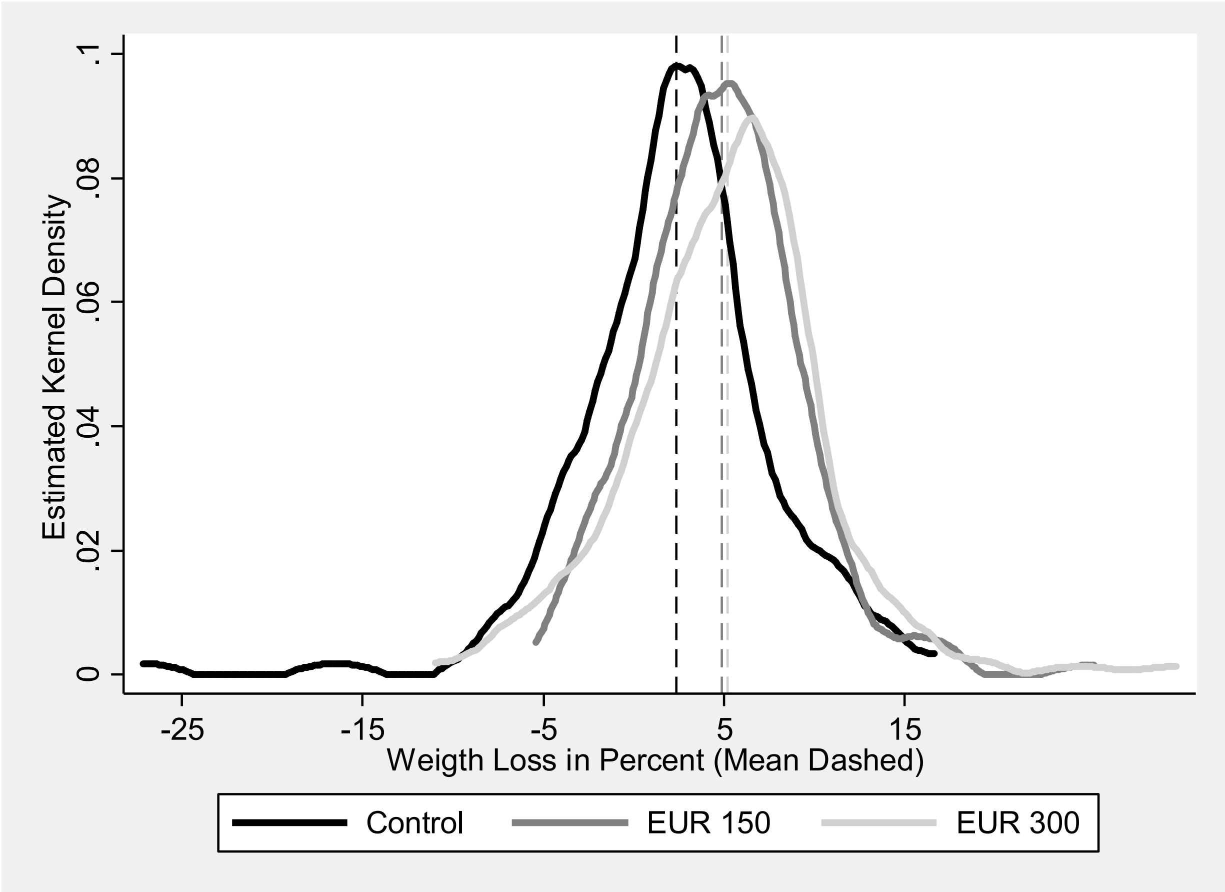


Figure 3: Distribution of Percentage Change in Body Weight by Phase-one Monetary Incentives (Months 1–4)

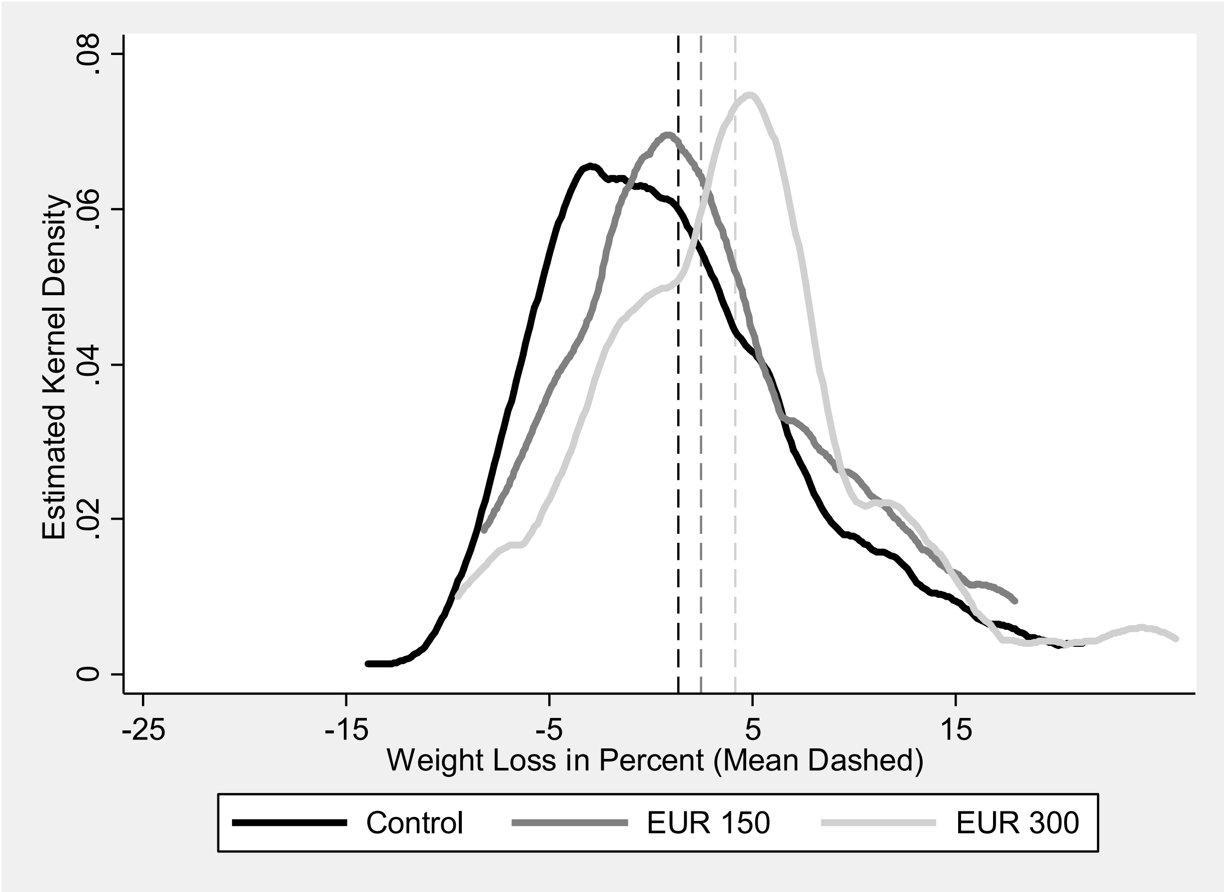


Figure 4: Distribution of Percentage Change in Body Weight by Phase One Monetary Incentives (Months 1–10)

Note: Inverse probability weights are used for the estimation of the kernel densities.

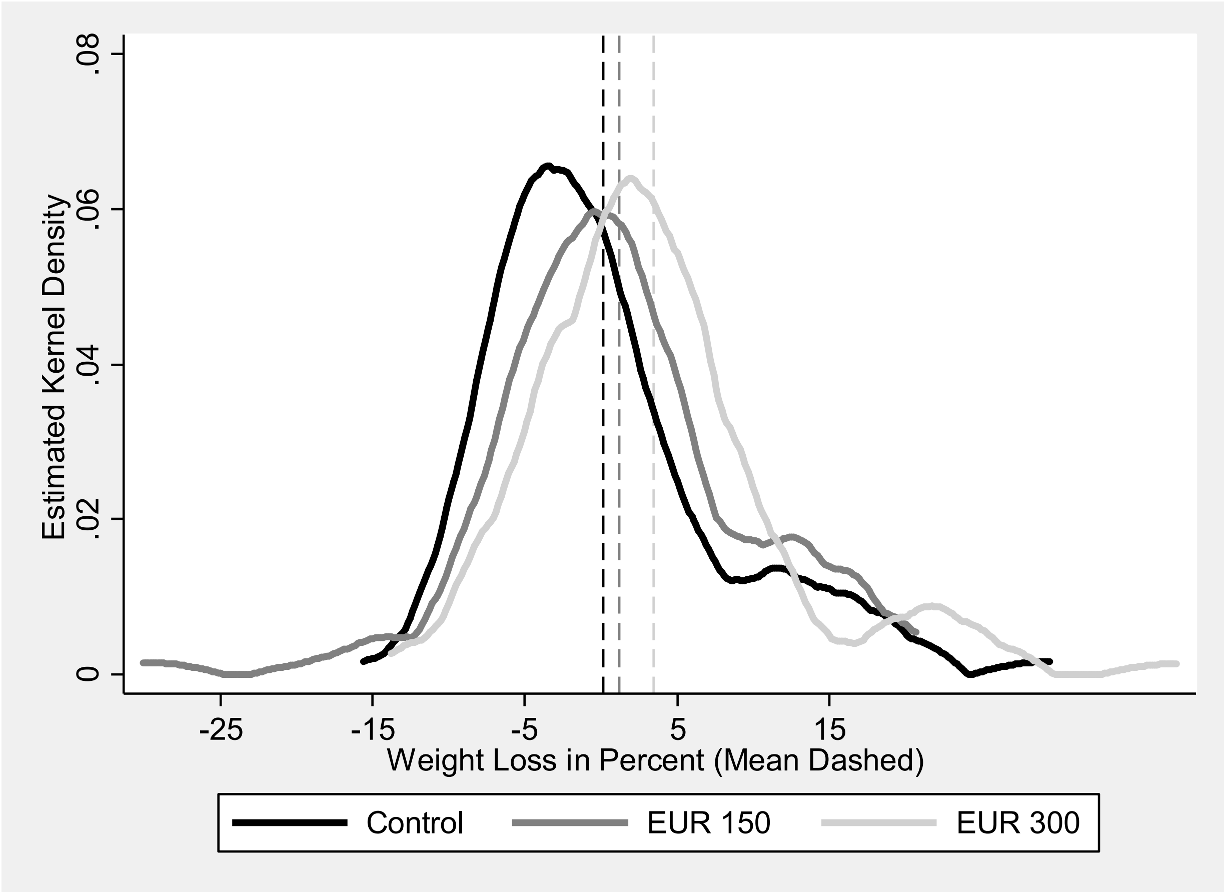


Figure 5: Distribution of Percentage Change in Body Weight by Phase-One Monetary Incentives (Months 1–22)

Note: Inverse probability weights are used for the estimation of the kernel densities.

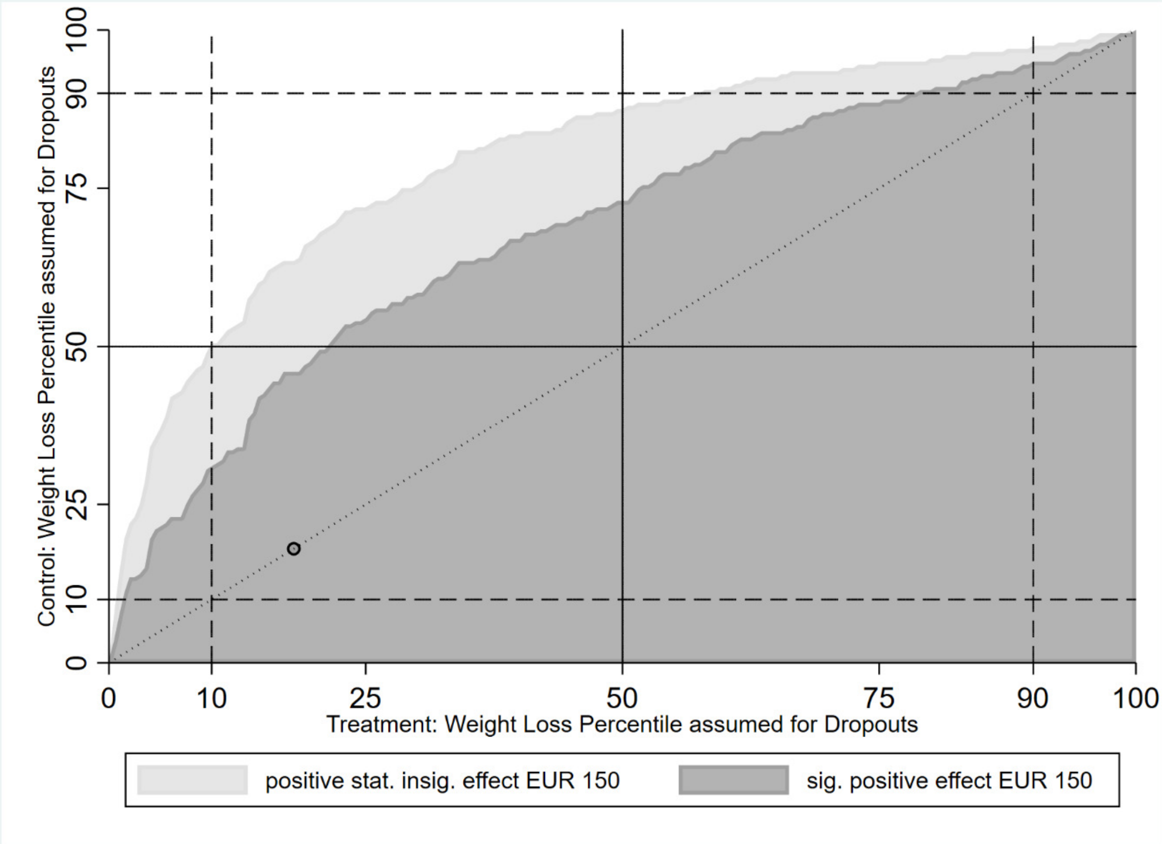


Figure 6a: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 150*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–4**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

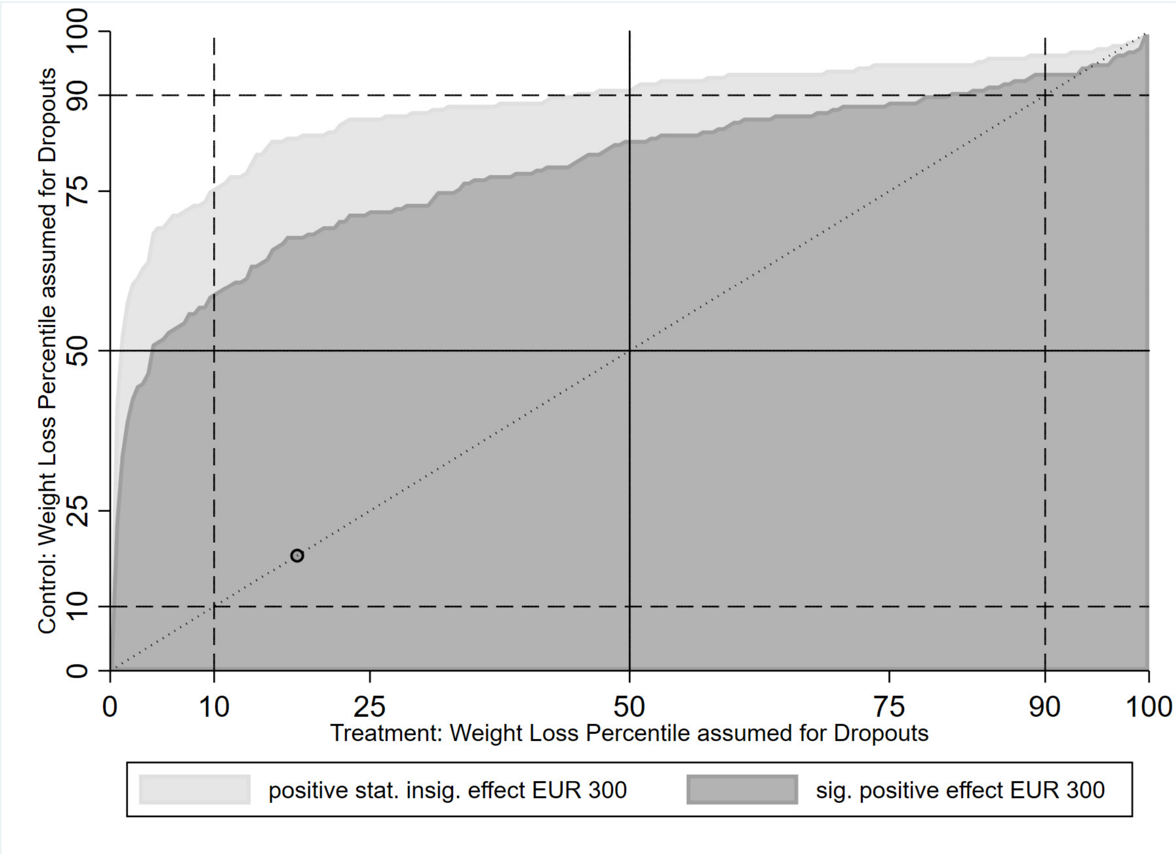


Figure 6b: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 300*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–4**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

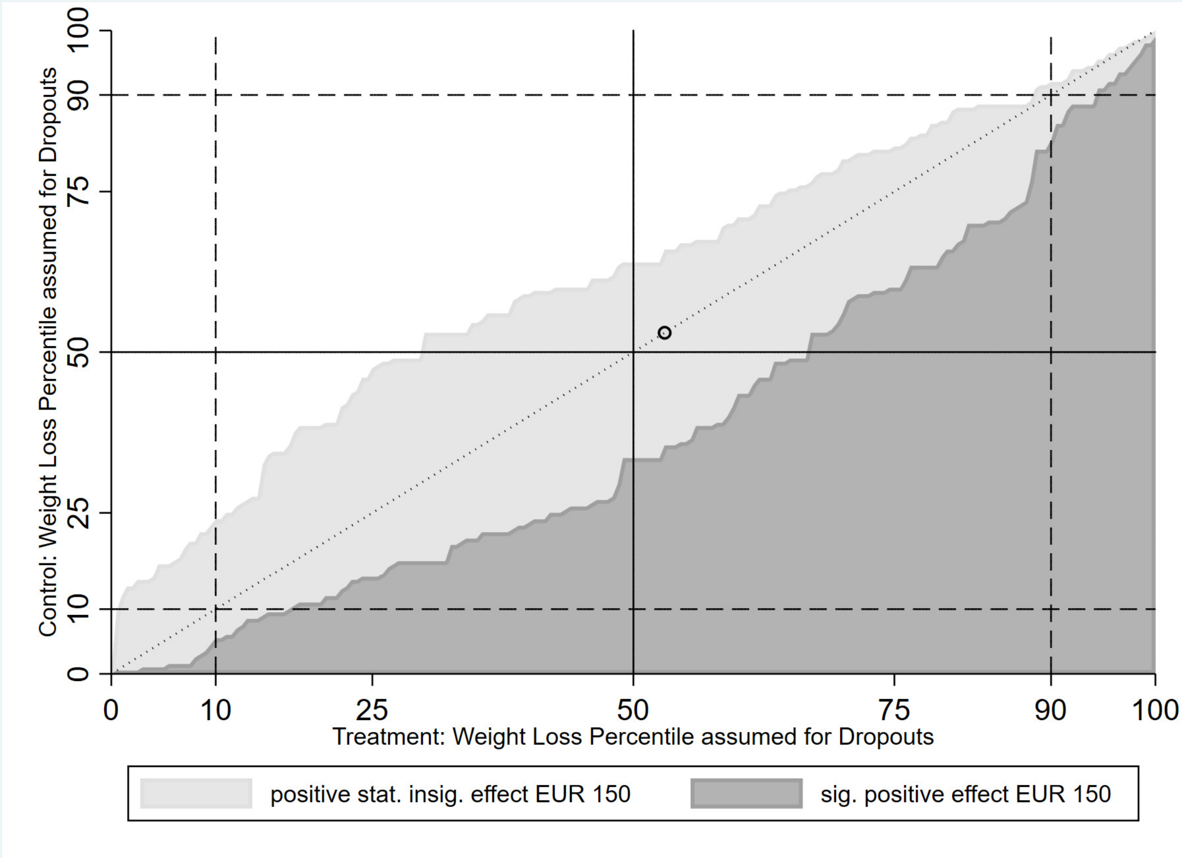


Figure 6c: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 150*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–10**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

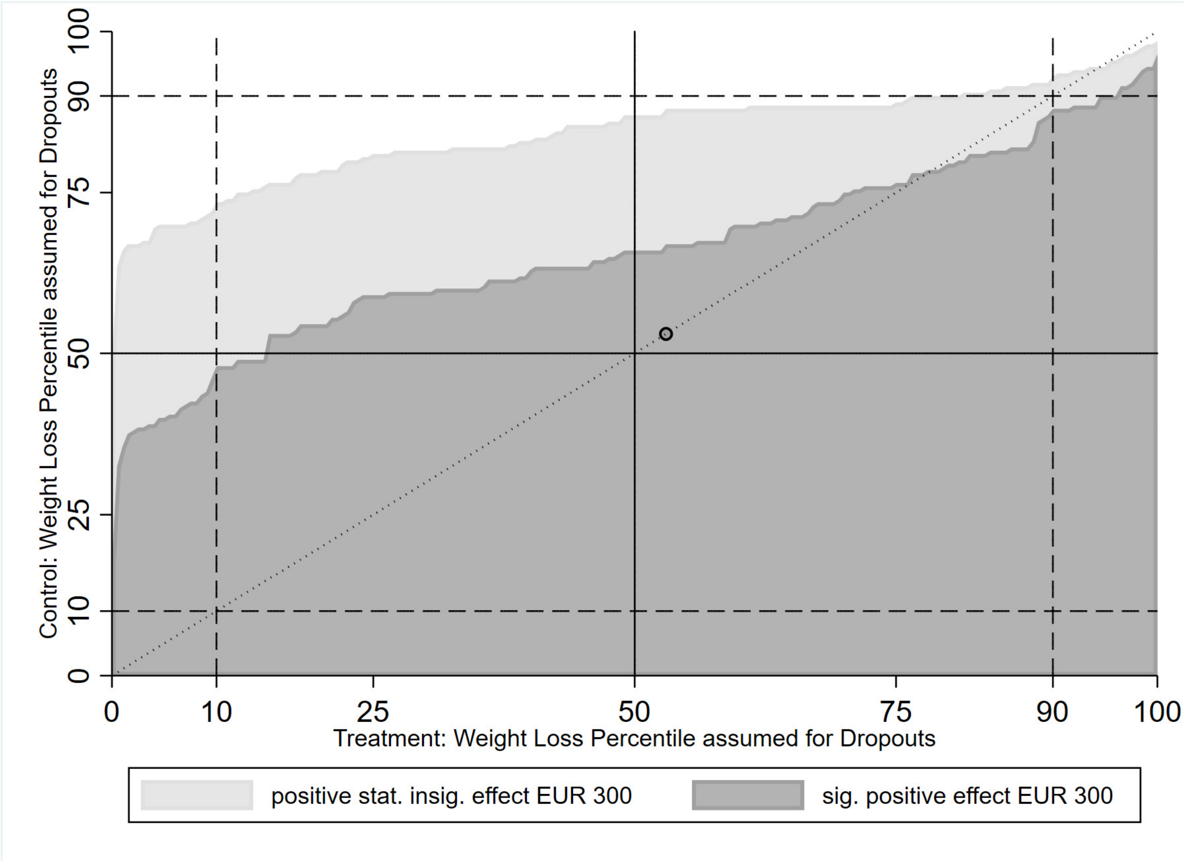


Figure 6d: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 300*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–10**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

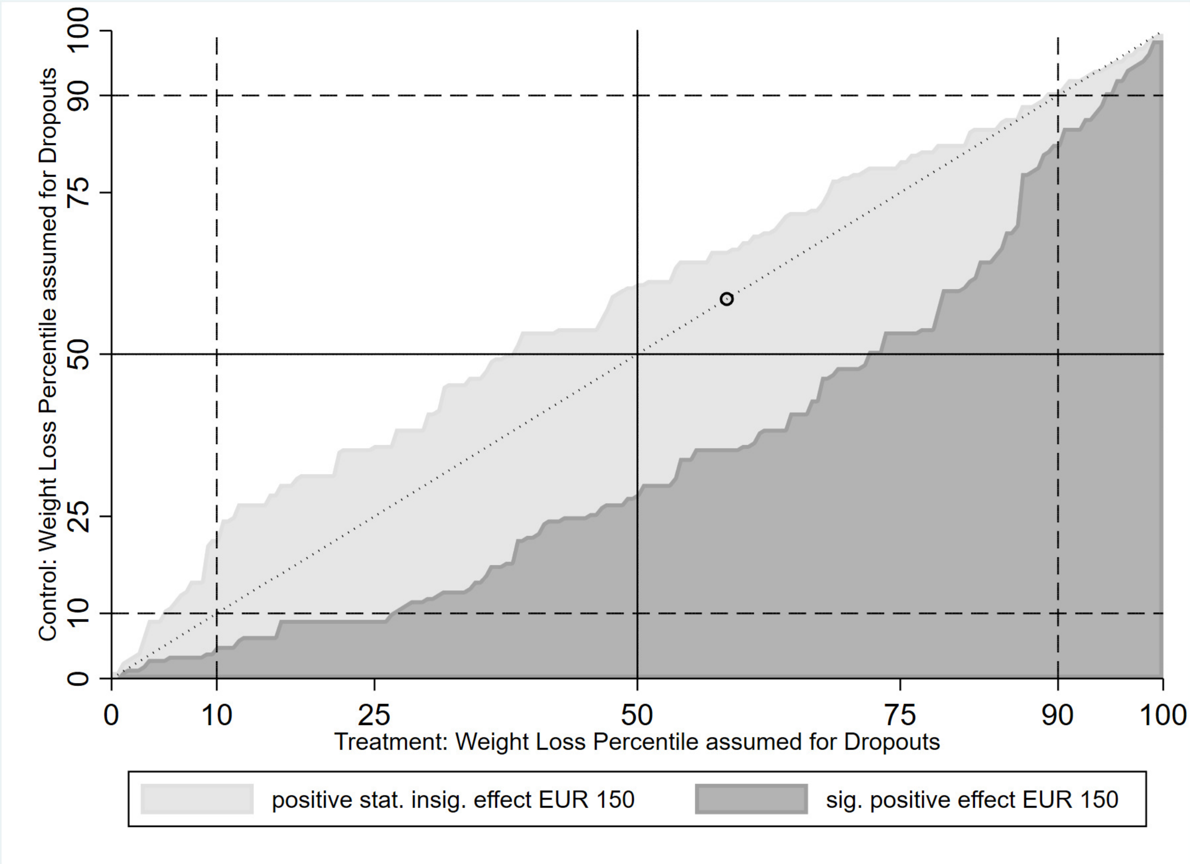


Figure 6e: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 150*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–22**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

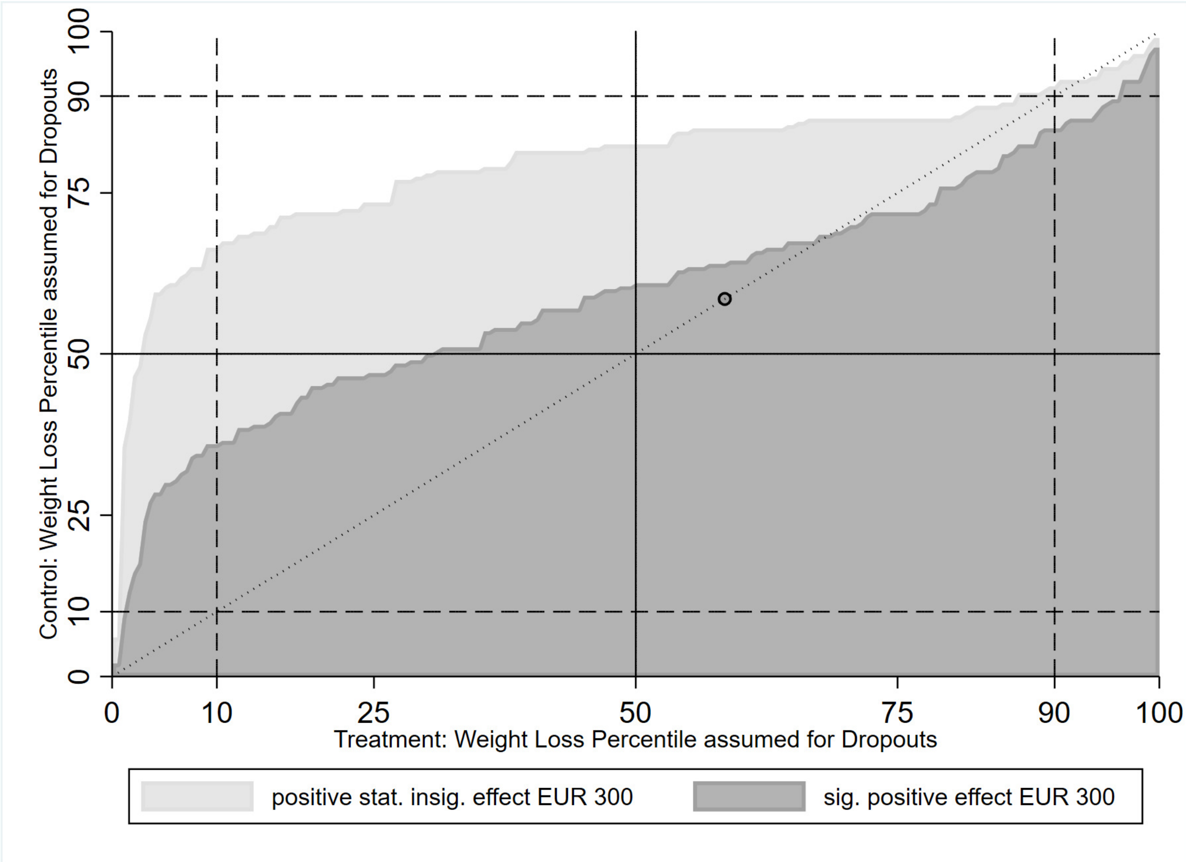


Figure 6f: Combinations of Weight Loss Assumed for Dropouts in the Control Group and ***Group 300*** Yielding a Positive/Stat. Significantly Positive Treatment Effect (**Months 1–22**)

Notes: Statistical significance at the 10% level (one-sided test); the dotted line indicates scenarios in which the same level of weight loss is assumed for dropouts of both groups. The small circle marks the result of the pre-intervention outcome carried forward method.

Figure 6: Statistical Significance of the Treatment Effect as a Function of the Weight Change of Dropouts in the Treatment and Control Groups

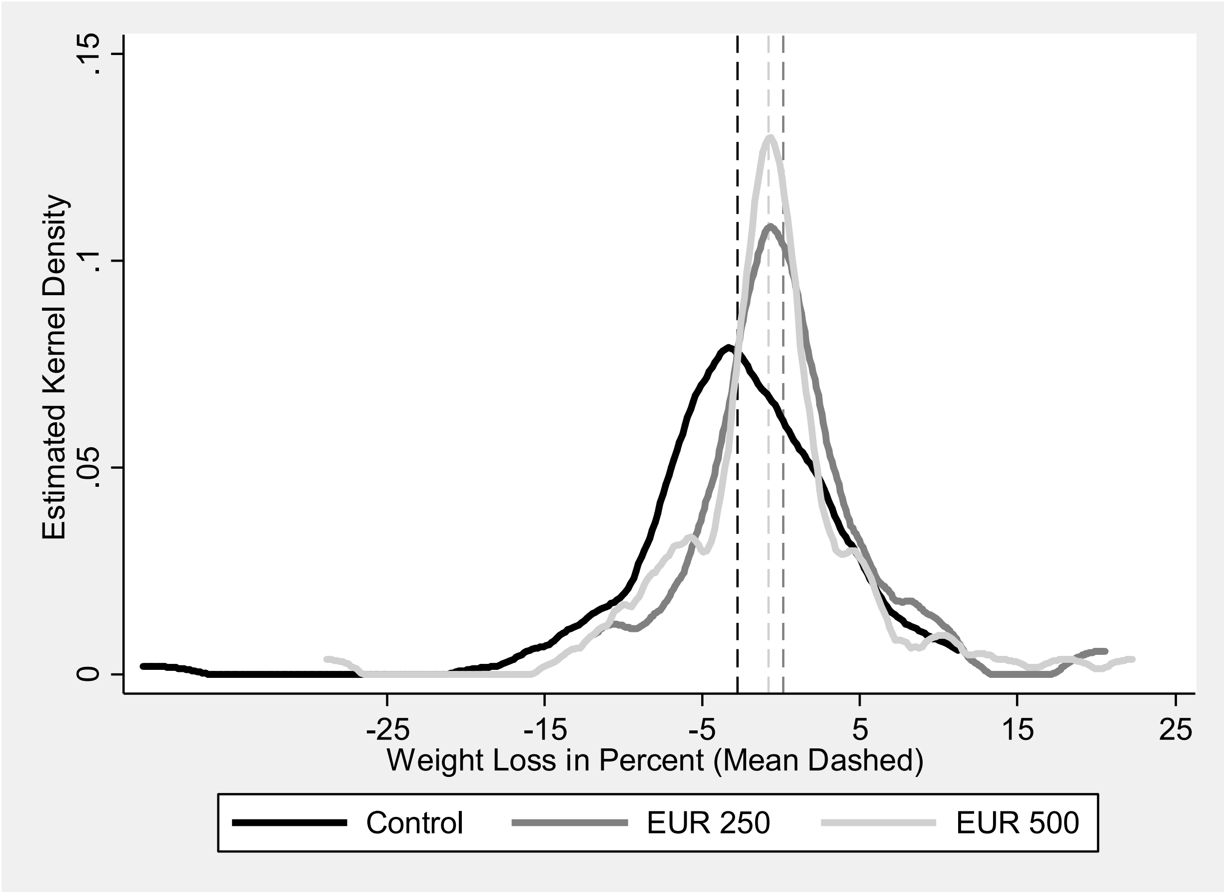


Figure 7: Distribution of Percentage Change in Body Weight by Phase Two Monetary Incentives (Months 5–10)

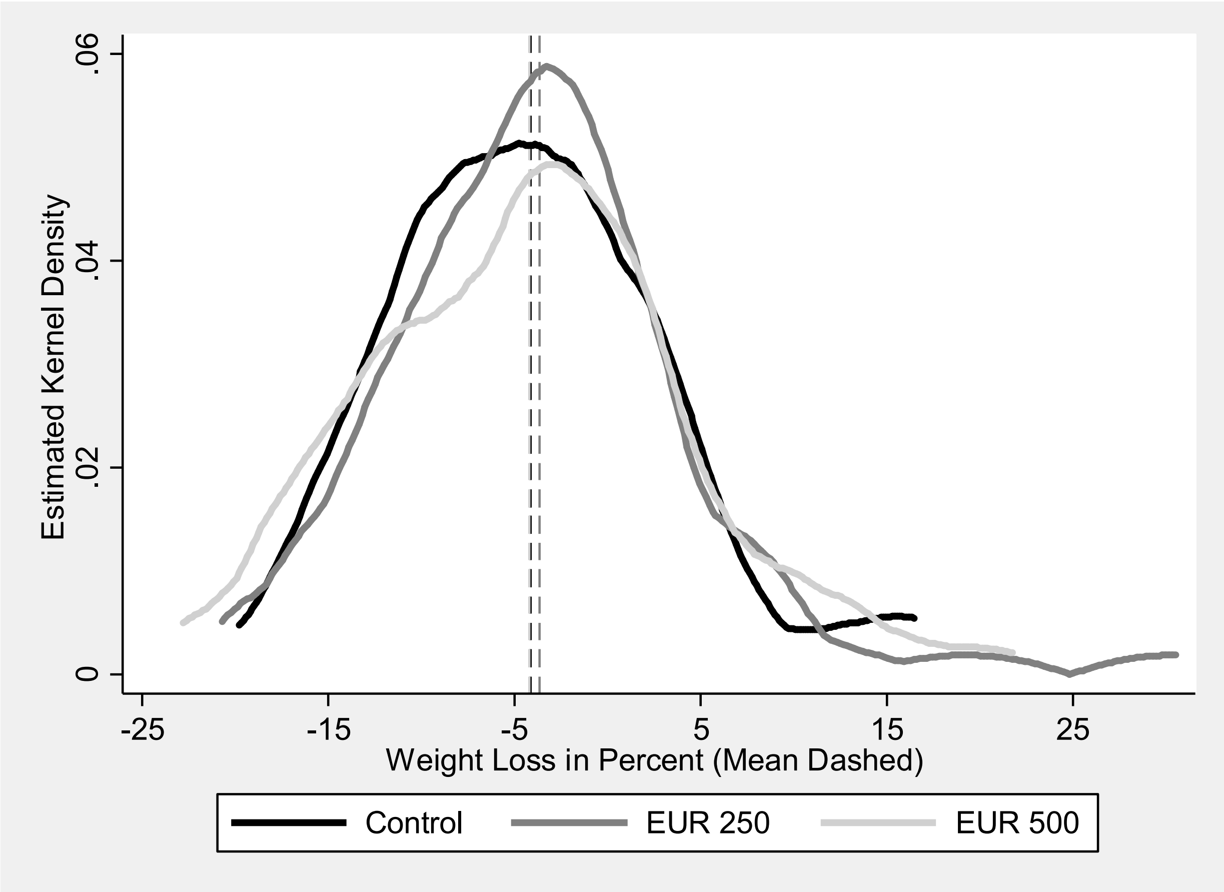


Figure 8: Distribution of Percentage Change in Body Weight by Phase Two Monetary Incentives (Months 5–22)

Table 1: Socioeconomic Background of the Study Population and the Obese in Germany

	Study Population	Patients of the Four Rehabilitation Clinics	Representative Obese in Germany (BMI ≥ 30)
Female (%)	32.23	34.17	39.98
Age (Years)	48.11	49.69	57.11
Married (%)	61.03	71.37	62.23
Resident of Baden-Württemberg (%)	100	94.99	11.84
Native (%)	78.89	82.67	86.30
Full-Time Employed[1,2] (%)	69.44	76.12	34.85
Part-Time Employed[1,2] (%)	9.04	11.01	14.27
Unemployed[3,2] (%)	13.20	8.23	6.90

Notes: The statistics relating to the patients of the four rehabilitation clinics are weighted averages. As the clinics' weights, we use the share of the participants recruited by them. [1] The remaining observations among those who reported being employed are marginally employed (2.15 percent) or did not provide information on the type of employment (1.72 percent). [2] Here we distinguish between the unemployed and the not-employed (4.45 percent). [3]The categories full-time employed, part-time employed, marginally employed, no information on the type of occupation, unemployed, and not-employed add up to one.
Source: Own data collection, German Federal Pension Fund (year 2011) and German Socio-economic Panel (SOEP, year 2011).

Table 2: Descriptive Statistics of Phase One Premium Groups
(Mean Values, *Full Unweighted Sample* and *Inverse Probability Weighting*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All		Control		EUR 150		EUR 300	
	all unw.†	inv. p. w.‡	all unw.†	inv. p. w.‡	all unw.†	inv. p. w.‡	all unw.†	inv. p. w.‡
Pre-treatment Values								
BMI before Rehab.	38.900	38.948	38.420	38.574	39.012	38.749	39.270	39.547
Baseline BMI	37.591	37.583	37.138	37.222	37.632	37.330	38.007	38.231
Target Weight Loss (Percentage)	6.511	6.501	6.442	6.436	6.579	6.601	6.511	6.473
Unhealthy Ingredients	0.041	0.041	0.041	0.039	0.035	0.019	0.046	0.064°
Healthy Eating	0.135	0.122	0.102	0.106	0.147	0.152	0.153	0.108
Good Eating Habits	0.129	0.130	0.118	0.136	0.140	0.103	0.129	0.149
Exercise	0.306	0.306	0.316	0.341	0.337	0.313	0.265	0.260
Bad Kissingen	0.250	0.336	0.250	0.290	0.251	0.367	0.249	0.356
Bad Mergentheim	0.249	0.421	0.248	0.453	0.248	0.394	0.249	0.413
Isny	0.251	0.053	0.250	0.057	0.251	0.062	0.251	0.040
Glottertal	0.251	0.190	0.252	0.200	0.250	0.177	0.251	0.191
Female	0.322	0.336	0.290	0.249	0.314	0.332	0.364	0.436
Age (Years)	47.515	47.080	47.749	47.686	47.538	46.084	47.254	47.422**
Native	0.783	0.796	0.788	0.810	0.744	0.767	0.817	0.810
Married	0.625	0.638	0.662	0.679	0.613	0.619	0.601	0.614
Cum. Dropout Rate after Month: 4	0.255	0.254	0.333	0.314	0.271	0.283	0.158	0.160**°°
10	0.407	0.420	0.489	0.498	0.403	0.451	0.329	0.302**°°
22	0.545	0.566	0.602	0.624	0.542*	0.588	0.491	0.480**
Marg. Dropout Rate Months: 5–10	0.153	0.165	0.156	0.184	0.131	0.168	0.171	0.142
11–22	0.138	0.147	0.113	0.127	0.140	0.137	0.162	0.178
# of Observations (Unweighted)	695	489	231	192	236	158	228	139

Notes: ** Deviation from the control group significant at 5%, * significant at 10%; °° deviation from EUR 150 group significant at 5%, ° significant at 10%; standard deviations omitted because most variables are binary. “Bad Mergentheim,” “Bad Kissingen,” “Isny” and “Glottertal” refer to the locations of the four rehabilitation clinics.
† Unweighted means for the full sample of participants; ‡ inverse probability-weighted means for the sample used in the analysis of medium-term effects.

Table 3: Mean Comparison Across Phase One Premium Groups

	# of Obs.	Experimental Groups			Δ to Control	
		Control	EUR 150	EUR 300	EUR 150	EUR 300
	(by Groups)	(1)	(2)	(3)	(4)	(5)
Unweighted Means (Months 1–4)						
Weight Loss (Percentage)	518 (154/172/192)	2.372** (0.439)	4.861** (0.336)	5.201** (0.393)	2.489** (0.569)	2.829** (0.554)
Fewer Unhealthy Ingredients	478 (143/160/175)	-0.007 (0.021)	-0.000 (0.009)	0.011 (0.020)	0.007 (0.025)	0.018 (0.025)
Healthier Eating	444 (130/150/164)	0.269** (0.046)	0.187** (0.044)	0.293** (0.045)	-0.083 (0.066)	0.023° (0.064)
Improved Eating Habits	505 (151/168/186)	0.093 (0.038)	0.125** (0.034)	0.097** (0.028)	0.032 (0.048)	0.004 (0.047)
More Exercise	399 (123/129/147)	0.016 (0.057)	-0.023 (0.050)	0.020 (0.052)	-0.040 (0.076)	0.004 (0.074)
Inverse Probability Weighting (Months 1–10)						
Weight Loss (Percentage)	234 (85/70/79)	1.343 (0.926)	2.468** (0.895)	4.155** (0.879)	1.125 (1.288)	2.812** (1.277)
Fewer Unhealthy Ingredients	217 (77/67/73)	-0.009 (0.016)	-0.008 (0.008)	0.055 (0.030)	0.001 (0.018)	0.064**°° (0.034)
Healthier Eating	206 (76/66/64)	0.188* (0.070)	0.225* (0.085)	0.258** (0.073)	0.037 (0.110)	0.071 (0.101)
Improved Eating Habits	229 (84/69/76)	0.000 (0.063)	0.114 (0.050)	0.074 (0.058)	0.114* (0.080)	0.074 (0.085)
More Exercise	164 (64/48/52)	-0.196 (0.074)	0.045 (0.078)	0.085 (0.085)	0.241** (0.108)	0.281** (0.137)

Inverse Probability Weighting (Months 1–22)						
Weight Loss (Percentage)	176 (64/53/59)	0.137 (1.194)	1.202 (1.271)	3.472** (1.135)	1.065 (1.744)	3.335** (1.648)
Fewer Unhealthy Ingredients	165 (60/49/56)	-0.011 (0.026)	-0.023 (0.016)	0.071 (0.039)	-0.012 (0.030)	0.083**°° (0.047)
Healthier Eating	151 (55/48/48)	0.152 (0.065)	0.233 (0.089)	0.359** (0.092)	0.081 (0.111)	0.207** (0.113)
Improved Eating Habits	173 (63/52/58)	0.088 (0.072)	0.152 (0.065)	0.132 (0.081)	0.064 (0.097)	0.044 (0.108)
More Exercise	124 (49/35/40)	-0.145 (0.086)	0.015 (0.116)	0.128 (0.107)	0.160 (0.144)	0.273** (0.138)

Notes: All the numbers of observations are unweighted. The numbers in parentheses indicate the unweighted size of each experimental group. For the periods after incentive removal (second and third panels of the table), besides dropout from the experiment, the probability of selection into the weight maintenance incentive groups affects the number of unweighted observations. For instance, for the variable "weight loss" between Month 0 and Month 10 (second panel of the table), the weighted number of observations is 404 (123/124/157). Applying the sample dropout rate in the inverse probability estimation sample (second column of Table 2) to the weighted number of observations yields the original number of observations (695). ** significant at 5% (two-sided columns 1–3, one-sided columns 4 and 5), * significant at 10% (two-sided columns 1–3, one-sided columns 4 and 5); °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means and for coefficients in parentheses. All the coefficients are obtained by inverse probability weighting OLS, regressing the respective outcome variable on the dummy variables indicating the two premium groups.

Table 4: Dropout Robust Effects of Phase One Premiums

	# of Obs. by Estimation Procedure	Self-Reported Weight†		Pre-intervention Outcome Carried Forward		Conservative Lee Bounds#	
		EUR 150	EUR 300	EUR 150	EUR 300	EUR 150	EUR 300
		(1)	(2)	(3)	(4)	(5)	(6)
Unweighted Means (Months 1–4)							
Weight Loss (Percentage)	557/695/695	1.819** (0.571)	2.156** (0.557)	1.962** (0.569)	2.798** (0.554)	2.119** [0.733++	1.003** [0.228++
Fewer Unhealthy Ingredients	654/654	-	-	0.005 (0.025)	0.014 (0.025)	-0.000 [-0.038	-0.028 [-0.072]
Healthier Eating	620/620	-	-	-0.039 (0.066)	-0.070° (0.064)	-0.181 [-0.323	-0.171 [-0.317
Improved Eating Habits	682/682	-	-	0.029 (0.048)	0.020 (0.046)	-0.031 [-0.176	-0.126 [-0.176
More Exercise	575/575	-	-	-0.026 (0.076)	0.006 (0.074)	-0.067 [-0.277	-0.158 [-0.312
Inverse Probability Weighting (Months 1–10)							
Weight Loss (Percentage)	350/411/488	0.760 (1.100)	2.137* (1.145)	0. 802 (1.288)	2.554** (1.277)	0.258 [-2.221	1.118 [-1.047
Fewer Unhealthy Ingredients	394/471	-	-	-0.000 (0.018)	0.050* (0.034)	0.001 [-0.036	0.009 [-0.022
Healthier Eating	383/460	-	-	0.036 (0.110)	0.088 (0.101)	-0.072 [-0.276	-0.010 [-0.210
Improved Eating Habits	406/483	-	-	0.075 (0.080)	0.060 (0.085)	0.045 [-0.138	0.031 [-0.138
More Exercise	341/418	-	-	0.133 (0.108)	0.170 (0.137)	0.203 [-0.077	0.179 [-0.108

Inverse Probability Weighting (Months 1–22)							
Weight Loss (Percentage)	215/353/488	0.757 (1.506)	3.394**° (1.461)	0. 638 (1.744)	2.581* (1.648)	-0.111 [-3.494	2.866 [-0.811
Fewer Unhealthy Ingredients	341/476	-	-	-0.007 (0.030)	0.060* (0.047)	-0.013 [-0.073	0.011 [-0.039
Healthier Eating	327/462	-	-	0.056 (0.111)	0.180* (0.113)	-0.022 [-0.236	0.104 [-0.129
Improved Eating Habits	350/485	-	-	0.042 (0.097)	0.052 (0.108)	0.006 [-0.204	-0.112 [-0.204
More Exercise	301/436	-	-	0.076 (0.144)	0.159 (0.138)	0.175 [-0.152	0.039 [-0.295

Notes: # Conservative, i.e., lower Lee bound [conservative, i.e., lower bound of the Lee bounds based on the 95 percent confidence interval for the estimated treatment effect]; ++ Lee bounds based on the 95 percent confidence interval for the treatment effect do not include non-positive values. † Only applicable to the outcome body weight, since for the behavioral outcomes only self-reported information is available; ** significant at 5% (one-sided test), * significant at 10% (one-sided test); °° difference between premium groups significant at 5%, ° significant at 10%; robust S.E.s for estimated means in parentheses.

Table 5: Descriptive Statistics by Phase Two Premium Groups (Mean Values)
(Incentivized Individuals in the First Phase Only)

	All (1)	Control (2)	EUR 250 (3)	EUR 500 (4)
Pre-intervention Values (Start of Phase Zero)				
BMI before Rehab.	38.596	38.587	39.014	38.180
Baseline BMI	37.275	37.173	37.756	36.883
Target Weight Loss (Percentage)	6.427	6.403	6.408	6.469
Unhealthy Ingredients	0.038	0.04	0.038	0.038
Healthy Eating	0.167	0.13	0.173	0.192
Good Eating Habits	0.149	0.133	0.167	0.146
Exercise	0.327	0.302	0.315	0.364
Bad Kissingen	0.324	0.364	0.274	0.337
Bad Mergentheim	0.377	0.351	0.405	0.373
Isny	0.049	0.026	0.036	0.084
Glottertal	0.25	0.260	0.286	0.205
Female	0.328	0.416	0.274*	0.301
Age (Years)	47.873	46.636	48.536	48.349
Native	0.796	0.814	0.813	0.762
Married	0.651	0.676	0.671	0.608
End of Phase One				
Percentage Change Body Weight	-7.514	-7.899	-7.479	-7.192
Target Weight Realized	0.623	0.597	0.619	0.651
Unhealthy Ingredients	0.005	-0.030	0.053**	-0.013°
Healthy Eating	0.291	0.381	0.256	0.250
Good Eating Habits	0.119	0.176	0.120	0.063
Exercise	-0.021	0.017	-0.029	-0.048
Dropout				
Cum. Dropout Rate after Month: 10	0.135	0.143	0.119	0.145
22	0.340	0.364	0.321	0.337
Marg. Dropout Rate Months: 10–22	0.205	0.221	0.202	0.193
# of Observations	244	77	84	83

Notes: ** Deviation from the control group significant at 5%, * significant at 10%; °° deviation from the EUR 250 group significant at 5%, ° significant at 10%; standard deviations omitted because most variables are binary. "Bad Mergentheim," "Bad Kissingen," "Isny" and "Glottertal" refer to the locations of the four rehabilitation clinics.

Table 6: Mean Comparison Across Phase Two Premium Groups
(Incentivized Individuals in the First Phase Only)

		Experimental Groups			Δ to Control	
		Control	EUR 250	EUR 500	EUR 250	EUR 500
	# of Obs. (by Groups)	(1)	(2)	(3)	(4)	(5)
Unweighted Means (Months 5–10)						
Weight Loss (Percentage)	211 (66/74/71)	-3.139** (0.843)	0.295 (0.652)	-0.149 (0.761)	3.435** (1.065)	2.990** (1.076)
Fewer Unhealthy Ingredients	188 (57/66/65)	0.000 (0.025)	0.015 (0.026)	0.000 (0.000)	0.015 (0.030)	0.000 (0.030)
Healthier Eating	181 (55/64/62)	-0.182** (0.082)	-0.047 (0.061)	0.000 (0.069)	0.135* (0.100)	0.182** (0.101)
Improved Eating Habits	199 (62/69/68)	-0.145** (0.056)	-0.029 (0.062)	-0.029 (0.055)	0.116* (0.083)	0.116* (0.083)
More Exercise	173 (49/60/64)	0.020 (0.080)	0.000 (0.067)	0.078 (0.081)	-0.020 (0.112)	0.058 (0.110)
Unweighted Means (Months 5–22)						
Weight Loss (Percentage)	161 (49/57/55)	-4.173** (0.970)	-3.730** (1.041)	-3.595** (1.179)	0.443 (1.534)	0.578 (1.547)
Fewer Unhealthy Ingredients	145 (41/53/51)	-0.024 (0.024)	0.038 (0.026)	-0.039 (0.027)	0.062*°° (0.038)	-0.015°° (0.039)
Healthier Eating	139 (43/52/44)	-0.093 (0.087)	-0.096 (0.079)	-0.068 (0.076)	-0.003 (0.113)	0.025 (0.118)
Improved Eating Habits	153 (47/56/50)	-0.085 (0.095)	-0.036 (0.057)	0.020 (0.061)	0.049 (0.100)	0.105 (0.103)
More Exercise	132 (44/44/44)	-0.023 (0.076)	0.023 (0.069)	0.000 (0.103)	0.045 (0.119)	0.023 (0.119)

Notes: ** Significant at 5% (two-sided columns 1–3, one-sided columns 4 and 5), * significant at 10% (two-sided columns 1–3, one-sided columns 4 and 5); °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means in parentheses; a deviation from the control group.

Appendix

1. Summary of Previous Related Experiments

There are earlier clinical trials involving monetary rewards for weight loss in the obese that examine post-treatment effects. First, Volpp et al. (2008) examine a lottery incentive program and a deposit contract and find that monetarily induced lifestyle changes are not sustainable. The participants no longer exhibited improved behaviors about half a year after incentive removal. As pointed out by the authors, the experimental design does not allow the isolation of the effect of the financial incentives themselves from daily weight monitoring, call-ins, and feedback about progress. Only members of the treatment groups had to weigh themselves each day and to share body weight information with the project staff. If this differential treatment between treatment and control groups is relevant to weight loss, the effects cannot be attributed solely to the monetary incentives.

Second, Kullgren et al. (2013) study individual and group-based financial incentive programs offered by a company to its employees. They find statistically significant effects twelve weeks after the intervention for the team rewards only. At the monthly weigh-ins, the weight loss goal was revised each time it would have required the participants to lose more weight over the remaining intervention period than would be considered to be healthy. This implies that weight loss success effectively determined the overall weight loss goal, i.e., weight loss goals were endogenous. It is likely that weight loss goals per se are relevant to weight loss, which makes the attribution of the observed effects to monetary incentives difficult. Similarly, although provided equally to the treatment and control groups, extensive weight monitoring and feedback about progress during the intervention period are factors that make the nature of the incentive scheme different from genuine price variation. For instance, they may increase salience and loss aversion. Third, Finkelstein et al. (2017) find significant effects of monetary incentives on weight loss and weight-related behavior during and four months after the incentive period. In line with Kullgren et al. (2013), the overall weight loss target is endogenous.

Volpp et al. (2009) examine monetary rewards for the completion of a smoking cessation program, smoking cessation, and continued abstinence from smoking. The study finds that the treated participants were significantly more likely to quit smoking and less likely to relapse. The post-treatment effect of monetary rewards for smoking cessation is difficult to disentangle from the lasting effect of the smoking cessation program and the effects of providing monetary incentives for continued abstinence from smoking.

Further studies exist that focus on the estimation of the short-term effects of weight loss rewards (as opposed to the post-treatment effects). We refer to Paloyo et al. (2014) and Sykes-Muskett et al. (2015) for comprehensive reviews of this literature.

2. Technical Description of Inverse Probability Weighting

Technically, the objective of the analysis is the estimation of $E\left(\Delta y \,\middle|t_1^{reward}\right) - E\left(\Delta y\middle|t_1^{control}\right)$, where Δy denotes the weight loss and improvements in healthy behaviors over the entire period (10 and 22 months, respectively) and t_1^{reward} and $t_1^{control}$ group memberships in the first phase. To simplify the notation, we introduce a vector of group membership indicators t_1. Following Wooldridge's (2002) notation and indexing observations with i, the original (biased) estimator can be written as:

$$\min_{\beta} \sum_{i=1}^{N} (\Delta y_i - \beta' t_{1i})^2.$$

The estimator calculates a vector of group means β, which can be interpreted as running a linear regression on three experimental group indicators. By conditioning the analysis on participants who were not promised any reward in the second phase, the estimator takes the following form:

$$\min_{\beta} \sum_{i=1}^{N} s_i(\Delta y_i - \beta' t_{1i})^2,$$

where s_i is an indicator of the absence of rewards in the second phase. This estimator yields inconsistent estimates if s_i is correlated with the error term of the regression. The estimator, which – under certain conditions – proves to be consistent in the presence of endogenous sample selection, is:

$$\min_{\beta} \sum_{i=1}^{N} \frac{s_i}{p_i}(\Delta y_i - \beta' t_{1i})^2.$$

Here p_i denotes the probability of entering the estimation sample conditional on Δy_i and a vector of further variables z_i, i.e., $p_i = E(s_i|\Delta y_i, z_i)$. It is apparent that p_i is a function of the endogenous variable Δy_i. This implies that the above estimator for β is consistent only if s_i is uncorrelated with Δy_i conditional on z_i, rendering the estimator an inappropriate approach most of the time.

In the present case, however, the inverse probability weighting estimator satisfies this requirement for the following reason: including a binary variable r_{1i}, indicating success in the weight loss phase, along with t_{1i} in z_i removes the dependence of p_i and Δy_i:

$$p_i = E(s_i|\Delta y_i, t_{1i}, r_{1i}) = E(s_i|r_{1i}) = \begin{cases} 1 \;\; if \; r_{1i} = 0 \\ 1/3 \;\; if \; r_{1i} = 1. \end{cases}$$

The equation states that unsuccessful individuals are never eligible for the additional rewards and successful participants are promised the second reward with a probability of two to three, irrespective of observable and non-observable factors. Hence, the selection is exclusively based on observable factors.[47]

This selection problem does not arise in the examinations of Hypothesis 2. Since the additional rewards only concern those individuals who have already lost sufficient weight during the first phase, conditioning on $r_{1i} = 1$ is logical. As the randomization of group membership is conditional on $r_{1i} = 1$, too, group membership is purely random. Hence, comparing group means across second-premium groups yields unbiased estimates.

[47] Moreover, the design of the experiment guarantees another essential condition for inverse probability weighting: $E(s_i|r_{1i} = 1) > 0$. In many studies selection into the estimation sample is a deterministic function of variables such as success. Consider, for instance, a weight maintenance incentive scheme in a non-experimental context, in which success in weight reduction deterministically makes an individual eligible for incentives. Hence, $E(s_i|r_{1i} = 1) = 0$ holds and inverse probability weighting becomes impossible.

3. Table and Figures

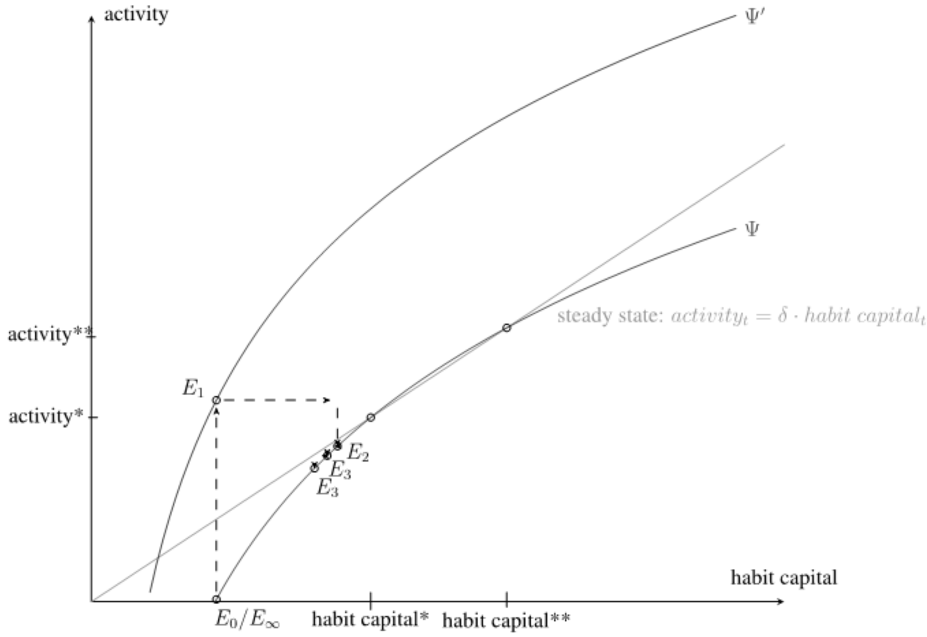


Figure A1: Temporary Financial Incentives are too Small to Induce Lasting Behavioral Change
Note: Ψ' denotes the shift of the activity curve by temporary cash incentives.

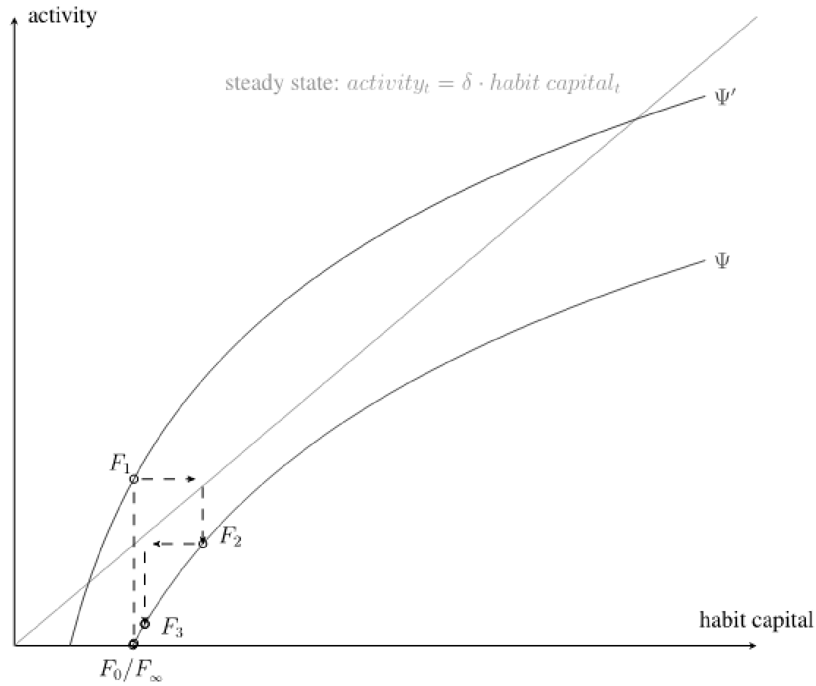


Figure A2: Temporary Financial Incentives of Any Size Will Not Induce Lasting Behavioral Change
Note: Ψ' denotes the shift of the activity curve by temporary cash incentives.

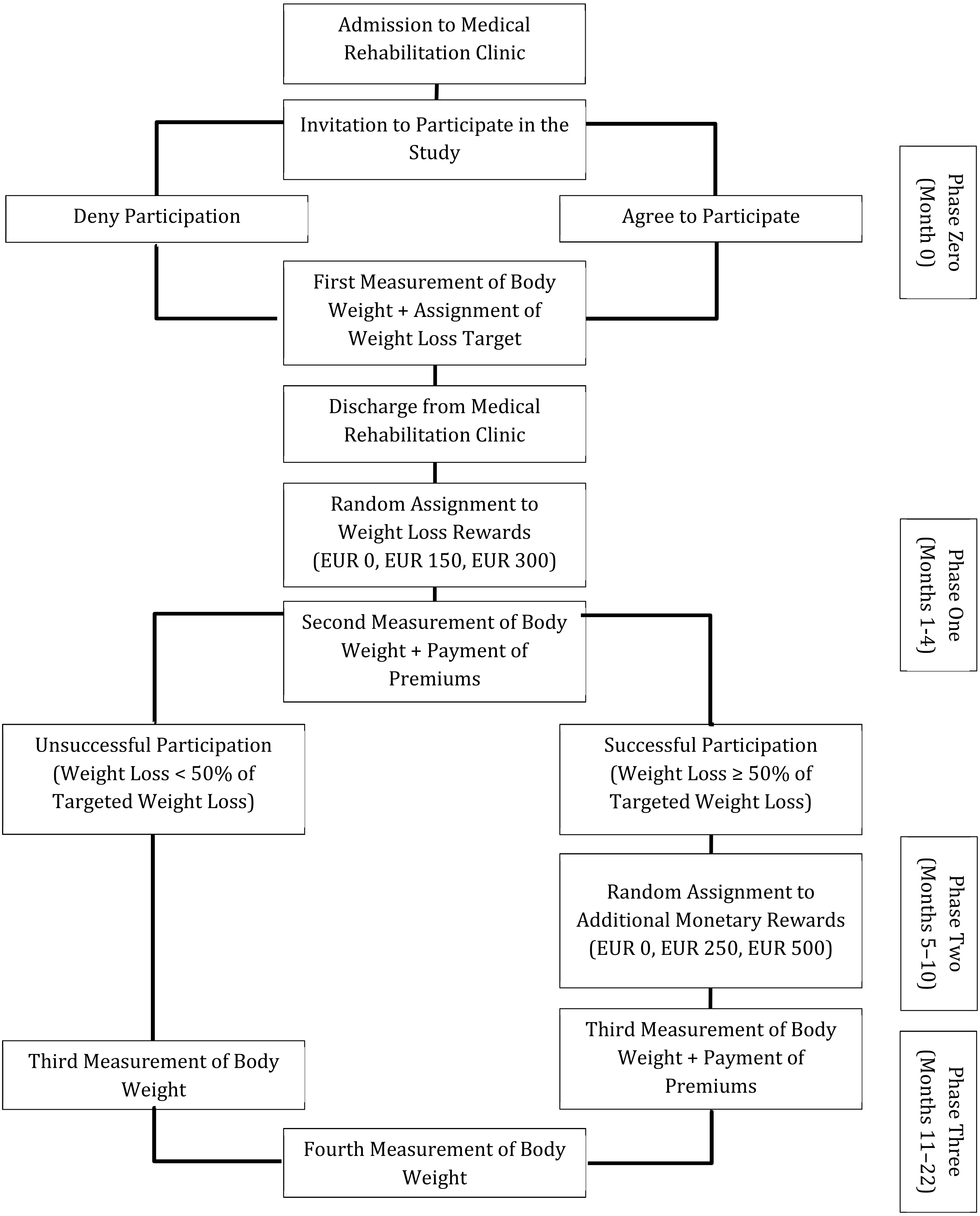


Figure A3: Flow Chart

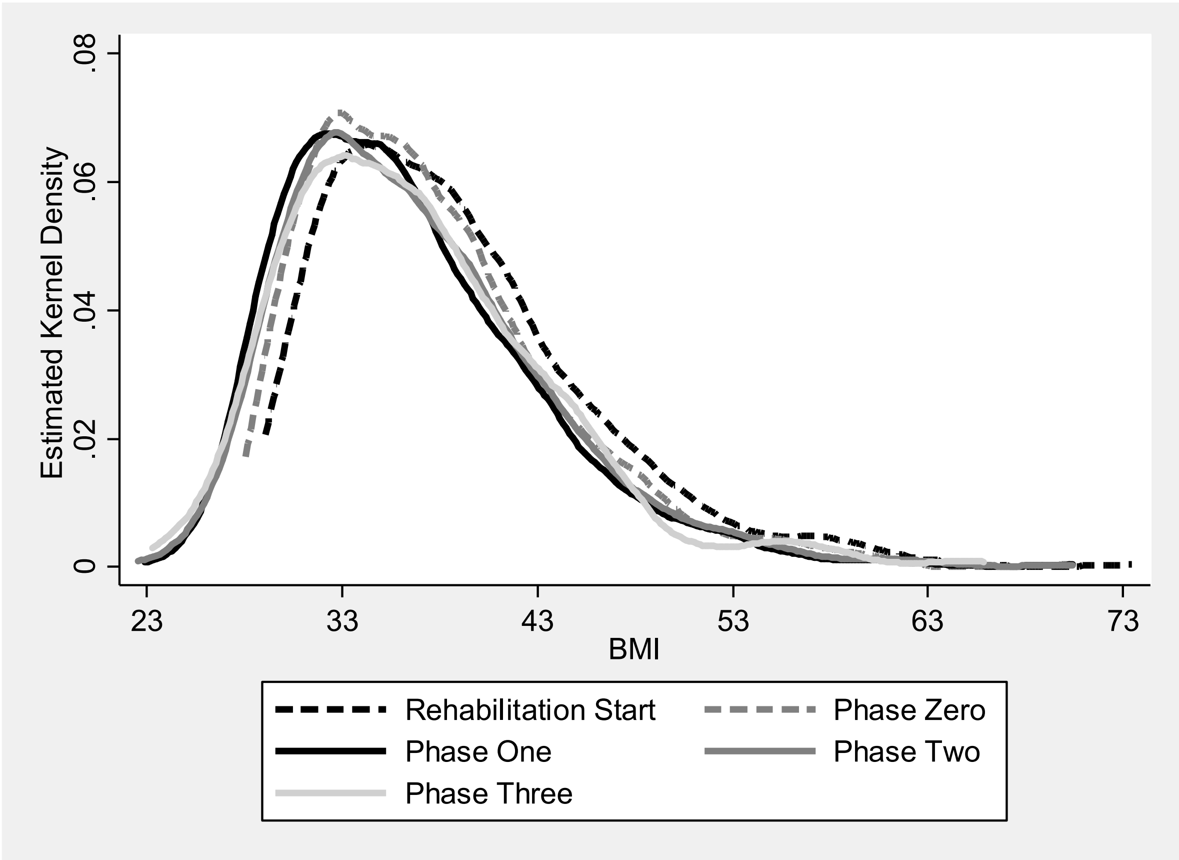


Figure A4: Distribution of BMI at the End of Each Phase

Notes: Missing values were imputed using the BMI at the start of the weight loss phase. The inclusion criterion of a BMI ≥ 30 refers to the day of clinic admission. Persons with a BMI ≥ 60 are often considered to be "super-super obese" (e.g., Stephens et al., 2008).

Table A1: Descriptive Statistics of Selected Samples
by Phase Two Premium Groups (Mean Values)

	All (1)	Control (2)	EUR 250 (3)	EUR 500 (4)
Pre-intervention Values of the Selected Sample after 10 Months				
BMI before Rehab.	38.383	38.622	38.774	37.750
Baseline BMI	37.112	37.268	37.570	36.491
Target Weight Loss (Percentage)	6.410	6.359	6.432	6.436
Bad Kissingen	-7.344	-7.841	-7.150	-7.062
Bad Mergentheim	0.582	0.576	0.578	0.591
Isny	0.331	0.341	0.289	0.364
Glottertal	0.369	0.388	0.367	0.352
Female	0.068	0.035	0.067	0.102*
Age (Years)	0.232	0.235	0.278	0.182
Native	0.308	0.365	0.311	0.250
Married	48.441	47.600	49.000	48.682
# of Observations	263	85	90	88
Pre-intervention Values of the Selected Sample after 22 Months				
BMI before Rehab.	38.014	38.230	38.585	37.203
Baseline BMI	36.783	36.842	37.478	35.993
Target Weight Loss (Percentage)	6.455	6.448	6.384	6.537
Bad Kissingen	-7.437	-8.119	-7.226	-7.029*
Bad Mergentheim	0.581	0.619	0.583	0.544
Isny	0.350	0.381	0.278	0.397
Glottertal	0.379	0.365	0.375	0.397
Female	0.069	0.048	0.069	0.088
Age (Years)	0.202	0.206	0.278++	0.118
Native	0.310	0.365	0.319	0.250
Married	48.576	47.159	49.194	49.235
# of Observations	203	63	72	68

Notes: ** Deviation from the control group significant at 5%, * significant at 10%; °° deviation from the EUR 150 group significant at 5%, ° significant at 10%; standard deviations omitted because most variables are binary. “Bad Mergentheim,” “Bad Kissingen,” “Isny” and “Glottertal” refer to the locations of the four rehabilitation clinics.

Table A2: Covariate-Adjusted Effects of Phase One Premiums

	# of Obs.	EUR 150	EUR 300
Months 1–4			
Weight Loss (Percentage)	518	2.460**	2.775**
		(0.567)	(0.552)
Fewer Unhealthy Ingredients	478	0.012	0.018
		(0.026)	(0.026)
Healthier Eating	444	-0.085	0.032
		(0.067)	(0.066)
Improved Eating Habits	505	0.018	0.003
		(0.049)	(0.048)
More Exercise	399	-0.044	0.001
		(0.080)	(0.078)
Months 1–10			
Weight Loss (Percentage)	234	1.618	3.028**
		(1.157)	(1.315)
Fewer Unhealthy Ingredients	217	-0.006	0.072*°°
		(0.027)	(0.043)
Healthier Eating	206	0.036	0.085
		(0.107)	(0.109)
Improved Eating Habits	229	0.092	0.097
		(0.084)	(0.084)
More Exercise	164	0.190	0.285**
		(0.125)	(0.135)
Months 1–22			
Weight Loss (Percentage)	176	0.548	2.789
		(1.671)	(1.754)
Fewer Unhealthy Ingredients	165	0.003	0.099*°
		(0.032)	(0.056)
Healthier Eating	151	-0.032	0.170
		(0.113)	(0.122)
Improved Eating Habits	173	0.076	0.057
		(0.103)	(0.103)
More Exercise	124	0.212	0.373**
		(0.140)	(0.170)

Notes: ** significant at 5%, * significant at 10%; °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means in parentheses.

Table A3: Covariate-Adjusted Effects of Phase Two Premiums

	# of Obs.	EUR 250	EUR 500
Months 5–10			
Weight Loss	211	2.809**	2.648**
(Percentage)		(1.229)	(1.200)
Fewer Unhealthy	188	0.044	0.039
Ingredients		(0.036)	(0.035)
Healthier Eating	181	0.138	0.176
		(0.122)	(0.120)
Improved Eating	199	0.173*	0.117
Habits		(0.097)	(0.095)
More Exercise	173	0.010	0.032
		(0.137)	(0.133)
Months 5–22			
Weight Loss	161	-0.421	0.518
(Percentage)		(1.920)	(1.860)
Fewer Unhealthy	145	0.072°°	-0.013°°
Ingredients		(0.049)	(0.048)
Healthier Eating	139	0.110	0.040
		(0.132)	(0.141)
Improved Eating	153	0.027	0.101
Habits		(0.126)	(0.125)
More Exercise	132	0.119	0.063
		(0.139)	(0.137)

Notes: ** significant at 5%, * significant at 10%; °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means in parentheses.

Table A4: Effects of Second-Period Premiums by Success in the First Phase

	# of Obs.	Target Weight Achieved in Weight Loss Phase		Target Weight Not Achieved in Weight Loss Phase	
		EUR 250 (1)	EUR 500 (2)	EUR 250 (3)	EUR 500 (4)
Months 5–10					
Weight Loss (Percentage)	154/109	3.528** (1.425)	1.687 (1.418)	2.110** (1.059)	2.441** (1.081)
Fewer Unhealthy Ingredients	138/97	-0.003 (0.036)	-0.024 (0.036)	0.003 (0.036)	0.032 (0.036)
Healthier Eating	132/94	0.107 (0.120)	0.020 (0.121)	0.165 (0.146)	0.159 (0.150)
Improved Eating Habits	147/104	0.024 (0.098)	0.065 (0.096)	0.202* (0.109)	0.166 (0.111)
More Exercise	129/90	-0.074 (0.126)	-0.008 (0.122)	0.090 (0.148)	0.245 (0.149)
Months 5–22					
Weight Loss (Percentage)	119/84	1.874 (1.943)	0.483 (1.992)	-1.377 (1.894)	-0.732 (1.894)
Fewer Unhealthy Ingredients	110/76	0.050 (0.049)	-0.083* (0.050)	0.048 (0.047)	0.011 (0.047)
Healthier Eating	104/71	0. 034 (0.127)	-0.048 (0.136)	0.105 (0.178)	-0.020 (0.184)
Improved Eating Habits	116/78	0.003 (0.118)	0. 137 (0.123)	0.018 (0.142)	0.010 (0.146)
More Exercise	101/66	0. 031 (0.135)	-0.063 (0.135)	-0.136 (0.169)	-0.056 (0.172)

Notes: # Lower absolute Lee bound; ** significant at 5%, * significant at 10%; °° difference between premium groups significant at 5%, ° significant at 10%; ++ difference in effects across target weight achievement in the weight loss phase significant at 5%, + significant at 10%; S.E.s in parentheses.