RUHR
ECONOMIC PAPERS

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

Does Money Burn Fat?

Evidence from a Randomized Experiment

RUHR UNIVERSITÄT BOCHUM RUB RWI #368

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, email: 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. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Joachim Schmidt
RWI, Phone: +49 (0) 201/81 49-292, e-mail: joachim.schmidt@rwi-essen.de

Ruhr Economic Papers #368

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-423-5

Ruhr Economic Papers #368

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

Does Money Burn Fat?

Evidence from a Randomized Experiment

RUHR UNIVERSITÄT BOCHUM RUB

RWI

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/10.4419/86788423
ISSN 1864-4872 (online)
ISBN 978-3-86788-423-5

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

Does Money Burn Fat? – Evidence from a Randomized Experiment

Abstract

We test whether financial incentives have an effect on weight reduction in a randomized controlled trial involving 700 obese persons assigned to three experimental groups. While two treatment groups obtain €150 and €300, respectively, for achieving an individually assigned target weight within four months, a control group receives no such premium. The results indicate that the weight losses for the treatment groups are 2.6 and 2.9 percentage points higher than that achieved by the control group, raising the average total weight loss for the incentivized groups to 5 percent of the initial weight. This percentage is typically regarded as a threshold to improve the health status of the obese. Further evidence indeed indicates some health improvements. The higher reward causes only the group of obese women to lose more weight. Overall, the results suggest that financial incentives can motivate people to lose weight significantly.

JEL Classification: I10, I18, H23, C93

Keywords: Randomized experiment; financial incentives for weight loss; obesity; non-random sample attrition; effect heterogeneity

September 2012

---

1 *Boris Augurzky, RWI; Thomas K. Bauer, RWI, Ruhr-Universität Bochum, IZA Bonn; Arndt R. Reichert, RWI; Christoph M. Schmidt, RWI, Ruhr-Universität Bochum, IZA Bonn, CEPR London; Harald Tauchmann, RWI, CINCH.* *– 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, Alfredo R. Paloyo, Adam Pilny, Margarita Pivovarova, Gisela Schubert, and Marlies Tepaß for research assistance. We also gratefully acknowledge the comments and suggestions of participants of the following: German Health Economics Association meeting in 2012, European Conference on Health Economics 2012, and Annual Congress 2012 of the German Economic Association. 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: arndt.reichert@rwi-essen.*

It is well established that obesity increases morbidity and reduces life expectancy (for a comprehensive overview, see Sassi, 2010). Further negative consequences of obesity include unhappiness (Oswald and Powdthavee, 2007; Forste and Moore, 2012), a reduced probability of being employed (Morris, 2007), and lower wages (Baum and Ford, 2004; Han et al., 2009; Morris, 2006). It is therefore not surprising that about 60 percent of the overweight and 70 percent of the obese report that they try to lose weight (Bish et al., 2005; Oswald and Powdthavee, 2007). However, only few are actually able to reduce their body weight and many even gain further weight (Crawford et al., 2000; Williamson et al., 1992).

Economic theory attributes the weight-loss difficulties of the obese to time-inconsistent preferences (Ikeda et al., 2010; Ruhm, 2012), which make them fail to consistently balance the benefits and the costs of a healthy body weight, causing physical activity to be postponed and food to be overconsumed (O'Donoghue and Rabin, 1999). Even if the obese were able to control their body weight and acted perfectly rationally, another economic problem would remain in the sense that the obese do not have to (fully) bear the costs associated with their excessive body weight. They would hardly have a financial incentive to lose weight because the actual individual costs of being obese are too low and the (few) benefits of a healthy body weight are too distant to justify the effort of losing weight.

In fact, due to risk-independent health insurance premiums, the incremental health care costs of the obese are passed on to the society (Bhattacharya and Sood, 2007), totaling between 0.7 and 2.8 percent of a country's health care expenditures (Withrow and Alter, 2011). Since these estimates do not take into account that obese people are likely to die younger, they probably overestimate health care costs of obesity. However, even taking this argument into account, Bhattacharya and Sood (2011) still report higher lifetime health care expenditures for the obese. Moreover, the financial consequences of a reduced employment probability and an increased risk of early retirement are neither fully borne by the obese due to unemployment benefits and pension systems (Narbro et al., 1996; Cawley, 2000; Houston et al., 2008). Further examples are sickness benefits that prevent the obese from bearing the increased costs of absenteeism (Cawley et al., 2007). Others discuss a reduced working productivity due to on-the-job illness, which is entirely

borne by the employer under the assumption of wage rigidity (Tunceli et al., 2006; Gates et al., 2008).

A promising economic solution to both time-inconsistent preferences and cost externalization would be to provide financial incentives contingent on weight loss. On the one hand, such financial rewards would counterbalance the externalization of the costs associated with obesity, thereby working as a Pigouvian subsidy for a healthy body weight. On the other hand, by directly increasing the short-term benefits from healthy behavior, they would bridge the gap between short-run costs and long-run benefits.[1]

Since weight loss may help to reduce the costs associated with obesity, employers and health insurances have become increasingly interested in confronting their employees or their insured with monetary incentives to lose weight. In fact, some have already pilot-tested bonus programs for weight loss by assessing enrollment, program attrition, and weight loss effectiveness. Relton et al. (2011) show in a before-after analysis that bonuses foster weight loss among obese NHS insured in the UK, while Cawley and Price (2011) report less optimistic results in a US workplace environment. However, these studies do not establish that the observed effects on weight loss are causal. Several recent randomized controlled studies have found obese individuals being rewarded for modest weight loss to be more successful in losing weight than without the rewards (e.g., Volpp et al., 2008; John et al., 2011). Nevertheless, due to serious limitations such as lacking methodological rigor, attrition bias, and weak inference caused by a low number of observations, the evidence on the effectiveness of monetary rewards remains unsettled. For a detailed evidence-based survey of the literature, see Paloyo et al. (2011).

This paper reports the results of a randomized experiment aiming to test whether financial incentives actually cause obese people to reduce weight. We improve upon the existing literature by conducting a large experiment with 700 obese medical rehabilitation patients who were either paid €0, €150 or €300 for achieving an individually assigned contractual target weight within four months.[2] The experimental design enables us to also analyze whether the effect varies by

[1] DellaVigna and Malmendier (2006) argue similarly when explaining observed consumer behavior related to health club attendance.

[2] In terms of purchasing power parities (PPP), the rewards correspond to $188 and $376. We used the purchasing power parities exchange rate of 2011 provided by OECD (2011).

the premium level and by gender. We made every effort to minimize sample attrition, eventually achieving a lower attrition rate than most experiments on financial incentives for weight loss, and to probe the sensitivity of the experimental estimates with respect to attrition. Furthermore, we investigate whether certain health indicators, including the measured cholesterol level and self-reported health, have developed differently across the experimental groups.

Our empirical results suggest that the individuals in the two treatment groups lost twice as much weight as the control group. The significant weight loss differential amounts on average to about 2.6 and 2.9 percentage points of initial body weight, depending on the amount of the reward. For the treatment groups, the probability of reaching the target weight is significantly higher by 18 and 25 percentage points, respectively. Despite the substantial quantitative difference involved, members of the group with the higher premium do not lose significantly more weight than those with the lower premium. However, considering obese women alone, there is a significant difference between the two treatment groups. These results prove to be robust to various efforts of accounting for non-random sample attrition. We also provide some corroborative evidence which shows that members of the treatment groups display greater effort to achieve their target weight by, e.g., avoiding to snack between regular meals. Female members of the group with the higher reward were also more aware about the composition of their diet. Further effects are found on the self-assessed health status.

The remainder of this paper is organized as follows. The subsequent section describes the experimental design and introduces the data, Section II discusses the estimation strategy, while Section III presents the estimation results. Section IV discusses the means facilitating to achieve weight loss and probes into the connection between weight loss and health outcomes. Section V summarizes our main findings and concludes.

I Experiment and Data

A Design and Implementation

Our experiment has been funded by the *Pakt für Forschung und Innovation*, which is part of the excellence initiative of the German government, and conducted by the Rheinisch-Westfälische Institut für Wirtschaftsforschung (RWI) in Essen. 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 cooperated in the project. The clinics invited obese patients to participate in the experiment in the final week of their rehabilitation stay. During their medical rehabilitation, the patients had received a weight loss program that varied from clinic to clinic. The participants were comprehensively informed about the procedures of the experiment by handouts and personal instructions at experiment initiation. The eligibility criteria for participation in the experiment were a BMI above 30 at admission, an age between 18 and 75 years, and being a registered resident in the German federal state of Baden-Württemberg. 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. The study protocol was approved by the ethics commission of the Chamber of Medical Doctors of Baden-Württemberg.

The medical staff of the rehabilitation clinics carried out the baseline measurements of several medical target variables, such as the body-mass-index (BMI), blood glucose level, and cholesterol level. The physician in charge assigned to the participants an individual weight loss target between six and eight percent of the current body weight which they were supposed to realize within the next four months. The target was always chosen to lie above the critical threshold associated with beneficial health effects (Vidal, 2002). Participants were told to consult their general physician or the physician in charge at the rehabilitation clinic as soon as they would face any health complaints during the four months of weight reduction. In addition to the measurement of the medical target variables, participants answered a detailed questionnaire related to their socioeconomic background, further health outcomes and preventive behavior.

Random assignment to the treatment and control groups took place after the discharge from the clinic. Randomization was carried out without replacement within blocks of 51 participants, stratified by clinic. 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, members of the treatment groups were promised to receive up to €150 ('group 150') and €300 ('group 300'), respectively. The premium levels and the length of the treatment period are in the range of previous studies. Jeffery et al. (1983), for instance, have premia of $30, $150, and $300, which correspond in terms of PPP to €54, €272, and €544 in prices of 2011.[3]

More specifically, members of the treatment groups received no premium if they had reached less than fifty percent of their targeted weight loss. Once the achieved weight loss exceeded fifty percent, they were rewarded proportionally to the maximum reward. The full bonus was paid if they had reached or even exceeded their weight loss target. 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. As a member of the control group she receives no premium. As a member of the treatment group she obtains €107 ($134 in PPP) in group 150 and €214 ($268 in PPP) in group 300. If she loses only 4.1 kg (9 lbs.), she receives no reward regardless of her group assignment. For a weight loss of 8.4 kg or more she receives the whole group specific premium. The payment mechanism is illustrated in Figure 1.

Participants were informed by regular mail about their maximum possible premium and of the week they were supposed to attend the weigh-in at a pharmacy.[4] Since the participants spent the treatment period at home and not in the 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.[5]

Two weeks prior to the end of the weight loss period, a reminder for the control measurement

[3] Converted into present values of € based on the US consumer price index and the purchasing power parities exchange rate of 2011 provided by OECD (2011).

[4] Participants could postpone the date of measurement or move it forward by means of an early phone call.

[5] See Angrist and Lavy (2009) for a similar argument in the context of a within-school randomized trial.

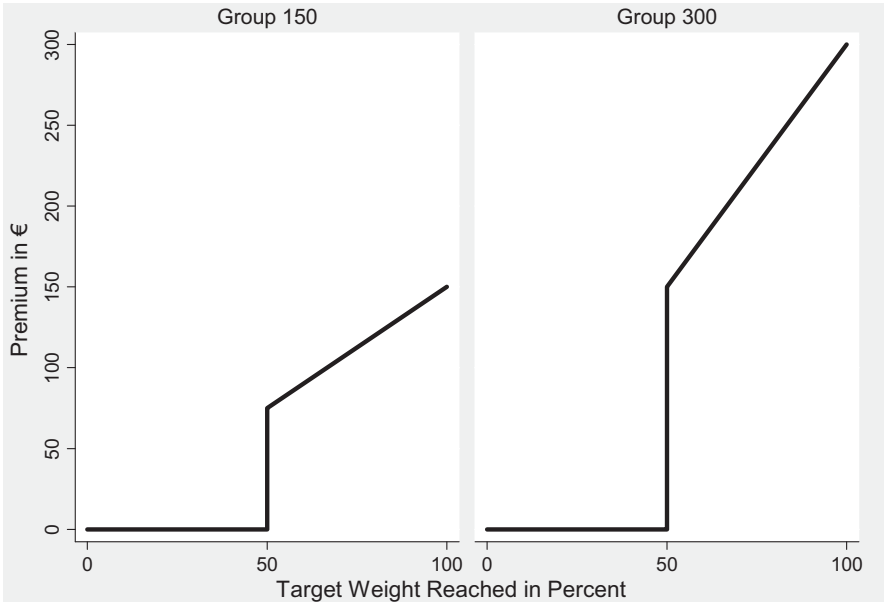

Figure 1: **Payment Mechanism**

of the body weight, blood glucose and cholesterol levels was sent to the participants. The letter contained a second 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, to figure out whether the necessary equipment for the measurements were available and to give brief instructions. Pharmacies without scales were not eligible. Pharmacies which had a scale but lacked devices for the measurement of the blood glucose and the cholesterol levels were eligible as second-best options in order to avoid excessive travel time for attending the weigh-in.

Some participants exited the study explicitly by means of a phone call. Yet, a much larger number dropped out by not showing up at the control measurement, without giving any notice. To reduce sample attrition, all participants whose documents were still pending three working days after the specified week of measurement were called by phone. Participants were questioned about the reason for not sending in the documents and were encouraged to make up for the

weigh-in. To make the effort worthwhile to everybody, all participants received €25 ($31 in PPP) if they attended the weigh-in and sent in the documents, regardless of the size of weight loss and group assignment. The premium was still paid if the date of measurement indicated by the pharmacist was within 14 days after the end of the supposed weigh-in week.

The advantage of assigning participants to specified pharmacies is twofold. First, treatment group participants cannot go from one pharmacy to the other in order to take advantage of probable measurement errors of the scales and, thereby, biasing the results. Second, this procedure generates exogenous variation in a determinant of sample attrition which is exploited to tackle selectivity bias (Section III).

B The Participants

The recruitment of a total number of 700 participants took place between March 2010 and August 2011.[6] Consequently, the last participant finished her individual four months of weight loss by the end of January 2012. Table 1 provides comprehensive descriptive statistics of the participants. 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, see Figure 2). The average weight loss target in percent of body weight at the time of experiment initiation amounts to 6.5 percent.

About 32 percent of the participants are women and 79 percent were born in Germany. These shares are substantially lower as the corresponding averages for the obese in Germany.[7] With a mean age of 48 years, the study population is also roughly ten years younger, on average. The share of employed individuals (82 percent) among participants is almost twice as large as in the obese German population, while the share of married participants does not deviate much from the corresponding share among the obese in Germany. The differences are due to the fact that for most patients of the co-operating clinics medical rehabilitation is paid by the German pension fund, which predominantly aims at avoiding unemployment or early retirement, hence, requiring that their ability to work is generally recoverable. Therefore, our study population

[6] Two individuals had to be excluded from the trial because of not further meeting the inclusion criteria by becoming pregnant and by developing cancer, respectively.

[7] Descriptive statistics for the average obese in Germany based on a representative German household panel (SOEP) are presented in column three of Table A1 in the Appendix. Column two of this table displays socioeconomic characteristics of the average patient of the rehabilitation clinics.

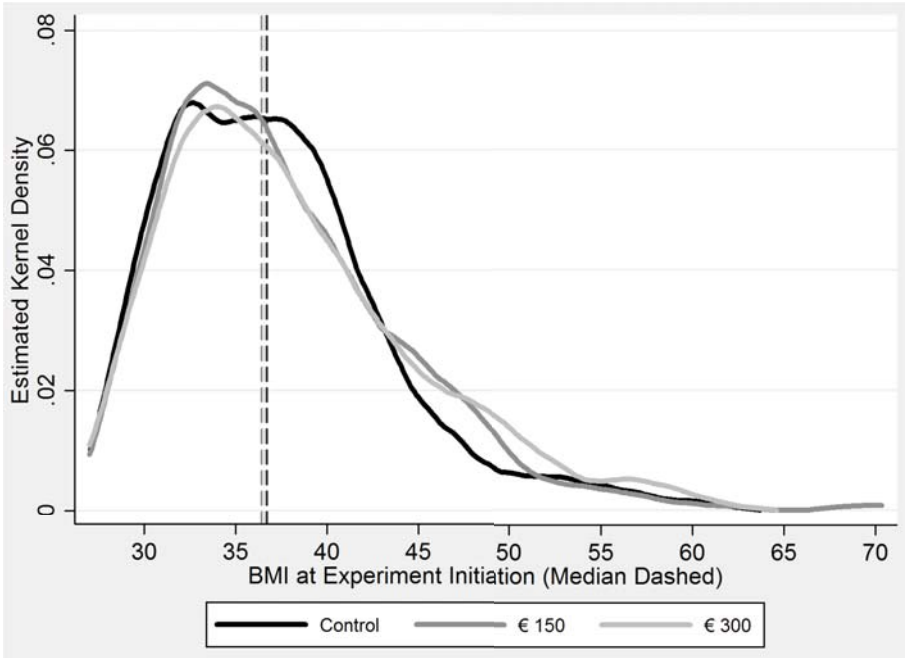

Figure 2: **Distribution of BMI at Experiment Initiation by Experimental Groups**

Notes: The inclusion criterion of a BMI $\geq$ 30 refers to the day of clinic admission. During the clinic stay (before random assignment), participants lost already 4 kg (8.8 lbs), eventually having a BMI lower than 30 at the start of the weight-loss phase. Persons with a BMI $\geq$ 60 are often considered as "super-super obese" (e.g., Stephens et al., 2008).

oversamples persons that are available for the labor market, while there are many obese in the overall population that are already retired.

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. The clinic in Bad Kissingen is specialized on gastroenterology and endocrinology, whereas the clinic in Glottertal treats patients with psychosomatic disorders. Most of the participants came to the clinics because of diagnoses other than adiposity. However, it often appears that their symptoms are related to their body weight and that a successful treatment implies weight reduction inter alia. For instance, many of the participants are diagnosed with some type of diabetes mellitus, which is often caused or reinforced by excessive weight.

Table 1 shows that the randomization algorithm yielded an even distribution of clinics across

Table 1: **Descriptive Statistics (Mean Values at Experiment Initiation)**

	All	Control	€150	€300	Compliers	Dropouts
Initial weight (before rehab, kg)	116.9	115.3	117.9	117.5	115.5	121.1**
Starting weight (after rehab, kg)	113.0	111.5	113.7	113.8	111.4	117.6**
Target weight loss (%)	6.49	6.44	6.53	6.52	6.44	6.65*
Clinic: Bad Mergentheim (%)	0.42	0.42	0.41	0.43	0.39	0.49**
Bad Kissingen (%)	0.33	0.32	0.36	0.32	0.34	0.33
Isny (%)	0.18	0.19	0.18	0.18	0.21	0.12**
Glottertal (%)	0.07	0.07	0.06	0.07	0.07	0.07
Female (%)	0.32	0.29	0.31	0.36	0.33	0.30
Age (years)	48.1	48.1	48.4	47.9	49.0	45.6**
Native (%)	0.79	0.79	0.75	0.82	0.79	0.77
Married (%)	0.61	0.63	0.61	0.59	0.64	0.53**
Population density ($km^2/10000$)	0.70	0.69	0.70	0.73	0.66	0.82**
City (%)	0.23	0.22	0.23	0.25	0.20	0.33**
Educ. attainment: very low (%)	0.15	0.18	0.13	0.14	0.14	0.17
low (%)	0.49	0.46	0.50	0.51	0.51	0.44
medium (%)	0.25	0.27	0.27	0.22	0.23	0.31**
high (%)	0.10	0.10	0.10	0.11	0.11	0.08
Employed (%)	0.82	0.86	0.81	0.79+	0.83	0.79
Self-assessed-health: good (%)	0.25	0.28	0.24	0.24	0.26	0.24
satisfactory (%)	0.43	0.41	0.43	0.45	0.43	0.41
poor (%)	0.24	0.23	0.25	0.25	0.25	0.24
bad (%)	0.05	0.05	0.06	0.04	0.04	0.07
Pharmacy in town (%)	0.64	0.65	0.62	0.65	0.66	0.56**
Blood sugar test (%)	0.83	0.86	0.83	0.79+	0.83	0.81
Cholesterol test (%)	0.57	0.61	0.56	0.55	0.57	0.59
# of observations	698	233	236	229	520	178

Notes: ++ deviation from control group significant at 5%, + significant at 10%; ** deviation from non-dropouts 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.

the three groups. Moreover, most covariates are balanced between the experimental groups. Even though a few individual covariates differ significantly between the control and the two treatment groups, the overall covariate balance is warranted by respective tests (Hansen and Bowers, 2008) indicating that the randomization procedure worked properly. In total, 178 participants did not attend the weigh-in at the end of the intervention period turning into dropouts. Among the main reason for not continuing the experiment (reported by dropouts who were reached by phone) was that they did not have time for it (27 percent), did not want to continue to participate in the study (16 percent), forgot about it (14 percent), were in holiday (11 percent), or were ill (9 percent). The resulting attrition rate amounts to 25.5 percent. This is considerably lower than in most field experiments on financial incentives for weight loss (Paloyo et al., 2011). Descriptive statistics (last two columns of Table 1) suggest that dropouts differ systematically from non-dropouts. For instance, they are significantly younger and have a significantly higher baseline weight (Figure A1 in the Appendix). Moreover, they are more likely to be members of the control group and of group 150 (see Table 2), suggesting that sample attrition is non-random.

Table 2: **Shares of Compliers (non-attriters) by Gender and Group**

	Experimental Groups						All Observations	
	Control		€150		€300			
Males	0.679	(165)	0.722	(162)	0.822	(146)	0.738	(473)
Females	0.632	(68)	0.743	(74)	0.880	(83)	0.760	(225)
All	0.665	(233)	0.729	(236)	0.843	(229)	0.745	(698)

Notes: Number of observations at the start of the experiment per cell in parentheses.

II Estimation Strategy

Since the present empirical analysis relies on data gathered from a randomized trial, one might infer the treatment effect by simply comparing the mean weight across the experimental groups after treatment. Since assignment to treatment is perfectly random, the treatment status is statistically independent from individual characteristics related to the body weight. However, we do not exclusively pursue this approach for three reasons.

First, comparing raw means does not account for random differentials in pre-treatment levels of body weight. While such differentials in levels will asymptotically vanish, they may still hamper the isolation of treatment effects in a finite sample. Indeed, though differences in mean pre-treatment weight between the control group and group 300 are statistically insignificant, amounting to 2.3 kg (5.1 lbs.), they are still relevant in medical terms, corresponding to roughly 31 percent of the mean weight loss target (see Table 1). In order to eliminate any level effects, we consider *changes* in body weight, using three measures for the dependent variable: (i) the percentage change in body weight, (ii) the absolute change in the BMI, and (iii) a binary variable indicating whether the target weight has been reached.

Second, we are able to reduce sampling error by basing the analysis on multivariate ordinary least squares (OLS) regressions (i.e., the linear probability model for the binary outcome variable). Besides the treatment indicators, we consider age, population density at the district level, dummy variables indicating the particular clinic, month of recruitment, gender, being a German native, whether participants live in a city[8], educational attainment[9], employment status, self-assessed

[8]The city dummy variable indicates individuals living in municipalities that cover more than one zip-code.

[9]We distinguish five educational levels. Individuals without any educational degree or who completed the 9th grade of secondary school but no vocational training (or missing information on vocational training) are defined to have a very low educational level. Those who finished the 9th grade of secondary school having a vocational training represent the low educated participants. Medium and high educated participants finished the 10th grade of secondary school and hold

health[10], and indicator variables that capture the condition at the control weigh-in. In particular, the pharmacists were asked to indicate whether the last food intake was more than half an hour or even more than two hours ago, whether the participants 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.[11] The latter allow us to capture possible ways of how participants could influence their measured body weight other than through weight loss.

Except for those related to the weigh-in condition, all variables enter the analysis as pre-treatment values. Following a standard approach (e.g., Morris, 2006; Spenkuch, 2012), we deal with missing values in covariates by replacing them 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 height, the set of educational indicators, and employment.[12] As a robustness check, we also estimated treatment effects excluding observations with missing information.[13]

Third and most important, non-random sample attrition may render the simple comparison of means to be a biased estimator of the treatment effect. Individuals who are members of one of the two treatment groups and have realized a body weight that entitles them to receive a reward, have a stronger incentive to stay in the program than the others.[14] Moreover, being successful in reducing weight – regardless of being rewarded or not – might positively affect further participation in the program because individuals may derive utility just from receiving credits for their success. Evidently, non-random sample attrition is a possible source of selection bias whose

a university-entrance diploma ('Abitur'), respectively.

[10] In the analysis, five dummy variables are used indicating a bad, poor, satisfactory, good, and very good (reference category) perceived health.

[11] 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.

[12] We prefer the 'missing indicator' method to simply excluding observations with missing information for two reasons: (i) item non-response occurs for several covariates. Hence, confining the analysis to complete cases reduces the sample size substantially, despite the fact that the share of missing values is rather low for most variables; (ii) just like the 'missing indicator' method, the 'complete case' approach is biased if the assumption that information is missing at random conditional on covariates does not hold (Jones, 1996), which is most likely the case in any survey.

[13] The results are available upon request. Although some specifications cannot be re-estimated because too many observations are excluded, the general pattern of results is robust with respect to the method of dealing with missing values.

[14] Opposed to the majority of evaluation studies (cf. Heckman et al., 1999), 'dropout' does not denote members of the treatment groups who endogenously do not receive treatment. In the present study, the treatment is *exposure* to a financial incentive. Because members of the treatment group cannot abstain from being exposed if they were informed about their treatment status, they are definitely treated. In this analysis, dropout denotes participants that do not provide full information at fourth months.

direction is not clear a priori (see Appendix A.1). In oder to tackle this potential bias, we pursue five estimation strategies.

Controlling for observed characteristics may already reduce the selectivity bias due to non-random sample attrition. Therefore, the above mentioned OLS regression is the first estimation strategy. The second approach is the classical Heckman (1976, 1979) selection correction (heckit) estimator. Due to greater numerical stability, we employ the classical two-step approach rather than joint maximum-likelihood estimation for the continuous dependent variables. For the lefthand-side variable 'target weight realized', we employ probit estimation with sample selection correction (Van de Ven and Van Pragg, 1981) instead, because the crucial assumption of joint normality in the regular heckit estimator is inconsistent with estimating the linear probability model. The design of the experiment renders several variables as valid exclusion restrictions: 'pharmacy in town', 'cholesterol test', and 'blood sugar test'. While 'pharmacy in town' indicates whether the assigned pharmacy lies within the same zip-code[15] as the participant's place of residence, the latter two indicate whether the respective blood tests are available at the pharmacy.[16]

The reasoning behind these exclusion restrictions is that participants cannot influence which pharmacy they are assigned to and that no obvious direct link to weight reduction can be assumed for the respective variables (during the weight loss period, participants do not know which particular pharmacy they will have to attend for the weigh-in). Yet, the exclusion restrictions may influence the decision to visit the pharmacy: The distance represents a rough measure of the cost the participants have to bear when attending the weigh-in. The effects of the blood tests on compliance may go in either direction. On the one hand, getting information on personal blood sugar and cholesterol levels can be regarded as an extra benefit of visiting the pharmacy. On the other hand, test persons may be frightened about getting punctured and, for this reason, may choose not to comply.

Our third approach is a generalization of the classical heckit model. This generalization is required if program continuation does not only depend on the treatment status and the realized

[15] Most (64 percent) participants were assigned to a pharmacy in the same ZIP code. A rough estimate of the average distance from the home address to the pharmacy is 2 kilometers with the maximum of about 30 kilometers and a minimum of a few meters. Around 500 pharmacies attended participants for control measurements.

[16] As displayed in Table 1, a large share (83 percent) of pharmacies had the measurement devices for the blood sugar test available. For the cholesterol test, the corresponding figure is lower (57 percent).

weight loss but also on the interaction of both. This is likely to hold for the present case as only both belonging to an treatment group and realizing a sufficient weight loss entitles for a reward, which sets strong incentives for attending the weigh-in. In econometric terms, this renders the first stage, i.e., the probit part of the heckit model, heterogeneous across treatment groups in terms of both coefficients and the error-term, where the latter becomes group-wise heteroscedastic (see Reichert and Tauchmann, 2012, for a detailed discussion).

In the present application, we use a simple approach for accommodating heterogeneity by estimating separate probit models for each experimental group. Although this approach does not exploit all available parameter restrictions and, hence, is not fully efficient, it proves to be more stable in computational terms than full information maximum likelihood. At the second stage of the estimation procedure, heterogeneity across experimental groups is accounted for by including group-specific inverse Mill's ratios. Here, the group-specific coefficients capture the group-specific heterogeneity in the error-correlation between the two parts of the model.

Though the availability of exclusion restrictions solves the identification problem in the heckit-type models, another problem remains. Selection-correction models are known to be vulnerable to specification error. In particular, a possible departure from normality (Yoo and Yang, 2000) and misspecification of the set of covariates that enter the selection equation may have a severe impact on the estimation result. For this reason, we apply two additional approaches that aim at estimating treatment effect bounds, rather than correcting for selection bias in a parametric framework.

The fourth approach is what in the medical literature is referred to as 'intention-to-treat' analysis. The intuition behind this concept is to regard all participants of a randomized trial as full-fledged members of that experimental group they have been randomly assigned to, irrespective of whether they drop out from the trial or deny required information.[17] No firm consensus on how to impute such missing data seems to exist in the medical literature (Hollis and Campbell, 1999). We follow the most common approach which assumes that the outcome variable returned to its

[17] Early applications of the intention-to-treat analysis aim to address selectivity bias that may arise from some participants that, e.g., for medical reasons, receive a treatment even though they were randomly assigned to the control group, and vice versa (Newell, 1992). For this particular problem, the understanding of the intention-to-treat approach largely coincides in the medical and economic literature (cf. Heckman et al., 1998). Yet, transferring it to the case of missing responses is a different problem because no information about the outcome variable is available for the relevant individuals.

baseline level, i.e., the outcome remained unaffected (John et al., 2011). In technical terms, this approach adjusts the OLS coefficient of the treatment effect towards zero. Thus, the estimates can be interpreted as being a worst case bound for the treatment effect, following a similar intuition as the approach of Horowitz and Manski (2000).[18]

The fifth approach follows Lee (2009) who suggests a method for estimating explicit lower and upper treatment effect bounds in the presence of non-random sample attrition. The bounds correspond to extreme assumptions about the missing information that are consistent with the observed data. A major advantage compared to parametric approaches is that this method rests on very few assumptions, i.e., random assignment of treatment and 'monotonicity', which means that the treatment status affects selection in just one direction.[19] Since Lee (2009) considers a single treatment group instead of two groups as relevant for the present experiment, besides considering both treatment groups separately, we also pool them when calculating treatment effect bounds.

In technical terms, the approach of Lee (2009) rests on a trimming procedure. Either from below or from above, the experimental group that suffers less from sample attrition is trimmed at the quantile of the outcome variable that corresponds to the share of 'excess observations' in this group. Calculating group differentials for the mean outcome yields the lower and the upper bound of the treatment effect, respectively, depending on whether trimming is from below or above.[20] Thus, the procedure yields bounds for the average treatment effect among always-compliers. The basic procedure does not condition on covariates. Yet, by conditioning on further variables that matter for selection, one may obtain tighter bounds. Calculating tightened Lee-bounds is based on splitting the sample into cells defined by a single or several covariates. Continuous covariates need to be converted into categorical variables. Then, trimmed means are

[18] In the present application, however, a different (yet unknown) baseline might be preferable as also the members of the control group – though not receiving a financial incentive – participate in a weight reduction program and, for this reason, may lose some weight. Nevertheless, we stick to what is standard in the medical science and report results using zero weight loss as baseline.

[19] This means receiving a treatment makes dropping out either more or less likely for *any* individual. For the present empirical application, this is a reasonable assumption because the exposition to monetary rewards gives incentives to continue experiment participation.

[20] To describe the calculation of the bounds in a more technical way, we denote the outcome variable by Y, the treatment indicator by T and the indicator for attending the weigh-in by W. Let q_T and q_C denote the shares of observations without missing information on Y for the treatment and the control group, respectively, define $q \equiv {}^{(q_T - q_C)}/_{q_T}$, and let y_q^T denote the qth quantile of the distribution of observed outcome in the treatment group. Then the lower bound and upper bound is estimated by $\frac{\sum_i 1\left(T_i=1, W_i=1, Y_i \leq y_{1-q}^T\right) Y_i}{\sum_i 1\left(T_i=1, W_i=1, Y_i \leq y_{1-q}^T\right)} - \frac{\sum_i 1(T_i=0, W_i=1) Y_i}{\sum_i 1(T_i=0, W_i=1)}$ and $\frac{\sum_i 1\left(T_i=1, W_i=1, Y_i \geq y_q^T\right) Y_i}{\sum_i 1\left(T_i=1, W_i=1, Y_i \geq y_q^T\right)} - \frac{\sum_i 1(T_i=0, W_i=1) Y_i}{\sum_i 1(T_i=0, W_i=1)}$, respectively, where i indexes participants and $1(\cdot)$ denotes the indicator function.

calculated separately for each cell where the trimming proportion is specific to each cell. Finally, a weighted average of trimmed means is calculated.

DiNardo et al. (2006) suggest that in the context of an experimental analysis such variables may be generated by randomly varying the effort made to prevent participants from dropping-out of the experiment. This suggestion nicely meets our earlier reasoning concerning the exclusion restrictions used in the heckit model. Hence, besides gender and rehabilitation clinic, we use these exclusion restrictions for tightening the estimated treatment effect bounds. Taking into account that the estimated bounds are subject to sampling error leads to calculating a confidence interval[21] for the treatment effect. This interval captures both uncertainty about the bias due to non-random sample attrition and uncertainty due to sampling error.

III Results

As a point of reference, we first compare the observed unconditional weight reduction across experimental groups. Figure 3 shows that the distribution of the percentage change in body weight is shifted towards lower values for both treatment groups if compared to the control group. According to Kolmogorov-Smirnov tests (Heathcote et al., 2010), the weight loss distributions of the treatment groups stochastically dominate the distribution of the control group. This is a very strong result as it means that receiving financial incentives increases the likelihood of realizing *any* level of weight reduction. However, when comparing the distributions of both treatment groups with each other, the Kolmogorov-Smirnov test is inconclusive.

These results are confirmed by Table 3, which indicates that also the other two measures of weight reduction considered, i.e., the mean change in the BMI and the share of participants being successful in realizing the target weight, are significantly larger in absolute terms for the treatment groups than for the control group. Even though the control group has been able to significantly reduce the body weight, the means of the three outcome variables are about twice as large for both treatment groups as for the control group. Comparing the two treatment groups yields no

[21] This interval is not simply the combination of confidence intervals for the bounds as it covers the treatment effect itself, rather than the bounds-defined interval, with probability $1 - \alpha$; see Lee (2009) and Imbens and Manski (2004).

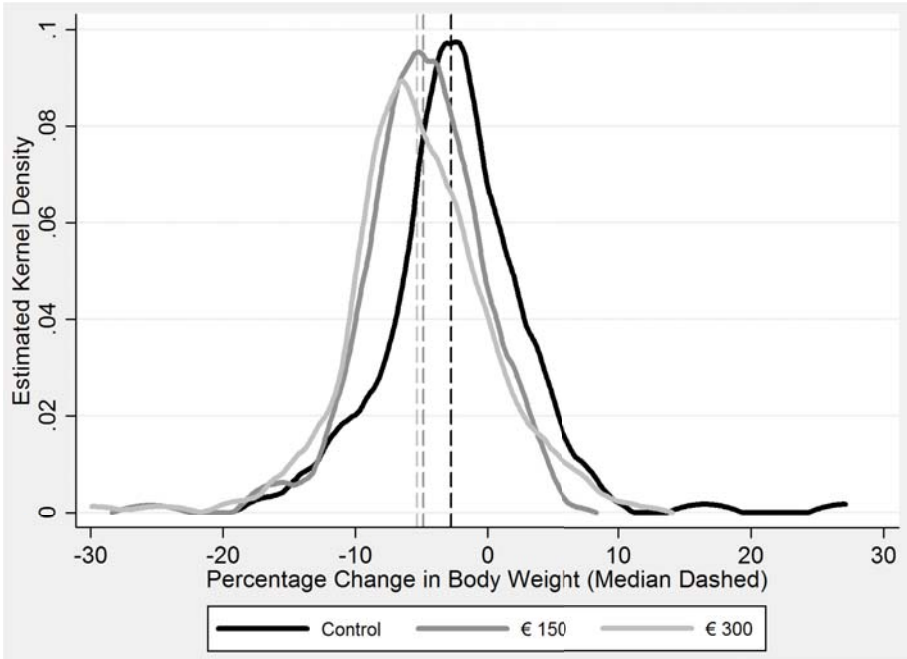

Figure 3: **Distribution of Percentage Change in Body Weight for Experimental Groups**

significant differential. Hence, the size of the reward seems to play a relatively minor role as compared to the pure existence of any reward.

The OLS results indicate that the estimated treatment effects are close to the simple comparison of mean weight loss (Columns 1 and 2 of Table 4). This has to be expected as the treatment status is not systematically related to any individual characteristic related to the participants' body weight, with the potential exception of those variables that control for the conditions at weigh-in (for comprehensive results, see Table A2 in the Appendix). Thus, the OLS estimates also indicate a pronounced effect of the monetary incentives on weight loss, while the differential effect of being rewarded by either €150 or €300 is only marginal. OLS results further reveal a much stronger treatment effect for women in the group 300 relative to men as well as to women in the group with the lower incentives. The difference between the two treatment groups is particularly large for the weight loss measure 'target weight realized'. The gender-specific heterogeneity in the effect of the higher reward may be explained by the share of employed being lower among women. Thus,

Table 3: **Unconditional Mean of Outcome Variables**

	Experimental Groups			Δ to Control[a]	
	Control	**€150**	**€300**	**€150**	**€300**
Percentage Change in Body Weight	-2.348^{**}	-4.861^{**}	-5.174^{**}	-2.513^{**}	-2.826^{**}
	(0.437)	(0.336)	(0.392)	(0.551)	(0.587)
Absolute Change in BMI	-0.879^{**}	-1.802^{**}	-1.957^{**}	-0.923^{**}	-1.079^{**}
	(0.155)	(0.131)	(0.154)	(0.204)	(0.219)
Target Weight Realized	0.194^{**}	0.378^{**}	0.456^{**}	0.184^{**}	0.262^{**}
	(0.032)	(0.037)	(0.036)	(0.049)	(0.048)
# of observations	155	172	193	-	-

Notes: $**$ significant at 5% (one-sided test); $*$ significant at 10% (one-sided test); S.E.s for estimated means in parentheses; [a] deviation from control group. All coefficients are obtained by OLS (and the linear probability model, respectively) regressing the respective outcome variable on the dummy variables indicating the two premium groups.

relatively more women than men may regard €300 as a relatively large amount of money.[22]

Concerning the regular heckit model (and the probit model with sample selection correction for the binary outcome 'target weight realized'), one may expect the estimated coefficients to be substantially smaller than those obtained from OLS, because a possible sample selection bias is now accounted for. Yet, as shown in Appendix A.1, a bias away from zero can likewise occur. In fact, heckit yields estimated treatment effects (Columns 3 to 4 of Table 4) which are close to those obtained from OLS. While for men the selection correction model yields somewhat larger estimated treatment effects relative to OLS, this does not hold for women. Compared to OLS, the effects for women are slightly smaller and the difference in weight loss between the two treatment groups is no longer significant. The small size and large standard errors of the coefficient of the inverse Mill's ratio (Table A2 in the Appendix) is in line with the impression that applying the heckit correction does not change the results substantially.[23] Just as the regular heckit estimator, its generalization yields effects that are similar in size to OLS (Table 4, Columns 5 and 6). Yet, the standard errors are rather large, in particular, if the coefficients are estimated separately by gender.

The general picture of the intention-to-treat analysis (Table 4, columns 7 and 8) is the same as of the OLS analysis. However, the estimated treatment effects are substantially smaller. This

[22]Since regression results may prove sensitive to the selection of covariates, we present results for two additional model specifications with a reduced number of covariates in the Appendix. The intermediate specification excludes the dummy variables capturing the participants' condition at the control weigh-in. A basic specification further excludes education, employment and self-assessed health. The results for the basic and extended specification are presented in Tables A3 and A4, which yield qualitatively the same and quantitatively very similar results to our preferred model specification.

[23]Table A5 in the Appendix displays comprehensive results for the selection equation. Tests on the joint significance of the explanatory variables that only enter the selection equations indicate that they are relevant for the decision to attend the weigh-in. In particular, participants who were assigned to a nearby pharmacy continued the experiment significantly more often.

Table 4: **Estimated Treatment Effects**

	OLS		Heckit[a]		General. Heckit[b]		Intention-to-Treat	
	€150	€300	€150	€300	€150	€300	€150	€300
Percentage Change in Body Weight								
All	-2.645^{**}	-2.898^{**}	-2.655^{**}	-2.929^{**}	-2.618^{**}	-2.294^{**}	-1.712^{**}	-2.096^{**}
	(0.549)	(0.611)	(0.566)	(0.743)	(0.972)	(0.923)	(0.385)	(0.459)
Males	-2.604^{**}	-2.325^{**}	-3.035^{**}	-3.318^{**}	-1.591	-1.508	-1.724^{**}	-1.670^{**}
	(0.643)	(0.733)	(0.776)	(1.005)	(1.335)	(1.197)	(0.445)	(0.547)
Females	-2.980^{**++}	-4.408^{**++}	-2.812^{**}	-3.570^{**}	-0.939^{++}	-3.275^{**++}	-1.744^{**++}	-3.078^{**++}
	(1.201)	(1.202)	(1.024)	(1.136)	(1.804)	(1.349)	(0.845)	(0.872)
Absolute Change in BMI								
All	-0.961^{**}	-1.081^{**}	-0.987^{**}	-1.159^{**}	-0.858^{**}	-0.697^{**}	-0.628^{**}	-0.788^{**}
	(0.203)	(0.227)	(0.215)	(0.282)	(0.370)	(0.355)	(0.142)	(0.170)
Males	-0.961^{**}	-0.912^{**}	-1.133^{**}	-1.309^{**}	-0.577	-0.443	-0.640^{**}	-0.655^{**}
	(0.245)	(0.283)	(0.301)	(0.390)	(0.515)	(0.467)	(0.171)	(0.210)
Females	-0.960^{**++}	-1.514^{**++}	-0.914^{**}	-1.288^{**}	-0.249^{++}	-1.133^{**++}	-0.560^{**++}	-1.078^{**++}
	(0.408)	(0.426)	(0.366)	(0.405)	(0.574)	(0.436)	(0.290)	(0.310)
Target Weight Realized								
All	0.181^{**+}	0.251^{**+}	0.175^{**++}	0.258^{**++}	-	-	0.117^{**++}	0.193^{**++}
	(0.051)	(0.050)	(0.047)	(0.043)	-	-	(0.034)	(0.038)
Males	0.224^{**}	0.238^{**}	-	-	-	-	0.151^{**}	0.186^{**}
	(0.065)	(0.064)	-	-	-	-	(0.044)	(0.048)
Females	0.070^{++}	0.312^{**++}	-	-	-	-	0.027^{++}	0.232^{**++}
	(0.086)	(0.085)	-	-	-	-	(0.061)	(0.064)

Notes: $**$ significant at 5% (one-sided test); $*$ significant at 10% (one-sided test); $++$ significant differential in treatment effects between treatment groups at 5% (one-sided test); $+$ significant differential in treatment effects between treatment groups at 10% (one-sided test); comprehensive set of covariates used. [a] two-step estimation for continuous outcome variables, marginal effects for Van de Ven and Van Pragg (1981) probit estimation with sample selection correction for the binary outcome 'target weight realized'; [b] not applied to the binary outcome 'target weight realized, reduced number of categories for self-assessed health (very good/good, satisfactory, poor/bad).

corresponds to the argument that this approach yields a worst-case effect bound. Although the intention-to-treat approach adjusts the estimated effects towards zero, the results do not put the existence of effects into question. With just a single exception (target weight realized, €150, female), all estimated treatment effects are statistically significant.

Table 5 finally reports the estimated treatment effect bounds (Lee, 2009). Column 1 and 2 display the lower and upper effect bounds of being member of an treatment group, irrespective of belonging to group 150 or group 300. Columns 3 to 6 present the results for both treatment groups separately. Overall, the treatment effect bounds also point at financial incentives effectively encouraging the obese to reduce body weight. None of the estimated effect intervals includes the value of zero. This does not fully apply to the corresponding confidence intervals. Here, for some cases (e.g., €300, men), we cannot firmly rule out a zero effect. Yet, even for these few cases concluding a zero treatment effect involves very strong assumptions. In particular, the bounds are based on the assumptions that dropping out results in an extreme upward bias of the group differential in the mean outcome, and, in addition, that sampling error results in overrating the

Table 5: **Estimated Treatment Effects Bounds**

	Treatment Groups					
	Joint		**€150**		**€300**	
	Lower	Upper	Lower	Upper	Lower	Upper
Percentage Change in Body Weight						
All	-3.833^{**}	-1.447^{**}	-2.991^{**}	-1.941^{**}	-4.623^{**}	-1.055^{**}
	[−4.900	−0.439]	[−4.232	−0.593]	[−5.770	−0.013]
Males	-3.245^{**}	-1.067^{*}	-2.864^{**}	-1.537^{**}	-3.817^{**}	−0.358
	[−4.468	0.059]	[−4.084	−0.478]	[−5.200	0.900]
Females	-5.501^{**}	-2.754^{**}	-3.837^{**}	-2.356^{**}	-6.294^{**}	-2.428^{**}
	[−7.431	−0.709]	[−5.936	−0.386]	[−8.369	−0.341]
Absolute Change in BMI						
All	-1.448^{**}	-0.499^{**}	-1.189^{**}	-0.558^{**}	-1.827^{**}	-0.334^{**}
	[−1.838	−0.118]	[−1.516	−0.239]	[−2.156	−0.042]
Males	-1.249^{**}	-0.391^{*}	-1.078^{**}	-0.563^{**}	-1.527^{**}	−0.194
	[−1.717	0.064]	[−1.621	−0.069]	[−2.064	0.277]
Females	-1.914^{**}	-0.803^{**}	-1.352^{**}	-0.775^{**}	-2.238^{**}	-0.870^{**}
	[−2.583	−0.125]	[−2.077	−0.088]	[−3.037	−0.068]
Target Weight Realized						
All	0.127^{**}	0.304^{**}	0.125^{**}	0.220^{**}	0.121^{**}	0.382^{**}
	[0.033	0.389]	[0.015	0.315]	[0.010	0.488]
Males	0.150^{**}	0.284^{**}	0.177^{**}	0.241^{**}	0.121^{*}	0.332^{**}
	[0.041	0.386]	[0.047	0.362]	[−0.011	0.458]
Females	0.091	0.366^{**}	0.008	0.183^{**}	0.139^{*}	0.451^{**}
	[−0.070	0.513]	[−0.211	0.335]	[−0.038	0.619]

Notes: 95% confidence interval for average treatment effect (*ATE*) in parentheses; $**$ significant at 5% (one-sided test); $*$ significant at 10% (one-sided test); for tightening bounds that set of covariates is selected from the variables 'gender', 'clinic', 'pharmacy in town', 'cholesterol test available', and 'blood glucose test available', which yields the narrowest confidence interval for *ATE*.

true treatment effect to an extent which occurs in just five out of hundred draws.

Interestingly, if only the smaller reward of €150 is considered, the estimated confidence intervals do not question the existence of an effect (except for the single case of females and the binary outcome variable). This can be explained by the relatively small attrition rate in this group. Relative to the group 300, the share of dropouts is close to the corresponding share of the control group (Table 2). This implies less extreme assumptions about the possible selection bias and, thus, results in tighter bounds. Since estimated treatment effect bounds clearly argue in favor of the €150 reward exerting an effect on weigh loss, it seems unjustified to cast doubts on the reward of €300 exerting any effect. Comparing the results for men and women, the gender differential for the effect of the €300 reward is even more pronounced than for the models discussed above.

Having established a statistically significant effect of financial incentives on weight loss, the question remains whether this effect is of medical relevance. Economic theory suggests an effect in the observed direction. However, only an impact of sufficient magnitude renders financial incentives a valuable tool in assisting the obese in reducing body weight.

While several statistical approaches yield similar point estimates for the effect under scrutiny, this does not apply to the intention-to-treat results and the estimated Lee (2009) bounds. Yet, the former is a priori known to adjust the estimate toward zero under assumptions with questionable validity for the present case and the latter yields lower bounds that in absolute terms exceed any point estimate for the average treatment effect. The upper bounds, in contrast, point towards smaller absolute effects, which, just like the intention-to-treat estimates, have to be considered as worst-case bounds. Since, the regular heckit model suggests that selection bias does not result in exaggerating the estimated effects, focusing on worst-case bounds in discussing the effect size does not seem to be appropriate. A rather conservative point estimate for the average effect is obtained by OLS. We therefore focus on the OLS results in the subsequent discussion.

OLS estimation yields an average treatment effect of 2.6 and 2.9 percentage points reduction in body weight. Compared to the control group, which already managed to lose 2.3 percent without being exposed to any financial incentive (Table 3), these figures indicate that the monetary rewards cause the obese to be more than twice as successful. This has to be regarded as a substantial impact. Medical scientists often consider five-percent weight loss as the relevant threshold for improving the health status of an obese person (Vidal, 2002). In the light of this, at least for the present weight loss program, financial incentives seem to make the difference for whether mean weight reduction is sufficient for translating into better health.

Turning to the absolute change in BMI, an average treatment effect of roughly one BMI unit also has to be regarded as significant in medical terms, though it may be insufficient for curing obesity for most participants. Here, the general picture of members of the treatment groups being roughly twice as successful as members of the control group applies, too. This also holds for the third measure of weight reduction, i.e., the probability to reach the target weight. The exposition to financial incentives increases the probability of realizing the contractual target weight by 18 and 25 percentage points, respectively. Taking into account that only one in five members of the control group is successful in his or her attempt, financial incentives act really strongly.

Splitting the sample by gender reveals that for men, the estimated effect of the €300 reward is somewhat – though insignificantly – smaller than the effect of the €150 reward. This pattern does

not apply to the regression for the probability to reach the contractual target weight, which yields an insignificantly larger effect of the €300 reward. In contrast, for women, the analysis shows a substantial and statistically significant differential effect for the two treatment groups. With respect to percentage weight loss, the effects amount to 4.4 points for the higher and 3 points for the lower reward. This provides evidence that an even stronger reduction in body weight can be achieved for obese women if the offered amount of money is increased.

IV Mediators and Health Outcomes

Financial incentives neither burn fat nor improve individual health directly, but they seem to be able to change the mind. They urge the obese to take the necessary means to lose weight. Weight loss is eventually achieved by the means of mediators such as the frequency of exercise or going on diet. Although at first sight, weight loss may represent the ultimate goal of the financial incentives under study, weight loss itself may turn into a mediator affecting other outcomes such as individual health.

In this section, we first show that the financial incentives have induced some behavioral changes and then investigate whether these behavioral changes have exerted any effect on health outcomes. While the former is important to explain the observed effects on weight loss, positive health effects would be of great interest to health policy makers. In the following analysis, we use the awareness of the composition of the consumed food, often snacking between meals, taking stairs instead of the elevator, frequency of exercise, self-assessed health, the cholesterol level, and the blood glucose level as dependent variables.

The blood test results were reported by the medical staff of the clinic and the pharmacist, respectively. The other six variables were retrieved from the questionnaires at the start and the end of the treatment period. The awareness of the diet composition, climbing stairs, frequency of exercise, and self-assessed health are measured on an ordinal scale. The variable snacking indicates whether participants report that they often snack while doing other tasks. Descriptive statistics of the dependent variables are displayed in Table A6 in the Appendix.[24]

[24] Descriptive statistics for self-assessed health are presented in Table 1.

In line with Section II, we compare changes rather than post-treatment levels across the experimental groups. In particular, we use the percentage change in the cholesterol and blood sugar level. The ordinal variables and the snacking variable are transformed into a variable with only three scales: increase (+1), unchanged (0), and decline (-1). For instance, individuals that reported a poor health status at the start of the treatment period but reported a good health at its end, are assigned the value +1. While the results will be useful to indicate the significance and direction of the effects, we abstain from discussing the size of the estimated effects, because the coefficients will be difficult to interpret due to the transformation of the dependent variables.

In contrast to the information on body weight, all these dependent variables have missing values among compliers due to either item non-response or some pharmacies lacking the necessary measurement devices (in case of the blood test information). While for the latter missing information is exogenously determined, for the purpose of interpretation we assume random non-response for the other outcome variables. Except of the generalized heckit estimation, the same statistical procedures as in the previous analysis are employed.

The results are presented in Table 6. At first sight, the treatment effects on the mediators and health outcomes appear to be less consistent across the different estimation approaches than the respective effects on the three measures of weight loss, i.e., non-random selection seems to play a bigger role. Hence, it can be misleading to focus only on the results of one specific estimation strategy. For this reason, we take into account the whole set of estimates in order to elaborate a conclusive general pattern for the treatment effects on each dependent variable.

Concerning the change in the frequency of climbing stairs, all four estimation approaches yield statistically significant treatment effects. Even though the underlying assumptions about the bias of the treatment effect and the sampling error are extreme, the lower treatment effect bounds (Lee, 2009) are positive and significant for both groups. This suggests that the obese take significantly more often the stairs instead of the elevator if they are financially rewarded for weight loss. As Table 6 shows, the four estimation strategies yield a quite similar pattern for the effect of the €150 incentive on snacking behavior, indicating that significantly more members of the respective group than of the control group stopped to snack frequently. The results for the €300 incentive

Table 6: **Treatment Effect on Mediators and Health Outcomes**

	OLS		Heckit		Intention-to-Treat		Lee-Bounds	
	€150	€300	€150	€300	€150	€300	€150	€300
Diet Composition	0.074	0.044	0.086	0.079	0.087	0.106	0.065	−0.165*
(aware)	(0.092)	(0.091)	(0.093)	(0.124)	(0.067)	(0.068)	[−0.113	[−0.341
Snacking	0.101*	0.085	0.113*	0.124*	0.089**	0.096**	0.044	−0.096*
(less)	(0.058)	(0.057)	(0.058)	(0.075)	(0.042)	(0.043)	[−0.067	[−0.211
Climbing Stairs	0.184*	0.179*	0.290**	0.415**	0.168**	0.205**	0.159**	0.154**
(more)	(0.096)	(0.097)	(0.138)	(0.168)	(0.070)	(0.072)	[−0.009	[0.006
Exercise	0.054	0.074	0.071	0.128	0.051	0.108*	−0.063	−0.142*
(more)	(0.080)	(0.078)	(0.081)	(0.106)	(0.057)	(0.058)	[−0.208	[−0.293
Self-assessed Health	0.117	0.106	0.121*	0.131	0.078	0.059	0.112*	−0.074
(better)	(0.075)	(0.073)	(0.073)	(0.095)	(0.056)	(0.056)	[−0.025	[−0.170
Cholesterol Level	0.008	−0.031	0.038	0.026	0.011	−0.013	0.048	0.036
(change,%)	(0.038)	(0.037)	(0.042)	(0.053)	(0.026)	(0.026)	0.132]	0.104]
Blood Sugar Level	−0.010	0.045	−0.035	0.032	−0.012	0.039	0.025	0.123**
(change,%)	(0.043)	(0.043)	(0.046)	(0.052)	(0.029)	(0.030)	0.104]	0.209]

Notes: ** significant at 5%; * significant at 10%; the extended set of covariates is used. [and] indicate the lower and upper limits of the confidence interval. The respective bounds are lower [and upper] bounds. For brevity, only the conservative bound for the test of the hypothesis of no improvement in the respective outcome variable is displayed, i.e., if a positive (negative) effect is expected, the lower (upper) bounds are reported.

are inconclusive.

Regarding the awareness of the diet composition, all four estimation strategies indicate an effect of the reward of €300 for the group of women only. Women of group 150 also increased their dietary awareness. This effect is, however, statistically insignificant. No effects are found for men. There is no evidence of an overall increase in the frequency of physical exercise either. Yet, all four estimation approaches suggest that women of group 300 have exercised significantly more often compared to the control group. Separate estimation results by gender are displayed in Table A7 in the Appendix.

Regarding the change in the self-assessed health status, the heckit approach yields a statistically significant treatment effect for group 150 at the 10 percent level. In contrast, the effects estimated by OLS and intention-to-treat are insignificant. However, the respective lower treatment effect bound is significantly positive, which strongly points at improvements in the self-assessed health status. For the higher reward, we cannot rule out a zero effect although we explain this lack of evidence by the significantly lower share of dropouts which increases the variance in the models accounting for sample selection. With respect to the cholesterol and the blood sugar levels, we find no coherent picture that would indicate any particular effect.

V Conclusion

This paper provides comprehensive evidence for the usefulness of financial incentives to encourage obese people to reduce body weight. The analysis has improved upon the existing literature by conducting a large randomized experiment, distinguishing between two different premium levels as well as gender, and accounting for non-random sample attrition. The results are highly relevant for both employers and health insurance companies that search for effective measures to reduce obesity among their employees or their insured. Financial incentives for weight loss could be easily integrated in existing bonus programs for healthy behavior, disease management programs, and other health promotion programs.

We find that a reward of € 150 doubles the success in reducing body weight. The control group who did not receive a reward for achieving an individual-specific weight-loss target of between 6 and 8 percent of initial body weight within four months has lost, on average, 2.4 percent. Patients who were rewarded by up to € 150 lost about 5 percent. The causal effect of the reward amounts to 2.6 percentage points, with the estimate being insensitive with respect to attrition. The size of the effect is in line with most recent medical trials on this topic, even though these are not able to rule out the possibility of biased estimates of the treatment effect due to non-random sample attrition.

For instance, Volpp et al. (2008) estimate the effect of a lottery contract with a total expected value of $365 (€ 306 in prices of 2011). In total, the lottery group lost 5.4 percent, significantly more (3.7 percentage points) than the control group. Participants had to achieve a weekly weight loss of 0.45 kg (1 lb.) during four months corresponding to an average target weight loss of 6.7 percent. In line with our results, Volpp et al. (2008) find that the monetary rewards cause the obese to take the five-percent weight-loss hurdle, which is often regarded as a threshold to be met in order to improve health status (Vidal, 2002). In fact, being the first who estimate the effect on health outcomes of financial incentives aimed at weight loss, our results indicate some health improvements.

We further tested whether the size of the monetary reward is relevant for weight loss. Looking at men and women jointly, the effect of the higher reward (amounting to 2.9 percentage points)

is not statistically different from the effect of the lower reward. This is in line with the existing evidence which suggests that higher financial incentives do not typically translate into a higher weight loss (see e.g., Jeffery et al., 1983) and, therefore, tend to be less cost-effective. In our study, the mean reward for members of group 150 amounted to €62 ($78 in PPP) and €150 for group 300. Thus, a one-percent-reduction of weight 'costs' €26 ($33 in PPP) in group 150, while in group 300 it costs €60 ($75 in PPP).

In the light of this observation, we would recommend the smaller bonus. However, splitting the sample by sex shows that the larger reward of €300 causes obese women to lose about 1.4 percentage points more weight than a reward of €150 – but not men. This heterogeneity in the effects by gender might be explained by women considering €300 as a relatively larger amount of money due to their lower employment share and, thus, lower average income. Although the lower reward is also more cost-effective when focusing on women alone, since there may be increasing returns to weight loss, we hesitate to recommend the smaller bonus in the case of women.

Despite the ample set of results derived in out study, some open questions remain. First, will the estimated effects be persistent over time? Paloyo et al. (2011) discuss that all surveyed studies on financial incentives for weight loss find large weight regains, i.e., the effects tend to vanish in the middle and long term. Since these studies face serious limitations such as, e.g., lack of methodological rigor, further research is needed. If large regains are confirmed, continuous financial incentives for weight maintenance might help the obese to keep the reduced body weight in the long term. Second, it is unclear whether a monetary penalty would yield similar results as a reward. If so, fat taxes might be a way to achieve similar effects on body weight.

Third, is there any interaction effect of the stay in the clinic and the financial incentive? The received medical advice and weight loss treatment during the clinic stay may help the financial incentive to be more effective. Irrespective of this possibility, we recommend to combine financial incentive schemes with courses of medical instructions to prevent participants from following unhealthy diet regimes or exaggerating weight reduction. Fourth, as in any experimental study, there is no perfect overlap between the study population and the target population. Further

research should address whether and what settings hinder or boost the effectiveness of monetary rewards for weight loss.

References

Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial, *American Economic Review* **99**(4): 1384–1414.

Baum, C. L. and Ford, W. F. (2004). The wage effects of obesity: a longitudinal study, *Health Economics* **13**(9): 885–899.

Bhattacharya, J. and Sood, N. (2007). Health insurance and the obesity externality, *Advances in Health Economics and Health Services Research* **17**(6): 279–318.

Bhattacharya, J. and Sood, N. (2011). Who pays for obesity?, *Journal of Economic Perspectives* **25**(1): 139–58.

Bish, C. L., Blanck, H. M., Serdula, M. K., Marcus, M., Kohl, H. W. and Khan, L. K. (2005). Diet and physical activity behaviors among americans trying to lose weight: 2000 behavioral risk factor surveillance system, *Obesity Research* **13**(3): 596–607.

Cawley, J. (2000). An instrumental variables approach to measuring the effect of body weight on employment disability, *Health Service Research* **35**(5): 1159–1179.

Cawley, J. and Price, J. A. (2011). Outcomes in a program that offers financial rewards for weight loss, *in* M. Grossmann and N. H. Mocan (eds), *Economic Aspects of Obesity*, 1 edn, University of Chicago Press, chapter 4, pp. 91–126.

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.

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.

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

DiNardo, J., McCrary, J. and Sanbonmatsu, L. (2006). Constructive Proposals for Dealing with Attrition: An Empirical Example, *University of Michigan Working Paper* .

Forste, R. and Moore, E. (2012). Adolescent obesity and life satisfaction: Perceptions of self, peers, family, and school, *Economics & Human Biology* **10**(4): 385 – 394.

Gates, D. M., Succop, P., Brehm, B. J., Gillespie, G. L. and Sommers, B. D. (2008). Obesity and presenteeism: the impact of body mass index on workplace productivity., *Journal of Occupational and Environmental Medicine* **50**(1): 39–45.

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

Hansen, B. B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies, *Statistical Science* **23**(2): 219–236.

Heathcote, A., Brown, S., Wagenmakers, E.-J. and Eidels, A. (2010). Distribution-free tests of stochastic dominance for small samples, *Journal of Mathematical Psychology* **54**(5): 454–463.

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.

Heckman, J. J., Lalonde, R. J. and Smith, J. A. (1999). The economics and econometrics of active labor market programs, Vol. 3, Part A of *Handbook of Labor Economics*, Elsevier, pp. 1865–2097.

Heckman, J. J., Smith, J. and Taber, C. (1998). Accounting for dropouts in evaluations of social programs, *The Review of Economics and Statistics* **80**(1): 1–14.

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**(449): 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.

Ikeda, S., Kang, M.-I. and Ohtake, F. (2010). Hyperbolic discounting, the sign effect, and the body mass index, *Journal of Health Economics* **29**(2): 268–284.

Imbens, G. and Manski, C. F. (2004). Confidence intervals for partially identified parameters, *Econometrica* **72**(6): 1845–1857.

Jeffery, R. W., Gerber, W. M., Rosenthal, B. S. and Lindquist, R. A. (1983). Monetary contracts in weight control: Effectiveness of group and individual contracts of varying size, *Journal of Consulting and Clinical Psychology* **51**(2): 242 – 248.

John, L., Loewenstein, G., Troxel, A., Norton, L., Fassbender, J. and Volpp, K. (2011). Financial incentives for extended weight loss: A randomized, controlled trial, *Journal of General Internal Medicine* **26**(6): 621–626.

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression, *Journal of the American Statistical Association* **91**(433): 222–230.

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. (2006). Body mass index and occupational attainment, *Journal of Health Economics* **25**(2): 347 – 364.

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

Narbro, K., Jonsson, B., Larsson, B., Waaler, H., Wedel, H. and Sjöström, L. (1996). Economic consequences of sick-leave and early retirement in obese swedish women, *International Journal of Obesity Related Metabolic Disorders* **20**(10): 895–903.

Newell, D. J. (1992). Intention-to-treat analysis: implications for quantitative and qualitative research, *International Journal of Epidemiology* **21**(5): 837–841.

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

OECD (2011). Oecd.stat. Accessed on 08 June 2012.
URL: */content/data/data-00285-en*

Oswald, A. J. and Powdthavee, N. (2007). Obesity, unhappiness and the challenge of affluence: Theory and evidence, *Economic Journal* **117**(521): 441–454.

Paloyo, A. R., Reichert, A. R., Reinemann, H. and Tauchmann, H. (2011). The causal link between financial incentives and weight loss: an evidence-based survey of the literature, *Ruhr Economic Papers* **290**.

Reichert, A. R. and Tauchmann, H. (2012). When outcome heterogeneously matters for selection: a generalized selection correction estimator, *RWI* **mimeo**.

Relton, C., Strong, M. and Li, J. (2011). The 'Pounds for Pounds' weight loss financial incentive scheme: an evaluation of a pilot in NHS Eastern and coastal kent, *Journal of Public Health* **33**(4): 536–542.

Ruhm, C. J. (2012). Understanding overeating and obesity, *Journal of Health Economics, 10.1016/j.jhealeco.2012.07.004* .

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

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

Stephens, D. J., Saunders, J. K., Belsley, S., Trivedi, A., Ewing, D. R., Iannace, V., Capella, R. F., Wasielewski, A., Moran, S., Schmidt, H. J. and Ballantyne, G. H. (2008). Short-term outcomes for super-super obese (bmi ¿ or =60 kg/m2) patients undergoing weight loss surgery at a high-volume bariatric surgery center: laparoscopic adjustable gastric banding, laparoscopic gastric bypass, and open tubular gastric bypass, *Surgery for Obesity and Related Diseases* **4**(3): 408 – 415.

Tunceli, K., Li, K. and Williams, L. K. (2006). Long-term effects of obesity on employment and work limitations among u.s. adults, 1986 to 1999, *Obesity* **14**(9): 1637–1646.

Van de Ven, W. P. and Van Pragg, B. M. (1981). The demand for deductibles in private health insurance: A probit model with sample selection, *Journal of Econometrics* **17**(2): 229–252.

Vidal, J. (2002). Updated review on the benefits of weight loss, *International Journal of Obesity Related Metabolic Disorders* **26**(Suppl 4): 25–8.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J. and Loewenstein, G. (2008). Financial incentive–based approaches for weight loss, *JAMA: The Journal of the American Medical Association* **300**(22): 2631–2637.

WHO (2006). *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia*, World Health Organization.

Williamson, D. F., Serdula, M. K., Anda, R. F., Levy, A. and Byers, T. (1992). Weight loss attempts in adults: goals, duration, and rate of weight loss., *American Journal of Public Health* **82**(9): 1251–1257.

Withrow, D. and Alter, D. A. (2011). The economic burden of obesity worldwide: a systematic review of the direct costs of obesity, *Obesity Reviews* **12**(2): 131–141.

Yoo, S.-H. and Yang, C.-Y. (2000). Dealing with bottled water expenditures data with zero observations: a semiparametric specification, *Economics Letters* **66**(2): 151–157.

Appendix

A.1 Possible Bias Implied by Naively Comparing Observed Group Means

For the sake of simplicity, consider a binary outcome variable Y, i.e. 'target weight realized' and a uniform treatment T. Here $Y = 1$ indicates success and $Y = 0$ indicates failure, while $T = 1$ indicates exposure to financial incentives and $T = 0$ indicates control group membership. The purpose of the empirical analysis is to estimate the expected value of Y conditional on treatment status T, i.e. $\mathrm{P}(Y = 1|T = 1)$, denoted by p_T, and $\mathrm{P}(Y = 1|T = 0)$, denoted by p_C, and ultimately the difference $p_T - p_C$, capturing the treatment effect under scrutiny. However, Y is only observed conditional on attending the weigh-in ($W = 1$), but unobserved for dropouts ($W = 0$). Hence group means of observed outcome do not consistently estimate $\mathrm{P}(Y = 1|T = 1)$ and $\mathrm{P}(Y = 1|T = 0)$. They rather are unbiased estimators for the probabilities $\mathrm{P}(Y = 1|T = 1 \cap W = 1)$ and $\mathrm{P}(Y = 1|T = 0 \cap W = 1)$, respectively.

Further consider the following identities that directly follow from the familiar axioms of probability theory and Bayes's theorem:

$$\mathrm{P}\left(Y = 1|T = 1 \cap W = 1\right) = \frac{\mathrm{P}\left(Y = 1 \cap T = 1 \cap W = 1\right)}{\mathrm{P}\left(T = 1 \cap W = 1\right)}, \tag{1}$$

$$\mathrm{P}\left(Y = 1 \cap T = 1 \cap W = 1\right) = \mathrm{P}\left(W = 1|Y = 1 \cap T = 1\right) \times \mathrm{P}\left(Y = 1|T = 1\right) \times \mathrm{P}\left(T = 1\right), \tag{2}$$

$$\mathrm{P}\left(T = 1 \cap W = 1\right) = \mathrm{P}\left(Y = 1 \cap T = 1 \cap W = 1\right) + \mathrm{P}\left(Y = 0 \cap T = 1 \cap W = 1\right). \tag{3}$$

It is key to our argument that the probability for attending the weigh-in may vary depending on both, treatment status T and being successful Y, see section II. Hence, we write the probabilities for attending the weigh-in conditional on failure relative to the probability for attending conditional on success as

$$\delta_T \equiv \frac{\mathrm{P}\left(W = 1|Y = 0 \cap T = 1\right)}{\mathrm{P}\left(W = 1|Y = 1 \cap T = 1\right)}, \tag{4}$$

$$\delta_C \equiv \frac{\mathrm{P}\left(W = 1|Y = 0 \cap T = 0\right)}{\mathrm{P}\left(W = 1|Y = 1 \cap T = 0\right)}. \tag{5}$$

Substituting (2), (3), and (4) into (1) and rearranging terms yields

$$\mathrm{P}\left(Y=1|T=1\cap W=1\right) = \frac{p_T}{p_T+\delta_T(1-p_T)}, \quad (6)$$

$$\mathrm{P}\left(Y=1|T=0\cap W=1\right) = \frac{p_C}{p_C+\delta_C(1-p_C)}, \quad (7)$$

where the second equation follows by analogy to the first. Thus the expected value of the group differential in mean observed outcome is $\frac{p_T}{p_T+\delta_T(1-p_T)} - \frac{p_C}{p_C+\delta_C(1-p_C)}$ rather than $p_T - p_C$.

From (7) and (6) it becomes obvious that comparing group means of observed outcome yields an unbiased estimate of the treatment effect if $\delta_T = \delta_C = 1$ holds. That is the case if the probability to attend the weigh-in is not related to success, neither for the treatment group nor for the control group. Yet, it is more plausible to assume $\delta_T < \delta_C \leq 1$, since only successful members of the treatment group are exposed to strong financial incentives for showing up at the weigh-in. Then, an upward bias is guaranteed for the limiting case $\delta_C = 1$. If we further assume $p_T \geq p_C$, i.e. a non-negative treatment effect, an upward bias is also guaranteed for the limiting case $p_T = p_C$.

Yet, the direction of bias is generally unknown for the general case $\delta_T < \delta_C < 1$ and $p_T > p_C$, albeit an upward bias will occur for a relatively wide region of parameter values. A downward bias may occur if the link between failure and dropping out is strong for both groups and the group differential in the probability for success is large.

A.2 Figures and Tables

Table A1: **Socioeconomic Background of the Study Population and the Obese in Germany**

	Study Population	Patients of the four rehabilitation clinics	Representative Obese in Germany (BMI≥30)
Female (%)	32.23	34.17	49.84
Age (years)	48.11	49.69	57.11
Married (%)	61.03	71.37	62.23
Resident of Baden-Württemberg (%)	100.00	94.99	11.84
Natives (%)	78.89	82.67	86.30
Full-time employed$^{\diamond,*}$ (%)	69.44	76.12	34.85
Part-time employed$^{\diamond,*}$ (%)	9.04	11.01	14.27
Unemployed$^{\diamond,*}$ (%)	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. $^{\diamond}$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). $^{\diamond}$Here, we distinguish between the unemployed and the not-employed (4.45 percent). * 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).

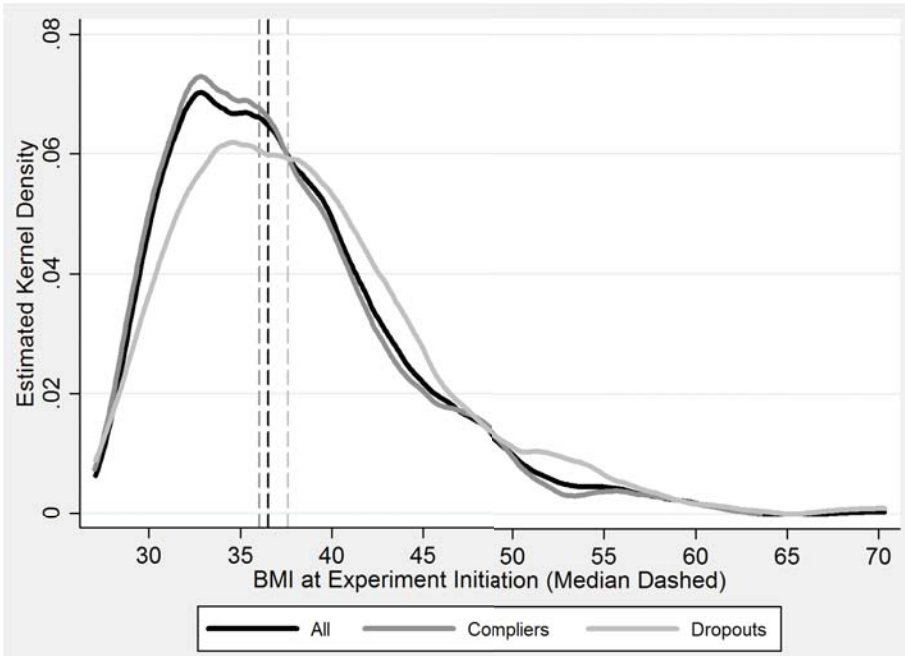

Figure A1: **Distribution of BMI at Experiment Initiation for Dropouts and Compliers**

Table A2: **Results for Percentage Change in Body Weight**

	OLS		Heckit		General. Heckit[a]		Intention-to-Treat	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
€150	−2.645**	0.549	−2.655**	0.566	−2.618**	0.972	−1.712**	0.385
€300	−2.898**	0.611	−2.929**	0.743	−2.294**	0.923	−2.096**	0.459
Clinic: Bad Mergentheim	−0.182	0.610	−0.178	0.554	−0.029	0.590	0.022	0.443
Glottertal	−0.649	0.969	−0.647	0.946	−0.636	0.952	−0.287	0.705
Isny	−0.559	0.585	−0.577	0.698	−0.665	0.659	−0.437	0.466
City	0.399	0.760	0.428	0.868	0.500	0.808	0.108	0.523
Population density	0.334	0.377	0.332	0.411	0.347	0.418	0.174	0.265
Female	1.391**	0.530	1.388**	0.489	1.362**	0.497	0.935**	0.392
Age	0.031	0.025	0.030	0.030	0.025	0.025	0.019	0.016
Native	0.365	0.532	0.369	0.585	0.352	0.585	0.219	0.384
Educ. attainment: very low	−1.369	0.872	−1.360	0.897	−1.276	0.896	−1.180*	0.674
low	−1.753**	0.798	−1.746**	0.747	−1.746**	0.745	−1.388**	0.621
medium	−1.818**	0.840	−1.799**	0.870	−1.721**	0.828	−1.295**	0.640
Employed	−0.915	0.631	−0.919	0.603	−0.783	0.617	−0.629	0.441
Self-assessed-health: good	1.631	1.397	1.619	1.531	1.675	1.523	0.885	0.878
satisfactory	1.570	1.354	1.566	1.494	1.692	1.504	0.907	0.856
poor	1.504	1.379	1.499	1.518	1.617	1.522	0.911	0.859
bad	4.987**	1.624	5.004**	1.856	5.160**	1.841	2.704**	1.077
1/2 h since last food intake	0.341	1.335	0.344	1.337	0.346	1.336	0.214	1.431
2 h since last food intake	−0.303	1.281	−0.301	1.270	−0.293	1.272	−0.371	1.380
Wearing shoes	2.839**	1.141	2.845*	1.582	2.719*	1.594	3.059**	1.038
Wearing shoes × heavy shoes	1.232*	0.743	1.233*	0.692	1.275*	0.729	1.199	0.738
Wearing pullover	0.701	0.490	0.702	0.499	0.714	0.509	0.603	0.463
Wearing long trousers	−0.001	0.784	−0.001	0.655	0.013	0.664	−0.035	0.760
Early	0.406	1.067	0.406	1.200	0.374	1.235	0.177	0.995
One week late	0.890	0.607	0.888*	0.535	0.930*	0.554	0.937	0.584
Two weeks late	1.693**	0.725	1.692**	0.738	1.756**	0.754	1.829**	0.682
Three weeks late	2.022*	1.132	2.018	1.237	1.948	1.231	2.006*	1.116
> three weeks late	2.309**	1.080	2.306**	1.100	2.299**	1.108	2.369**	1.009
Constant	−6.253**	2.904	−6.146*	3.553	−6.193*	3.190	−6.054**	2.323
Inverse Mill's ratio (λ)	-	-	−0.127	2.130	-	-	-	-
λ_{control}	-	-	-	-	0.089	1.505	-	-
$\lambda_{€150}$	-	-	-	-	0.097	1.503	-	-
$\lambda_{€300}$	-	-	-	-	−2.310	2.094	-	-
σ	4.902		4.656		-		4.288	
# of observations	520		520		520		698	

Notes: ∗∗ significant at 5% (two-sided test); ∗ significant at 10% (two-sided test); comprehensive set of covariates used; monthly dummies included. [a] reduced number of categories for self-assessed health (very good/good, satisfactory, poor/bad).

Table A3: **Estimated Treatment Effects (Basic Specification)**

	OLS		Heckit[a]		General. Heckit[b,c]		Intention-to-Treat	
	€150	€300	€150	€300	€150	€300	€150	€300
Percentage Change in Body Weight								
All	-2.850^{**}	-3.189^{**}	-3.045^{**}	-3.765^{**}	-3.087^{**}	-2.464^{**}	-2.132^{**++}	-2.937^{**++}
	(0.548)	(0.584)	(0.593)	(0.771)	(1.042)	(1.032)	(0.409)	(0.462)
Males	-2.717^{**}	-2.556^{**}	-3.305^{**}	-3.953^{**}	-2.182^{*}	-1.729^{*}	-2.107^{**}	-2.458^{**}
	(0.644)	(0.696)	(0.846)	(1.143)	(1.338)	(1.230)	(0.499)	(0.578)
Females	-4.257^{**}	-4.923^{**}	-4.193^{**}	-4.420^{**}	-2.138^{*+}	-4.081^{**+}	-2.992^{**++}	-4.151^{**++}
	(1.060)	(1.122)	(1.005)	(1.106)	(1.620)	(1.260)	(0.730)	(0.785)
Absolute Change in BMI								
All	-1.042^{**}	-1.204^{**}	-1.125^{**+}	-1.451^{**+}	-1.142^{**}	-0.776^{**}	-0.787^{**++}	-1.110^{**++}
	(0.204)	(0.219)	(0.226)	(0.294)	(0.396)	(0.403)	(0.152)	(0.175)
Males	-1.012^{**}	-1.020^{**}	-1.254^{**}	-1.594^{**}	-0.816^{*}	-0.510	-0.792^{**}	-0.968^{**}
	(0.244)	(0.269)	(0.333)	(0.449)	(0.525)	(0.484)	(0.190)	(0.224)
Females	-1.500^{**}	-1.768^{**}	-1.483^{**}	-1.633^{**}	-0.736^{*+}	-1.466^{**+}	-1.062^{**++}	-1.510^{**++}
	(0.372)	(0.406)	(0.370)	(0.407)	(0.562)	(0.438)	(0.259)	(0.286)
Target Weight Realized								
All	0.209^{**+}	0.291^{**+}	0.196^{**++}	0.288^{**++}	-	-	0.163^{**++}	0.269^{**++}
	(0.049)	(0.048)	(0.048)	(0.042)	-	-	(0.036)	(0.039)
Males	0.241^{**}	0.254^{**}	0.231^{**}	0.262^{**}	-	-	0.187^{**}	0.239^{**}
	(0.062)	(0.061)	(0.058)	(0.051)	-	-	(0.046)	(0.050)
Females	0.208^{**++}	0.380^{**++}	-	-	-	-	0.156^{**++}	0.330^{**++}
	(0.084)	(0.078)	-	-	-	-	(0.062)	(0.062)

Notes: $**$ significant at 5% (one-sided test); $*$ significant at 10% (one-sided test); $++$ significant differential in treatment effects between treatment groups at 5% (one-sided test); $+$ significant differential in treatment effects between treatment groups at 10% (one-sided test); comprehensive set of covariates used. [a] two-step estimation for continuous outcome variables, marginal effects for Van de Ven and Van Pragg (1981) probit estimation with sample selection correction for binary outcome 'target weight realized'; [b] not applied to binary outcome 'target weight realized'; [c] reduced set of covariates for selection equations used, if only one gender is considered.

Table A4: **Estimated Treatment Effects (Extendend Specification)**

	OLS		Heckit[a]		General. Heckit[b,c]		Intention-to-Treat	
	€150	€300	€150	€300	€150	€300	€150	€300
Percentage Change in Body Weight								
All	-2.965^{**}	-3.242^{**}	-3.057^{**}	-3.505^{**}	-3.263^{**}	-2.852^{**}	-2.153^{**++}	-2.959^{**++}
	(0.538)	(0.594)	(0.575)	(0.717)	(0.921)	(0.864)	(0.404)	(0.468)
Males	-2.761^{**}	-2.465^{**}	-3.218^{**}	-3.494^{**}	-2.654^{**}	-1.768^{*}	-2.093^{**}	-2.371^{**}
	(0.634)	(0.720)	(0.774)	(0.974)	(1.317)	(1.212)	(0.493)	(0.584)
Females	-4.103^{**+}	-5.252^{**+}	-4.022^{**}	-4.679^{**}	-2.263^{*++}	-4.487^{**++}	-2.833^{**++}	-4.370^{**++}
	(1.041)	(1.134)	(0.969)	(1.049)	(1.561)	(1.267)	(0.786)	(0.841)
Absolute Change in BMI								
All	-1.078^{**}	-1.214^{**}	-1.136^{**}	-1.378^{**}	-1.109^{**}	-0.949^{**}	-0.793^{**++}	-1.114^{**++}
	(0.201)	(0.222)	(0.221)	(0.276)	(0.351)	(0.332)	(0.150)	(0.176)
Males	-1.019^{**}	-0.967^{**}	-1.220^{**}	-1.420^{**}	-0.984^{**}	-0.534	-0.783^{**}	-0.923^{**}
	(0.242)	(0.278)	(0.308)	(0.388)	(0.523)	(0.480)	(0.188)	(0.225)
Females	-1.450^{**+}	-1.873^{**+}	-1.429^{**}	-1.726^{**}	-0.783^{*++}	-1.633^{**++}	-1.013^{**++}	-1.585^{**++}
	(0.360)	(0.404)	(0.353)	(0.382)	(0.529)	(0.421)	(0.276)	(0.300)
Target Weight Realized								
All	0.217^{**+}	0.285^{**+}	0.200^{**++}	0.277^{**++}	-	-	0.161^{**++}	0.265^{**++}
	(0.050)	(0.049)	(0.047)	(0.042)	-	-	(0.037)	(0.039)
Males	0.238^{**}	0.238^{**}	0.229^{**}	0.238^{**}	-	-	0.181^{**}	0.231^{**}
	(0.063)	(0.062)	(0.056)	(0.053)	-	-	(0.047)	(0.050)
Females	0.186^{**++}	0.397^{**++}	-	-	-	-	0.131^{**++}	0.342^{**++}
	(0.083)	(0.082)	-	-	-	-	(0.061)	(0.065)

Notes: $**$ significant at 5% (one-sided test); $*$ significant at 10% (one-sided test); $++$ significant differential in treatment effects between treatment groups at 5% (one-sided test); $+$ significant differential in treatment effects between treatment groups at 10% (one-sided test); comprehensive set of covariates used. [a] two-step estimation for continuous outcome variables, marginal effects for Van de Ven and Van Pragg (1981) probit estimation with sample selection correction for binary outcome 'target weight realized'; [b] not applied to binary outcome 'target weight realized'; [c] reduced set of covariates for selection equations used, if only one gender is considered.

Table A5: **Estimated Coefficients for Selection Equations**

	Regular Heckit		**Generalized Heckit**[a]					
			Control		**€ 150**		**€ 300**	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
€ 150	0.196	0.131	-	-	-	-	-	-
€ 300	0.688**	0.144	-	-	-	-	-	-
Clinic: Bad Mergentheim	−0.102	0.133	0.200	0.237	0.032	0.240	−0.492*	0.297
Glottertal	0.054	0.241	0.373	0.409	0.011	0.438	−0.035	0.530
Isny	0.427**	0.180	0.398	0.314	0.639*	0.350	0.713	0.466
City	−0.562**	0.177	−0.394	0.309	−0.848**	0.337	−0.462	0.355
Population density	0.009	0.103	−0.278	0.199	0.087	0.189	−0.020	0.216
Female	0.044	0.128	0.077	0.242	−0.073	0.242	0.103	0.282
Age	0.024**	0.006	0.022**	0.010	0.039**	0.012	0.020*	0.011
Native	−0.101	0.153	−0.228	0.296	−0.048	0.262	0.199	0.364
Educ. attainment: very low	−0.236	0.239	−0.748	0.419	0.330	0.475	−0.471	0.484
low	−0.174	0.207	−0.352	0.386	−0.193	0.368	−0.111	0.420
medium	−0.398	0.215	−0.774	0.394	−0.100	0.386	−0.467	0.441
Employed	0.111	0.150	0.230	0.287	0.224	0.276	−0.351	0.337
Self-assessed-health: good	0.343	0.385	-	-	-	-	-	-
satisfactory	0.158	0.377	−0.158	0.253	0.051	0.256	−0.383	0.310
poor	0.157	0.382	−0.493*	0.276	0.037	0.292	−0.389	0.335
bad	−0.294	0.449	-	-	-	-	-	-
Cholesterol test	−0.163	0.137	−0.533**	0.259	0.132	0.247	−0.226	0.301
Blood sugar test	0.343*	0.175	0.665*	0.342	0.140	0.340	0.425	0.359
Pharmacy in town	0.437**	0.121	0.429**	0.219	0.730**	0.232	0.213	0.272
Constant	−1.120*	0.621	1.087	1.018	−2.660**	0.922	0.239	1.088
Wald-test on joint sig. of variables excluded from main equation (p-value)	0.001		0.041		0.010		0.608	
# of observations	698		233		236		229	

Notes: ∗∗ significant at 5% (two-sided test); ∗ significant at 10% (two-sided test); comprehensive set of covariates used; monthly dummies included. [a] reduced number of categories for self-assessed health (very good/good, satisfactory, poor/bad).

Table A6: **Descriptive Statistics for Mediators and Health Outcomes (Mean Values at Experiment Initiation)**

	All	Control	€150	€300	Compliers	Dropouts
Blood sugar (mg/dl)	100.88	99.73	102.02	100.85	99.78	104.09*
Cholesterol (mg/dl)	187.71	187.58	185.83	189.79	187.1	189.5
Snacking (%)	0.46	0.44	0.49	0.44	0.45	0.47
Frequency of exercise: never (%)	0.20	0.17	0.22	0.22	0.18	0.28**
occasionally (%)	0.35	0.41	0.32++	0.32++	0.35	0.35
once a week (%)	0.23	0.25	0.21	0.23	0.24	0.21
several times a week (%)	0.19	0.16	0.22+	0.21	0.21	0.14**
daily (%)	0.02	0.01	0.03+	0.02	0.02	0.02
Awareness caloric content: very low (%)	0.22	0.22	0.22	0.22	0.22	0.22
	0.19	0.22	0.19	0.17	0.21	0.14**
.	0.23	0.25	0.21	0.22	0.23	0.22
.	0.20	0.16	0.22	0.21	0.19	0.21
	0.10	0.11	0.11	0.08	0.08	0.15**
very high (%)	0.07	0.05	0.05	0.10++	0.07	0.06
Awareness dietary composition: very low (%)	0.18	0.17	0.19	0.18	0.18	0.20
	0.15	0.18	0.14	0.13	0.17	0.09**
.	0.23	0.23	0.24	0.23	0.22	0.26
.	0.19	0.18	0.21	0.19	0.19	0.20
	0.15	0.16	0.14	0.14	0.13	0.18
very high (%)	0.10	0.09	0.08	0.13	0.11	0.08
Frequency of climbing stairs: never (%)	0.12	0.13	0.13	0.10	0.10	0.18**
	0.14	0.15	0.13	0.13	0.15	0.12
	0.24	0.20	0.24	0.28+	0.25	0.22
.	0.17	0.15	0.16	0.20	0.17	0.15
	0.16	0.16	0.15	0.16	0.17	0.12*
always (%)	0.18	0.21	0.19	0.14+	0.17	0.22
# of observations	698	233	236	229	520	178

Notes: ++ deviation from control group significant at 5%, + significant at 10%; ** deviation from non-dropouts significant at 5%, * significant at 10%; standard deviations omitted because of most variables being binary. The values of the cholesterol level and blood glucose level were higher at the start of the clinic stay. The WHO (2006) suggests a blood glucose level above 126 milligram per deciliter as a diagnostic criterion for diabetes. A high total serum cholesterol level is defined as 240 milligrams per deciliter.

Table A7: **Treatment Effect on Mediators and Health Outcomes by Gender**

	OLS		Heckit		Intention-to-Treat		Lee-Bounds	
	€150	€300	€150	€300	€150	€300	€150	€300
Females								
Diet Composition	0.189	0.595**	0.188	0.573**	0.131	0.519**	0.237	0.331**
(aware)	(0.169)	(0.159)	(0.144)	(0.159)	(0.123)	(0.119)	[−0.090	[0.046
Snacking	0.009	0.110	0.014	0.139	0.045	0.159*	−0.079	−0.052
(less)	(0.118)	(0.112)	(0.104)	(0.111)	(0.086)	(0.082)	[−0.279	[−0.185
Climbing Stairs	0.419**	0.223	0.422**	0.242	0.336**	0.318**	0.313*	0.214
(more)	(0.199)	(0.206)	(0.164)	(0.193)	(0.139)	(0.142)	[−0.009	[−0.117
Physical Exercise	−0.010	0.237	0.027	0.458**	0.039	0.262**	−0.147	0.042
(more)	(0.167)	(0.166)	(0.178)	(0.206)	(0.122)	(0.117)	[−0.457	[−0.206
Self-assessed Health	0.123	0.172	0.113	0.107	−0.010	0.037	0.016	−0.043
(better)	(0.153)	(0.146)	(0.136)	(0.153)	(0.116)	(0.111)	[−0.262	[−0.299
Cholesterol Level	0.090	0.008	0.075	−0.023	0.110**	0.004	0.236**	0.109**
(change,%)	(0.073)	(0.075)	(0.057)	(0.064)	(0.046)	(0.045)	0.375]	0.197]
Blood Sugar Level	0.036	0.041	0.077	0.132	0.076	0.128**	0.097*	0.167**
(change,%)	(0.070)	(0.069)	(0.086)	(0.097)	(0.061)	(0.061)	0.196]	0.284]
Males								
Diet Composition	0.016	−0.157	0.026	−0.138	0.057	−0.046	−0.026	−0.293**
(aware)	(0.115)	(0.114)	(0.120)	(0.146)	(0.083)	(0.085)	[−0.209	[−0.500
Snacking	0.146**	0.050	0.191**	0.169*	0.116**	0.060	0.087	−0.097
(less)	(0.070)	(0.069)	(0.079)	(0.101)	(0.051)	(0.052)	[−0.046	[−0.223
Climbing Stairs	0.105	0.148	0.248	0.407*	0.118	0.168*	0.118	0.081
(more)	(0.122)	(0.123)	(0.204)	(0.238)	(0.09)	(0.092)	[−0.076	[−0.142
Physical Exercise	0.097	−0.012	0.105	0.006	0.094	0.028	−0.026	−0.191**
(more)	(0.097)	(0.096)	(0.098)	(0.126)	(0.068)	(0.070)	[−0.193	[−0.356
Self-assessed Health	0.127	0.049	0.158*	0.157	0.095	0.042	0.180**	−0.055
(better)	(0.091)	(0.090)	(0.096)	(0.123)	(0.067)	(0.068)	[0.021	[−0.244
Cholesterol Level	−0.036	−0.053	0.071	0.194	−0.013	−0.013	−0.035	0.013
(change,%)	(0.049)	(0.048)	(0.130)	(0.197)	(0.034)	(0.034)	0.064]	0.111]
Blood Sugar Level	−0.014	0.059	−0.022	0.071	−0.012	0.022	−0.023	0.091*
(change,%)	(0.059)	(0.058)	(0.058)	(0.070)	(0.037)	(0.039)	0.086]	0.204]

Notes: ** significant at 5%; * significant at 10%; the extended set of covariates is used. [and] indicate the lower and upper limits of the confidence interval. The respective bounds are lower ([) and upper (]) bounds. For brevity, only the conservative bound for the test of the hypothesis of no improvement in the respective outcome variable is displayed, i.e., if a positive (negative) effect is expected, the lower (upper) bounds are reported.