Who responds to financial incentives for weight loss? Evidence from a randomized controlled trial

Alfredo R. Paloyo
University of Wollongong, apaloyo@uow.edu.au

Arndt R. Reichert
World Bank, USA

Monika Reuss-Borst
Rehazentren Baden-Wurttemberg

Harald Tauchmann
Friedrich-Alexander-Universitat Erlangen-Nurnberg

Publication Details
Who responds to financial incentives for weight loss? Evidence from a randomized controlled trial

Abstract
There is a paucity of evidence on the heterogeneous impacts of financial incentives on weight loss. Between March 2010 and January 2012, in a randomized controlled trial, we assigned 700 obese persons to three experimental arms. We test whether particular subgroups react differently to financial incentives for weight loss. Two treatment groups obtained a cash reward (€150 and €300 with 237 and 229 participants, respectively) for achieving an individually-assigned target weight within four months; the control group (234 participants) was not incentivized. Participants and administrators were not blinded to the intervention. We find that monetary rewards effectively induced obese individuals to reduce weight across all subgroups. However, there is no evidence for treatment-effect heterogeneity for those groups that were incentivized. Among those who were in the €300 group, statistically significant and large weight losses were observed for women, singles, and those who are not working (all above 4 kg in four months). In addition, the magnitude of the reward matters only for women and migrants. The effectiveness of financial incentives to reduce weight nevertheless raises sensitive ethical issues that should be taken into consideration by policymakers.

Keywords
trial, who, controlled, randomized, evidence, loss, weight, incentives, financial, responds

Disciplines
Business

Publication Details

This journal article is available at Research Online: http://ro.uow.edu.au/buspapers/881
WHO RESPONDS TO FINANCIAL INCENTIVES FOR WEIGHT LOSS? EVIDENCE FROM A RANDOMIZED CONTROLLED TRIAL∗

Alfredo R. Paloyo
University of Wollongong

Arndt R. Reichert
World Bank

Monika Reuss-Borst
Rehaklinik am Kurpark

Harald Tauchmann
University of Erlangen-Nuremberg

August 2015

Abstract

In a randomized controlled trial involving 700 obese persons assigned to three experimental arms, we test whether financial incentives have heterogeneous effects over subgroups on weight reduction. Two treatment groups obtained a cash reward (€150 and €300 with 237 and 229 participants, respectively) for achieving an individually-assigned target weight within four months; the control group (234 participants) was not incentivized. Participants and administrators were not blinded to the intervention. We find that monetary rewards effectively induced obese individuals to reduce weight across all subgroups. However, there is no evidence for treatment-effect heterogeneity for those groups that were incentivized. Among those who were in the €300 group, statistically significant and large weight losses were observed for women, singles, and those who are not working (all above 4 kg in four months). In addition, the magnitude of the reward matters only for women and migrants.

∗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, Adam Pilny, Margarita Pivovarova, Gisela Schubert, Michael Tamminga, and Marlies Tepä for research assistance. We 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 Hoferberth, Marina Humburg, Silke Kohlenberg, Thomas Krohm, Max Lux, Tatjass Meier, Constanze Schaal, and Wolfgang Stiels. Thanks are also due to Thomas K. Bauer, Christoph M. Schmidt, and Boris Augurzky, whose support proved invaluable for the execution of the experiment. We also thank the editor and two anonymous reviewers. A previous version of this manuscript was circulated as "Heterogeneous Causal Effects of Financial Incentives on Weight Loss", Ruhr Economic Paper No. 383. Correspondence: Alfredo R. Paloyo, Centre for Human and Social Capital Research, University of Wollongong, RWI, IZA; alfredo@paloyo.net.
1 Introduction

There is increasing interest on the part of policymakers, health professionals, and social scientists to evaluate the impact of potential modifiers of behavior such as financial incentives. This kind of incentive is commonly used. For example, negative financial incentives are widely applied in the form of taxes on products such as alcohol or tobacco [Grossman et al. 1993]. Incentive programs are not unusual for health insurers, who would like their customers to engage in healthy behavior [Stock et al. 2010]. However, the use of financial incentives in the treatment of obesity and overweight still raises ethical concerns [Halpern, Madison, and Volpp 2009], and its effectiveness is not well-established with respect to the issue of weight maintenance [Jeffery 2012; John et al. 2011; Augurzky et al. 2015].

A number of previous studies have issues with respect to sample size and methodology (see Paloyo et al. [2014] for an evidence-based survey of the literature), consequently making a causal interpretation of previous results unwarranted. Endogeneity and selection issues have not always been addressed adequately in the earlier literature. More recent evidence, however, shows that financial incentives can induce weight loss, at least in the short term [Augurzky et al. 2012; Finkelstein et al. 2007; Volpp et al. 2008].

However, it remains unknown whether subgroups of patients respond more than others to financial incentives [John, Loewenstein, and Volpp 2012]. It may turn out that money or the prospect thereof "nudges" only people with existing financial problems or special social or cultural backgrounds toward a healthier lifestyle. If this is the case, financial incentives could be allocated more appropriately to probable responders to make disbursement more cost-effective. For subgroups where financial incentives do not matter, an alternative approach or combination of approaches may be necessary to induce positive behavioral changes that would lead to weight loss and weight maintenance in the long run.

Ethical issues pervade the use of financial incentives to change behavior. For instance, incentives could increase socioeconomic and health inequities, undermine the doctor–patient
therapeutic relationship, usurp personal autonomy, and—at least for those who are economically disadvantaged—be construed as coercive [Lunze and Paasche-Orlow 2013]. Issues of promoting “mercenary” behavior (incentivizing behavior that ought to be done anyway), privacy concerns (tracking sensitive medical information to evaluate compliance), and fairness (paying Peter for what Paul does for free) have also been raised [Halpern, Madison, and Volpp 2009]. Shaw [2007] flatly argues that it is not acceptable to pay for people to ensure compliance. More generally, ethical issues arise out of using nudges in modifying behavior, especially when designed to improve health [Blumenthal-Barby and Burroughs 2012].

An ethical issue that is especially relevant for this paper is that which relates to heterogeneous effects. If a particular segment of the population responds to a financial incentive—say, those who are genetically predisposed to more easily achieve target weights and those who are located in an environment which allows for healthier food choices—then an incentive program may reward members of society who need less institutional help as opposed to providing support to those who need it most [Halpern, Madison, and Volpp 2009]. Policymakers may, however, recognize an ethical obligation to allocate scarce resources such that medical benefits are maximized, even if this results in differential treatment of individuals who differ with respect to their responsiveness to financial incentives. We discuss these ethical concerns later when we conclude.

A remaining institutional issue is the extent to and ease with which policies may be targeted to responding subgroups. It may be the case that certain clinical measures (such as fasting blood sugar or low- or high-density lipoproteins) or psychological measures (such as the locus of control or the “Big Five” personality traits [Cobb-Clark and Schurer 2011; Cobb-Clark et al. 2014]) are highly correlated with the success of financial incentives at inducing weight loss. However, administrative data are typically based on broad socioeconomic indicators, such as sex, employment status, educational attainment, citizenship, and place of residence. Collecting more detailed clinical and psychological indicators is a costly undertaking (and is rife with thorny privacy issues). From a policy perspective, therefore, it is administratively simpler to
deploy a broad-based financial-incentives program—one that is based on observable and easily verifiable socioeconomic characteristics, if we are to engage in any policy targeting at all.

We contribute to the literature by estimating the impact of financial incentives across a number of relevant subgroups. The estimated impacts can guide policymakers and health professionals in choosing the most appropriate policy instrument for a particular group of people. Although there are a number of papers that estimate the impact of cash or near-cash incentives on weight loss, there has been no attempt thus far to examine seriously the potential for treatment-effect heterogeneity across a number of baseline characteristics (John, Loewenstein, and Volpp [2012] provide some exploratory evidence). Even in the absence of treatment-effect heterogeneity, such a result would nevertheless provide policymakers and health professionals the license to apply an incentive system over a broad base, making its disbursement administratively simpler. This can contribute to its success in settings where the institutional mechanisms, including the number of qualified program administrators, are deficient.

This paper seeks to address this gap in the literature by analyzing a large-scale, prospective, randomized, controlled field experiment. Based on experimental data of 700 obese medical rehabilitation patients, the present analysis was conducted to test the primary hypothesis that randomly assigned positive, output-oriented financial incentives exhibit a heterogeneous treatment effect on weight—measured in kilograms lost—as a function of baseline characteristics. Participants were assigned an individualized weight-loss target, and a financial reward was given to those who were able to achieve it or partially achieve it within four months.

In addition, we also contribute to the discussion on the ethical dimension of offering financial incentives for health-related behavior and outcomes, as well as to the use of nudges in general to modify behavior. A number of ethical issues are immediately raised if people with certain characteristics respond to nudges more than others. These ethical issues typically have a direct link to anti-discriminatory legislation that aims to prevent groups of people being unduly prejudiced by an incentive scheme that does not take into account differing abilities.
Our objective is to examine whether monetary rewards for weight loss are more influential on patients with certain baseline characteristics that are unaffected by the intervention. We are then able to identify responders and non-responders. In determining which characteristics are relevant for our analysis, we take the point of view of public administrators (e.g., social health insurers) who have data that are often limited to basic socioeconomic variables such as gender and employment status, and without direct information on income, wealth, or clinical indicators.

The global increase of obesity and its comorbidities is considered to be one of the toughest challenges confronting health-care systems worldwide. Between 1980 and 2008, the worldwide prevalence of obesity more than doubled [Finucane et al. 2001]. By 2015, it is predicted that 2.3 billion adults will be overweight, with more than 700 million of them obese. In Germany in particular—the site of the experiment discussed in this paper—66 percent of the male and about 50 percent of the female adult population are overweight according to a large epidemiological study with almost 20,000 participants aged between 14 and 80 years. The percentage of German adults being obese amounts to more than 20 percent [Max-Rubner Institut 2008].

The associated economic burden of obesity is tremendous [Bhattacharya and Sood 2011]. Incremental medical expenditures and the value of lost productivity, including absenteeism and presenteeism (health-related limitations at work), as a result of overweight and obesity are estimated by Finkelstein et al. [2010] to range from $322 for overweight to $6,087 for grade III obese men (morbidly obese or having a body-mass index (BMI) $\geq$ 40 kg/m$^2$). For women, estimates range from $797 for overweight to $6,694 for grade III. In the US, it has been estimated that the total annual cost attributable to obesity among full-time employees is $73.1 billion. Individuals with a BMI above 35 kg/m$^2$ represent 37 percent of the obese population, but are responsible for 62 percent of excess costs. Gates et al. [2008] report that workers with a BMI equal or above 35 kg/m$^2$ experience the greatest health-related work limitations, being significantly less productive than mildly obese workers.
From an ethical perspective, the economic burden of obesity is a nontrivial matter. Within the context of a national or public health insurance system or where an employer-based scheme is in place (generally, where risk pooling exists), the costs associated with unhealthy individuals are already partially borne by those who are healthy or those who engage in healthier activities. Cognizant of this, Halpern, Madison, and Volpp [2009] argue that it makes sense for a financial incentive scheme to be broadly implemented if it is cost-effective since it reduces the burden borne by those who are otherwise healthy.

The causes for the upsurge of obesity are well-known [Rosin 2008]. The increase of energy-dense convenience or fast foods, the decrease in physical activity due to more sedentary jobs, motorized transport, increasing urbanization, labor-saving devices, but also changes in leisure activities (television, computer games, etc. over outdoor, physical play), result in an imbalance of calorie intake and expenditure, leading to a calorie surplus for many people. As a consequence of these mainly environmental, social, and partly political changes, obesity is not only a problem in well-developed, high-income countries but is also increasing in developing countries.

Empirically, one can observe that obesity follows a strong socioeconomic gradient [Baum and Ruhm 2009]. For example, obesity rates are increasing with age and are inversely correlated with high educational level and income. Moreover, evidence from Germany suggests that single individuals more often have normal weight than married, divorced, or widowed persons, and adolescents with a migration background have a greater risk of becoming obese than natives [Kurth and Schaffrath Rosario 2007]. This suggests that targeted policy instruments could potentially be used to achieve a more cost-effective allocation of resources.

The effective treatment of obesity is of crucial importance not only for the obese individuals themselves, but also for the social system and employers. Treatment programs mainly consist of advice on diet and physical exercise tailored to the individual patient's needs and may be best characterized as behavioral weight-loss intervention programs. The fundamental requirement for success is to achieve an energy balance between calories spent and calorie input for
weight maintenance, and a state of calorie deficit for weight loss. Although behavioral interventions have been proven to be effective in various studies [Appel et al. 2011], most programs are, at best, only modestly successful and do not address all patients adequately. In particular, some patients do not respond to particular interventions, and an attempt to separately identify at least the broad characteristics of the non-responders has not been adequately undertaken before.

We test for treatment-effect heterogeneity over the following six socioeconomic sub-groups: (1) male/female, (2) single/in a relationship, (3) high educational attainment/low educational attainment, (4) native/migrant, (5) employed/not employed, and (6) urban/rural. Since the members of the control group who did not receive a financial incentive were also asked to lose weight by their physicians during their medical rehabilitation stay, a secondary objective is to determine which subgroup would not have lost weight without the addition of the financial incentive.

The reason for this approach is that a large-scale monetary incentive scheme will likely be managed by public entities or, if not, at least by organizations that will not have access to detailed personal health information due to privacy restrictions or some other institutional constraint. However, assuming that there is substantial effect heterogeneity with respect to individual characteristics, it is a priori unclear whether the "publicly observable" characteristics, such as sex and employment status, are actually relevant for targeted interventions. It is entirely possible that other characteristics play a much stronger role for the effectiveness of monetary rewards. For instance, genetic endowment and household-specific eating and cooking conventions may be better predictors of weight loss, and, hence, may influence the effectiveness of the monetary rewards more than any other variable. Yet, policymakers will hardly have any authentic information about such household and individual characteristics, and they are thus rendered ill-suited to serve as criteria for a differential treatment in an intervention.

The following describe our expectations with respect to the variation in treatment effects across baseline covariates:
(1) Females may respond more to the financial incentive because it typically constitutes a larger share of their labor income. They may also be rewarded more in the labor market for their weight loss, since the cost of obesity for females has been shown to be higher (e.g., Morris [2007], Hebl and Turchin [2005], and Puhl and Heuer [2009]). Recent research from Reichert [forthcoming] indicates that women also benefit more from weight loss if we consider their employment probabilities post-weight loss.

(2) Single individuals may be more flexible with the use of their leisure time and be less constrained by intra-family conventions, which enables them to better regulate both their physical activity and their diets. However, singles also lack the emotional support of a partner [Saccone and Israel 1978].

(3) A high cognitive ability may be necessary to understand the benefits of losing weight, so we expect those with a high educational attainment to lose more weight [Halkjær et al. 2003]. Moreover, better educated individuals may have easier access to information that is relevant for successfully losing body weight. The robust correlation between education and health is documented elsewhere, such as the survey by Grossmann [2006] and the natural experiment analyzed by Braakmann [2011].

(4) Language barriers may hinder the benefits of rehabilitation for obese migrants. However, since migrants earn, on average, less than natives (e.g., Lehmer and Ludsteck [2011]), the relatively higher share of the reward in their income may induce them to lose more weight than natives. Migrants are also observed to change their obesogenic behaviors upon arrival, perhaps due to acculturation [Delavari et al. 2013].

(5) The employed may be better equipped to lose weight, since they are more likely to be better educated and can afford to enroll in fitness classes. However, their leisure time constitutes a smaller part of their overall time budget compared to those who are not employed, which limits the opportunities for them to expend calories outside work. Moreover, the extra financial reward may not be seen as a significant amount to earn by losing weight. Obesity itself, of course, could potentially have an impact on employment prospects [Morris 2007; Reichert, forthcoming].
Nutritional habits exhibit a substantial rural–urban differential. Although inhabitants of urban areas may be exposed to unhealthier food choices (e.g., fast food), the diet in rural areas is still more meat-oriented in Germany (the study site). In addition, urban areas afford better educational opportunities, which may mitigate the effect of the prevalence of junk food. This difference between rural and urban areas has been documented elsewhere as well [Reeder et al. 1997; Zimmet et al. 1983].

2 Methods

The randomized controlled trial was funded by the Pakt für Forschung und Innovation and administered by the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI). Participants were invited to enroll in the experiment during the final week of their stay in a rehabilitation clinic in the state of Baden-Württemberg (BW) in Germany. The study protocol was approved by the ethics commission of the Chamber of Medical Doctors of BW. A detailed description of the trial is available elsewhere [Augurzky et al. 2012], but we briefly describe it here.

Patients that were eligible for recruitment had a BMI of 30 or above at admission, were aged between 18 and 75 years, and were residents of BW. The exclusion criteria included pregnancy, psychological and eating disorders, tumor disease within the last five years, and alcohol or drug abuse. Between March 2010 and August 2011, a total of 700 participants were recruited, though two were later excluded because of pregnancy and cancer. The last weigh-in occurred in January 2012. The flow of the study is depicted in Figure 1.

Outcomes were ascertained by trained medical personnel, who measured the height and weight of participants. A detailed questionnaire was administered to collect information on socioeconomic background and other health-related variables. Upon clinical discharge, the responsible physician set an individual weight-loss target that had to be realized within the subsequent four months. This was a weight reduction of about 6 to 8 percent of the patient's weight, which was defined by the medical directors of the four rehabilitation clinics after discussions with the medical staff. Participants were stratified by clinic, and computer-aided randomization was car-
ried out within each stratum. Random assignment to the experimental group took place after the weight-loss target was set by the physician in charge. Hence, the medical staff could not discriminate between the different experimental groups.

The three experimental arms include a control group and two treatment groups, with the latter consisting of a group facing a €150 reward and another group with a corresponding €300 reward for achieving the individualized weight-loss target. The amounts are in line with previous studies (e.g., Jeffery et al. [1983]) to facilitate comparison. All experimental arms received medical advice and counseling. Patients assigned to the treatment groups received no reward if their weight loss was less than 50 percent of their target weight loss. Once the threshold 50 percent has been reached, the financial reward was disbursed proportionally, with the full amount paid out only if the weight-loss target was achieved or exceeded. For example, for a study participant in the €300 group who achieved a weight loss of 80 percent of the target, he or she would receive €240 (or 80 percent of €300). Once the experiment was underway, weigh-ins were conducted in designated pharmacies using a weighing scale operated by a pharmacist to avoid self-reported weight measures, which could lead to systematic measurement error.

The nonlinear disbursement scheme was intended to prevent participants from “gaming” the system, which is also relevant to prevent harm to the study participants. If they could achieve the full amount after achieving a particular target, participants just around the cutoff may engage in unhealthy or unethical behavior to be just above the cutoff right before the weigh-in. For example, they could refuse to eat and drink a day before the weigh-in to lose just enough weight to be above the cutoff and then receive the full amount. With a proportional disbursement, the likelihood to game the system this way is reduced.

At baseline, the study staff was blinded to the treatment group. In the pharmacy, the pharmacists were blinded but participants could have told the pharmacist about their treatment status. To prevent strategic behavior (choosing a “cooperative” pharmacist), participants were assigned to a specific pharmacy. Although the characteristics of dropouts were statistically dif-
ferent from those who completed the study [Augurzky et al. 2012], the intention-to-treat analy-

sis performed below does not generate the bias that would otherwise occur from a per-protocol
analysis (ignoring selection effects).

The financial incentive is positive and output-based. It is positive in the sense that partic-
ipants are rewarded for having achieved a particular goal. Negative incentives, in contrast, typi-
cally include participants paying a bond at the start of the experiment, which they can then for-
feit if they are unable to meet the criteria set by the researchers. Furthermore, it is output-based
because the reward is conditional on having achieved the desired weight loss, which is the out-
put of behavioral decisions that an individual can take. Input-based incentives are typically tied
to better diet and an active physical lifestyle, which are generally harder or at least more costly
to observe than a simple measure of one's weight on a particular date. (A taxonomy of incentive
designs is provided in Paloyo et al. [2014].)

The choice of a positive, output-based incentive is guided both by efficiency considera-
tions as well as fairness. First, a negative incentive could represent a credible threat of harm to
the study participants. We avoid that with a positive incentive since the choices of study partici-
pants are less likely to be driven by coercive forces. However, to the extent that some individuals
are more able to lose weight relative to others, it may also be unfair to those who put in the same
amount of effort but do not achieve the desired outcome (say, for physiological reasons). Second,
an output-based incentive reinforces personal autonomy since it allows the participant to
choose the optimal mix of inputs or processes to achieve their desired outcomes. Specifically
incentivizing particular activities (gym membership, food diary, etc.) would not only be harder
to monitor, one may also argue that it is overly paternalistic. Third, previous studies have shown
that output-contingent incentives are more effective than process-based ones [Schmidt, Asch,
and Halpern 2012].

Mean weight loss (in kg) was calculated for the two treatment arms and the control arm.
The result for the control group is interpreted as baseline weight loss that cannot be attributed
to financial rewards. The difference between average weight loss in the treatment arms and baseline weight loss is regarded as the incentive-induced weight loss. That is to say, we account for the weight loss that would occur in any case, with or without the financial incentive. This weight loss is likely attributable to the combination of setting a personalized weight-loss target and the medical advice and counseling that was afforded all study participants.

Outcomes are examined based on the intention-to-treat principle with the assumption that dropouts return to baseline. The ITT approach is a solution to deal with imperfect compliance and missing data in experimental studies which still ultimately delivers unbiased estimates [Fisher et al. 1990; Gupta 2011]. These estimates obtained from ITT analyses are typically conservative (i.e., in a sense, they are the lower-bound estimates) since they are usually scaled up by the compliance rate. Of course, when compliance is perfectly assured, then the ITT equals the average treatment effect.

However, it is important to consider the nature of the program to see the relevance of the ITT estimates. A rewards system such as the one used here will likely be administered by policymakers on a voluntary basis. Forcing compliance would be a costly and an ethically tenuous proposition. It is arguably the more policy-relevant parameter to estimate, as suggested by Bloom [2008]. Indeed, even if the intended program is made compulsory, any attempt to ensure perfect compliance is unlikely to succeed, so the relevance of the ITT parameter survives.

Treatment effects are estimated via ordinary least squares as the difference in weight loss between the treatment arms and the control arm of the experiment. Consequently, in assessing the effect heterogeneity of the financial incentives on weight loss according to baseline characteristics, the estimation of baseline and incentive-induced weight loss was conducted over the subgroup categories motivated above. All estimates are presented with their associated 95-percent confidence intervals.
3 RESULTS

3.1 Baseline characteristics

Of the 700 participants that were randomized, 234 were assigned to the control group, 237 to the €150 group, and 229 to the €300 group; two were eventually excluded from the sample, while a total of 178 missed the follow-up weigh-in (i.e., 25 percent dropped out); see Figure 1 in Section 2. The average target weight loss is 6.5 percent from a starting average weight of 113 kg. The distribution of weight loss by experimental groups does not demonstrate the presence of outliers.

We regard participants as employed if they are either full- or part-time employed, marginally employed, or have not provided information on the type of their employment. Highly educated participants either finished the 10th grade of secondary school or hold a university-entrance diploma (Abitur). Individuals with a good self-reported health status answered the question on perceived health with good or very good (the other response options were satisfactory, poor, and bad). The city variable indicates individuals living in municipalities that cover more than one postcode.

Variables measured at baseline were shown to be balanced between groups (Table 1). About a third of the sample consists of women and about 80 percent were born in Germany (natives). The mean age is 48 years and most of the participants are employed. All mean comparisons between groups except the percentage of German natives and the percentage of the employed for the €300 group indicate covariate balance at baseline using the method proposed by Hansen and Bowers [2008] for stratified experimental studies.

3.2 Intention-to-treat effect

The results in Figure 2 show the mean weight loss for each experimental arm in the trial. Replacing the missing values for the dropouts with their initial values, one can see that the control group lost almost 2 kg even without the prospect of a financial reward. In the case of the two
treatment arms, the weight loss amounted to about 4 to 5 kg (left panel of the figure). Weight losses in both treatment arms were significantly higher than in the control arms, revealing that financial incentives can motivate people to lose weight significantly. Using information only on those who completed the study (completed-case or per-protocol analysis) shows a similar weight-loss pattern, though the mean weight loss is slightly higher for all groups (right panel of the figure).

The two panels of the graph illustrate the well-known selection bias one expects to observe in per-protocol analyses. This can occur since it is probable that those who foresee themselves as benefitting the most from the treatment will have a higher likelihood of completing the study. This endogeneity in sample attrition makes the analysis based on the intention-to-treat principle more reliable. The companion paper Augurzky et al. [2012] discusses endogeneity in more detail, with a more thorough exposition on the average treatment effect, which they demonstrate to be robust to a number of methods that account for sample attrition, including the intention-to-treat analysis. Our paper concentrates on the evidence for treatment-effect heterogeneity that underlies what is observed for the whole study population.

### 3.3 Treatment heterogeneity

The estimation results for mean baseline weight change (denoted as $\alpha$) and for mean incentive-induced weight change (denoted as $\delta$), i.e., additional weight loss attributable to the monetary rewards, are presented in Table 2. The table further displays the statistical significance of these effects and the equality of baseline as well as the incentive effects between the subgroups of interest. The 95% confidence intervals are also presented.

Overall, the estimates for the baseline weight change range from $-0.179$ kg (95% CI $[-1.843, +1.485]$) for those who are not employed to $-2.241$ kg (95% CI $[-2.993, -1.489]$) for men. The estimates for the incentive-induced weight change range from $-0.962$ kg (95% CI $[-2.669, +0.745]$) for urban residents (€150) to $-4.272$ kg (95% CI $[-6.385, -2.159]$) for singles (€300). Only in two instances (migrants and those living in urban areas) do we not observe
the €150 to induce a statistically significant weight loss (indicated by *). Increasing the financial incentive to €300, however, results in a significant weight loss for all these groups. Being in the €150 group induced the highest weight loss for those who are not working (−3.326 kg, 95% CI [−5.733, −0.919]), while being in the €300 group translated to maximum incentive-induced weight loss for singles (−4.272 kg, 95% CI [−6.385, −2.159]).

One can see that people who were given the standard treatment of weight-loss counseling and accompanying medical advice also lost weight, except for the subgroup of women, singles, and those who are not employed. Similar to what we can observe in the overall results, adding a financial incentive on top of the standard treatment generated even more substantial weight losses than what would have occurred in the absence of the incentive. The chance to earn an extra €150 at the end of the four-month weight-loss period, for example, generates an average of 3.3 kg lost for those who are not working.

Higher financial rewards only correspond to significantly higher observed weight loss (indicated by +) for two specific subgroups (for women and migrants). The largest difference is for migrants: a 100-percent increase in financial incentives corresponds to a 285-percent increase in weight loss (from −1.141 kg, 95% CI [−2.690, 0.408] to −3.261 kg, 95% CI [−4.810, −1.712]). Although the size of the estimated effect of €300 is higher for all subgroups as compared to the effect of the €150 reward, the difference between the two effects is only statistically significant for the aforementioned groups. Therefore, while there is evidence of a positive relationship between the size of the financial incentive and the resulting amount of weight loss, the results indicate that the presence of a financial incentive itself is more important than its size.

The result for women and migrants are especially relevant considering that, first, women are likely to benefit the most from weight loss; and, second, that migrants (from low- and middle-income countries to high-income countries) deteriorate in health, as measured by BMI, as they acculturate in their new environments. As obese women lose weight, they find it more likely to find employment, and the benefit to them seems to be higher than the corresponding bene-
fit to men [Morris 2007; Reichert, forthcoming]. At least for migrant children, we have evidence from a natural experiment that migration increases both the BMI and the likelihood of being obese for 3- to 5-year olds [Stillman et al. 2012].

We do not observe any significant general treatment effect heterogeneity (indicated by #) for the financial incentives across subgroups. However, there is some heterogeneity in the baseline weight change. Even though they were not motivated by financial rewards, men lost more weight than women, participants in a relationship lost more weight than singles (at 10-percent significance level), and the employed lost more weight than those who are not employed. It is also notable that in each of these categories, the counterpart subgroup did not exhibit any significant baseline weight reduction. However, when faced with a financial reward (irrespective of the amount), these groups of people are able to significantly lose weight. This implies that financial incentives also work for people who do not benefit from conventional intervention programs. Indeed, these groups are exactly those which exhibit the most substantial weight loss when given a bonus for losing weight.

Take the case of those who are not working. Without the financial incentive, they showed no significant decrease in their baseline weight when subjected to an individualized target weight loss and medical advice and counseling alone. However, those unemployed which ended up in either of the incentivized arms lost quite a substantial amount (3.3 kg and 4.2 kg for an incentive of €150 and €300, respectively). In this case, where conventional methods of inducing weight loss are unsuccessful, financially incentivizing the unemployed to lose weight is a promising alternative for the relevant authorities.

The pattern seems to follow an income gradient, as these groups of people are more likely to have an economically disadvantaged position relative to their counterparts (e.g., male vs. women or employed vs. not working). In fact, if we take the observations from people who belong to the first quartile of the income distribution (regression results are not presented but are available upon request), we find a similar pattern—that is, being prompted to realize a certain
weight-loss target alone does not generate a statistically significant change in weight, but financial incentives do, irrespective of the amount.

4 DISCUSSION

We deployed a large-scale, prospective, randomized, controlled field experiment designed to test the effectiveness of financial incentives in inducing weight loss among obese patients. The experiment demonstrated that a positive, output-based financial incentive is an effective instrument to achieve a significant weight loss among the obese, consistent with what has been found in a number of previous studies [Paloyo et al. 2014]. However, a general heterogeneity in the treatment effect of financial incentives is neither observed for socioeconomic characteristics nor for exhibited health-related behaviors in the study. Heterogeneity is observed for the impact of traditional weight-loss counseling and medical advice coupled with a personalized weight-loss target.

In addition, there is evidence for particular subgroups (for women and migrants) that increasing the amount of financial incentive has a directly proportional effect on the amount of weight loss. Moreover, for migrants and urban citizens, the chance to achieve €150 is not enough to induce further weight loss in addition to that achieved via ordinary medical advice and counseling. Only the higher monetary reward motivated them to lose significantly more weight as compared to baseline. This means that the effect of promising more money for weight loss generally depends on specific individual characteristics.

Another important result is the capacity of financial incentives to induce people to lose weight when they would otherwise have failed to do so when exclusively facing conventional treatments. This is observed for women, singles, and those who are not employed. These characteristics are observed for those who are lower in the income distribution. For these groups, it is not sufficient simply to give them a personalized weight-loss target, medical advice, and weight-loss counseling. It is precisely for these people where a financial incentive could have a relevant impact.
Since the other groups exhibited weight loss anyway (though, admittedly, the magnitude is smaller) even without the marginal incentive, this is a particularly relevant result for resource-constrained policymakers. Considering a fixed budget constraint, if the objective of the policy based on financial incentives is to ensure that everyone loses weight, a larger part of the monetary amount might be better allocated to these groups. Therefore, the publicly observable personal characteristics are informative enough for a potentially more efficient allocation of resources.

It is not obvious, however, that such a scenario should be the goal of public policy. For instance, while inducing every obese person to lose weight is desirable, ensuring that the weight loss is medically significant nevertheless remains a relevant factor. If that were the case, the obese facing higher financial incentives are most likely to lose the most weight and therefore reap the health benefits of such a weight loss. Although ensuring that everyone loses, say, at least 1 kg is an attractive proposition, it is not necessarily better than increasing the likelihood of a particular group of people to lose at least 5 percent of their current weight, which is often regarded as a threshold to improve the health status of the obese [Vidal 2002]. In other words, it is not altogether clear that inching everyone closer to the target is more beneficial than increasing the number of people who hit the target. In fact, the medical literature suggests that the latter should probably take precedence, since it is not likely that health benefits will accrue to an obese person who loses 1 kg, implying that the financial incentive used to achieve that would have been wasted.

Notably, public support for financial incentives to improve health outcomes has been shown to be contingent on the effectiveness of the incentives to generate the desired results [Promberger et al. 2012]. To the extent that its effectiveness at inducing weight loss for obese people has been demonstrated here, it is much more likely that a generalized or scaled-up policy similar to the one designed in our experiment could be put forward by national or local governments, or perhaps even private organizations, would be palatable to the eventual (tax)payer.
The results of the trial raise certain ethical considerations that deserve further discussion. First, there is the idea that the financial incentive could be construed as coercive since the amount matters for particular subgroups. Second, since there is potential for resources to be diverted from those who would lose weight anyway without the financial incentive to those who would only lose weight with the incentive, it creates a situation where an individual is rewarded for an outcome that another individual is able to achieve without the additional incentive. Third, resources may also be reallocated from those who are unresponsive to the financial incentive simply because they are physically, economically, or environmentally incapable of achieving the target outcome, raising the issue of discrimination. There are other ethical considerations apart from these three (e.g., the promotion of mercenary social values, stigmatization, and the undermining of the doctor–patient relationship) but these are general enough to the use of financial incentives for behavioral change. A number of them are discussed in Halpern, Madison, and Volpp [2009] and Lunze and Paasche-Orlow [2013].

Large financial incentives could be very influential for those who are economically disadvantaged. A “nudge” for people falling under the right tail of the income distribution may well represent a “shove” for those who fall under the other tail. This could be construed as coercive since it potentially undermines a person’s freedom of choice. However, at least with respect to weight loss, the evidence available shows that people want to lose weight but are simply unable to do so because of, among other things, self-control issues [John, Loewenstein, and Volpp 2012]. Providing them with an instrument to achieve that goal is hardly coercive or punishing because there is no “credible threat of negative ramifications” [Lunze and Paasche-Orlow 2013:660]. Indeed, what merely happens is that the incentives simply allow individuals to overcome “self-defeating tendencies toward immediate gratification without constraining [their] options” [Halpern, Madison, and Volpp 2009:515, emphasis supplied].

This approach has been termed “asymmetric paternalism” because people are indeed nudged toward the ultimately beneficial choice (which would have been inaccessible to them for various reasons) but the incentive does not eliminate the choice to lead a more unhealthy life-
style if the person chooses to do so, anyway [Loewenstein, Brennan, and Volpp 2007]. Of course, more research has to be conducted with respect to the optimal size of the incentive, i.e., identifying the point where the marginal benefit is just offset by the marginal cost, and this is fertile ground for future researchers.

One may consider it unfair to pay individuals to lose weight while other individuals are able to do it without the additional incentive. However, whenever there is risk pooling, healthy individuals are already subsidizing the unhealthy individuals in the first place [Halpern, Madison, and Volpp 2009]. This means that when unhealthy individuals are induced to engage in healthier activities, part of the benefits also accrue to those who were already healthy. This translates into, for example, lower health insurance premiums or a lower tax burden. The social cost of paying individuals to engage in healthier lifestyles may be offset by the benefits that accrue directly to those individuals and the indirect benefits that healthier individuals may receive.

Financial incentives may also be construed as discriminatory against those who face difficulties changing their behavior for a variety of reasons. Some individuals live in “food deserts” where barriers that prevent access to healthy foods exist [Cummins and Macintyre 2002]. Others may simply not have the economic or indeed genetic capacity to change their behavior or to reap the benefits of their healthier activities. To make the financial reward contingent on achieving the desired outcome would presumably prejudice those who are less able to comply given their circumstances. While a negative incentive program would unduly punish those who are unable to achieve the outcomes or participate in the incentivized activity, positive incentives at least do not impinge on the opportunity to do so [Halpern, Madison, and Volpp 2009]. This potential for discrimination to occur raises certain legal issues as well [Gostin 2007; Schmidt, Asch, and Halpern 2012]. For instance, an obese and disabled person may not have the physical capacity to engage in an activity tied to a financial incentive. In some jurisdictions (e.g., US), an alternative activity may be required by law [Lunze and Paasche-Orlow 2013].
While previous studies have shown that financial incentives are effective at inducing weight loss [Paloyo et al. 2014], the evidence for its effectiveness at maintaining the lower weight is limited and not definitive [Reuss-Borst et al. 2015; Augurzky et al. 2015; Jeffery 2012]. The present study is unable to shed light on this issue. Parallel to the pursuit of establishing the effectiveness of financial incentives for weight maintenance, researchers should also consider whether treatment-effect heterogeneities exist in this space, much as we have considered heterogeneity in the initial weight-loss phase. The results of these studies will inform on the optimal design of the incentive scheme which takes into account both issues of efficiency and equity or fairness. It may be left up to some democratic mechanism to decide over policies that are, on one hand, effective but unfair and, on the other hand, ineffective but fair when both cannot be simultaneously achieved.

In summary, the present paper shows that financial incentives effectively help obese individuals to reduce weight across all subgroups, whereas the magnitude of the reward seems to be relevant only for certain subgroups. Individual characteristics seem to matter more in the absence of monetary rewards since the paper reveals that control-group members of some subgroups did not significantly lose weight. We provide evidence that monetary rewards have the capacity to induce people to lose significant weight even if their individual characteristics imply unfavorable weight-loss prospects, i.e., people for whom the conventional weight-loss intervention programs are unsuccessful. If the objective of a policy based on financial incentives is to ensure that everyone loses weight, a larger part of the monetary amount might be better allocated to these groups. However, any incentive-based policy should take into consideration the ethical issues that automatically arise out of making the incentive contingent on participation in healthy activities or achieving healthy outcomes.
Literature


<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Control</th>
<th>EUR 150</th>
<th>EUR 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial weight (before rehab, kg)</td>
<td>117 ± 23</td>
<td>115 ± 22</td>
<td>118 ± 24</td>
<td>118 ± 24</td>
</tr>
<tr>
<td>Starting weight (after rehab, kg)</td>
<td>113 ± 22</td>
<td>111 ± 20</td>
<td>114 ± 23</td>
<td>114 ± 23</td>
</tr>
<tr>
<td>Target weight loss (%)</td>
<td>6.5 ± 1.2</td>
<td>6.4 ± 0.9</td>
<td>6.5 ± 1.3</td>
<td>6.5 ± 1.4</td>
</tr>
<tr>
<td>Bad Kissingen (%)</td>
<td>33.4</td>
<td>32.2</td>
<td>35.6</td>
<td>32.3</td>
</tr>
<tr>
<td>Bad Mergentheim (%)</td>
<td>41.7</td>
<td>41.6</td>
<td>40.7</td>
<td>42.8</td>
</tr>
<tr>
<td>Glottertal (%)</td>
<td>6.60</td>
<td>7.30</td>
<td>5.90</td>
<td>6.60</td>
</tr>
<tr>
<td>Isny (%)</td>
<td>18.3</td>
<td>18.9</td>
<td>17.8</td>
<td>18.3</td>
</tr>
<tr>
<td>Female (%)</td>
<td>32.2</td>
<td>29.1</td>
<td>31.3</td>
<td>36.3</td>
</tr>
<tr>
<td>Age (years)</td>
<td>48 ± 9</td>
<td>48 ± 10</td>
<td>48 ± 9</td>
<td>47 ± 9</td>
</tr>
<tr>
<td>Single (%)</td>
<td>20.4</td>
<td>18.5</td>
<td>20.9</td>
<td>21.8</td>
</tr>
<tr>
<td>Married (%)</td>
<td>61.1</td>
<td>63.1</td>
<td>61.3</td>
<td>59.0</td>
</tr>
<tr>
<td>Native (%)</td>
<td>78.9</td>
<td>79.5</td>
<td>75.1</td>
<td>82.2*</td>
</tr>
<tr>
<td>Employed (%)</td>
<td>82.4</td>
<td>85.8</td>
<td>81.7</td>
<td>79.5+</td>
</tr>
<tr>
<td>High educational attainment (%)</td>
<td>34.5</td>
<td>35.2</td>
<td>36.0</td>
<td>32.3</td>
</tr>
<tr>
<td>Population density (1000/km²)</td>
<td>7.5 ± 7.3</td>
<td>6.9 ± 7.1</td>
<td>7.0 ± 7.8</td>
<td>7.3 ± 7.1</td>
</tr>
<tr>
<td>City (%)</td>
<td>23.4</td>
<td>22.1</td>
<td>23.0</td>
<td>25.1</td>
</tr>
<tr>
<td>Blood Sugar Level (mmol/l)</td>
<td>188 ± 37</td>
<td>188 ± 37</td>
<td>186 ± 37</td>
<td>200 ± 38</td>
</tr>
<tr>
<td>Cholesterol Level (mmol/l)</td>
<td>101 ± 24</td>
<td>100 ± 23</td>
<td>102 ± 25</td>
<td>101 ± 24</td>
</tr>
</tbody>
</table>

* Deviation from control group significant at 5%. + Deviation from EUR 150 group significant at 5%. ± Standard deviations of binary variables omitted. Bad Kissingen, Bad Mergentheim, Isny, and Glottertal refer to the locations of the four rehabilitation clinics. Covariate balance is warranted by respective tests (Hansen and Bowers 2008).
Table 2 – Baseline Weight Change and Incentive-induced Weight Change by Selected Subgroups

<table>
<thead>
<tr>
<th></th>
<th>Baseline Weight Change</th>
<th>Incentive-induced Weight Change</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>CI</td>
<td>δ₁ (€150)</td>
<td>CI</td>
<td>δ₂ (€300)</td>
</tr>
<tr>
<td>Women</td>
<td>-0.592*</td>
<td>[-1.737, 0.553]</td>
<td>-2.290**</td>
<td>[-3.684, -0.897]</td>
<td>-4.007**</td>
</tr>
<tr>
<td>Single</td>
<td>-0.560</td>
<td>[-2.040, 0.919]</td>
<td>-3.115*</td>
<td>[-5.228, -1.002]</td>
<td>-4.272*</td>
</tr>
<tr>
<td>Migrant</td>
<td>-1.404*</td>
<td>[-2.446, -0.363]</td>
<td>-1.141*</td>
<td>[-2.690, 0.408]</td>
<td>-3.261**</td>
</tr>
<tr>
<td>Not employed</td>
<td>-0.179*</td>
<td>[-1.843, 1.485]</td>
<td>-3.326*</td>
<td>[-5.733, -0.919]</td>
<td>-4.240*</td>
</tr>
<tr>
<td>Urban</td>
<td>-1.457*</td>
<td>[-2.727, -0.187]</td>
<td>-0.962</td>
<td>[-2.669, 0.745]</td>
<td>-2.832*</td>
</tr>
</tbody>
</table>

*Weight change significant at 5%, ** (δ₂ − δ₁) significant at 5%, # deviation in weight change between the two subgroups significant at 5%. Two-sided t-tests used. The estimates and standard errors are obtained from subgroup-specific ordinary least-square regressions of weight change on the two treatment group indicators and a constant term.
Figure 1: Flow of the Study

700 eligible patients underwent randomization

- 234 were assigned to € 0
  - 155 had weight measured at 4 months
  - 78 missed weigh-in
  - 1 developed tumor
  - 233 were included in primary analysis

- 237 were assigned to € 150
  - 172 had weight measured at 4 months
  - 64 missed weigh-in
  - 1 withdrew from study
  - 236 were included in primary analysis

- 229 were assigned to € 300
  - 193 had weight measured at 4 months
  - 36 missed weigh-in
  - 229 were included in primary analysis
Figure 2: Mean Weight Change in Body Weight by Experimental Group