RUHR
ECONOMIC PAPERS

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

Small Cash Rewards for Big Losers – Experimental Insights Into the Fight Against the Obesity Epidemic

#530

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Jochen Kluve
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #530

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-607-9

Ruhr Economic Papers #530

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

Small Cash Rewards for Big Losers – Experimental Insights Into the Fight Against the Obesity Epidemic

UNIVERSITÄT DUISBURG ESSEN

RWI

RUHR UNIVERSITÄT BOCHUM RUB

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788607
ISSN 1864-4872 (online)
ISBN 978-3-86788-607-9

Boris Augurzky, Thomas K. Bauer, Arndt R. Reichert,
Christoph M. Schmidt, and Harald Tauchmann[1]

Small Cash Rewards for Big Losers – Experimental Insights Into the Fight Against the Obesity Epidemic

Abstract

We complement the empirical evidence on the sustainability of weight loss achieved through cash rewards and, for the first time, rigorously examine the potential of cash rewards to prevent weight cycling. In a three period randomized controlled trial, about 700 obese persons were first assigned to two treatment groups, which were promised cash contingent on the achievement of an individually assigned target weight, and to a control group. Successful participants were subsequently allocated to two treatment groups offered cash rewards for confirming the previously achieved target weight and to a control group. This is the first experiment of this kind that finds effects of weight loss rewards up to 18 months after they were removed. Additional rewards only significantly improve the sustainability of weight loss while they are in place.

JEL Classification: I12, I18, D03, C93

Keywords: field experiment; weight cycling; sustainability; incentives

December 2014

1 Boris Augurzky, RWI; Thomas K. Bauer, RWI, Ruhr-Universität Bochum, IZA Bonn; Arndt R. Reichert, The World Bank, RWI; Christoph M. Schmidt, RWI, Ruhr-Universität Bochum, IZA Bonn, CEPR London; Harald Tauchmann, RWI, CINCH, University of Erlangen-Nuremberg. *– The authors are grateful to the "Pakt für Forschung und Innovation" for funding, and to Rüdiger Budde, Viktoria Frei, Karl-Heinz Herlitschke, Klaus Höhner, Julia Jochem, Mark Kerßenfischer, Lionita Krepstakies, Claudia Lohkamp, Thomas Michael, Carina Mostert, Stephanie Nobis, Margarita Pivovarova, Gisela Schubert, and Marlies Tepaß for research assistance. We also gratefully acknowledge the comments and suggestions of participants of the following: participants of the research workshop "Empirical Economics" at the Ludwig Maximilian Universität München. 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 correspondence to Arndt R. Reichert, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany,e-mail: Thomas.Bauer@rwi-essen*

I. Introduction

People often make individual choices which differ from those that would maximize social welfare and even their own long-run utility. Monetary incentives that seek to change this kind of behavior have become increasingly popular. In fact, behavioral interventions across a wide range of areas–from contributions to public goods to education and health–nowadays include financial incentive schemes.

The emerging literature shows that financial incentives for healthy behavior clearly work in the short run. Yet, it is unclear whether induced changes disappear–or even reverse–when incentives are removed. There are two conflicting theories concerning the long run effects of cash rewards: The "motivation crowding theory" says that offering incentives may reduce helpful other motives of behavior, i.e., intrinsic motivation, and, therefore, increase unhealthy behavior through, for instance, the signal that replacing this behavior by a better one is difficult or not attractive. The alternative hypothesis ("habit-formation") explains a sustained change in habitual behavior by a positive correlation between past and current consumption, i.e., the development of behavioral automaticity. A literature overview is presented in Gneezy et al. (2011).

Based on data from a randomized experiment, the present paper investigates the longer run effects of financial incentives for weight reduction in obese people. Targeting body weight is particularly interesting because many obese individuals fail in their weight loss attempts and the majority among those who succeed soon regains the weight (Crawford et al., 2000). Finding that financially induced weight loss is sustainable would provide strong evidence in favor of the "habit formation" theory.

We furthermore examine the effect of financial incentives for confirming a previously achieved target weight. We are the first to rigorously test monetary rewards to sustain healthy behavior. Besides being important itself, stretching short run success could be important for the sustainability of the intervention effects because habit formation may take time. This seems to be especially relevant for eating habits (and perhaps exercise habits) given the widespread phenomenon of weight cycling.

The importance of finding effective means to fight the obesity pandemic is difficult to overestimate. On the one hand, obesity increases morbidity, reduces life expectancy, and deteriorates life satisfaction (for a comprehensive overview, see Sassi, 2010). On the other hand, through negative effects on the probability of being employed (Morris, 2007) and wages (Han et al., 2009) as well as positive effects on the risk of early retirement (Houston et al., 2008), costs of absenteeism (Cawley et al., 2007) and lifetime health care expenditures (Bhattacharya and Sood, 2011), obesity represents a significant burden for welfare systems.[1] The experiment involved 700 participants of four medical rehabilitation clinics and was conducted between spring 2010 and summer 2013. The experimental groups were first offered EUR 0, EUR 150 (USD 188 in PPP) or EUR 300 (USD 376 in PPP) for achieving an individually assigned contractual target weight loss between 6 and 8 percent within 4

[1] Based on data gathered from the same randomized experiment, Reichert (2012) reports a positive causal effect of weight loss on employment prospects.

months.[2] After completion of the weight loss phase, participants who had achieved at least 50 percent of the contractual weight loss were randomly assigned to three experimental groups which were promised EUR 0, EUR 250 (USD 313 in PPP), and EUR 500 (USD 627 in PPP) for confirming a body weight below the target weight 10 months after enrollment. Eventually, body weight was measured once more at the end of a 12 months period following the end of the second intervention.

As documented elsewhere (Augurzky et al., 2012), we find strong effects of both monetary rewards for weight loss at the end of the weight loss phase. Importantly, even though reward groups partially regain weight after removal of the incentives, the effects are persistent (results not presented elsewhere). This is the first experiment involving monetary rewards that achieved lasting effects on body weight.

Promising successful losers an additional cash reward to keep the healthier body weight is also an effective strategy. While the control group of the second intervention significantly regained weight, both reward groups were similarly successful in preventing the "yo-yo effect". At the end of the experiment, however, these differences were no longer observable.

This evidence survives a series of robustness checks. For instance, the estimated effects of monetary rewards for weight loss are robust with respect to non-random sample attrition. Moreover, potential strategic behavior of treated participants to achieve their targets, e.g., wearing light clothing, is not able to explain the results of the second intervention.

All in all, our results for the weight loss rewards are not in line with motivation crowding out because we do not observe that they worsen the degree of obesity in the longer term. This would be expected in the presence of remarkable negative effects on intrinsic motivation once the opposing relative price effect of monetary rewards stops due to incentive removal. On the contrary, our finding of lasting effects provides evidence in favor of the habit formation theory. However, it may well be that habit formation just dominates motivational effects. Results for financial incentives for confirming a previously achieved target weight are not perfectly in line with the habit formation theory. Alternative explanations are discussed.

Our work adds to a small but growing literature on the longer run effects of monetary incentives to encourage health preventive behavior. Prominently, Charness and Gneezy (2009) report the encouraging result that financial incentives to exercise are successful in creating a positive habit in people who formerly did not regularly exercise. Acland and Levy (2013) confirm their finding of increased exercise levels after the removal of the incentive but show that people eventually give up the acquired habit a few months later. Experimental studies on smoking cessation (e.g., Volpp et al. 2006) and weight loss (e.g., Volpp et al., 2008) do not find that monetarily induced lifestyle changes are sustainable in the sense that people exhibit improved behaviors even after incentive removal.

[2] We use the purchasing power parities exchange rate of 2011 provided by OECD (2012).

This paper also contributes to the most recent literature on financial incentives to sustain health-related behavioral change. We found only three studies that investigate the role of financial incentives to encourage maintenance of recently acquired positive behaviors. In a randomized experiment, Volpp et al. (2009) examine monetary rewards for completion of a smoking-cessation program, for smoking cessation, and, importantly, for continued abstinence from smoking, finding that treated participants were significantly more likely to quit smoking and less likely to lapse. A limitation of the study is that the causal effect of monetary rewards for continued smoking abstinence cannot be separated from the long-term effects of participation in a smoking-cessation program and financial incentives for smoking cessation.[3]

Royer et al. (2012) analyze the effectiveness of a self-funded commitment contract in improving the lasting effect of monetary incentives for exercise. Experiment participants were encouraged to deposit money, which was refundable contingent on the continuation of regular exercise. This approach has been previously examined in a randomized experiment on weight maintenance after substantial weight loss by Kramer et al. (1986). Participants in a treatment group paid a deposit of USD 120, which they were refunded if they did not regain weight within one year and attended several discussion meetings about weight maintenance progress and problems. While Royer et al. find that deposit contracts produced lasting effects, Kramer et al. do not observe significant differences in weight development between the deposit group and a control group. Since participants in Kramer et al. (1986) received not only the financial-incentive treatment, but also interacted with each other during discussion meetings, the effect of the deposit cannot be singled out. Further limitations of the study are discussed in Paloyo et al. (2013).

We extend the existing knowledge on monetary rewards for sustained health-related behavioral change by singling out their causal effect from other factors that may confound the results as in Volpp et al. (2009) and Kramer et al. (1986). In doing so, we concentrate on "carrots" rather than "sticks" as compared to Royer et al. (2012). We argue that analyzing carrots is simply more relevant in the face of a remarkable tendency in modern legal systems to increasingly use carrots (De Geest and Dari-Mattiacci, 2013). Moreover, we focus on obese individuals, who compared to healthy weight people may respond differently to behavioral interventions due to self-control problems.

The remainder of this paper is organized as follows. The subsequent section describes the experimental design and introduces the data, Section III discusses the estimation strategy, while Section IV presents the estimation results and Section V concludes.

[3] It is evident that smoking cessation programs still have an effect several months after their completion (Zhu et al., 2000; Quist-Paulsen and Gallefoss, 2003). Similarly, financial incentives for smoking cessation may have lasting effects. Volpp et al. (2006), for instance, find a positive–yet insignificant–long-term effect of monetary rewards for attendance of a smoking cessation program and for smoking cessation. They cannot reject these effects due to lack of statistical power.

II. Experiment and Data

IIA. Design and Implementation

As part of the excellence initiative of the German government, this research project has received financial support by the *Pakt für Forschung und Innovation*. We cooperated 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. Obese patients were invited to participate in the experiment in the final week of their rehabilitation stay. Medical rehabilitation consisted of a weight loss program that varied from clinic to clinic. Only patients with a BMI above 30 at admission, an age between 18 and 75 years, and who were a registered resident in the German federal state of Baden-Württemberg were invited to the experiment. Exclusion criteria were considerable language barriers, pregnancy, psychological and eating disorders, tumor disease within the last five years, abuse of alcohol and drugs, and serious general diseases. We informed participants about the procedures of the experiment by 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.

Baseline measurements of several medical variables such as the body-mass-index (BMI) were carried out in the clinics. Participants answered a detailed questionnaire related to their socioeconomic background, further health outcomes and preventive behavior. Moreover, they were assigned an individual weight loss target between six and eight percent of the current body weight for the weight-loss period by the physician in charge. Individual weight loss targets had to lie above the critical threshold associated with beneficial health effects (Vidal, 2002).

Aimed to assess the effectiveness of financial incentives to motivate obese rehabilitation patients to lose weight, the experiment consisted of four phases. After the discharge from the clinic (clinical phase), participants entered the weight loss phase (four months), which was followed by a six months weight maintenance and 12 months follow-up phase. Two randomizations took place: one at the start of the weight loss phase and an unannounced second one at the start of the weight maintenance phase. Figure I summarizes the experimental design.

Random assignment to the treatment and control groups in the weight loss phase took place after the discharge from the clinic. Stratified randomization by the clinics was carried out without replacement within blocks of 51 participants. Based on this randomization procedure, the participants were equally assigned to one of three groups: either the control group or to one of two treatment groups. While members of the control group were not promised to receive any reward for achieving their weight loss target, as described below, members of the treatment groups were promised to receive up to EUR 150 (‘group 150’) and 300 (‘group 300’) respectively.[4]

[4] The premium levels and the length of the treatment period are in the range of previous studies. Jeffery (1983), for instance, have premiums of US$ 30, US$ 150, and US$ 300, which correspond in terms of PPP to

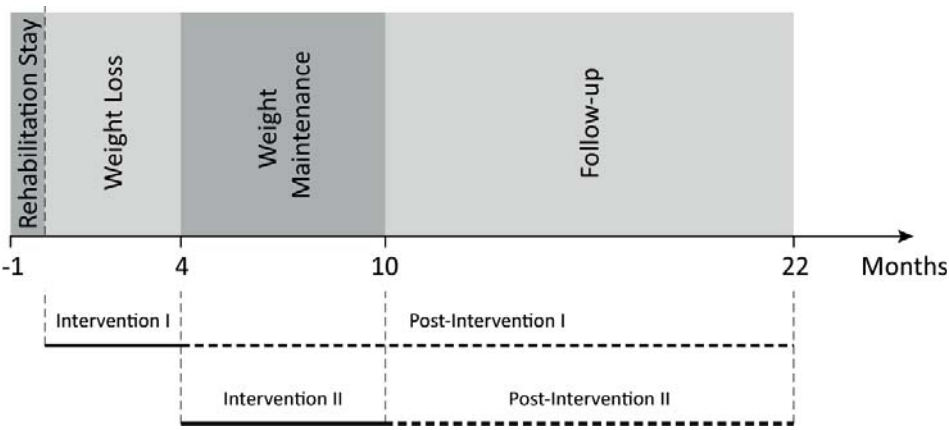

Figure I: Experimental Design

All successful participants (irrespective of group assignment in the weight loss phase) were randomized again at the start of the weight-maintenance phase. Participants were considered as successful if their achieved weight loss exceeded fifty percent of targeted weight loss. Randomization (without replacement and stratification by the clinics) was used to produce three weight-maintenance experimental groups with equal shares of participants. Here, the two weight maintenance premium groups were promised to receive up to EUR 250 ('group 250') and 500 ('group 500'), respectively. Participants assigned to the control group were not informed about weight maintenance randomization. All participants were told to assure that their weight does not exceed the individually assigned target weight during weight maintenance and follow-up.

Members of the premium groups were paid the full bonus if they had reached or even exceeded their weight loss target at the end of the respective phase. Once the achieved weight loss exceeded fifty 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 the reduced body weight during the weight maintenance phase. As a member of the control group in both phases, she receives no premium. As a member of the treatment group in the weight loss phase, she obtains EUR 107 (USD 134 in PPP) in the lower reward group and EUR 214 (USD 268 in PPP) in the higher reward group. At the end of weight maintenance, she receives another EUR 179 (USD 224 in PPP) and EUR 357 (USD 477 in PPP), depending on being 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 confirming the target weight later on) she receives the whole group specific premiums.

Participants were informed by regular mail about their maximum possible premium (does not apply to members of the control groups) and of the week they were supposed to

EUR 54, EUR 272, and EUR 544 in prices of 2011 (converted into present values of EUR based on the US consumer price index and the purchasing power parities exchange rate of 2011 provided by OECD (2012c)).

attend the weigh-in at a pharmacy.[5] Since the participants spent the intervention periods outside the medical rehabilitation clinic, interactions between participants are very unlikely. Thus, we do not expect a perception of unfairness that may be associated with randomization. Most importantly, control group participants should not be affected by the treatment status of other participants.[6]

We asked 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 the body weight was sent to 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. In order not to rely on self-reported weight, the reminder indicated to the participants a nearby pharmacy for the control measurement. Pharmacies had been called by project staff beforehand in order to ask for participation. By assigning participants to specified pharmacies we ruled out that treatment group participants go from one pharmacy to the other in order to take advantage of probable measurement errors of the scales, i.e., strategic behavior to achieve their targets.

Experiment attrition occurred in two ways. First, some participants left the experiment by actively canceling their participation. Second, a larger number did not send the required documents at the end of an experimental phase. To reduce sample attrition, 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 sent in the documents, regardless of 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.

IIB. The Participants

The recruitment of a total number of 700 participants took place between March 2010 and August 2011. Five individuals had to be excluded from the trial because of a missing consent form, becoming pregnant, developing cancer, and internal documentation problems.[7] The last participant finished follow-up by the end of July 2013.

The average body weight at the start of the experiment (after rehab) is 113.0 kg (249.4 lbs.) or 37.6 in terms of BMI (for the distribution of the BMI over time, see Figure AI). Table I provides a description of the study population. About 68 percent of the participants are men and 21 percent have a migration background. These shares are substantially lower as the corresponding averages for the obese in Germany.[8]

[5] Participants could postpone the date of measurement or move it forward by means of an early phone call.
[6] See Angrist and Lavy (2009) for a similar argument in the context of a within-school randomized trial.
[7] Results are robust with respect to treating these individuals as dropouts in sensitivity checks described in Section III.
[8] Descriptive statistics for the average obese in Germany based on a representative German household panel (SOEP) are presented in column three of the Table. Column two of this table displays socioeconomic characteristics of the average patient of the rehabilitation clinics.

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

	Study Population	Patients of the four rehabilitation	Representative Obese in Germany (BMI
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
Natives (%)	78.89	82.67	86.30
Full-time employed$^{\lozenge,+}$ (%)	69.44	76.12	34.85
Part-time employed$^{\lozenge,+}$ (%)	9.04	11.01	14.27
Unemployed$^{\circ,+}$ (%)	13.20	8.23	6.90

Notes: Statistics relating to the patients of the four rehabilitation clinics are weighted averages. As the clinics' weights, we use the shares of participants recruited by the clinics. $^{\lozenge}$The remaining observations among those who report to be employed are marginally employed (2.15 percent) or have not provided information on the type of employment (1.72 percent). $^{+}$Here, we distinguish between the unemployed and the not-employed (4.45 percent). $^{\circ}$The categories full-time employed, part-time employed, marginally employed, no information on type of occupation, unemployed, and not-employed add up to one.
Source: Own data collection, German Federal Pension Fund, German Socio-economic Panel (SOEP).

The mean age of the study population (48 years) lies about ten years below the average age of the obese in Germany, while the share of employed participants (82 percent) is almost twice as large. Only the share of married participants does not considerably deviate from the respective share of Germany's obese. The main reason for these differences is that for most patients of the co-operating clinics medical rehabilitation is paid by the German pension fund, whose predominant goal is to avoid work inability and early retirement. Since there are many obese in the overall population that are already retired, our study population oversamples persons that are available for the labor market.

Four rehabilitation clinics, located in different towns, have been involved in the trial. 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 are specialized on gastroenterology as well as endocrinology and patients with psychosomatic disorders, respectively. Many participants came to the clinics because of diagnoses other than adiposity although their symptoms are related to their body weight. All participants are medically indicated to lose weight.

III. Hypotheses and Methods

The analysis aims to estimate causal effects of financial incentives on body weight. While short term effects are analyzed in detail by Augurzky et al. (2013), the present paper concentrates on the estimation of medium and longer term effects. Two main hypotheses are analyzed: (i) Financial incentives for weight loss have an effect on body weight after their removal, i.e., post-intervention effects; (ii) Monetary rewards for confirming a previously achieved target weight prevent weight regain during the intervention and after the intervention has ended. Here, we additionally analyze effect heterogeneity across the

degree of target weight achievement in the weight-loss phase and across treatment status in the first intervention. As secondary outcome variable, we use a dummy variable which indicates whether the individually assigned target weight is met.

In the first hypothesis, we examine whether individuals who were exposed to financial incentives during the weight loss phase (group 150 and group 300) have lost more weight compared to the control group 10 and 22 months after the start of the experiment. If we find significant differentials across weight loss experimental groups in weight change between the start of the intervention period and 6 as well as 18 months after the intervention, estimates for the effect of monetary rewards for weight loss presented in Augurzky et al. (2013) are persistent, i.e., financially incentivized obese people do not regain more body weight lost than members of the control group do.

In order to examine the second hypothesis, we compare mean weight loss over the weight maintenance phase between individuals who were promised rewards for confirming a previously achieved target weight (group 250 and group 500) and the control group. Only individuals who were eligible for randomization prior to the weight maintenance phase, i.e., those who successfully reduced their body weight during the weight-loss phase, are considered in the analysis. We also address lasting effects here. Hence, besides considering weight change between months 5–10, we look at the change in body weight between months 5–22. The longer period allows us to investigate the effects of the two monetary rewards after they have been removed for about one year.

Analyzing effect heterogeneity regarding the degree of target weight achievement addresses the question whether it makes a difference for the effectiveness of the second intervention that individuals, who only partially achieved their target weight in the weight loss phase, have to continue to lose weight in order to obtain the full premium. This means that they actually receive a hybrid reward: While they receive some money for confirming the previously achieved body weight, they may obtain some additional money if they achieve their target weight in full. To address this question, the analysis is conducted separately for participants who partially and who fully achieved their weight loss target in the weight loss phase.

Effect heterogeneity across treatment status in the preceding weight loss phase, on the other hand, enables us to indirectly test the theory of motivation crowding out, which suggests that intrinsic motivation is reduced by extrinsic rewards. Due to the eligibility criterion for the second intervention, all participants who were included in the randomization are successful losers. Yet, members of the weight loss control group were successful in the absence of extrinsic rewards. They reduced their body weight in the weight loss phase based on intrinsic motivation alone. In contrast, members of weight loss reward groups reduced their body weight based on either intrinsic motivation, extrinsic motivation or the combination of the two. Hence, this group, on average, achieved weight loss relatively more often based on extrinsic motivation. According to the motivation crowding theory, monetary rewards for confirming the previously achieved body weight should be less effective in members of the weight loss control group than members of the weight loss

treatment groups. We argue that, in the latter group, there is simply less intrinsic motivation that can be destroyed by extrinsic rewards. Here, the analysis is conducted separately for premium group members and control group members in the weight loss phase.

Since the present analysis rests on data generated in the course of a randomized trial, simply comparing means across treatment and control groups yields an unbiased estimate of the causal effect. The reason is that randomization ensures that experimental groups do exclusively differ in terms of receiving the treatment. Aimed to address sensitivity of our results with respect to random imbalance of individual characteristics and potential strategic behavior of participants to achieve their target, we use multivariate ordinary least squares regressions (i.e., the linear probability model for the binary indicator). As covariates we include age, gender, month of recruitment, and variables that relate to the weigh-in. We asked the pharmacists to indicate whether the participants' last food intake was more than half an hour or even more than two hours ago, whether they were wearing shoes (and if so whether these were heavy), a pullover, long trousers, and whether they attended the control weigh-in within the specified time. An additional set of dummy variables captures whether participants attended the control weigh-in prior to the specified date of measurement, within the right week (reference category), two weeks, three weeks or more than three weeks after this date. Variables that describe the condition at the control weigh-in allow us to capture possible ways of how participants could influence their measured body weight other than through weight loss. This may be particularly relevant for the analysis of monetary rewards for confirming a previously achieved target weight (second hypothesis) since, at the end of the second intervention, members of the treatment groups may behave strategically to achieve their target, i.e., to increase the bonus.

Except for those related to the weigh-in, all variables enter the analysis as pretreatment values. Following a standard approach (e.g., Morris, 2006; Spenkuch, 2012), we deal with missing values in covariates by replacing them by zero and including additional dummy variables indicating missing values. Only sex is imputed using the prediction from a probit regression of the variable on relevant individual characteristics. Imputation is preferred to the complete case method, i.e., 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.

In subsequent paragraphs, we describe two selection problems that we need to tackle in the analysis of the first hypothesis. While the first selection problem requires the use of the inverse probability weighting estimator throughout the analysis, the second selection problem is addressed in the course of sensitivity checks.

Addressing Selection Problem Induced by the Second Intervention (Hypothesis 1 only)

The design of the experiment involves financial incentives at different stages of the experiment. This challenges the isolation of lasting effects of the monetary rewards for weight loss because successful participants in the weight loss phase were eligible for incentives in the weight maintenance phase, whereas those who failed to reach the target weight were automatically excluded from the second intervention. Provided that success in

the weight loss phase was greater in the weight loss premium groups and assuming effectiveness of the second intervention, a simple comparison of weight development across weight loss experimental groups yields biased effect estimates. In specific, these comparisons will most likely exaggerate the effects because they partially capture the negative body weight effect of the weight maintenance rewards.

Excluding individuals who are promised monetary rewards for confirming a previously achieved target weight invokes another identification problem. Since eligibility for these rewards is endogenous, a comparison of weight development across weight loss experimental groups gives a disproportionate high weight to individuals who failed to reach the target weight. However, a simple inverse-probability weighting estimator as, for instance, the one suggested by Wooldridge (2002) is able to solve the endogeneity problem. The estimator weighs observations in such a way that the original distribution of observations across weight loss experimental groups is restored. Importantly, due to the experimental design, we know the exact probability that an individual is excluded from the sample (which in observational data needs to be estimated) and therefore dispose of correct information regarding the weight for each observation. We refer to Section A.I in the Appendix for a more technical description of the problem and the estimation method.

Addressing Sample Attrition (Hypothesis 1 only)

As mentioned in Section IIA., several participants drop out from the experiment, despite substantial efforts to keep attrition rates low. In detail, from the initial 695 participants 177 already dropped out during the weight loss phase, another 106 did not attend the weigh-in at the end of the weight maintenance phase, and further 96 participants dropped out during follow-up. If sample attrition was random, our estimates for the effects of the financial incentives would be unbiased. However, we are concerned that termination of experiment participation is endogenous because members of the premium group tend to have lower cumulative attrition rates than members of the control group 10 and 22 months after the start of the experiment (Table II). A lower marginal dropout rate cannot be explained by financial motives because weight loss rewards are already removed. Yet, the significantly lower attrition rate of group 300 at the end of the weight loss phase, which may well be attributed to financial motives, is likely to be carried forward due to the fact that participants who dropped out of the trial were no longer asked to attend weigh-ins at the pharmacy. We therefore employ several estimation methods that address this selection problem.

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. In the course of the phone call, these participants were interviewed regarding the current body weight. We argue that participants had no financial incentive to misreport body weight as monetary rewards for weight loss were no longer promised (and participants who belong to the premium groups in the weight maintenance phase are excluded from the analysis as discussed below). Using self-reported

weight for individuals without ordinarily measured weight information substantially reduces the attrition rate in the estimation of the incentive effects.

Two further methods aim to estimate treatment effects under extreme assumptions about the distorting effect induced by non-random sample attrition. The second method is the so called intention-to-treat approach,[9] which represents best practice in the medical literature. The intention-to-treat approach aims to consider all participants in the analysis irrespective whether they actually dropped out from the trial. More specifically, the method consists of missing information imputation, whereby no consensus on the imputation algorithm seems to exist (Hollis and Campbell, 1999). Like most medical studies, we make the extreme assumption that body weight of dropouts remained at the baseline level, i.e., zero reduction in body weight is assumed. This assumption–in some sense–is not perfectly consistent with our data, as in any phase of the experiment there are individuals who reduce weight and individuals who gain weight.[10]

The third method is the trimming procedure proposed by Lee (2009). It trims the distribution of the outcome variable for the experimental group (treatment or control) that suffers less from sample attrition (i.e., that has more participants with information on the outcome variable, i.e., `excess observations') at 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 not-trimmed sample of the other yields the estimated treatment effect bound. Here, we make the extreme assumptions that 'excess observations' in one group are those with the most favorable and least favorable weight development. This yields the lower and the upper bound of the treatment effect, respectively, depending on whether trimming is from below or above. According to Lee (2009), the procedure yields the bounds for the average treatment effect among always compliers.

IV. Results

We first analyze the lasting effects of weight loss rewards and subsequently address the mean impact of monetary rewards for confirming a previously achieved target weight as well as their impact on relevant subgroups.

IVA. Effect of premiums for weight loss

Randomization Check

Before investigating lasting effects of the weight loss premiums, we give reassurance that the randomization worked properly and that the inverse-probability weighting estimator properly restores the original distribution of observations across weight loss experimental groups. The upper panel of Table II provides an overview of relevant individual characteristics for the study population used in the analysis of the first intervention (Column 1) and each experimental group separately (Columns 2-4). Most variables are balanced

[9] In the present context, rather than selection into treatment, selection into the estimation sample is the relevant problem. Yet, we stick to this terminology as it is standard in the medical literature (see e.g. Hollis and Campbell, 1999).

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

Table II: Descriptive Statistics by Weight-loss-premium Groups
(Mean Values, *Inverse-probability Weighting)*

	All	Control	EUR 150	EUR 300
Pretreatment Values				
BMI Before Rehab	38.948	38.574	38.749	39.547
Baseline BMI	37.583	37.222	37.330	38.231
Target Weight Loss (Percent)	6.501	6.436	6.601	6.473
Bad Kissingen	0.336	0.290	0.367	0.356
Bad Mergentheim	0.421	0.453	0.394	0.413
Isny	0.053	0.057	0.062	0.040
Glottertal	0.190	0.200	0.177	0.191
Female	0.336	0.249	0.332	0.436**
Age (years)	47.080	47.686	46.084	47.422
Native	0.796	0.810	0.767	0.810
Married	0.638	0.679	0.619	0.614
Posttreatment Values				
Percentage Change Body Weight (Months 0-4)	-4.489	-2.998	-5.135**	-5.260**
Target Weight Realized After 4 Months	0.337	0.214	0.370**	0.418**
Total Dropout Rate After Month: 4	0.254	0.314	0.283	0.160**°°
10	0.420	0.498	0.451	0.302**°°
22	0.566	0.624	0.588	0.480**
# of Observations (Unweighted)	489	192	158	139

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

between the experimental groups, including body weight at the start of the medical rehabilitation stay and at baseline. Average target weight loss within the first four months of the trial amounts to 6.5 percent, which is well above the critical threshold for health improvements in the obese of 5 percent (Vidal, 2002).

The lower panel of Table II describes weight loss during the first four months of the experiment. As previously reported in Augurzky et al. (2012), all experimental groups were able to reduce their body weight. Weight loss of the control group may be attributable to lasting effects of the clinic weight loss program or the effect of receiving a specified weight loss target by a physician. Weight loss premium groups lost, on average, significantly more weight than the control group. Likewise, they were more likely to achieve the individually assigned target weight. Group 300 was furthermore significantly less likely to drop out of the experiment after four months. The same attrition pattern is observable at the end of the weight maintenance and follow-up phases, motivating the use of methods that deal with non-random sample selection in sensitivity checks.

Body Weight After 10 Months (6 Months After Intervention I Ended)

Table III: Mean Comparison Across Weight-loss-premium Groups
(Inverse-probability Weighting)

	Experimental Groups			Δ to Control	
	Control	EUR 150	EUR 300	EUR 150	EUR 300
Months 0-10					
Percentage Change in Body Weight	-1.343 (0.926)	-2.468** (0.895)	-4.155** (0.879)	-1.125 (1.288)	-2.812** (1.277)
Target Weight Realized	0.203** (0.060)	0.258** (0.065)	0.236** (0.057)	0.055 (0.088)	0.032 (0.083)
# of Observations (Unweighted)	85	70	79	-	-
Months 0-22					
Percentage Change in Body Weight	-0.137 (1.194)	-1.202 (1.271)	-3.472** (1.135)	-1.065 (1.744)	-3.335** (1.648)
Target Weight Realized	0.174** (0.065)	0.258** (0.074)	0.239** (0.066)	0.084 (0.098)	0.065 (0.093)
# of Observations (Unweighted)	64	53	59	-	-

Notes: ** significant at 5%, * significant at 10% ; °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means and for coefficients in parentheses. All coefficients are obtained by inverse-probability weighting OLS, regressing the respective outcome variable on the dummy variables indicating the two premium groups.

At the end of the weight maintenance phase, we observe that all experimental groups, on average, weigh less than at the start of the experiment. As displayed in the upper panel of Table III (Columns 1-3), the control group, group 150, and group 300 lost about 1.3, 2.5, and 4.2 percent, respectively. Weight loss is only significant in the two reward groups. The differences between each treatment group and the control group amount to 1.1 and 2.8 percentage points (Columns 4-5). This means that group 150 and group 300 lost, on average, 1.1. and 2.8 percentage points more weight than the control group. Weight loss of group 300 (but not of group 150) was significantly higher than weight loss of the control group. Nevertheless, we do not observe any statistically significant difference in weight loss across the two premium groups. Pooling both treatment groups together yields a difference to the control group of about 2.1 percentage points, which is significant at the 7 percent level (not displayed in the table). Figure II displays the distribution of weight loss by experimental group indicating that the effects of the rewards are not primarily due to a small number of participants with very large changes in body weight.

Considering the binary indicator for target weight achievement, we observe that both reward groups were more likely to be successful than the control group. On average, group 150 and group 300 had a 5.5 and 3.2 percentage points higher likelihood to achieve their

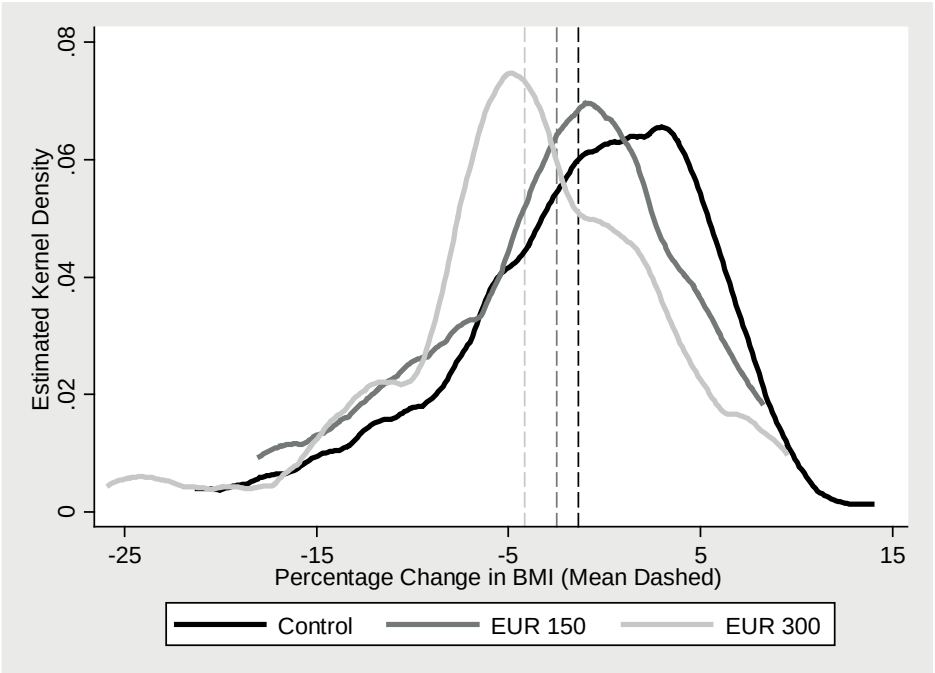

Figure II: Distribution of Percentage Change in Body Weight by Experimental Groups of Intervention I (Months 0-10)

Notes: Inverse-probability weights used for the estimation of kernel densities.

individual target weight than the control group. Remarkably, the share of successful participants is lower in group 300 than in group 150. Yet, all these differences are statistically insignificant.

Body Weight After 22 Months (18 Months After Intervention I Ended)

We observe a very similar pattern at the end of the follow-up phase as at the end of the weight maintenance phase (lower panel of Table III). While weight loss diminishes in all experimental groups (Columns 1-3), weight loss is still larger in the two premium groups. The difference between group 150 and the control group remains roughly the same as after 10 months. The difference between group 300 and the control group slightly increases from 2.8 to 3.3 percentage points and is statistically significant (Column 5). Yet, the inter incentive-group differential remains statistically insignificant. The difference between both treatment groups pooled together and the control group amounts to 2.3 percentage points (*p*-value of 0.11, not displayed in the table). Figure III illustrates the distribution of weight loss by experimental group.

The intergroup differences in the share of participants who achieve their individually target weight increases from month 10 to month 22. The group 150 and group 300 now has a 8.4 and 6.5 percentage points higher likelihood to be successful than the control group.

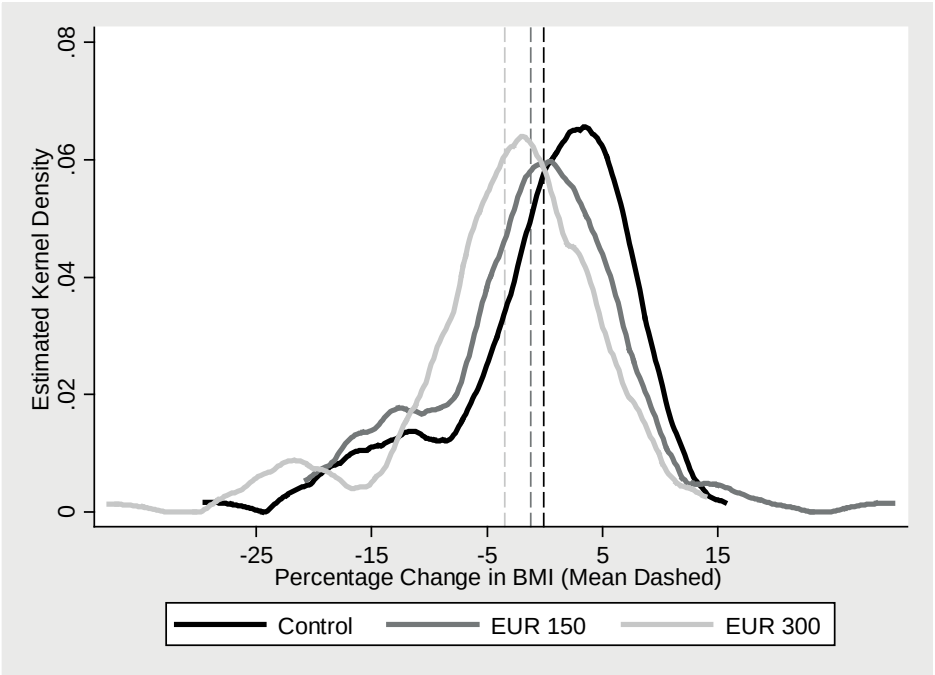

Figure III: Distribution of Percentage Change in Body Weight by Experimental Groups of Intervention I (Months 0-22)

Notes: Inverse-probability weights used for the estimation of kernel densities.

Again, the share of successful participants is lower in group 300 than in group 150. All these differences are statistically insignificant.

Sensitivity Analyses

Results proof to be robust with respect to taking into account individual characteristics and variables related to the weigh-in in a multivariate regression. The coefficients for the binary indicator for group 300 are statistically significant at the 2 and 6 percent level after 10 and 22 months, respectively (Table AI in the Appendix). Only the coefficient of the indicator for group 150 turns positive after 22 months. The coefficient of a pooled indicator for both treatment groups is statistically significant after 10 months and insignificant after 22 months (*p*-value of 27 percent, results not displayed in the table). Results are also robust for the binary indicator `target weight realized'.

Results for the three different approaches that deal with non-random sample attrition are presented in Table IV. Including observations with self-reported information on body weight into the analysis does not remarkably alter the results (Column 1). The differences between group 300 and the control group are significant at the 6 and 2 percent levels after 10 and 22 months, respectively. After 22 months, there is a weight loss differential among the two premium groups that is significant at the 8 percent level. Results from simple mean comparisons of group means are also confirmed if the two treatment

Table IV: Attrition-robust Effects of Weight-loss Premiums
(Inverse-probability Weighting)

	Self-reported Weight		Intention-to-Treat		Lee-Bound[#]	
	EUR 150	EUR 300	EUR 150	EUR 300	EUR 150	EUR 300
Months 0-10						
Percentage Change in Body Weight	-0.760 (1.100)	-2.137* (1.145)	-0.680 (0.702)	-2.225**° (0.807)	0.334 (1.155)	0.391 (0.869)
Target Weight Realized	0.007 (0.076)	-0.005 (0.074)	0.040 (0.050)	0.062 (0.053)	-0.014 (0.091)	-0.203 (0.036
# of Observations (Unweighted)	305		488		488	
Months 0-22						
Percentage Change in Body Weight	-0.757 (1.506)	-3.394**° (1.461)	-0.443 (0.693)	-1.754** (0.770)	0.464 (1.969)	0.462 (1.175)
Target Weight Realized	0.052 (0.084)	0.055 (0.083)	0.041 (0.043)	0.059* (0.046)	0.013 (0.113)	-0.174 (0.040)
# of Observations (Unweighted)	215		488		488	

Notes: [#] Lower Absolute Lee-Bound; ** significant at 5%, * significant at 10% (one-sided test); °° difference between premium groups significant at 5%, ° significant at 10%; S.E.s for estimated means in parentheses.

groups are pooled together (not displayed in the table). Considering the secondary outcome variable, we find qualitatively the same results at the end of the follow-up period as before. Yet, after 10 months, both reward groups are only about as successful as the control group.

Also the intention-to-treat analysis generally confirms simple mean comparison results (Column 2). The differences in weight loss between group 300 and the control group after 10 and 22 months are statistically significant. Also group 150 has a higher average weight loss than the control group, yet, the differences after 10 and 22 months are not statistically significant. We observe statistically significant differences in weight loss between group 300 and group 150 at the 6 percent level after 10 months and at the 11 percent level after 22 months. The intention-to-treat analysis yields significant differences for the pooled treatment group after 10 and 22 months (*p*-values of 2 and 7 percent).

In contrast, the trimming procedure proposed by Lee (2009) does not confirm previous results (Column 3). The very conservative lower effect bounds yield that, at the time of the post-treatment weigh-ins, *lower* weight loss of both reward groups relative to the control group can still be regarded consistent with the data. The optimistic effect bounds (not displayed in the table), on the contrary, point to the possibility of a highly increased weight loss in both reward groups relative to the control group. For instance, the estimated upper absolute Lee-bound indicates an increased weight loss by 2.1 and 5.1 percentage

points due the EUR 150 and EUR 300 reward, respectively. Results for the secondary outcome variable are qualitative the same.

Discussion

Our analysis reveals that monetary incentives for weight loss have lasting effects on the body weight of obese individuals. We observe a statistically significant effect of about 2.8 and 3.3 percentage points for the EUR 300 reward 6 and 18 months after its removal. The lower reward is also positively related to weight loss, yet, its effect is individually insignificant at the end of both post-treatment periods. There is not statistically significant difference between the effects of the higher and the lower reward. In fact, if we pool both treatment groups together, we observe a significant treatment effect at significance levels of around 10 percent.

The evidence survives a series of sensitivity checks. Yet, the estimated lower absolute Lee-bounds call the general finding of lasting effects of financial incentives into question. Nevertheless, among approaches that probe the sensitivity of the results with respect to non-random sample attrition, we pay less attention to the Lee-procedure because we believe that it relies on assumptions that are too extreme for the present case. The reason is that successful members of the reward groups do not have stronger financial incentives to continue with the experiment because weight loss does no longer imply any additional cash payments. The assumption that excess observations are those with the most favorable weight development after the intervention has ended (cf. Section III) is just difficult to justify. Without doubts, the assumptions underlying the upper absolute Lee-bound are also unrealistic. Hence, we rather consider the estimated lower and upper treatment effect bounds as a very broad interval that includes the true treatment effect.

Our finding that the higher reward does not cause significantly more weight loss after 10 and 22 months relative to the lower reward is quite sensitive with respect to both including covariates and accounting for non-random sample attrition. Hence, we cannot rule out that lasting effects increase with reward size. It is important to mention that the estimated treatment effects after pooling the two treatment groups together survive sensitivity checks (except the trimming procedure which yields a lower bound rather than a point estimate).

Results of the first intervention argue in favor of the habit formation theory. Even though treated participants slightly regain weight after incentives are removed, we observe a positive lasting effect on weight loss. The development of a behavioral automaticity that operates against the general tendency of weight regain is best able to explain the results. In turn, our results argue against monetary rewards crowding out intrinsic motivation in the present study population. Yet, it may well be that motivational effects are also present but just not large enough to notably oppose beneficial effect of developed habits.

Our findings are well in line with Charness and Gneezy (2009) who show that financial incentives to exercise have positive lasting effects, arguing that there is scope for monetary intervention in health-related habit formation. While previous experimental studies on financial incentives for weight loss, such as Volpp et al. (2008), do not find evidence of a

backfire effect either, we are the first to report a positive effect of cash rewards for weight loss after the intervention has ended.

IVB. Effect of premiums for confirming a previously achieved target weight

Analysis of the Mean Effect

Randomization Check

With the purpose of giving reassurance that the randomization worked properly, we present descriptive statistics for the study population used in the analysis of the second intervention (Column 1) and each experimental group (Columns 2-4) in the upper panel of Table V. Except one clinic indicator, all variables are balanced between the experimental groups. Importantly, body weight at the start of the medical rehabilitation stay and at the start of the first intervention is uncorrelated with treatment. Moreover, body weight after 4 months does not significantly differ across experimental groups either (middle panel of the table). The average participant in the second intervention lost about 7.3 percent weight during the weight loss phase. About 59 percent of the participants achieved their individually assigned target weight after 4 months.

In the lower panel of Table V, we show the average attrition rates. Attrition among participants of the second intervention is lower than among participants of the first intervention (cf. Table II). Importantly, we do not observe any structural attrition pattern for the second intervention, i.e., the reward groups are not significantly more likely to comply than the control group. For this reason, we abstain from extensively discussing the sensitivity of the results for the second intervention with respect to sample attrition.[11]

Body Weight After 10 Months (Period of Intervention II)

While members of the control group had reduced their body weight by 7.7 percent during the weight loss phase (Table V), they significantly regained about 2.8 percent during the weight maintenance phase as displayed in Table VI (Column 1). The reward groups, which similarly reduced their body weight in the first four months, in contrast, did not significantly regain weight during the intervention period (Columns 2-3). Group 250 slightly lost further weight and Group 500 only regained roughly 0.8 percent. These weight changes translate into significant differences between both reward groups and the control group (Column 4-5). Weight change in group 250 and group 500 was 2.9 and 2.0 percentage points more favorable than in the control group. There is no significant difference between the two reward groups. As a matter of course, the difference in weight change between both treatment groups pooled together and the control group is significant. Figure IV displays the distribution of weight change during the intervention period by experimental group.

Concerning the secondary outcome, we find that group 250 and group 500 were about 16 and 18 percentage points more likely to confirm the weight loss target than the control group, respectively. Hence, here, group 500 was, on average, about 2 percentage points more likely to be successful than group 150. While the differences to the control

[11] All results are robust with respect to sample attrition and available upon request.

Table V: Descriptive Statistics by Weight-maintenance-premium Groups (Mean Values)

	All	Control	EUR 250	EUR 500
Pretreatment Values				
BMI Before Rehab	38.661	38.768	38.940	38.278
Baseline BMI	37.324	37.308	37.725	36.944
Target Weight Loss (Percent)	6.424	6.403	6.446	6.425
Bad Kissingen	0.305	0.308	0.282	0.327
Bad Mergentheim	0.389	0.404	0.388	0.375*
Isny	0.058	0.029	0.058	0.087
Glottertal	0.248	0.260	0.272	0.212
Female	0.305	0.337	0.291	0.288
Age (years)	47.981	47.010	48.806	48.135
Native	0.800	0.832	0.788	0.782
Married	0.667	0.700	0.670	0.630
Values After Intervention I				
BMI after 4 Months	34.585	34.454	34.968	34.338
Percentage Change Body Weight (Months 0-4)	-7.316	-7.709	-7.180	-7.059
Target Weight Realized after 4 Months	0.585	0.558	0.583	0.615
Posttreatment Values				
Total Dropout Rate after Month: 10	0.154	0.183	0.126	0.154
22	0.347	0.394	0.301	0.346
# of Observations	311	104	103	104

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

group are statistically significant, we do not find a significant difference among the two reward groups.

Body Weight After 22 Months (12 Months after Intervention II Ended)

At the end of the follow-up period, we observe pronounced weight regain in all three experimental groups (lower panel of Table VI). Within 18 months after successful weight loss, the average participant of the second intervention regained about 4 percent of previously lost body weight. Nevertheless, average weight loss throughout the entire experiment still amounts to more than 3 percent (comp. Table V with Table VI). After 22 months, we no longer observe statistically significant differences across experimental groups. Figure V displays the distribution of weight change at the end of the follow-up phase by experimental group.

Regarding the share of participants that confirmed their target weight after 22 months, we do not find any statistically significant differences between experimental groups either.

Table VI: Mean Comparison Across Weight-maintenance-premium Groups

	Experimental Groups			Δ to Control	
	Control	EUR 250	EUR 500	EUR 250	EUR 500
Months 5-10					
Percentage Change in Body Weight	2.800** (0.730)	-0.130 (0.571)	0.792 (0.946)	-2.930** (0.940)	-2.008** (0.946)
Target Weight Realized	0.353** (0.052)	0.511** (0.053)	0.534** (0.053)	0.158** (0.075)	0.181** (0.075)
# of Observations	85	90	88	-	-
Months 5-22					
Percentage Change in Body Weight	4.157** (0.935)	3.650** (0.958)	4.231** (1.037)	-0.507 (1.391)	-0.074 (1.410)
Target Weight Realized	0.333** (0.060)	0.333** (0.056)	0.324** (0.057)	-9*10-17 (0.082)	-0.010 (0.083)
# of Observations	63	72	68	-	-

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; a deviation from control group. All coefficients are obtained by OLS, regressing the respective outcome variable on the dummy variables indicating the two premium groups.

Sensitivity Analysis

Taking into account individual characteristics and variables related to the weigh-in, a multivariate regression confirms previous results (Table AII in the Appendix). A single exception is that the control group and the group 250 do no longer exhibit a significant difference in the share of participants who confirm their target weight at the end of the intervention period .This points to strategic behavior adopted by treated participants to achieve their targets.

Discussion

Our analysis shows that monetary incentives for confirming a previously achieved target weight have a temporary effect on the body weight of obese individuals. We observe a statistically significant effect of more than 2 percentage points on weight change for both rewards during the intervention period. Their effects on the likelihood to realize the target weight exceed 15 percentage points. This means that the rewards were successful in preventing weight regain. The higher reward did not proof to be more effective than the lower reward, i.e., any differences among the two reward groups may due to pure chance. Our evidence survives several sensitivity checks.

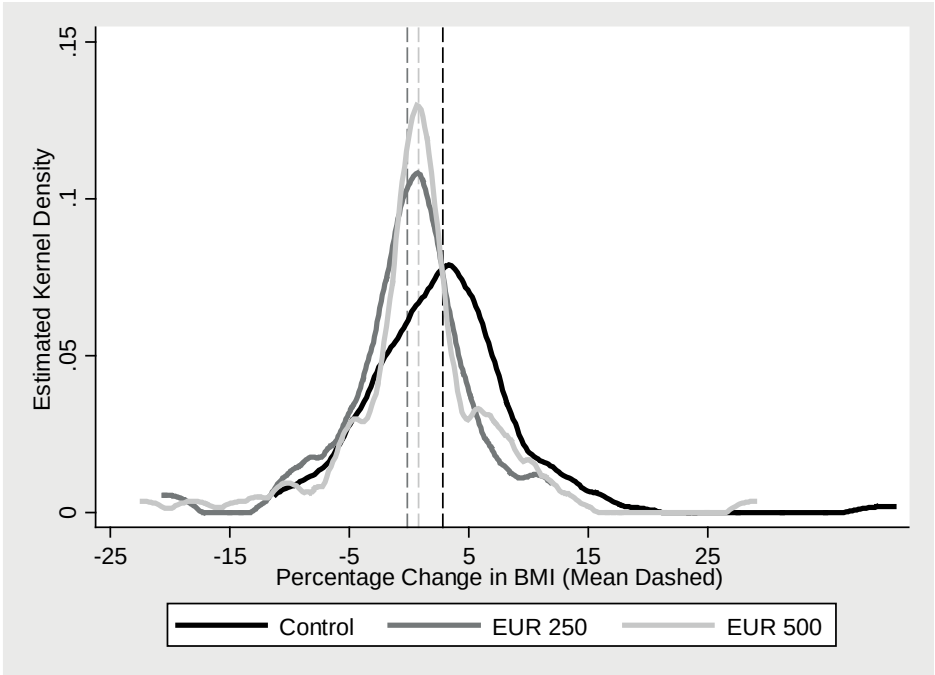

Figure IV: Distribution of Percentage Change in Body Weight by Experimental Groups of Intervention II *(Months 5-10).*

Due to its clearly temporary nature, we attribute the effectiveness of the rewards first and foremost to the standard price effect, which makes weight regain less attractive. We do not rule out that the second intervention causes participants to improve healthy behavior acquired during the preceding weight loss phase through, for instance, learning

effects. Improved behavioral automaticity arguably affects body weight in the same direction as the relative price effect. Evidently, the relative price effect, perhaps backed by the beneficial effects of continuous habit formation, dominates motivational effects. Results are also in line with the notion that there are no treatment effects on intrinsic motivation of the study population at all.

Given our previous result of monetary rewards for weight loss inducing lasting effects, the finding of significant effects of monetary rewards for confirming a previously achieved target might seem a puzzle. Yet, one has to keep in mind that, despite these lasting effects, even treated participants, on average, regained some body weight after the intervention period. Moreover, the study populations of the two interventions do not perfectly overlap. In fact, we do not condition on merely being exposed to financial incentives in the weight loss phase but on weight loss success.

Within 12 months after the intervention period, the effects of monetary incentives for confirming a previously achieved target weight vanish. Importantly, however, incentivized participants are not worse off at the end of the follow-up phase than

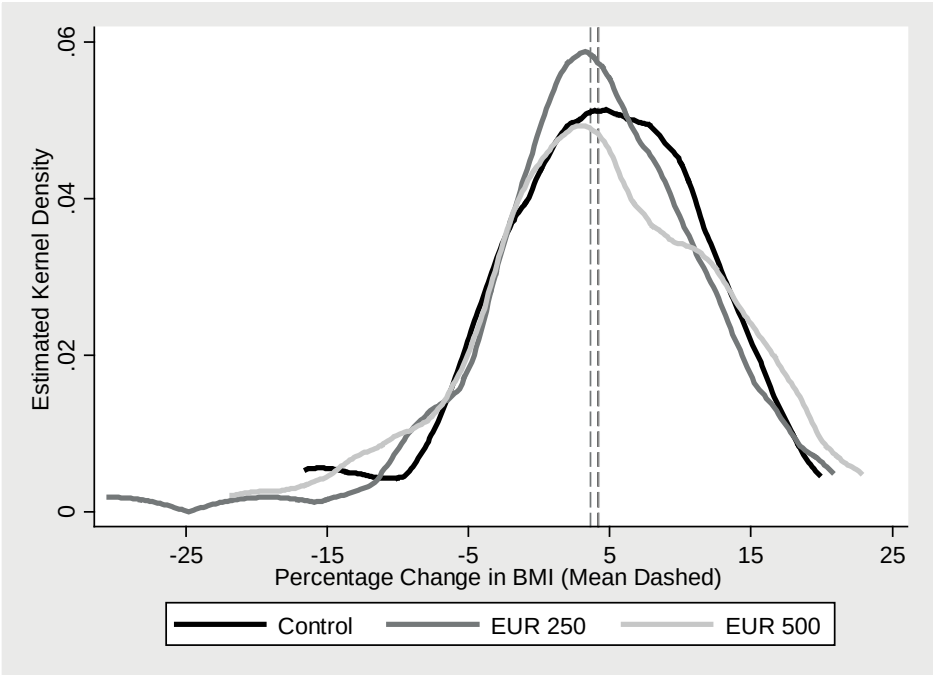

Figure V: Distribution of Percentage Change in Body Weight by Experimental Groups of Intervention II *(Months 5-22)*

participants of the control group. Hence, results confirm our findings from the first intervention that there is no complete motivation crowding out due to monetary incentives in the present study population, i.e., extrinsic rewards do not backfire.

The absence of lasting effects of rewards for confirming a previously achieved target weight seems to be at odds with the habit formation theory which was confirmed in the first analysis. According to this theory, the rewards should have contributed to the solidification of behaviors.

The finding that only the weight loss rewards formed healthy habits may be explained by the sequential nature of the two interventions. Successful participants make their experience with physical exercise and healthy diets during the weight loss phase. Some of them already have developed some behavioral automaticity at the time that the second intervention begins while others may have not. In the subsequent phase, there may well be a margin for change in behavior for both those who had previously adopted behavioral patterns and those who have not. Yet, in either case, such change is most likely induced by hook or by crook and, therefore, is not sustainable. In other words, after the weight reduction phase is completed, the chance for permanent behavioral change has either be sized or missed.

An alternative explanation is based on the interplay between the theories of habit formation and motivation crowding out. One may argue that a developed behavioral automaticity may countervail negative motivational effects of extrinsic rewards, which

only–or more strongly–arise in the second intervention. Monetary rewards for confirming a previously achieved target weight, as opposed to the weight loss rewards, were not announced in advance. Participants who were promised the rewards in the weight maintenance phase had all been successful in the previous four months and were most likely proud of their achievement, irrespective of prior group membership. Against this background, monetary rewards during the weight maintenance phase may signal that weight maintenance is even more difficult than weight loss and, therefore, reduce initial motivation to stay thinner.[12] While the relative price effect overcompensates the negative effect of the rewards on intrinsic motivation during the weight maintenance phase, weight regain occurs once the incentives are removed. The developed behavioral automaticity may prevent participants from being worse off due to the second intervention. The main argument for the deceptive contradiction that there are only negative motivational effects of the monetary rewards for confirming a previously achieved target weight is that weight loss rewards may be perceived as supportive while rewards in the weight maintenance phase are perceived as rather controlling, i.e., impairing self-determination and self-esteem. Explaining the results by the interplay between both theories reflects that our experiment neither provides ultimate evidence in favor of the one nor against the other.

Analysis of Effect Heterogeneity

The Role of Target Weight Achievement at the End of the Weight Loss Phase

The design of the experiment allows us to investigate whether monetary incentives for weight maintenance in previously fully successful participants is as effective as monetary incentives to fully achieve a target weight in previously partially successful participants. In Table VII, we present the effects of the second intervention estimated separately for participants who fully achieved their target weight (Columns 2-3) and participants who partially achieved their target weight at the end of the weight loss phase (Columns 4-5). The effects at the end of the weight loss phase and follow-up phase do neither significantly differ across the two subgroups for the primary nor for the secondary outcome variable. This indicates that there is no effect heterogeneity across the degree of target weight achievement. Interestingly, during the intervention period, the monetary rewards did not significantly increase the likelihood to achieve the target weight of participants who had only partially achieved it in the previous period.

The results discussed in this paragraph are not sensitive with respect to controlling for individual characteristics and variables that relate to the weigh-in (Table AII in the Appendix). Moreover, they are robust with respect to excluding observations with the largest and lowest 2.5 percent weight change in each subsample (results available upon request).[13]

[12] See Gneezy et al. (2011) for a similar argumentation.

[13] The positive point estimates for the effect of the two rewards on percentage change in body weight at the end of the follow-up period for participants who had only partially achieved the target weight in the previous period (lower panel of the table, Columns 3-4) are sensitive with respect to both covariate adjustment and trimming. Hence, we abstain from interpreting them.

Table VII: Effects of Weight-Maintenance Premiums by Success in Weight Loss Phase

	Target Weight Achieved in Weight-Loss Phase		Target Weight Not Achieved in Weight-Loss Phase	
	EUR 250	EUR 500	EUR 250	EUR 500
Months 5-10				
Percentage Change in Body Weight	-3.528** (1.425)	-1.687 (1.418)	-2.110** (1.059)	-2.441** (1.081)
Target Weight Realized	0.240** (0.093)	0.188** (0.093)	0.045 (0.094)	0.147 (0.096)
# of Obs. (Unweighted)	154		109	
Months 5-22				
Percentage Change in Body Weight	-1.874 (1.943)	-0.483 (1.992)	1.377 (1.894)	0.732 (1.894)
Target Weight Realized	0.040 (0.112)	0.011 (0.115)	-0.033 (0.101)	$2*10^{-16}$ (0.093)
# of Obs. (Unweighted)	119		84	

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.

The Role of Group Membership in the Weight Loss Phase

Estimated effects of the second intervention conditional on premium group (Columns 1-2) and control group membership in the weight loss phase (Column 3-4) are presented in Table VIII. The two subpopulations differ with respect to the degree of intrinsic motivation for weight loss that they had achieved in the first four months of the experiment. We observe that monetary incentives for confirming the target weight highly affect weight development of previously incentivized participants during the intervention period (upper panel of the table). Effects on both outcome variables are statistically significant. Participants who were not previously incentivized, in contrast, do not exhibit a significant and intended behavioral response to financial incentives for confirming a target weight. Group 500 even gains more weight than the control group during the intervention period, i.e., the EUR 500 reward worsens weight development in this particular subgroup. Even though this difference is insignificant, which in the face of the low number of observations may be due to insufficient statistical power, it points to motivation crowding out. Hence, promising weight maintenance rewards in the context of a program aimed to achieve weight loss which relies on measures other than monetary rewards may not be very effective.

While the finding of monetary incentives for confirming a target weight being effective contingent on group membership in the weight loss phase lacks statistical significance for the EUR 250 reward, it is statistically confirmed for the EUR 500. This is indirect evidence for the importance of intrinsic motivation for weight loss.

The results discussed in this subsection are robust with respect to controlling for individual characteristics and variables that relate to the weigh-in. Moreover, this evidence is

Table VIII: Effects of Weight-Maintenance Premiums by Weight-loss Experimental Group

	Premium Group in Weight-Loss Phase		Control Group in Weight-Loss Phase	
	EUR 250	EUR 500	EUR 250	EUR 500
Months 5-10				
Percentage Change in Body Weight	-3.435**	-2.990**	-0.987	1.855^{++}
	(1.065)	(1.076)	(1.954)	(1.923)
Target Weight Realized	0.221**	0.244**	-0.109	-0.068^{+}
	(0.083)	(0.084)	(0.168)	(0.165)
# of Obs. (Unweighted)	211		52	
Months 5-22				
Percentage Change in Body Weight	-0.443	-0.578	-0.756	2.820
	(1.534)	(3.304)	(3.304)	(3.425)
Target Weight Realized	0.007	0.019	-0.024	-0.126
	(0.093)	(0.094)	(0.177)	(0.184)
# of Obs. (Unweighted)	119		42	

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 weight loss experimental groups significant at 5%, + significant at 10%; S.E.s in parentheses.

not sensitive with respect to excluding observations with the largest and lowest 2.5 percent weight change in each subsample (results available upon request).

The effects for the post-intervention period are all statistically insignificant, confirming the absence of lasting effects of the second intervention. We find no longer relevant effect heterogeneity across weight loss experimental groups at the end of the follow-up-period (lower panel of the table). This evidence is sensitive to covariate adjustment, wherefore we abstain from considering the results for the post-intervention effects of monetary rewards for confirming a previously achieved target weight in the general conclusion.

V. Conclusion

In this paper, we present unique evidence from a large randomized experiment. Two main research questions are addressed: first, do effects of monetary rewards for weight loss persist after incentives are removed? Second, can financial incentives promised for the maintenance of a reduced body weight prevent weight cycling?

Our work adds to a growing literature on the longer run effects of monetary incentives to encourage health preventive behavior and generates novel knowledge about coupling financial rewards with sustained health-related behavioral change. The study is motivated by an increased popularity of using monetary incentives in the design of policy

interventions across a wide range of areas such education and health. While there is generally too little knowledge on effective interventions to improve public health, this is 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.

The experiment involved 700 obese participants of four medical rehabilitation clinics who were randomly assigned to three experimental groups. The control group and two treatment groups were offered EUR 0, EUR 150 or EUR 300 for achieving an individually assigned contractual target weight loss. Successful participants were eligible for a subsequent intervention that was not announced in advance. Here, participants were offered rewards amounting to EUR 0, EUR 250, and EUR 500 for confirming a body weight below the target weight. Body weight was measured 4, 10, and 22 months after baseline.

It is unclear whether monetary incentives backfire once they are removed or whether they induce persistent weight lost. The experiment allows us to analyze whether motivation crowding out or habit formation prevails. While the former phenomenon consists of negative effects on helpful other motives of behavior, the latter rests on the development of behavioral automaticity.

Our results show that monetary rewards for weight loss have a lasting effect on the body weight of obese individuals. The effect of the EUR 300 reward, for instance, amounts to roughly 2.8 and 3.3 percentage points about 6 and 18 months after the intervention period, respectively. There is no statistically significant difference between the effects of the higher and the lower reward.

Previous studies did not find significant post-intervention effects. One possible explanation for the success of our intervention is the absence of frequent feedback by the experimenter. In the experiments of Volpp et al. (2008) and John et al. (2010), for instance, members of the treatment groups had to weigh themselves each day and call in their weight to the project staff. Moreover, they received daily feedback about weight loss progress. By the end of the intervention period, this elevated attention stopped and treated participants may have felt alone with their weight problem.

Our analysis furthermore shows that monetary incentives for confirming a previously achieved target weight have a temporary effect of more than 2 percentage points on the body weight of obese individuals. Again, the higher reward was not more effective than the lower reward. The effects vanish within 12 months after incentives to prevent weight cycling are removed. This is the first study that presents results for weight maintenance rewards.

While our results are more in line with the habit formation theory, it may well be that habit formation just dominates motivational effects, i.e., effects on intrinsic motivation are existent but not visible. One explanation for the absence of lasting effects of the weight maintenance rewards is that the negative motivational effects, which are arguably stronger in the second intervention because rewards were not announced in advance and therefore perceived as more controlling, offset beneficial effects of a developed behavioral automaticity.

Another result of our experiment is that the effects of the monetary rewards are heterogeneous across initial motivation for weight loss. Separate estimations for weight loss

premium and control group members yield that, during the period of the second intervention, the reward effect is larger among previously incentivized participants. To the best of our knowledge, this is the first−though indirect− evidence for the importance of intrinsic motivation for weight loss. The mere observation of a backfire effect of extrinsic rewards may be attributable to other factors than intrinsic motivation such as the utilization of measures to lose weight that solely focus on short term success.

Our randomized trial shows that financial incentives can have positive lasting effects on weight reduction in the obese, arguing that there is scope for monetary interventions in health policy. In the face of our results, it may be important to announce monetary rewards well in advance. This requires further research. We cannot give a clear answer with regards to the meaningfulness of weight maintenance rewards. In fact, at the end of the experiment, individuals who received both rewards are not worse off than individuals who received only one weight loss reward. Policymaker need to evaluate whether temporary effects on body weight are worth the money.

A considerable limitation of our study−as in any experimental paper−is related to the external validity of the results. Our participants may not satisfy the requirements on representativeness because experiment participation was voluntary. Indeed, we do not know to what extent our results are transferable to the general obese population. The effects on intrinsic motivation may, in fact, be very different in the general population. However, we argue that most policy interventions consist of programs that involve voluntary participation, too. We are also concerned about general equilibrium effects, which threaten the success of monetary interventions. It seems reasonable that some people gain weight in order to qualify for payments.

Since we see, nonetheless, substantial scope for more monetary interventions in health policy and related areas, we advocate a larger scale implementation of a similar program in combination with a subsequent−or in time−scientific evaluation.

References

Acland, D. and Levy, M. (2013): Naiveté, Projection Bias, and Habit Formation in Gym Attendance. Working paper

Augurzky, B., Bauer T. K., Reichert, A. R., Schmidt, C. M. and Tauchmann, H. (2012): Does Money Burn Fat? - Evidence from a Randomized Experiment, Ruhr Economic Papers #368.

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*, 1 edn, University of Chicago Press, chapter 2, pp. 35–64.

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* 49(12): 1317–1324.

Charness, G. and Gneezy, U. (2009): Incentives to Exercise, *Econometrica* 77(3): 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* 24(9): 1107–1110.

De Geest, G., and Dari-Mattiacci, G. (2013): The Rise of Carrots and the Decline of Sticks, *University of Chicago Law Review* 80(1): 341−392.

Gneezy, U., Meier S., and Rey-Biel, P. (2011): When and Why Incentives (Don`t) Work to Modify Behavior, *Journal of Economic Perspectives* 25(4): 191-210.

Han, E., Norton, E. C. and Stearns, S. C. (2009): Weight and wages: fat versus lean paychecks, *Health Economics* 18(5): 535–548.

Heckman, J. J. (1976): The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of Economics and Social Measurement* 5(4): 475–492.

Heckman, J. J. (1979): Sample selection bias as a specification error, *Econometrica* 47(1): 153−161.

Hollis, S. and Campbell, F. (1999): What is meant by intention to treat analysis? Survey of published randomised controlled trials, *British Medical Journal* 319(7211): 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* 95: 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* 17(1): 143–149.

Kramer, F. M., Jeffery, R. W., Snell, M. K. and Forster, J. L. (1986): Maintenance of successful weight loss over 1 year: effects of financial contracts for weight maintenance or participation in skills training. *Behavior Therapy* 17(3): 295–301.

Lee, D. S. (2009): Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects, *Review of Economic Studies* 76(3): 1071–1102.

Morris, S. (2007): The impact of obesity on employment, *Labour Economics* 14(3): 413–433.

OECD (2012). Statistics*, Organisation for Economic Co-operation and Development*. Accessed on June 8, 2012. **URL***: http://stats.oecd.org*

Paloyo, A. R., Reichert, A. R., Reinermann, H. and Tauchmann, H. (2013): The Causal Link Between Financial Incentives and Weight Loss: An Evidence-Based Survey of the Literature, *Journal of Economic Surveys*. doi: 10.1111/joes.12010.

Quist-Paulsen, P. and Gallefoss, F. (2003): Randomised controlled trial of smoking cessation intervention after admission for coronary heart disease, *British Medical Journal* 327(7426): 1254.

Reichert, A. (2012): Obesity, Weight Loss, and Employment Prospects – Evidence from a Randomized Trial, Ruhr Economics Paper No. 381.

Royer, H., Stehr, M., and Syndor, J. (2012): Incentives, commitments and habit formation in exercise: evidence from a field experiment with workers at a Fortune-500 Company. NBER Working Paper No. 18580.

Sassi, F. (2010): *Obesity and the Economics of Prevention: Fit not Fat*, OECD Publishing, Paris.

Spenkuch, J. (2012). Moral hazard and selection among the poor: Evidence from a randomized experiment, *Journal of Health Economics* 31(1): 72–85.

Volpp K. G., Gurmankin, L. A., Asch, D.A, Berlin, J. A., Murphy, J. J., Gomez, A., Sox, H., and Zhu, J. (2006): A randomized controlled trial of financial incentives for smoking cessation. *Cancer Epidemiology, Biomarkers & Prevention* 15(1):12−18.

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* 300(22): 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* 360(7):699–709.

Wooldridge, J. M. (2002): Inverse probability weighted M-estimators for sample selection, attrition and stratification, *Portuguese Economic Journal* 1(2), 117–139.

Zhu, S.-H., Melcer, T., Sun, J., Rosbrook, B., and Pierce, J.P. (2000): Smoking cessation with and without assistance: A population-based analysis, *American Journal of Preventive Medicine* 18(4): 305–311.

Appendix

I. Technical Description of Selection Problem in Analysis of the First Hypothesis

Technically, the objective of the analysis is the estimation of $E(\Delta y \,|t_1^{reward}) - E(\Delta y|t_1^{control})$, where Δy denotes weight change over the entire considered period (10 and 22 month, respectively) and t_1^{reward} and $t_1^{control}$ denote group memberships in the weight loss phase. To simplify 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 weight maintenance phase, the estimator takes on the following form:

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

where s_i is an indicator for the absence of rewards in the weight maintenance phase. This estimator yields inconsistent estimates if s_i is correlated with the error term of the regression. The estimator, which–under certain conditions–proofs 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 $\Delta y_i$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 times.

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} into 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 weight maintenance rewards and successful participants are promised weight maintenance rewards with probability 2/3, irrespective of observable and non-observable factors. Hence, selection is exclusively based on observable factors.[14]

[14] 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. There, 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.

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

II. Tables and Figures

Table AI: Covariate Adjusted Effects of Weight-loss Premiums
(Inverse-probability Weighting)

	EUR 150	EUR 300
Start of Weight-loss Phase to End of Weight-maintenance Phase (10 Months)		
Percentage Change in Body Weight	-1.618 (1.143)	-2.984** (1.310)
Target Weight Realized	0.106 (0.079)	0.062 (0.078)
# of Observations (Unweighted)	234	
Start of Weight-loss Phase to End of Follow-up Phase (22 Months)		
Percentage Change in Body Weight	0.185 (1.636)	-3.171* (1.671)
Target Weight Realized	0.085 (0.082)	0.121 (0.093)
# of Observations (Unweighted)	176	

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.

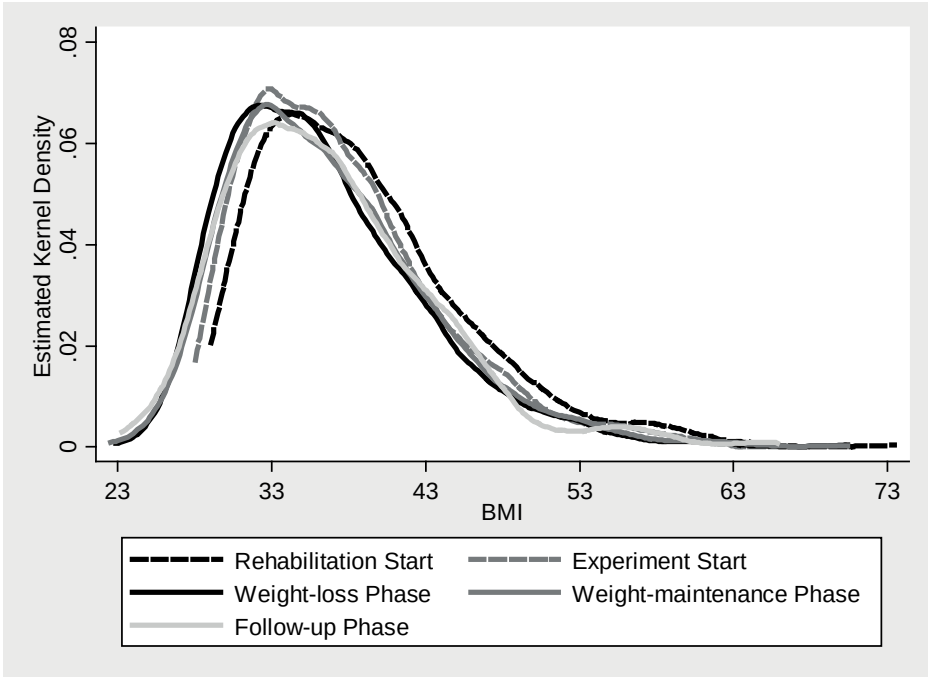

Figure AI: Distribution of BMI at the End of Each Phase

Notes: Missing values were imputed by BMI at start of 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 as "super-super obese" (e.g., Stephens et al., 2008).

Table AII: Covariate Adjusted Effects of Weight-Maintenance Premiums by Characteristics at End of Weight Loss Phase

End of Weight-Loss Phase	All		Target Weight Achieved		Target Weight Not Achieved		Premium Group		Control Group	
	EUR 250	EUR 500	EUR 250	EUR 500	EUR 250	EUR 500	EUR 250	EUR 500	EUR 250	EUR 500
Start of Weight-loss Phase to End of Weight-maintenance Phase (10 Months)										
Percentage Change in Body Weight	-2.681**	-2.107**	-3.673**	-2.863	-2.190*	-2.259*	-2.814*	-2.675**	-2.154	1.850^{++}
	(0.991)	(0.983)	(1.682)	(1.757)	(1.285)	(1.289)	(1.220)	(1.188)	(2.221)	(1.976)
Target Weight Realized	0.127	0.195**	0.207*	0.192*	-0.011^{+}	0.100	0.164*	0.238**	-0.102^{++}	0.085
	(0.080)	(0.080)	(0.109)	(0.114)	(0.114)	(0.115)	(0.097)	(0.094)	(0.224)	(0.199)
# of Observations (Unweighted)	263		154		109		211		52	
Start of Weight-loss Phase to End of Follow-up Phase (22 Months)										
Percentage Change in Body Weight	0.151	-0.070	-1.047	-0.801	-0.943	-1.206	0.434	-0.681	-2.706	-7.086^{+}
	(1.544)	(1.613)	(2.521)	(2.606)	(2.369)	(2.426)	(1.884)	(1.822)	(3.395)	(4.061)
Target Weight Realized	-0.007	0.002	0.019	0.007	-0.062	-0.019	-0.004	0.020	0.164	0.425^{++}
	(0.090)	(0.094)	(0.145)	(0.150)	(0.131)	(0.134)	(0.110)	(0.106)	(0.261)	(0.313)
# of Observations (Unweighted)	203		119		84		161		42	

Notes: # Lower Absolute Lee-Bound; ** significant at 5%, * significant at 10% ; °° difference between premium groups significant at 5%, ° significant at 10%; ++ difference across subsamples (full vs. partial target weight achievement in weight loss phase and weight loss premium vs. control group, respectively) significant at 5%, + significant at 10%; S.E.s for estimated means in parentheses.