HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 13/20

Obesity, Weight Loss, and Employment

Arndt Reichert

August 2013

york.ac.uk/res/herc/hedgwp

Obesity, Weight Loss, and Employment Prospects: Evidence from a Randomized Trial*

Arndt Reichert

RWI†

April 26, 2013

Abstract

This study presents credible estimates for the causal effect of a variation in obesity on employment. By exploring random assignment of a weight loss intervention based on monetary rewards, I provide convincing evidence that weight loss positively affects the employment prospects of obese women but not of obese men. Consistent with this, significant effects of weight loss on proxy variables for labor productivity are found only for obese women.

JEL codes: *I10, I18, J24, J21*

Keywords: *Obesity, weight loss intervention, IV estimation, sample selection, labor productivity, employment*

*I am grateful to to 'Pakt für Forschung und Innovation' for funding, Ronald Bachmann, Thomas K. Bauer, Alfredo R. Paloyo, Sandra Schaffner, Christoph M. Schmidt, Hendrik Schmitz, Harald Tauchmann, and Nicolas Ziebarth for helpful comments. I also gratefully acknowledge the comments and suggestions of participants of the German Health Economics Association meeting in 2012 and the European Conference on Health Economics 2012. I further acknowledge research assistance of 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, and Marlies Tepaß. I want to express my particular acknowledgements to 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, I thank Susan Eube, Michael Falentin, Ina Hofferberth, Marina Humburg, Silke Kohlenberg, Thomas Krohm, Max Lux, Tatjass Meier, Monika Reuss-Borst, Constanze Schaal, and Wolfgang Stiels.

†Address for correspondence: Arndt Reichert, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany. Email address: reichert@rwi-essen.de.

1 Introduction

Obesity rates are rapidly increasing in almost all industrialized countries and in numerous emerging economies (e.g., Prentice, 2006). This is not only extremely worrisome for public health reasons but also from economic perspectives since obesity is found to turn social welfare contributors into recipients. Several studies, for instance, present empirical results that point to the obese having a substantially lower employment probability than healthy-weight people (e.g., Morris, 2007; Han et al., 2009). However, the evidence on the employment effects of obesity has not yet been settled, which is reflected by the growing number of research articles on this topic.

There is an evident difficulty of establishing a causal relationship between being obese and employment because both states are most likely correlated with unobserved factors. Unobserved time preferences, for example, are related to obesity through food consumption and physical activity as well as to employment through health capital investments. Another problem is that employment itself may affect body weight. One argument is that junk food is relatively cheap and, therefore, consumed more by the unemployed relative to employed people. Recent studies have aimed to solve these identification problems by employing instrumental variable (IV) estimation. However, concerns remain with respect to the validity of the employed instruments such as the obesity status of biological relatives because this may affect employment outcomes through other pathways like common household environment factors (Lindeboom et al., 2010).

This study is the first to use a randomized controlled source of variation in body weight to overcome potential reverse causality and endogeneity problems in the estimation of employment effects. The analysis is based on instrumental variables generated in the course of a randomized experiment which was conducted with the primary purpose of examining the effectiveness of monetary rewards for weight loss. Obese medical rehabilitation patients were randomly assigned to a control group and two treatment groups, which were both financially rewarded contingent on weight loss. As shown in Augurzky et al. (2012a), the two rewards were very effective in motivating the obese to lose weight,[1] which qualifies them as instrumental variables for variation in body weight. Compared to the instrumental variables that have been employed in the relevant literature, the significant advantage of using the monetary rewards as instruments is that they undoubtedly operate on employment through weight loss alone, dispelling any concerns about the validity.

A further novelty of the study is the presentation of separate estimates for the effects on the job finding probability of obese unemployed and the job retaining probability of obese employees

[1] There are further studies which provide evidence for the effectiveness of monetary rewards to induce individuals to exhibit desired healthy behaviors (e.g., Charness and Gneezy, 2009; Augurzky et al., 2012b).

for Europe's largest economy. The latter effect has been less in focus although obese workers may have a considerably increased risk of layoff or early retirement relative to healthy-weight workers. The paper additionally presents estimates for the reduced-form effect which captures the causal impact of the weight-loss intervention on employment. It allows me to examine whether the intervention has direct benefits for the welfare system that are relevant for its cost-benefit assessment. Moreover, it enables me to be the first to address the question of whether the employment prospects of the particular group of obese people may be improved by public intervention. If this turns out to be true, labor market programs explicitly designed for the obese would represent promising policy options and should include a financial incentive scheme for weight loss. Finally, the paper helps to identify the appropriate target of labor market interventions. For the present study population, it is reasonable to assume that excessive weight and its associated health problems represent the factors that most limit labor productivity and, by implication, employment prospects. Thus, in the case of obese individuals, interventions designed to combat overweight and obesity have the potential to outperform classical labor market programs such as training.

Despite the fairly short period under study (four-month weight-loss intervention), I find evidence of a positive employment effect of weight loss for women. The effect applies to both currently employed and unemployed obese women, although results point at larger effects for the latter. For men, I find no effects on employment prospects. Sex-specific heterogeneity in the labor-market effects of obesity has been found previously (e.g., Morris, 2006, 2007; Bhattacharya and Bundorf, 2009). The analysis further shows that weight loss is significantly associated with improvements in proxy variables for labor productivity in obese women. Specifically, I find a significant positive effect of weight loss on physical well-being and a significant negative effect on the probability of having limitations at daily tasks due to health complaints. Although it is beyond the scope of the analysis to check whether the direction of causation is from labor productivity to employment (and not the other way around), my findings of productivity gains support the finding of employment effects for women. Consistent with no employment effect, the labor productivity of men is unaltered by weight loss. A possible explanation for the observed heterogeneity in the effects by sex is that, as compared to obese men, obese women benefit significantly more from weight loss in terms of physical health, which has been identified to be the major cause for the obesity-wage penalty (Lundborg et al., 2010) and is seemingly the predominant productivity barrier of medical rehabilitation patients.

The remainder of this paper is organized as follows. The subsequent section describes the experimental design and introduces the data, Section 4 explains the estimation strategy, and Section 5 presents the estimation results. Sections 8 and 9 discuss the main findings and conclude.

2 Background

Obesity and, in turn, weight loss may affect employment prospects in various ways.[2] An obvious argument for weight loss among the obese is the associated positive health effects.[3] In fact, in the accompanying paper, Augurzky et al. (2012a) find some evidence of health improvements in the obese due to weight loss. These are likely to translate into an increase in working productivity (and a reduction in the risk of work incapacity). Another argument is that obese persons who successfully reduce their weight have consequently expanded the set of potential tasks and occupations with which they could be matched. For instance, Everett (1990) and Puhl and Brownell (2001) demonstrate that employers consider obese workers as unfit for public sales positions. Moreover, by losing weight, obese people signal a healthier lifestyle which indicates that they are more willing to invest in human capital, implying expected improvements in future working productivity.[4]

Since weight loss represents a difficult task and often an important goal for the obese (Crawford et al., 2000), those who achieve it are likely to be more optimistic about achieving other goals, too. The improvement in self-confidence may cause them to express their ideas in group meetings more often or to be more convincing in job interviews.[5] Actually, Mocan and Tekin (2011) provide evidence of the impact of self-esteem on wages. A similar argument is that they gain physical attractiveness which was found to yield better employment prospects (e.g., Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998). Mobius and Rosenblat (2006) connect the beauty premium with self-confidence, finding that while physically attractive workers are more confident and better equipped with respect to oral skills, both attributes are causally related to higher wages. A further reason is that obese individuals who lose weight are likely to suffer less from labor market discrimination, which is an evident problem faced by the obese (O'Brien et al., 2012; Rooth, 2009; Roehling et al., 2007).[6]

The link between obesity and employment prospects has been extensively analyzed in the literature. While the majority of the studies focuses on the effect of obesity on wages (e.g., Register

[2] A model on the potential mediators of obesity on labor market outcomes is presented by Baum and Ford (2004) as well as by Bhattacharya and Bundorf (2009).

[3] For instance, Wing et al. (2011) find improvements in indicators of cardiovascular risk factors. Christensen et al. (2007) show that weight loss significantly reduces physical disability in patients with knee osteoarthritis. Hooper et al. (2007) provide evidence for a decrease in painful musculoskeletal conditions. Brown et al. (2011) reports mental health improvements in obese individuals with serious mental illness. Reviewing the literature, Blackburn (1995) confirms positive health effects of weight loss.

[4] For the importance of lifestyle for health, see Kenkel (1995) and Contoyannis and Jones (2004).

[5] The huge importance of personality psychology in economics has been extensively shown (Bowles et al., 2001; Heckman and Rubinstein, 2001; Borghans et al., 2008).

[6] Rooth (2009) sent fictitious applications with weight-manipulated photos to real job openings showing that the obese receive significantly fewer job interview invitations. This differential treatment in hiring was more pronounced for women. See also Klesges et al. (1990).

and Williams, 1990; Averett and Korenman, 1996; Pagán and Dávila, 1997; Cawley, 2004; Baum and Ford, 2004; Morris, 2006; Bhattacharya and Bundorf, 2009; Han et al., 2009; Cawley et al., 2009; Lundborg et al., 2010), there also exists an extensive literature on the effect of obesity on employment (e.g., Morris, 2007; Norton and Han, 2008; Han et al., 2009; Cawley et al., 2009; Lindeboom et al., 2010; Caliendo and Lee, 2011).

The majority of the studies have aimed to solve the endogeneity problem and reverse causality by employing instrumental variable estimation.[7] Here, the validity of the employed instrument is crucial for consistent identification of the causal effect of interest. Cawley (2004) and Lindeboom et al. (2010) use the obesity status of biological relatives although the latter cast doubt on their suitability as instruments after extensive validity tests. Norton and Han (2008) employ genetic individual information linked to obesity. Yet, although there seem to be some genes that are more and some that are less correlated with body weight, there is no ”fat” gene or ”skinny” gene (Norton and Han, 2008). This implies that genes may also be correlated with several personal characteristics other than body weight. In turn, it has to be expected that they are correlated with unobservable variables such as further health risks, being likely to affect employment through other channels than obesity and, hence, to confound the IV estimation (Wehby et al., 2008; Lawlor et al., 2008a,b). Finally, Morris (2007) uses the prevalence of obesity in the area of residence which seems to be the most convincing approach. However, despite extensively controlling for area characteristics, he is not able to rule out endogenous regional selection due to the cross-sectional nature of the used data (see also Caliendo and Lee, 2011).

So far, the empirical evidence has argued in favor of a negative effect of the body-mass index (BMI) (or being obese) on wages of women but not of men. Regarding the probability to be in employment, Morris (2007) reports a statistically significant negative effect, which is smaller for men. Caliendo and Lee (2011) report a significant effect only for women. Both analyses are based on data from continental European countries.[8] Most of the papers that use US data find no effect of obesity on employment. A plausible reason for the intercontinental discrepancy in the results may be a stronger wage rigidity in European countries that inhibit wages to adequately reflect labor productivity differentials. This, in turn, would imply that the adjustment mechanism primarily works through the employment decision (Blau et al., 2001). In the light of the critique on the employed instrumental variables and the ambiguity of the results, Lindeboom et al. (2010) state the need to ”...analyze the corresponding effects in other countries, where the labor markets look different, and using different sources of variation in obesity.”

[7]By using logistic regression and matching methods, Han et al. (2009), Cawley et al. (2009), and Caliendo and Lee (2011) make the critical assumption that all variables that affect the likelihood of being obese are accounted for in their analyses.

[8]In contrast, Lindeboom et al. (2010) find no significant employment effect of being obese in the UK.

3 Experiment and Data

The data were generated by a field experiment, which is briefly discussed below. A more extensive discussion of the experiment is provided by Augurzky et al. (2012a). The dataset and descriptive statistics of the study population are subsequently presented.

3.1 The Experiment

The experiment was carried out between March 2010 and January 2012. In total, 700 patients of four rehabilitation clinics in Southwestern Germany were recruited for participation in the experiment. Two individuals had to be excluded from the trial because of pregnancy and cancer. For most patients of these clinics, medical rehabilitation is paid for by the German pension fund, which predominantly aims at avoiding unemployment or early retirement, requiring that the patient's ability to work is generally recoverable. Therefore, a joint characteristic of the participating patients may be that, besides being obese, they were available for the labor market. A major heterogeneity among participants is assumed with respect to their health complaints. The clinic in Bad Kissingen specializes in gastroenterology and endocrinology. Those in Bad Mergentheim and Isny primarily focus on orthopedics, whereas the clinic in Glottertal treats patients with psychosomatic disorders. Most of the participants are admitted on the basis of a diagnosis other than adiposity. However, it often turns out that their symptoms are related to their weight. Therefore, a successful treatment, *inter alia*, implies weight reduction.

At the end of the clinic stay, which usually lasted for 3 or 4 weeks, the physician in charge measured height as well as weight and set an individual weight-loss target for the following four months ranging between 6 and 8 percent. The participants were asked to fill in a detailed questionnaire related to family and socioeconomic background, including employment status and health status. To be eligible, the patients were required to have had a BMI above 30 at the start of the clinic stay. Further inclusion criteria were an age between 18 and 75 years, residence in the state of Baden-Württemberg, and a suitable health condition.[9,10]

After the clinic discharge, participants were randomly assigned to one of three experimental groups.[11] While the control group was not promised any rewards for losing weight, one incentive

[9] The exclusion criteria were pregnancy, psychological and eating disorders, tumor diseases within the last five years, abuse of alcohol or drugs, and serious general diseases.

[10] The study protocol was approved by the ethics commission of the Chamber of Medical Doctors of Baden-Württemberg.

[11] The information of group assignment was then sent by means of a letter to the home address of the participants marking the beginning of the intervention period. This guaranteed that the participants were not affected by (the treatment status of) other patients (see Angrist and Lavy, 2009, for a similar argumentation). Moreover, this excludes that diverging weight-loss targets set by different physicians produces any structural differences between the experimental groups.

group could receive up to €150 and the other up to €300. In terms of purchasing power parities (PPP), the rewards correspond to \$188 and \$376.[12] In particular, both groups were rewarded proportionally to the maximum reward conditional on the achieved weight loss exceeding 50 percent of the weight-loss target. They received the full bonus only if they met the weight-loss target. The payment mechanism is illustrated in Figure A1 in the Appendix.

To control weight loss, participants were weighed again after four months. In a 14-week reminder letter, the participants were directed to a nearby pharmacy for the control weigh-in. Attached was a second detailed questionnaire which contained the same questions on time-varying variables as the one at experiment initiation. Around 75 percent of the participants complied with control weigh-in attendance and filled out the questionnaire who received €25 (\$31 in PPP) as a promised fringe benefit.

3.2 Variable Description and Descriptive Statistics

The outcome variable is a binary indicator of whether the participant is in employment after the intervention period, i.e., it takes on the value 1 if the participant reports to be employed and 0 otherwise. Participants who report to be temporarily incapable of work but nevertheless have an employment contract are coded as employed.[13] Individuals in vocational training are defined as employees as well. Eight cases without information on the outcome variable are discarded, which reduces the number of observations to 690.

The explanatory variable of primary interest is weight loss measured in BMI. The BMI is calculated at the start and the end of the intervention period based on measured, not self-reported, height and weight. The dataset also contains indicator variables for the two incentive groups which play a key role in the identification of the causal effect of weight loss on employment as described in Section 4.

As control variables, I use socioeconomic characteristics such as age, being married, being single, being born in Germany (native), and having children. I further distinguish participants with respect to the educational level. Participants with a 10th grade of secondary school ("Realschule") are defined to have a medium educational level. Those with at least a tertiary education entry certificate are defined as highly educated. I construct two dummy variables indicating the two educational levels, while the reference category (low education) consists of individuals that have

[12]I used the purchasing power parities exchange rate of 2011 provided by OECD (2012).

[13]This applies to only two percent of the participants. The results are qualitatively robust with respect to excluding these observations from the analysis. Moreover, they are qualitatively robust to excluding part-time employees from the analysis and with respect to using the first difference of a weighted employment variable, which I generated by assigning the full-time employed the value 2, the part-time employed the value 1, and the marginally employed the value 0, as dependent variable.

completed the 9th grade, have a lower degree, no degree, or that do not know their highest educational degree.

I also control for the self-assessed health status using two dummy variables that indicate individuals with a satisfactory and good self-assessed health. Further covariates are dummy variables for the rehabilitation clinics, time dummy variables indicating the quarter and year of experiment initiation, and a city indicator, which turns one for individuals living in municipalities that cover more than one ZIP code.

In order to give an idea about the study population, descriptive statistics at the start of the intervention period are presented in Table 1. Less than one third (32 percent) of the participants are women. Most participants (76 percent of women and 85 percent of men) were employed before the intervention. The average BMI of female participants was 38.4, which was a little higher than the BMI of male participants (37.2). The distribution of the BMI by sex before the weight-loss intervention is displayed in Figure A2 in the Appendix.

About 33, 37, 17, and 14 percent of the female participants were recruited by the clinic in Bad Kissingen, Bad Mergentheim, Isny, and Glottertal, respectively. The shares of male participants recruited by the clinic in Bad Mergentheim and Isny are considerably higher. In contrast, the clinic in Glottertal recruited significantly less male than female patients. Around 25 percent of the observations finished the 10th grade of secondary school. The share of participants with at least a high school degree amounts to 10 percent. All other participants have a lower (or unknown) educational attainment. The average age is about 48 years. Most participants (78 percent) were born in Germany. This group is significantly larger among men. More than 25 percent of the participants reported a good health status.

Table 2 shows that the randomization algorithm yielded an even distribution across the three groups which indicates that the randomization procedure having worked properly. Most importantly, before the weight-loss intervention, there are neither significant group differences with respect to employment nor to BMI. Note that most covariates are balanced between the experimental groups (a more comprehensive table is presented in Augurzky et al., 2012a). Even though a few covariates differ significantly between the control and the two incentive groups, the covariate balance is warranted by respective tests (Hansen and Bowers, 2008).

During the intervention period, however, significant differences between the control and the two incentive groups among both female and male participants have arisen (Table 2). With respect to weight loss measured in percent of BMI, female members of the control group lost, on average, 0.6 percentage points, while women being rewarded with € 150 and € 300 lost 3.9 and 5.1 percentage points, respectively. The difference in weight loss between the two premium groups is

Table 1: **Descriptive Statistics (Means)**

	Women	Men	Min	Max
Dependent variable				
employed (lag ≡ start of weight-loss phase)	0.757	0.853	0	1
employed$^{\diamond}$ (end of weight loss phase)	0.786	0.860	0	1
change in employment$^{\diamond}$	0.018	-0.003	-1	1
	(0.319)	(0.291)		
Endogenous regressor				
BMI (start of weight-loss phase)	38.416	37.246	27	70.402
	(6.585)	(6.215)		
BMI$^{\diamond}$ (end of weight-loss phase)	36.657	35.172	22.857	63.296
	(6.546)	(5.850)		
change in BMI$^{\diamond}$ (in percent)	-3.618	-4.509	-29.978	27.161
	(5.244)	(5.294)		
Instruments				
€ 150	0.329	0.34	0	1
€ 300	0.374	0.31	0	1
Controls				
city	0.234	0.235	0	1
age (in years)	48.301	48.034	20	68
native	0.676	0.845	0	1
single	0.171	0.218	0	1
married	0.590	0.620	0	1
child	0.778	0.682	0	1
medium education	0.264	0.249	0	1
high education	0.102	0.106	0	1
satisfactory self-assessed health	0.369	0.456	0	1
good self-assessed health	0.258	0.254	0	1
Bad Mergentheim	0.365	0.440	0	1
Glotterbad	0.135	0.034	0	1
Isny	0.167	0.188	0	1
year 2010	0.284	0.312	0	1
second quarter	0.144	0.162	0	1
third quarter	0.261	0.333	0	1
fourth quarter	0.392	0.359	0	1
dropout	0.243	0.265	0	1
Exclusion restrictions for sample selection				
pharmacy in town	0.631	0.638	0	1
cholesterol test	0.590	0.568	0	1
blood glucose test	0.851	0.818	0	1
number of observations	222	468		

Notes: Standard deviations for binary variables omitted. No statistics for reference category reported. Means do not include observations with missing values. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics. $^{\diamond}$ For compliers only. Number of observations: 168 (Women) and 344 (Men).

insignificant (p-value of 11 percent). Male members of the incentive groups also lost more weight than men in the control group. The distribution of weight loss by experimental groups and gender is displayed in Figure A3 in the Appendix. Moreover, the shares of participants being in employment developed differently across experimental groups (only among women). While the share of employed women declined in the control group, it remained unchanged in the group with the € 150 reward and increased in the group with the € 300 reward. The difference between the latter group and the control group is statistically significant indicating improved employment perspec-

Table 2: **Dependent and Main Explanatory Variable by Experimental Groups**

	Women	Men	Women: Control	Women: €150	Women: €300	Men: Control	Men: €150	Men: €300
At Experiment Initiation								
BMI	38.416	37.246	38.162	38.309	38.714	36.77	37.318	37.705
	(6.585)	(6.215)	(6.521)	(5.918)	(7.23)	(5.641)	(6.600)	(6.401)
employed	0.757	0.853++	0.773	0.795	0.711	0.890	0.824	0.841
number of observations	222	468	66	73	83	164	159	145
During Treatment Period								
change in BMI (in percent)	-3.618	-4.509	-0.588	-3.903**	-5.109**	-2.949	-5.325**	-5.183**
	(5.244)	(5.294)	(6.787)	(3.332)	(4.742)	(4.789)	(4.818)	(5.870)
change in employment	0.018	-0.003	-0.073	0.000	0.082**	-0.027	0.026	-0.008
	(0.319)	(0.291)	(0.346)	(0.336)	(0.277)	(0.285)	(0.311)	(0.276)
number of observations	168	344	41	54	73	111	114	119

Notes: ++ deviation from women significant at 5%, + deviation from women significant at 10%; ** deviation from control group significant at 5%, * deviation from control group significant at 10%. ◇◇ deviation from group with €150 reward significant at 5%, ◇ deviation from group €150 reward significant at 10%. Standard errors for binary variable ommitted. No statistics for reference categroy reported.

tives due to the higher reward . Observing (significant) differences between the premium groups and the control group with respect to both weight loss and employment prospects among female participants suggests a causal link between obesity and employment of women. Estimating its magnitude is the aim of the paper.

4 Estimation Strategy

In order to do so, I set up a regression model that explains the change in employment, $E_{i,t} - E_{i,t-1}$, by a percentage change in BMI[14], and a vector of control variables, $X_{i,t-1}$:

$$E_{i,t} - E_{i,t-1} = X'_{i,t-1}\beta + \gamma \left[\left(BMI_{i,t} - BMI_{i,t-1} \right) / BMI_{i,t-1} \right] + \epsilon_{i,t}.$$

Subscript i and t indicate the individual and time, ϵ represents a random error term, while γ and β are coefficient vectors subject to estimation.

The change in weight may suffer from endogeneity due to unobserved heterogeneity, rendering the regression results biased. For instance, individual time preference may be simultaneously correlated with difficulties in timing food intake and task completion on the job. In this case, γ may capture the effect of a high discount rate on employment prospects. Moreover, finding or retaining a job may affect weight loss. A more detailed argumentation of potential sources of bias is provided by Norton and Han (2008).

One way to control for endogeneity and reverse causality is by employing instrumental vari-

[14]Percentage change in BMI levels weight loss achievement across initial BMI values. Using the change in BMI or weight loss in kilograms yields qualitatively the same results.

ables. Intuitively, IV estimation solves this problem by using only the part of the variability in weight loss orthogonal to ϵ to estimate its pure effect on employment. IV estimation requires variables that are correlated with weight loss but otherwise unrelated to employment. I use the two premium group indicators as instruments because they were shown to be significantly correlated with weight loss (Augurzky et al., 2012a) but are uncorrelated with unobserved factors due to the experimental design. Hence, they represent perfect instruments for weight loss.

Theoretically, one would not need to include $X_{i,t-1}$. However, randomization may yield a random imbalance of some covariates in a finite sample (Lock, 2011). Another source of bias may arise due to sample attrition. Table 1 shows that several participants dropped out of the study. If dropping out of the sample is highly correlated with receiving financial incentives for weight loss, IV estimation may yield inconsistent estimates although the direction and magnitude of the bias would not be a priori clear.[15]

The inclusion of covariates may remove both sources of bias because, on the one hand, it eliminates a possible random correlation with the instruments and a potential random correlation of the instruments with unobserved variables on the portion of their association with the covariates (Imai et al., 2008). On the other hand, the covariates control for potential non-random sample attrition under the assumption that the decision to drop out from the experiment is entirely determined by the observable factors. This means that it is required that conditional on $X_{i,t-1}$ sample attrition is purely random. In Section 6, I test this assumption. A further attractive feature of including the regressors is that they reduce the variance of the employment effect of weight loss, hence, increasing the efficiency of the estimate (Vance and Ritter, 2012).

Note that due to the relatively low number of observations, a limited number of covariates is preferably employed in order to guarantee numerical stability of the maximum likelihood estimation. Covariates with a missing value were assigned the value of zero (there are no missing items for the BMI) and a binary indicator controlling for the corresponding observations is included. The sex of six participants is imputed based on height and employment information.[16]

[15] Assessing the possible bias implied by naively comparing observed group means of weight-loss success for the same financial incentive scheme, Augurzky et al. (2012a) prove that both an upward bias and a downward bias may occur. For the present empirical analysis, this implies that the direction of a possible bias due to sample selection is a priori unclear in the instrumental equation of the IV estimation. The direction of a possible attrition bias in the reduced-from, i.e., the numerator of the IV coefficient, is similarly difficult to predict. For instance, it may be that individuals who lose the job are more likely to continue with the experiment due to time constraints that fall away. On the other hand, they may have lost their daily routine which makes it difficult for them to meet the weigh-in deadline and to stay in the experiment, respectively. Hence, the direction of a potential sample-selection bias is a priori unclear for the IV coefficient as a whole. To recall the formula of the IV coefficient, see Footnote 18.

[16] Morris (2006) and Spenkuch (2012) take a similar procedure for dealing with missing values. I 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, 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. All results are qualitatively robust to excluding observations with a missing value for some covariate from the analysis, see Footnote 17.

For all covariates, I use the baseline values, which are unaffected by the experiment. The basic model specification accounts for the rehabilitation clinics, living in a city, time dummy variables indicating the quarter and year of experiment initiation, and basic socio-economic variables. In my preferred model specification, dummy variables indicating educational attainment and self-assessed health are additionally included.

5 Results

First, I briefly discuss OLS results as reference. The upper panel of Table 3 displays that only few covariates have a significant effect. Importantly, the percentage change in BMI is insignificant for both sexes irrespective of which covariate set is considered in the regression. The only significant result among women is that patients treated in the clinic of Bad Mergentheim have better employment prospects than patients treated in the clinic of Bad Kissingen (reference category) which points to the type of disease being relevant for labor market prospects. Among men, only the coefficient of age is significant, pointing to worse employment prospects of older men.

Concerning the IV estimation results, I find a statistically significant negative effect of percentage change in BMI on employment for women in both model specifications. This means that obese women who lose weight improve their employment prospects. For men, in contrast, the negative coefficient of weight loss is insignificant and considerably smaller.[17] Compared to OLS, IV results yields larger coefficients for percentage change in BMI in absolute terms. This is indication that unobserved factors or reversed causation mask the causal effect of weight loss on employment.

Turning to the results for the instrumental equation (lower panel of the tables), I observe significant coefficients for age among both women and men pointing at a lower weight loss success for older participants. The same pattern is observed for women living in urban areas and men with high educational attainment. Irrespective of the sex considered, the instrumental variables exhibit the expected negative sign and generally confirm the results of Augurzky et al. (2012a). Women who are promised €150 for weight loss reduce their body weight by about 3.7 percentage points more than the control group. If they are promised €300 for achieving the target weight they even lose 5.1 percentage points more than the control group. The differential effect of the higher reward is statistically significant. In the regression for men, the coefficients of the instru-

[17] As a robustness test, both model specifications are re-estimated excluding observations with a missing value for some covariate. The estimated effects for women remain virtually unchanged. For men, the results are qualitatively robust, too (Table A2 in the Appendix). Moreover, in Table A3 in the Appendix, I show the robustness of the results with respect to the inclusion of month-year dummy variables and clustering at the clinic-month-year level to allow for unobserved group errors (Donald and Lang, 2007).

Table 3: **IV Effects on Employment Prospects by Gender**

	Women				**Men**			
	OLS		IV		OLS		IV	
Covariate Set	Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended
structural equation: *change in employment*								
change in BMI (in percent)	-0.001	-0.002	-0.026**	-0.026**	-0.005	-0.005	-0.012	-0.010
	(0.005)	(0.005)	(0.013)	(0.012)	(0.003)	(0.003)	(0.013)	(0.014)
age	-0.003	-0.003	-0.001	-0.001	-0.005**	-0.005**	-0.005**	-0.004**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
native	-0.006	0.020	0.011	0.045	-0.039	-0.030	-0.046	-0.036
	(0.056)	(0.059)	(0.059)	(0.060)	(0.051)	(0.050)	(0.054)	(0.053)
married	-0.030	-0.031	-0.034	-0.029	-0.011	-0.011	-0.010	-0.010
	(0.056)	(0.057)	(0.061)	(0.059)	(0.045)	(0.047)	(0.045)	(0.046)
single	-0.036	-0.023	-0.040	-0.015	0.083	0.089	0.088	0.093
	(0.105)	(0.100)	(0.113)	(0.102)	(0.077)	(0.077)	(0.075)	(0.075)
having kids	-0.075	-0.057	-0.097	-0.070	0.099	0.101	0.089	0.093
	(0.062)	(0.061)	(0.076)	(0.068)	(0.063)	(0.062)	(0.066)	(0.065)
city	0.084	0.089	0.116	0.124	0.016	0.022	0.019	0.025
	(0.099)	(0.096)	(0.094)	(0.092)	(0.035)	(0.036)	(0.035)	(0.035)
Bad Mergentheim	0.122*	0.129*	0.142*	0.147**	-0.038	-0.040	-0.044	-0.044
	(0.070)	(0.071)	(0.073)	(0.072)	(0.037)	(0.038)	(0.038)	(0.038)
Glotterbad	0.012	0.011	0.027	0.029	-0.031	-0.022	-0.033	-0.024
	(0.103)	(0.100)	(0.107)	(0.098)	(0.043)	(0.042)	(0.042)	(0.041)
Isny	0.009	0.019	-0.019	-0.006	-0.037	-0.033	-0.045	-0.039
	(0.063)	(0.068)	(0.054)	(0.057)	(0.041)	(0.041)	(0.043)	(0.043)
medium education		-0.024		-0.033		0.035		0.037
		(0.062)		(0.065)		(0.038)		(0.037)
high education		0.083		0.136		-0.021		-0.012
		(0.078)		(0.088)		(0.053)		(0.057)
satisfactory self-assessed health		-0.021		-0.003		-0.003		-0.006
		(0.060)		(0.060)		(0.047)		(0.046)
good self-assessed health		-0.016		0.005		-0.015		-0.018
		(0.059)		(0.057)		(0.044)		(0.042)
constant	0.180	0.137	0.010	-0.068	0.187	0.141	0.144	0.110
	(0.178)	(0.197)	(0.233)	(0.252)	(0.147)	(0.153)	(0.172)	(0.177)
instrumental equation: *change in BMI (in percent)*								
€ 150			-3.834**	-3.652**			-2.489**	-2.521**
			(1.155)	(1.126)			(0.634)	(0.618)
€ 300			-5.002**	-5.134**			-2.412**	-2.359**
			(1.234)	(1.208)			(0.715)	(0.721)
age			0.077*	0.079*			0.062*	0.056*
			(0.044)	(0.041)			(0.032)	(0.033)
native			0.971	1.306			-1.081	-0.998
			(0.997)	(0.867)			(0.846)	(0.827)
married			0.055	0.236			-0.005	0.077
			(0.939)	(0.952)			(0.765)	(0.799)
single			0.229	0.543			0.735	0.715
			(1.704)	(1.554)			(1.075)	(1.097)
having kids			-0.815	-0.588			-1.279	-1.308
			(1.435)	(1.335)			(0.891)	(0.909)
city			1.756*	1.948**			0.644	0.600
			(0.932)	(0.898)			(0.835)	(0.826)
Bad Mergentheim			0.857	0.958			-1.018	-0.897
			(1.023)	(0.990)			(0.641)	(0.625)
Glotterbad			0.071	0.259			-0.555	-0.512
			(1.336)	(1.221)			(1.174)	(1.104)
Isny			-1.412	-1.326			-1.147	-1.065
			(0.943)	(0.985)			(0.799)	(0.767)
medium education				-0.280				0.410
				(0.826)				(0.661)
high education				2.195				1.773*
				(1.409)				(0.918)
satisfactory self-assessed health				0.765				-0.709
				(0.942)				(0.641)
good self-assessed health				0.189				-0.630
				(0.978)				(0.809)
constant			-4.455	-6.002*			-4.521*	-4.085
			(3.382)	(3.086)			(2.441)	(2.569)
number of observations			168				344	

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

mental variables are also significant and slightly stronger than in the regressions for women. The joint tests on instrument relevance yield *F*-statistics of 9.0 (women) and 9.5 (men) which are below the critical threshold of 10 suggested by Staiger and Stock (1997) and, hence, not very strong. However, since they are orthogonal to employment they do not incur a meaningful bias.[18] Weak-instrument robust tests (Finlay and Magnusson, 2009) confirm the finding of weight loss significantly improving employment prospects of women and no effects for obese men. This finding is further warranted by using a joint treatment-group indicator (grouping treatment-group members irrespective of being promised €150 or €300 for weight loss) as single instrument for weight loss, which turns out to be sufficiently strong (*F*-statistics above 16) in the IV regressions (Table A4 in the Appendix).

Regressing the outcome variable directly on the instruments for weight loss (but not weight loss itself, i.e., estimating the reduced form) including the same regressors, yields significant effects on employment prospects for women, too. The test of joint significance of the two group indicators reveal that financial rewards contingent on achieving a weight loss target of between 6 and 8 percent within four months significantly improve employment prospect of obese women. The coefficients of the two rewards do not significantly differ (*p*-value of 17.4 percent). Financially rewarding obese men, in contrast, is not a promising policy option. The results are presented in Columns 1 and 2 of Table A1 in the Appendix.

6 Sample Selection

In order to test the hypothesis of no selection problem (because of the dropouts) in the presence of an endogenous explanatory variable, I follow the test suggested by Wooldridge (2010) which is based on the inverse Mills ratio in a Heckman (1976, 1979) selection model.[19] Sample selection is then a problem if the *t*-statistic for the coefficient of the inverse Mills ratio is significant in an IV regression.

The design of the experiment renders several variables well suited as exclusion restrictions. In particular, I use variables not associated with weight reduction that relate to the distance from the assigned pharmacy ("pharmacy in town") and the availability of cholesterol as well as blood

[18]This is nicely illustrated by the IV formula $\hat{\beta}_{iv} = \frac{z'y}{z'x} = \beta + \frac{z'\epsilon}{z'x}$, where y is the dependent variable, z the instrumental variable, ϵ the error term, and x the vector of covariates; see also Murray (2006).

[19]Wooldridge (2010) suggests running a probit regression of the selection indicator on the instruments of the endogenous explanatory variables, the exogenous explanatory variables, and at least one variable that affects selection but is otherwise uncorrelated with the outcome variable using all observations. Then, the estimated inverse Mills ratio is obtained. As a second step, he proposes to carry out an IV regression of the outcome variable on the exogenous and instrumented endogenous explanatory variables including the estimated inverse Mills ratio as an additional regressor. For the second step, the selected subsample (individuals that did not drop out) is used.

glucose tests in the pharmacy.[20] Being assigned a nearby pharmacy is assumed to be positively correlated with program continuation because time and transportation costs are low. The availability of the respective blood tests may be negatively associated with sample attrition due to gains of being informed, although participants may likewise consider participation in the blood tests as wasting time or fear the prick that is involved with the blood test.

The probit results for the selection equation, where I use a dummy variable for dropping out of the sample as the dependent variable, are presented in Columns 3 and 6 of Table A5 (women) and Table A6 (men) in the Appendix. As expected, being assigned a pharmacy in town significantly reduces the dropout likelihood of women and men. With respect to the blood tests, the availability of a cholesterol test and a blood sugar test is a significant correlate of study continuation for women and men, respectively. For brevity, I do not discuss full IV regression results but mention only the results for the test on the coefficient on the estimated inverse Mills ratio. I find no evidence of a selection bias due to sample attrition among women. In particular, the p-value of the coefficient of the inverse Mills ratio for women is 0.80 (basic covariate set) and 0.73 (extended covariate set). Moreover, the estimated weight-loss effects remain virtually the same across the specifications. For men, the test statistics do not indicate a sample selection problem either.

Further sample selection checks presented in the Appendix focus on the estimation of worst-case reduced-form-effect bounds and show the robustness of the results with respect to accounting for sample selection. This is evidence for sample attrition being of minor relevance for the numerator of the IV coefficient. Since extensive empirical examinations of Augurzky et al. (2012a) point at sample attrition being also of minor relevance for the denominator of the IV coefficient, sample-selection bias represents a minor concern for the IV coefficient as a whole. This is in line with theoretical consideration from the outset and warrants sample-selection-test results.

7 Effect Size

Having established a statistically significant effect which is robust to sample attrition, the question remains whether this effect is of economic relevance. Since the coefficient capturing the effect of the change in body weight on the change in employment is difficult to interpret, I split the sample into participants that were employed and participants that were unemployed at the start of the experiment. While for the former, IV regression estimates yield the effect of weight loss

[20]Note that in the aforementioned 14-week control weigh-in reminder letter, the participants were informed about whether these blood tests were available in the assigned pharmacy. Participants did not have to pay for the measurements in the pharmacy. Moreover, during the weight-loss phase, the participants were unaware of the pharmacy they were assigned to.

on the job retaining probability, for the latter, they can be interpreted as the effect on the job finding probability. Likewise, reduced-form estimates yield the respective effects of the monetary rewards. Results are presented in Table 4. More comprehensive results are presented in Tables A7 (women) and A8 (men) in the Appendix.

The upper panel of the table shows that weight loss significantly (p-value of 6.7 percent) increases the job retaining probability of obese women. In my preferred specification, a one-percentage-point decrease in BMI increases the job retaining probability by 2.1 percentage points compared to a baseline of 75.7 percent. The estimated reduced-form effects indicate that only the higher monetary reward significantly increases the employment probability. The coefficient is significantly higher than the coefficient of the lower reward. Hence, while the reward of €150 does not exert any significant effect, promising obese women that are employed €300 for moderate weight loss increases their job retaining probability by 11.8 percentage points. This is an effect of economically meaningful size.

The effect on the job finding probability of obese unemployed women – though insignificant – tends to be even larger. Particularly large coefficients are found in the reduced-form estimation. The effect of the reward of €150 is even statistically significant (p- value of 5.9 percent), whereas the differential effect of the higher reward is not. However, due to the very small sample size, these estimates have to be interpreted with caution. In the light of the estimates based on the joint female sample, I would therefore consider -2.1 percentage points to be an appropriate conservative estimate for the effect of a percentage change in BMI on the job finding probability of obese women that are unemployed. Regarding the effects for obese men, I neither find significant effects on the job retaining probability of the employed nor on the job finding probability of the unemployed.

8 Potential Mediators of Weight Loss

In this section, I provide further support for the positive effect of weight loss on the employment prospects of obese women. The idea is to look at the effect of weight loss on potential mediators which eventually determine employment. Although the literature also suggests a direct link between weight loss and employment perspectives due to labor-market discrimination, weight loss is supposed to act on employment through positive effects on labor productivity. Thus, finding positive effects of weight loss on proxy variables for labor productivity for women (but not for men) would lead credence to the aforementioned findings.

The dataset contains two suitable measures for labor productivity. One relates to physical

Table 4: **Effects on Employment Prospects by Initial Employment Status**

	Employed		Unemployed	
Covariate Set	Basic	Extended	Basic	Extended
WOMEN				
IV Estimation				
change in BMI (in percent)	-0.021*	-0.021*	-0.067	-0.062
	(0.012)	(0.011)	(0.058)	(0.065)
Reduced-Form Estimation				
€ 150	0.027	0.028	0.334*	0.390*
	(0.070)	(0.073)	(0.173)	(0.207)
€ 300	0.115**	0.118**	0.292	0.388
	(0.056)	(0.057)	(0.181)	(0.262)
number of observations	129	129	39	39
MEN				
IV Estimation				
change in BMI (in percent)	-0.005	-0.006	-0.023	-0.020
	(0.012)	(0.012)	(0.028)	(0.028)
Reduced-Form Estimation				
€ 150	0.012	0.017	0.155	0.181
	(0.031)	(0.031)	(0.157)	(0.156)
€ 300	0.010	0.010	-0.116	-0.153
	(0.032)	(0.032)	(0.160)	(0.150)
number of observations	297	297	48	48

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for the control variables omitted.

well-being and the other to limitations at daily tasks due to health complaints. Physical well-being is thought of as a proxy variable for self-esteem and self-confidence which are considered as contributors to working productivity[21] and is on an ordinal scale, taking on values from 1 (very good) to 5 (bad). Limitations at daily tasks due to health complaints indicate how well individuals may be used for physical work or more generally as a proxy for the health status which is correlated with labor productivity at least through sickness absence. The participants were asked whether any of the following limitations applies: walking quickly or a long distance, lifting heavy objects, bending, kneeling, and stooping. The respondents could also tick that none of the limitations applies. Descriptive statistics of these variables at experiment initiation are displayed in Table A9 in the Appendix.

In order to eliminate random group differentials in the dependent variables at the start of the experiment, the change in the two variables rather than the post-treatment levels enter linear IV regressions as dependent variables. The estimation results are presented in Table 5. For women, the effect of a change in BMI (in percent) on self-assessed physical well-being is negative and significant (p-value of 5.1 percent). This is evidence for weight loss improving self-confidence of obese women. A significant negative effect is found for the change in BMI on limitations at daily

[21]The variable may likewise reflect the health status. However, the specific German formulation of the question in the questionnaire rather relates to how comfortable the participants feel about their bodies. The question in German is: "Wie würden Sie Ihr gegenwärtiges körperliches Wohlbefinden beschreiben?"

Table 5: **Estimated IV Effects on Proxy Variables for Labor Productivity**

	Women		**Men**	
Covariate Set	Basic	Extended	Basic	Extended
Self-assessed Physical Well-Being				
change in BMI (in percent)	-0.031**	-0.030*	-0.032	-0.033
	(0.016)	(0.016)	(0.020)	(0.020)
Limitations at Daily Tasks				
change in BMI (in percent)	-0.060	-0.059*	-0.014	-0.008
	(0.038)	(0.035)	(0.048)	(0.048)
number of observations	151	151	327	327

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Results for the control variables omitted.

tasks due to health complaints (*p*-value of 9.3 percent), i.e., weight loss also improves physical fitness of women. For men, no significant effects are found, neither on physical well-being nor on the limitations at daily tasks. Thus, the effect on both proxy variables for labor productivity supports the idea that weight loss induces labor productivity gains among obese women, while for men this seems not to be the case. The absence of productivity gains of losing weight for male participants is in line with the results of no effects on their probability to be in employment. Giving a sound explanation of the gender-specific heterogeneity in productivity returns to weight loss is beyond the scope of the paper.

Since the effects of weight loss on self-assessed physical well-being are not statistically different across both sexes, results for the effects on the mediators point at improvements in physical health being the major channel through which weight loss affects employment. This is in line with Lundborg et al. (2010) who find that physical fitness explains the major part of the obesity-wage penalty. In turn, I explain the gender heterogeneity in the employment effects of weight loss by weight loss exclusively improving physical health of obese women. However, it could likewise be that (physical) self-confidence is a characteristic that translates into better employment prospects in obese women but not in obese men due to gender differences in the types of occupation and, hence, occupational requirements. Moreover, as found by Rooth (2009), (statistical) discrimination of employers is predominantly directed against female workers and applicants, respectively.

9 Discussion and Concluding Remarks

This study presents credible estimates for the causal effect of a variation in obesity on employment. By exploring a unique exogenous source of weight loss, the analysis improves upon the existing empirical literature on this topic. In a randomized controlled experiment, obese rehabilitation patients were successfully motivated to lose weight by monetary rewards of €150 and

€ 300, respectively. This produced ideal instrumental variables because random assignment assured that the monetary rewards exclusively operated on the probability to be in employment through the BMI, dispelling any concerns about their validity.

I find convincing evidence for weight loss positively affecting the employment prospects of obese women but not of obese men. This result is robust across different model specifications and with respect to accounting for sample attrition. Considering obese women that were employed at experiment initiation alone, a one-percentage-point reduction in the BMI within four months increases the job retaining probability by 2.1 percentage points compared to a baseline of 75.7 percent. The effect on the job finding probability of obese unemployed women tends to be even larger, however due to the small number of respective observations, it has to be interpreted with caution. I explain the heterogeneity in the effects across gender by positive health effects of weight loss, which I only find for women. The great importance of physical health for labor productivity is unquestioned and applies even more to the study population consisting of medical rehabilitation patients with symptoms attributable to obesity. For the same reason, the estimates may be regarded as upper-bound effects.

The results fit well with previous findings for Germany. Caliendo and Lee (2011) report an employment probability differential between obese and healthy weight unemployed German women who are currently searching for a job of 13 percentage points. For obese unemployed German men, they find no effect on employment prospects. Analysis for other countries yield even stronger estimated effects. For instance, Morris (2007) reports that obese women in the UK are 23 percentage points less likely to be in employment than normal weight women. Contrary to my results, he finds a significant negative effect for obese men which, nevertheless, is lower than the respective effect for obese women. Lindeboom et al. (2010), in contrast, do not find any evidence of labor market effects of obesity for the UK. For the US, several research articles find either any significant link between obesity and employment irrespective of gender. These studies do not necessarily conflict with the results presented here. The divergence may be explained by the employed instruments estimating treatment effects for different compliant subpopulations. It may likewise reflect that the employment effects of obesity vary across labor market regulations.

The paper also presents the results for the reduced form which yields the causal employment effect of financially rewarding the obese for weight loss. I find that the two monetary rewards significantly improved the employment prospects of obese women. This shows that the weight-loss intervention had indirect economic benefits that give a strong argument in favor of its usage. Furthermore, it proves that the employment prospects of obese women may be effectively improved by policy interventions. My estimates can serve as a benchmark for labor market programs that specifically target the obese. Further research should question how long the estimated employment effects for obese women persist over time.

A rough estimate (Appendix) of the fiscal effects of the intervention yields that the reward of €300 has already a benefit-to-cost ratio larger than one after 40 and 46 days of effect persistence depending on the underlying scenario. This means that the reward is already amortized by its fiscal effects alone if the employment effects persist over slightly less than one and a half months. Since the effects of the reward of €150 are mostly insignificant, it is outperformed by the reward of €300. Assuming that the employment effects persist over one year, financially rewarding an obese woman for achieving a weight-loss target by €300 causes average expected fiscal net savings of between €2,106 ($2,639 in PPP) and €2,417 ($3,028 in PPP).

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.

Augurzky, B., Bauer, T. K., Reichert, A. R., Schmidt, C. M. and Tauchmann, H. (2012a). Does Money Burn Fat? Evidence from a Randomized Experiment, *IZA Discussion Papers 6888*, Institute for the Study of Labor (IZA).

Augurzky, B., Reichert, A. R. and Schmidt, C. M. (2012b). The effect of a bonus program for preventive health behavior on health expenditures, *Ruhr Economic Papers 373*, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI).

Averett, S. and Korenman, S. (1996). The economic reality of the beauty myth, *Journal of Human Resources* **31**(2): 304–330.

Bauer, T. K., Kluve, J., Schaffner, S. and Schmidt, C. M. (2009). Fiscal Effects of Minimum Wages: An Analysis for Germany, *German Economic Review* **10**(2): 224–242.

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 Bundorf, M. K. (2009). The incidence of the healthcare costs of obesity, *Journal of Health Economics* **28**(3): 649–658.

Biddle, J. E. and Hamermesh, D. S. (1998). Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre, *Journal of Labor Economics* **16**(1): 172–201.

Blackburn, G. (1995). Effect of degree of weight loss on health benefits, *Obesity research* **3**(Suppl. 2): 211–216.

Blau, F. D., Blau, F. D., Kahn, L. M. and Kahn, L. M. (2001). Comparative Analysis of Labor Market Outcomes: Lessons for the US from International Long-Run Evidence, *in* A. B. Krueger and R. Solow (eds), *The Roaring Nineties: Can Full Employment Be Sustained?*, Russell Sage Foundation, pp. 159–218.

Borghans, L., Duckworth, A. L., Heckman, J. J. and ter Weel, B. (2008). The economics and psychology of personality traits, *Journal of Human Resources* **43**(4): 972–1059.

Bowles, S., Gintis, H. and Osborne, M. (2001). The determinants of earnings: A behavioral approach, *Journal of Economic Literature* **39**(4): 1137–1176.

Brown, C., Goetz, J. and Hamera, E. (2011). Weight Loss Intervention for People With Serious Mental Illness: A Randomized Controlled Trial of the RENEW Program, *Psychiatric Services* **62**(7): 800–802.

Caliendo, M. and Lee, W.-S. (2011). Fat Chance! Obesity and the Transition from Unemployment to Employment, *IZA Discussion Papers 5795*, Institute for the Study of Labor (IZA).

Cawley, J. (2004). The impact of obesity on wages, *Journal of Human Resources* **39**(2): 451–474.

Cawley, J., Han, E. and Norton, E. C. (2009). Obesity and labor market outcomes among legal immigrants to the United States from developing countries, *Economics & Human Biology* **7**(2): 153–164.

Charness, G. and Gneezy, U. (2009). Incentives to exercise, *Econometrica* **77**(3): 909–931.

Christensen, R., Bartels, E. M., Astrup, A. and Bliddal, H. (2007). Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis, **66**(4): 433–439.

Contoyannis, P. and Jones, A. M. (2004). Socio-economic status, health and lifestyle, *Journal of Health Economics* **23**(5): 965–995.

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.

Donald, S. G. and Lang, K. (2007). Inference with Difference-in-Differences and Other Panel Data, *The Review of Economics and Statistics* **89**(2): 221–233.

Everett, M. (1990). Let an overweight person call on your best customer? Fat chance, *Sales and Marketing Management* **142**(4): 66–70.

Finlay, K. and Magnusson, L. M. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models, *Stata Journal* **9**(3): 398–421.

Hamermesh, D. S. and Biddle, J. E. (1994). Beauty and the labor market, *American Economic Review* **84**(5): 1174–1194.

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.

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. and Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program, *American Economic Review* **91**(2): 145–149.

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.

Hooper, M. M., Stellato, T. A., Hallowell, P. T., Seitz, B. A. and Moskowitz, R. W. (2007). Musculoskeletal findings in obese subjects before and after weight loss following bariatric surgery, *International Journal of Obesity* **31**(1): 114–120.

Imai, K., King, G. and Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference, *Journal of the Royal Statistical Sociecty, Series A (Statistics in Society)* **171**(2): 481–502.

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.

Kenkel, D. S. (1995). Should you eat breakfast? Estimates from health production functions, *Health Economics* **4**(1): 15–29.

Klesges, R., Klem, M., Hanson, C., Eck, L., Ernst, J., O'Laughlin, D., Garrott, A. and Rife, R. (1990). The effects of applicant's health status and qualifications on simulated hiring decisions, *International Journal of Obesity* **14**(6): 527–535.

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. and Davey Smith, G. (2008a). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology, *Statistics in Medicine* **27**(8): 1133–1163.

Lawlor, D. A., Windmeijer, F. and Smith, G. D. (2008b). Is Mendelian randomization 'lost in translation?': comments on 'Mendelian randomization equals instrumental variable analysis with genetic instruments' by Wehby et al., *Statistics in Medicine* **27**(15): 2750–2755.

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

Lindeboom, M., Lundborg, P. and van der Klaauw, B. (2010). Assessing the impact of obesity on labor market outcomes, *Economics and Human Biology* **8**(3): 309–319.

Lock, K. F. (2011). *Rerandomization to Improve Covariate Balance in Randomized Experiments*, Harvard University, Cambridge, MA.

Lundborg, P., Nystedt, P. and Rooth, D.-O. (2010). No Country for Fat Men? Obesity, Earnings, Skills, and Health among 450,000 Swedish Men, *IZA Discussion Papers 4775*, Institute for the Study of Labor (IZA).

Mobius, M. M. and Rosenblat, T. S. (2006). Why beauty matters, *American Economic Review* **96**(1): 222–235.

Mocan, N. and Tekin, E. (2011). Obesity, self-esteem and wages, *in* M. Grossmann and N. H. Mocan (eds), *Economic Aspects of Obesity*, 1 edn, University of Chicago Press, chapter 12, pp. 349–380.

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.

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments, *Journal of Economic Perspectives* **20**(4): 111–132.

Norton, E. C. and Han, E. (2008). Genetic information, obesity, and labor market outcomes, *Health Economics* **17**(9): 1089–1104.

O'Brien, K. S., Latner, J. D., Ebneter, D. and Hunter, J. A. (2012). Obesity discrimination: the role of physical appearance, personal ideology, and anti-fat prejudice, *International Journal of Obesity* **URL**. http://dx.doi.org/10.1038/ijo.2012.52.

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

Pagán, J. A. and Dávila, A. (1997). Obesity, occupational attainment, and earnings, *Social Science Quarterly* **78**(3): 756–770.

Prentice, A. M. (2006). The emerging epidemic of obesity in developing countries, *International Journal of Epidemiology* **35**(1): 39–99.

Puhl, R. and Brownell, K. D. (2001). Bias, discrimination, and obesity, *Obesity Research* **9**(12): 788–805.

Register, C. A. and Williams, D. R. (1990). Wage effects of obesity among young workers, *Social Science Quarterly* **71**(1): 130–141.

Roehling, M. V., Roehling, P. V. and Pichler, S. (2007). The relationship between body weight and perceived weight-related employment discrimination: The role of sex and race, *Journal of Vocational Behavior* **71**(2): 300–318.

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring, *Journal of Human Resources* **44**(3): 710–735.

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

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments, *Econometrica* **65**(3): 557–586.

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/m^2) 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.

Vance, C. and Ritter, N. (2012). The phantom menace of omitted variables - a comment, *Conflict Management and Peace Science* **29**(2): 233–238.

Wehby, G. L., Ohsfeldt, R. L. and Murray, J. C. (2008). 'Mendelian randomization' equals instrumental variable analysis with genetic instruments, *Statistics in Medicine* **27**(15): 2745–2749.

Wing, R. R., Lang, W., Wadden, T. A., Safford, M., Knowler, W. C., Bertoni, A. G., Hill, J. O., Brancati, F. L., Peters, A. and Wagenknecht, L. (2011). Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes., *Diabetes Care* **34**(7): 1481–1486.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2 edn, The MIT Press.

Appendix

A.1 Worst-Case Reduced-Form Effect Bounds

For the estimation of worst case reduced form effect bounds, I follow two different approaches which are both employed and described in more detail in Augurzky et al. (2012a).

First, I carry out an "intention-to-treat" analysis by regressing the employment indicator on the exogenous explanatory variables and the two IVs aiming to use all observations irrespective of whether they drop out of the experiment. Since there is no information about employment after the weight loss period for dropouts, missing data have to be imputed. There exists no consensus on one particular imputation algorithm (Hollis and Campbell, 1999). Therefore, I follow the most common procedure which assumes that the outcome variable takes on previous values, i.e., the values before the weight loss phase. By doing so, it is implicitly assumed that the outcome remains unaffected (John et al., 2011), implying a zero reduced form effect for dropouts. In technical terms, this approach substantially adjusts the raw group differential towards zero. The analysis yields reduced-form effects for women that are very similar – though slightly smaller – to those obtained for the selected subsample in terms of size and significance. This is displayed in the first four columns of Table A1. Regarding men, the intention-to-treat results confirm the absence of effects. Thus, for both sexes, the intention-to-treat analysis substantiates previous sample selection test results.

As a second approach, I consider bounds that are derived by Lee (2009) in the context of randomized experiments. The method consists of trimming the experimental group that suffers the least from sample attrition at the quantile of the outcome variables that corresponds to the share of "excess observations". The procedure yields upper and lower reduced-form-effect bounds which differ with respect to whether the trimming is from below or from above. Here, tightened Lee-bounds based on splitting the sample into cells defined by covariates are estimated. The results are presented in the last two columns of Table A1. The estimated lower bound for the group with the €150 reward includes a zero effect for women confirming OLS results (first two columns of the table). The respective bound for female members receiving the €300 rewards contingent on the achievement of the target weight does not include a zero effect despite extreme assumptions about the bias of the treatment effect. Even the confidence interval excludes negative values, wherefore, I can conclude that the significance of the effect of the reward of €300 is robust with respect to sample attrition. For male participants of the incentive groups, the lower bounds for both incentive groups point at zero treatment effects which is also in line with previous results.

Table A1: **Reduced-Form Effects on Employment Prospects**

			Sample-Selection Checks			
	Linear Regression		Intention-to-Treat		Lee (2009)-Bounds◇	
	Basic Set	Extended Set	Basic Set	Extended Set	Lower	Upper
Women						
€ 150	0.057	0.062	0.045	0.046	0.054	0.121*
	(0.066)	(0.068)	(0.049)	(0.050)	[-0.078	0.234]
€ 300	0.140**	0.136**	0.121**	0.116**	0.110*	0.186**
	(0.057)	(0.057)	(0.044)	(0.042)	[0.013	0.295]
number of observations	168	168	222	222	222	222
Men						
€ 150	0.043	0.040	0.026	0.028	0.035	0.053
	(0.039)	(0.039)	(0.028)	(0.028)	[-0.034	0.132]
€ 300	0.013	0.006	0.011	0.010	0.008	0.013
	(0.037)	(0.037)	(0.028)	(0.028)	[-0.073	0.116]
number of observations	344	344	468	468	468	468

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for the control variables omitted. ◇For tightening bounds, the set of covariates is selected from the variables 'female', indicators for the clinics, 'pharmacy in town', 'cholesterol test', and 'blood glucose test', which yields the narrowest confidence interval for the ATE.

A.2 Fiscal Effects of the Weight-Loss Intervention

In order to obtain a rough and conservative estimate of the fiscal effects of financially rewarding obese German women for achieving a set weight-loss target, I assume that workers who are prevented from losing their job and unemployed who gain employment receive an average gross labor income of € 1,500 (Scenario A) and € 2,000 (Scenario B), which corresponds to $1,879 and $2,506 in terms of PPP.[22] Following the assumptions of Bauer et al. (2009), they pay an income tax of 3.4 percent and a solidarity tax of 5.5 percent of the income tax. The social security contributions add up to 40.35 percent of gross income, consisting of 19.9 percent for the pension system, 3 percent for the unemployment insurance, 15.5 percent for social health insurance, and 1.95 percent for long-term care insurance.[23] In total, monthly fiscal revenues per worker are € 659 ($826 in PPP) in Scenario A and € 879 ($1,101 in PPP) in Scenario B. Further fiscal effects relate to public expenditures for unemployment benefits and active labor market programs which have to be paid if obese women had become or remained unemployed. Assuming that they received the social benefits (and not the higher unemployment benefits), the annual public expenditures would total € 12,480 ($15,637 in PPP) per women.

Regarding the effects of the monetary rewards on employment, I entirely rely on the estimates for the job retaining probability in order to keep calculations simple (Table 4). This means that the estimates for the group of obese employed women are applied to the group of obese unem-

[22]The dataset provides information on the household gross income. The average household gross income of single-income households amounts to about € 2000. However, this information is not very reliable due to implausible responses or non-response.

[23]The slight deviations to Bauer et al. (2009) are due to several reforms that took place in the meantime.

ployed woman, irrespective of the fact that the effects on their job finding probability tend to be larger. According to early findings, the lower reward of €150 exerts no effect on the employment probability, i.e., monthly fiscal effects of the reward of €150 are €0, which is a conservative consideration. For the higher reward of €300, I take an employment effect of 11.8 percentage points which yields expected monthly fiscal effects of about €200 per woman in Scenario A and €226 in Secenario B. Hence, while the reward of €150 will never amortize based on fiscal effects alone, the higher reward amortizes during the second month. Depending on the scenario, it has a benefit-to-cost ratio of 1.34 and 1.51 at the end of the second month after the weight-loss intervention. If the employment effect of the reward of €300 persists over twelve months, the expected net fiscal effect lies between €2,106 ($2,639 in PPP) and €2,417 ($3,028 in PPP) per incentivized obese woman.

A.3 Figures and Tables

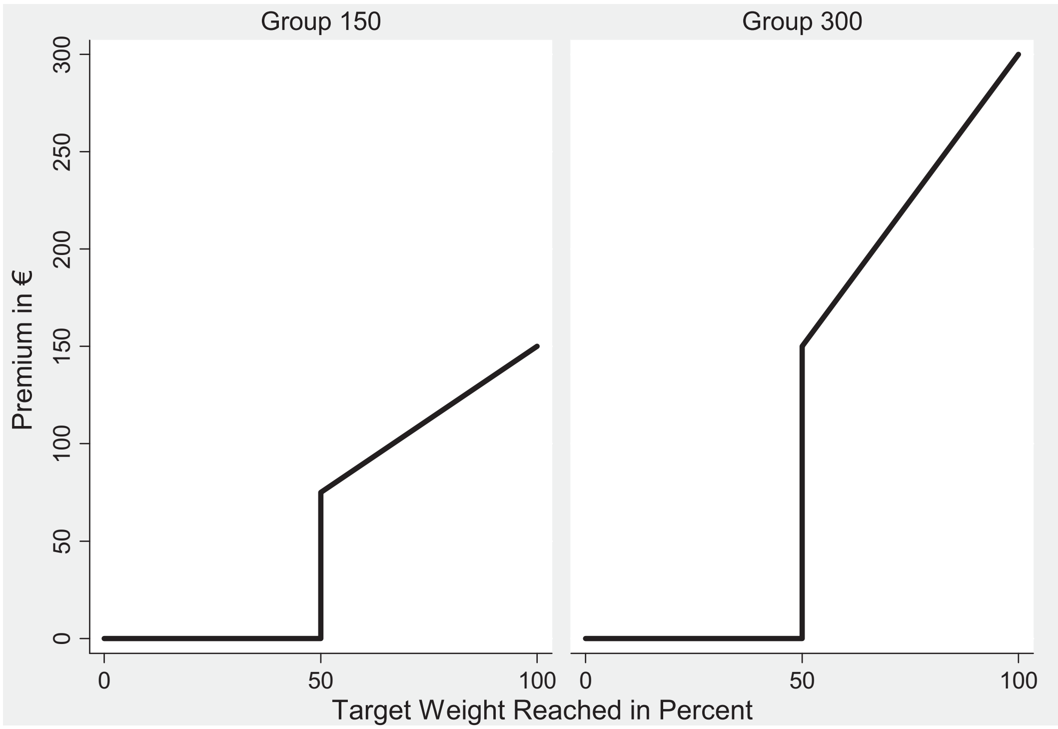

Figure A1: **Payment Mechanism**

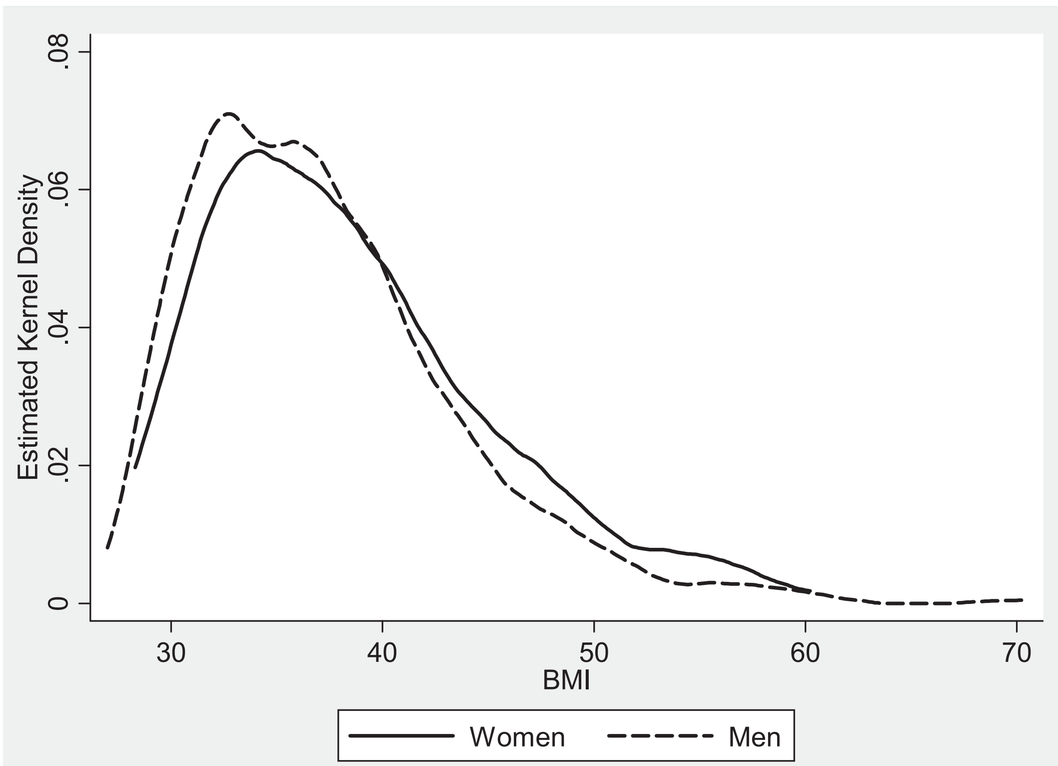

Figure A2: **Distribution of the BMI at the Start of Weight-Loss Phase by Gender**

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).

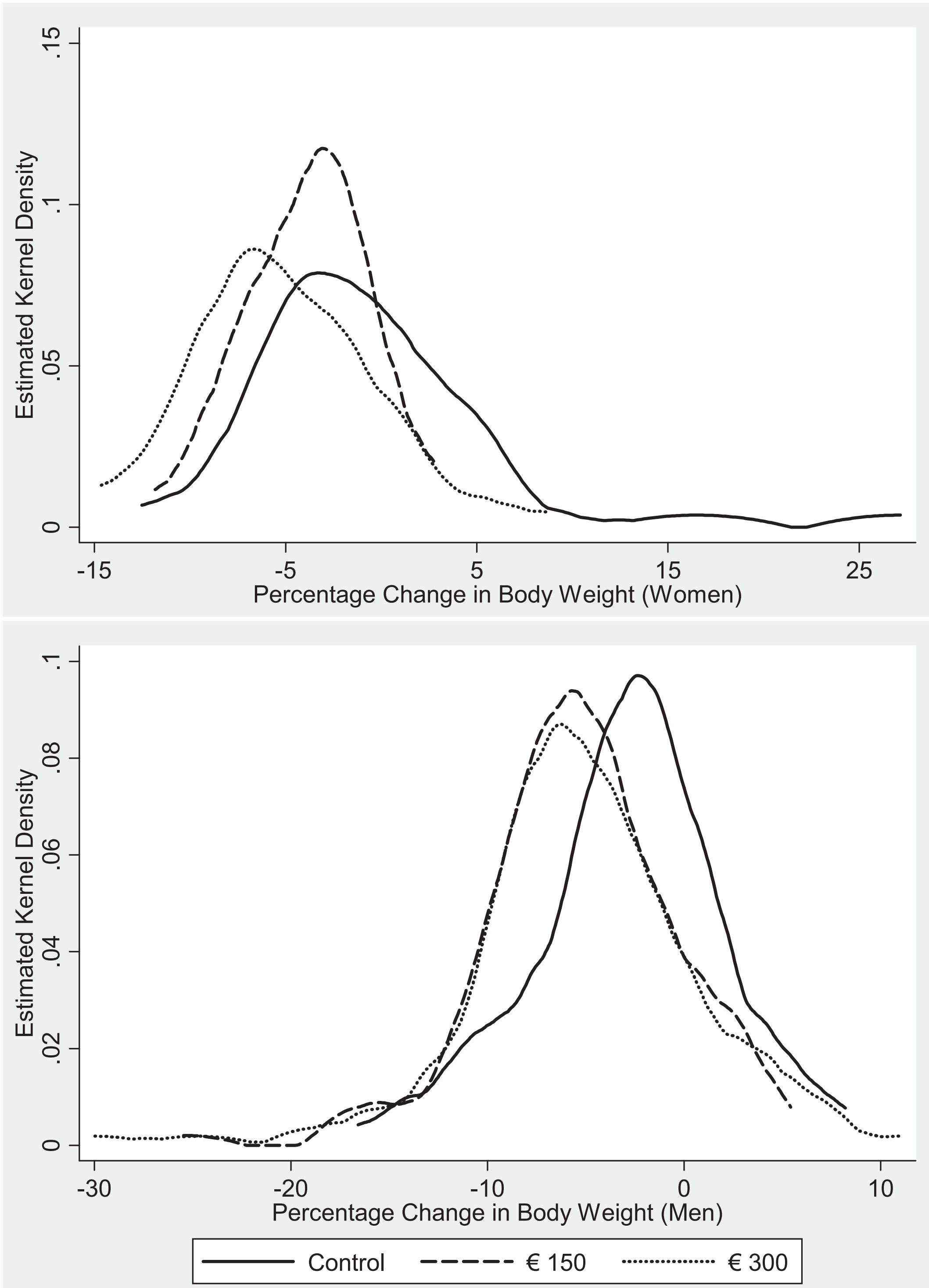

Figure A3: **Distribution of Change in Body Weight by Experimental Groups and Gender**

Table A2: **IV Effects on Employment Prospects: Missing Observations Excluded**

	Women		Men	
Covariate Set	Basic	Extended	Basic	Extended
structural equation: *change in employment*				
change in BMI (in percent)	-0.029**	-0.026**	-0.009	-0.006
	(0.014)	(0.013)	(0.013)	(0.013)
age	-0.001	-0.004	-0.007**	-0.006**
	(0.004)	(0.004)	(0.003)	(0.003)
native	0.005	0.020	-0.062	-0.054
	(0.061)	(0.064)	(0.059)	(0.057)
married	-0.020	0.000	-0.007	0.012
	(0.063)	(0.065)	(0.046)	(0.046)
single	-0.043	-0.060	0.060	0.074
	(0.120)	(0.103)	(0.076)	(0.072)
having kids	-0.095	-0.045	0.074	0.074
	(0.085)	(0.077)	(0.063)	(0.060)
city	0.121	0.139	0.015	0.023
	(0.096)	(0.095)	(0.035)	(0.036)
Bad Mergentheim	0.164**	0.174**	-0.036	-0.038
	(0.082)	(0.082)	(0.038)	(0.038)
Glotterbad	0.023	-0.030	-0.035	-0.028
	(0.112)	(0.096)	(0.044)	(0.042)
Isny	-0.021	-0.007	-0.043	-0.048
	(0.056)	(0.062)	(0.042)	(0.043)
medium education		-0.036		0.037
		(0.071)		(0.037)
high education		0.120		-0.010
		(0.094)		(0.055)
satisfactory self-assessed health		0.012		-0.007
		(0.061)		(0.044)
good self-assessed health		0.010		-0.006
		(0.063)		(0.046)
constant	0.016	0.094	0.275	0.226
	(0.269)	(0.255)	(0.182)	(0.195)
instrumental equation: *change in BMI (in percent)*				
€ 150	-3.670**	-3.342**	-2.446**	-2.242**
	(1.207)	(1.182)	(0.640)	(0.622)
€ 300	-4.849**	-5.068**	-2.667**	-2.821**
	(1.278)	(1.302)	(0.727)	(0.738)
age	0.082	0.072	0.055	0.059
	(0.053)	(0.051)	(0.040)	(0.042)
native	0.711	0.672	-1.587*	-0.980
	(1.027)	(0.885)	(0.867)	(0.834)
married	0.343	0.809	0.310	0.367
	(0.970)	(0.976)	(0.769)	(0.807)
single	-0.054	0.352	0.929	0.890
	(1.743)	(1.684)	(1.080)	(1.155)
having kids	-0.822	-0.611	-1.301	-1.393
	(1.561)	(1.496)	(0.896)	(0.944)
city	1.862**	2.083**	0.670	0.601
	(0.935)	(0.952)	(0.839)	(0.855)
Bad Mergentheim	1.612	1.957*	-1.106*	-0.976
	(1.074)	(1.051)	(0.651)	(0.638)
Glotterbad	0.250	0.571	-0.501	-0.334
	(1.373)	(1.326)	(1.211)	(1.182)
Isny	-1.056	-0.939	-0.980	-0.729
	(0.966)	(1.026)	(0.775)	(0.797)
medium education		-0.214		0.560
		(0.849)		(0.667)
high education		1.688		1.819*
		(1.428)		(0.930)
good self-assessed health		0.187		-0.879
		(0.989)		(0.821)
satisfactory self-assessed health		0.693		-0.947
		(0.956)		(0.652)
constant	-4.851	-6.110*	-4.133	-4.859*
	(3.674)	(3.527)	(2.834)	(2.939)
number of observations	157	152	331	319

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parenthesis. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A3: **IV Effects on Employment Prospects: Month-Year Dummies**

	Women		Men	
Covariate Set	Basic	Extended	Basic	Extended
structural equation: *change in employment*				
change in BMI (in percent)	-0.026*	-0.027*	-0.015	-0.013
	(-0.014)	(-0.014)	(-0.014)	(0.014)
age	-0.002	-0.002	-0.004**	-0.004**
	(-0.004)	(-0.004)	(-0.002)	(0.002)
native	0.007	0.030	-0.054	-0.045
	(-0.055)	(-0.061)	(-0.061)	(0.060)
married	-0.050	-0.038	-0.027	-0.026
	(-0.058)	(-0.055)	(-0.036)	(0.038)
single	-0.056	-0.029	0.104	0.108
	(-0.101)	(-0.086)	(-0.071)	(0.071)
having kids	-0.078	-0.046	0.108*	0.113*
	(-0.078)	(-0.066)	(-0.061)	(0.061)
city	0.125*	0.140*	0.016	0.020
	(-0.071)	(-0.075)	(-0.037)	(0.035)
Bad Mergentheim	0.157**	0.155**	-0.052*	-0.051*
	(-0.069)	(-0.069)	(-0.027)	(0.026)
Glotterbad	-0.003	0.011	-0.009	-0.002
	(-0.092)	(-0.092)	(-0.032)	(0.029)
Isny	-0.021	-0.012	-0.039	-0.032
	(-0.063)	(-0.067)	(-0.034)	(0.037)
medium education		-0.072		0.040
		(-0.077)		(0.038)
high education		0.099		-0.001
		(-0.083)		(0.048)
good self-assessed health		0.011		-0.013
		(-0.073)		(0.037)
satisfactory self-assessed health		0.018		-0.025
		(-0.052)		(0.033)
instrumental equation: *change in BMI (in percent)*				
€ 150	-3.479**	-3.193**	-2.397**	-2.432**
	(-1.139)	(-1.184)	(-0.654)	(0.657)
€ 300	-4.267**	-4.375**	-2.324**	-2.271**
age	0.058	0.059	0.065*	0.059*
	(-0.039)	(-0.042)	(-0.035)	(0.035)
native	0.740	1.197	-1.323	-1.240
	(-1.016)	(-1.008)	(-0.821)	(0.787)
married	0.411	0.601	-0.482	-0.405
	(-0.857)	(-0.861)	(-0.588)	(0.600)
single	0.061	0.488	0.703	0.703
	(-1.959)	(-1.730)	(-1.181)	(1.179)
having kids	-1.104	-0.637	-0.831	-0.831
	(-1.540)	(-1.364)	(-0.952)	(1.012)
city	1.511*	1.706**	0.657	0.599
	(-0.904)	(-0.849)	(-0.912)	(0.888)
Bad Mergentheim	0.833	0.891	-0.927*	-0.808*
	(-1.058)	(-1.039)	(-0.508)	(0.481)
Glotterbad	-0.364	-0.080	0.010	0.025
	(-1.100)	(-1.057)	(-1.290)	(1.179)
Isny	-1.801*	-1.536	-0.911	-0.855
	(-0.927)	(-0.926)	(-0.721)	(0.623)
medium education		-0.738		0.553
		(-0.743)		(0.833)
high education		2.078		1.883**
		(-1.464)		(0.940)
good self-assessed health		-0.074		-0.635
		(-1.126)		(0.778)
satisfactory self-assessed health		0.194		-0.850
		(-1.017)		(0.653)
contant	-6.234**	-6.634**	-3.951*	-3.760*
	(-2.676)	(-3.054)	(-2.183)	(2.165)
number of observations	168	168	334	334

Notes: ** significant at 5% * significant at 10%. Standard errors clustered at the clinic-month-year level in parenthesis. Missing values set to zero. 'Missing' indicator included. Results for month-year dummy variables not reported. Constant (and month-year dummy variables) partialled out in the structural equation. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A4: **IV Effects on Employment Prospects by Gender: Joint Treatment Group**

	Women		Men	
Covariate Set	Basic	Extended	Basic	Extended
structural equation: *change in employment*				
change in BMI (in percent)	-0.024*	-0.024*	-0.011	-0.009
	(0.013)	(0.013)	(0.013)	(0.014)
age	-0.001	-0.001	-0.005**	-0.004**
	(0.003)	(0.003)	(0.002)	(0.002)
native	0.009	0.043	-0.046	-0.035
	(0.059)	(0.061)	(0.054)	(0.053)
married	-0.033	-0.029	-0.010	-0.010
	(0.060)	(0.058)	(0.045)	(0.046)
single	-0.039	-0.016	0.088	0.092
	(0.111)	(0.100)	(0.075)	(0.075)
having kids	-0.095	-0.068	0.090	0.094
	(0.073)	(0.066)	(0.066)	(0.065)
city	0.112	0.121	0.019	0.024
	(0.095)	(0.093)	(0.035)	(0.035)
Bad Mergentheim	0.140*	0.145**	-0.043	-0.044
	(0.072)	(0.071)	(0.037)	(0.038)
Glotterbad	0.025	0.027	-0.033	-0.024
	(0.105)	(0.097)	(0.042)	(0.041)
Isny	-0.016	-0.003	-0.044	-0.038
	(0.055)	(0.059)	(0.043)	(0.042)
medium education		-0.032		0.037
		(0.064)		(0.037)
high education		0.131		-0.013
		(0.087)		(0.057)
satisfactory self-assessed health		-0.004		-0.005
		(0.059)		(0.045)
good self-assessed health		0.003		-0.018
		(0.057)		(0.042)
constant	0.029	-0.048	0.146	0.113
	(0.231)	(0.256)	(0.172)	(0.176)
instrumental equation: *change in BMI (in percent)*				
€ 150 or € 300	-4.542**	-4.569**	-2.451**	-2.439**
	(1.135)	(1.123)	(0.577)	(0.572)
age	0.083*	0.090**	0.061*	0.055*
	(0.043)	(0.040)	(0.032)	(0.033)
native	1.016	1.409	-1.067	-0.967
	(1.000)	(0.871)	(0.855)	(0.831)
married	0.007	0.163	-0.008	0.070
	(0.943)	(0.962)	(0.765)	(0.800)
single	0.330	0.646	0.728	0.697
	(1.705)	(1.572)	(1.076)	(1.097)
having kids	-0.617	-0.400	-1.285	-1.320
	(1.423)	(1.336)	(0.881)	(0.900)
city	1.661*	1.824**	0.651	0.612
	(0.920)	(0.898)	(0.841)	(0.829)
Bad Mergentheim	0.743	0.798	-1.022	-0.905
	(0.998)	(0.980)	(0.637)	(0.621)
Glotterbad	-0.034	0.094	-0.556	-0.516
	(1.315)	(1.213)	(1.176)	(1.110)
Isny	-1.549*	-1.531	-1.151	-1.075
	(0.935)	(0.991)	(0.793)	(0.759)
medium education		0.030		0.410
		(0.842)		(0.660)
high education		2.215		1.776*
		(1.416)		(0.916)
satisfactory self-assessed health		0.258		-0.621
		(0.997)		(0.815)
good self-assessed health		0.750		-0.700
		(0.944)		(0.645)
constant	-4.715	-6.568**	-4.519*	-4.083
	(3.361)	(3.103)	(2.436)	(2.564)
number of observations	168	168	334	334

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A5: **Sample Selection Test for Women (Wooldridge, 2010)**

Covariate	Basic Set			Extended Set		
	Structural Equation	Instrumental Equation	Selection Equation	Structural Equation	Instrumental Equation	Selection Equation
Endogenous regressor						
change in BMI (in percent)	-0.029			-0.030		
	(0.019)			(0.019)		
Instruments						
€ 150		-3.691**	-0.055		-3.566**	-0.060
		(1.104)	(0.057)		(1.075)	(0.056)
€ 300		-4.367**	-0.197**		-4.781**	-0.210**
		(1.261)	(0.053)		(1.328)	(0.054)
Exclusion restrictions						
cholesterol test			0.125**			0.122**
			(0.057)			(0.058)
blood glucose test			-0.104			-0.095
			(0.076)			(0.074)
pharmacy in town			-0.126**			-0.126**
			(0.046)			(0.046)
Controls						
age	0.00007	0.109	-0.010**	0.001	0.096	-0.010**
	(0.006)	(0.078)	(0.002)	(0.006)	(0.079)	(0.002)
native	0.010	0.865	0.020	0.043	1.215	0.029
	(0.059)	(0.989)	(0.056)	(0.060)	(0.901)	(0.057)
married	-0.033	0.046	0.008	-0.027	0.231	0.010
	(0.062)	(0.940)	(0.057)	(0.062)	(0.949)	(0.058)
single	-0.044	0.072	0.033	-0.020	0.461	0.030
	(0.117)	(1.696)	(0.099)	(0.106)	(1.571)	(0.094)
having kids	-0.117	-1.415	0.147*	-0.097	-0.911	0.152*
	(0.111)	(1.899)	(0.085)	(0.104)	(1.895)	(0.081)
city	0.104	1.164	0.203**	0.112	1.652	0.199**
	(0.101)	(1.428)	(0.051)	(0.097)	(1.382)	(0.050)
Bad Mergentheim	0.143*	0.775	0.018	0.148**	0.903	0.024
	(0.075)	(0.983)	(0.056)	(0.075)	(0.953)	(0.055)
Glotterbad	0.030	0.134	-0.037	0.034	0.294	-0.032
	(0.110)	(1.359)	(0.077)	(0.102)	(1.248)	(0.080)
Isny	-0.002	-0.726	-0.254**	0.015	-0.980	-0.251**
	(0.076)	(1.537)	(0.086)	(0.075)	(1.480)	(0.087)
medium education				-0.028	-0.154	-0.084
				(0.071)	(0.926)	(0.057)
high education				0.146	2.203	-0.036
				(0.096)	(1.416)	(0.086)
satisfactory self-assessed health				0.016	0.866	-0.040
				(0.073)	(1.154)	(0.061)
good self-assessed health				0.007	0.313	-0.064
				(0.067)	(1.113)	(0.064)
inverse Mill's ratio (λ)	-0.025	-0.911		-0.033	-0.459	
	(0.098)	(1.612)		(0.094)	(1.589)	
constant	0.013	-4.151		-0.087	-5.952*	
	(0.241)	(3.331)		(0.284)	(3.066)	
number of observations	168	168	222	168	168	222

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parenthesis. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A6: **Sample Selection Test for Men (Wooldridge, 2010)**

Covariate	Basic Set			Extended Set		
	Structural Equation	Instrumental Equation	Selection Equation	Structural Equation	Instrumental Equation	Selection Equation
Endogenous regressor						
change in BMI (in percent)	-0.006			-0.013		
	(0.013)			(0.019)		
Instruments						
€ 150		-3.103**	-0.062		-3.404**	-0.056
		(0.736)	(0.045)		(0.982)	(0.044)
€ 300		-4.541**	-0.186**		-4.188**	-0.183**
		(1.368)	(0.048)		(1.728)	(0.047)
Exclusion restrictions						
cholesterol test			0.010			0.015
			(0.046)			(0.046)
blood glucose test			-0.099*			-0.112*
			(0.058)			(0.058)
pharmacy in town			-0.083**			-0.082**
			(0.040)			(0.040)
Controls						
age	-0.005**	0.010	-0.005**	-0.003	0.103	-0.004**
	(0.003)	(0.043)	(0.002)	(0.005)	(0.074)	(0.002)
native	-0.042	-1.455*	-0.043	0.025	1.399	-0.035
	(0.054)	(0.880)	(0.057)	(0.070)	(0.932)	(0.055)
married	-0.013	-1.099	-0.089	-0.061	0.624	-0.084
	(0.048)	(0.947)	(0.054)	(0.068)	(1.530)	(0.054)
single	0.084	0.739	0.012	-0.002	0.293	0.022
	(0.077)	(1.086)	(0.074)	(0.098)	(1.604)	(0.073)
having kids	0.097	-1.434	-0.008	-0.045	-0.775	0.001
	(0.065)	(0.892)	(0.058)	(0.066)	(1.414)	(0.057)
city	0.018	1.542*	0.084*	0.125	1.555	0.086*
	(0.034)	(0.874)	(0.047)	(0.090)	(1.222)	(0.046)
Bad Mergentheim	-0.037	-0.641	0.035	0.145**	0.764	0.043
	(0.039)	(0.661)	(0.044)	(0.068)	(0.911)	(0.044)
Glotterbad	-0.028	0.403	0.065	0.057	-0.255	0.083
	(0.045)	(1.292)	(0.112)	(0.104)	(1.585)	(0.112)
Isny	-0.040	-2.001**	-0.073	-0.014	-1.043	-0.065
	(0.042)	(0.895)	(0.057)	(0.057)	(1.109)	(0.057)
medium education				0.012	-0.760	0.086*
				(0.070)	(1.393)	(0.045)
high education				0.096	2.368*	-0.061
				(0.097)	(1.420)	(0.071)
satisfactory self-assessed health				-0.011	0.850	-0.026
				(0.060)	(1.027)	(0.046)
good self-assessed health				-0.027	0.511	-0.066
				(0.058)	(1.212)	(0.054)
inverse Mill's ratio (λ)	0.014	4.674*		0.146	-2.064	
	(0.065)	(2.381)		(0.174)	(4.441)	
constant	0.171	-6.970**		-0.103	-4.538	
	(0.170)	(2.593)		(0.235)	(3.915)	
number of observations	344	344	468	168	168	468

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parenthesis. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A7: **IV Effects on Employment Prospects by Initial Employment Status for Women**

	Employed		Unemployed	
Covariate Set	Basic	Extended	Basic	Extended
structural equation: *change in employment*				
change in BMI (in percent)	-0.021*	-0.021*	-0.067	-0.062
	(0.012)	(0.011)	(0.058)	(0.065)
age	-0.002	-0.001	-0.007	0.003
	(0.002)	(0.002)	(0.015)	(0.019)
native	0.008	0.026	0.055	0.440*
	(0.052)	(0.054)	(0.198)	(0.227)
married	-0.049	-0.037	-0.202	-0.016
	(0.052)	(0.050)	(0.347)	(0.376)
single	-0.069	-0.044	-0.191	-0.066
	(0.064)	(0.058)	(0.441)	(0.377)
having kids	0.009	0.018	-0.228	-0.158
	(0.036)	(0.040)	(0.310)	(0.355)
city	0.005	0.002	0.215	0.390
	(0.077)	(0.077)	(0.279)	(0.256)
Bad Mergentheim	0.060	0.064	0.100	-0.139
	(0.073)	(0.072)	(0.332)	(0.388)
Glotterbad	-0.058	-0.050	0.124	0.006
	(0.089)	(0.079)	(0.373)	(0.274)
Isny	0.034	0.038	-0.525	-0.852
	(0.047)	(0.054)	(0.492)	(0.541)
medium education		-0.003		0.343
		(0.051)		(0.292)
high education		0.103*		0.337
		(0.053)		(0.301)
satisfactory self-assessed health		-0.001		0.216
		(0.054)		(0.359)
good self-assessed health		0.022		0.520
		(0.054)		(0.616)
constant	0.983**	0.885**	0.692	-0.527
	(0.201)	(0.212)	(0.809)	(1.140)
instrumental equation: *change in BMI (in percent)*				
€ 150	-3.719**	-3.721**	-3.565	-2.257
	(1.236)	(1.207)	(3.246)	(3.322)
€ 300	-5.267**	-5.414**	-4.652	-5.060
	(1.280)	(1.250)	(3.419)	(3.333)
age	0.063	0.069	-0.002	0.023
	(0.049)	(0.048)	(0.147)	(0.161)
native	1.177	1.344	-0.456	2.825
	(1.014)	(1.016)	(2.734)	(2.281)
married	0.276	0.216	-3.998	0.234
	(1.070)	(1.045)	(2.498)	(2.783)
single	0.252	0.586	-2.625	-4.056
	(1.775)	(1.760)	(4.484)	(4.272)
having kids	0.048	-0.010	-1.350	-3.214
	(1.231)	(1.238)	(4.216)	(3.917)
city	2.197**	2.245**	0.390	1.720
	(1.103)	(1.080)	(2.369)	(2.216)
Bad Mergentheim	0.892	1.067	4.235	-2.175
	(1.411)	(1.396)	(2.972)	(3.399)
Glotterbad	-0.660	-0.451	4.077	0.305
	(1.367)	(1.294)	(3.562)	(2.547)
Isny	-1.081	-0.880	0.725	-0.094
	(0.990)	(1.023)	(5.241)	(5.821)
medium education		-0.342		-0.281
		(0.974)		(3.446)
high education		0.430		2.185
		(1.464)		(3.100)
satisfactory self-assessed health		-0.332		6.991**
		(1.033)		(3.210)
good self-assessed health		-1.116		7.842**
		(0.960)		(3.667)
constant	-4.815	-5.316	2.449	-4.267
	(3.395)	(3.346)	(10.513)	(10.204)
number of observations	129	129	39	39

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A8: **IV Effects on Employment Prospects by Initial Employment Status for Men**

	Employed		Unemployed	
Covariate Set	Basic	Extended	Basic	Extended
structural equation: *change in employment*				
change in BMI (in percent)	-0.005	-0.006	-0.023	-0.020
	(0.012)	(0.012)	(0.028)	(0.028)
age	-0.004**	-0.003**	-0.012	-0.012**
	(0.002)	(0.002)	(0.008)	(0.006)
native	0.013	0.017	-0.193	-0.187
	(0.046)	(0.045)	(0.186)	(0.221)
married	-0.002	-0.002	0.060	-0.002
	(0.039)	(0.041)	(0.169)	(0.190)
single	0.054	0.059	0.043	-0.008
	(0.062)	(0.063)	(0.184)	(0.170)
having kids	0.110*	0.105*	0.058	0.136
	(0.059)	(0.057)	(0.215)	(0.259)
city	0.026	0.027	0.026	0.021
	(0.030)	(0.029)	(0.170)	(0.183)
Bad Mergentheim	-0.027	-0.036	0.082	0.096
	(0.030)	(0.030)	(0.131)	(0.130)
Glotterbad	0.014	0.019	-0.299	-0.466
	(0.037)	(0.038)	(0.212)	(0.330)
Isny	0.00003	-0.007	-0.258	-0.185
	(0.036)	(0.036)	(0.171)	(0.170)
medium education		0.013		0.091
		(0.027)		(0.188)
high education		-0.005		-0.140
		(0.053)		(0.199)
satisfactory self-assessed health		0.060		-0.071
		(0.037)		(0.163)
good self-assessed health		0.069*		0.190
		(0.037)		(0.252)
constant	1.021**	0.940**	0.700	0.720*
	(0.120)	(0.128)	(0.471)	(0.429)
instrumental equation: *change in BMI (in percent)*				
€ 150	-2.189**	-2.240**	-6.191**	-6.126*
	(0.652)	(0.643)	(2.801)	(3.146)
€ 300	-2.293**	-2.286**	-4.478	-5.171*
	(0.765)	(0.770)	(2.679)	(2.864)
age	0.043	0.037	0.110	0.091
	(0.034)	(0.037)	(0.088)	(0.080)
native	-1.190	-1.107	-0.786	-0.095
	(0.909)	(0.890)	(2.940)	(3.101)
married	-0.045	0.071	-1.527	-2.247
	(0.793)	(0.817)	(3.205)	(3.617)
single	1.041	0.993	-0.999	-1.831
	(1.107)	(1.124)	(3.147)	(3.207)
having kids	-1.327	-1.336	-0.398	-0.095
	(0.960)	(0.967)	(3.012)	(3.458)
city	0.338	0.350	3.486	3.682
	(0.896)	(0.889)	(2.406)	(2.819)
Bad Mergentheim	-1.241*	-1.082	2.548	2.364
	(0.671)	(0.658)	(2.653)	(3.057)
Glotterbad	-0.303	-0.254	-1.176	-3.362
	(1.578)	(1.499)	(3.027)	(5.607)
Isny	-1.358	-1.175	-0.082	0.831
	(0.849)	(0.825)	(2.533)	(2.573)
medium education		0.340		0.547
		(0.693)		(3.273)
high education		1.599		1.498
		(1.041)		(3.258)
satisfactory self-assessed health		-0.972		0.280
		(0.693)		(2.271)
good self-assessed health		-0.981		3.911
		(0.843)		(3.773)
constant	-3.654	-2.968	-5.880	-5.375
	(2.714)	(2.940)	(5.936)	(5.776)
number of observations	297	297	48	48

Notes: ** significant at 5% * significant at 10%. Robust standard errors in parentheses. Missing values set to zero. 'Missing' indicator included. Results for dummy variables indicating the year and the quarter not reported. Bad Mergentheim, Glotterbad, and Isny refer to the locations of the rehabilitation clinics.

Table A9: **Descriptive Statistics: Proxy Variables for Labor Productivity (Means)**

	Women	Men	Min	Max
physical well-being: very good	0.020	0.015	0	1
good	0.186	0.204	0	1
satisfactory	0.368	0.404	0	1
poor	0.343	0.308	0	1
bad	0.083	0.068	0	1
limitations at daily tasks	0.856	0.888	0	1
number of observations	222	468		

Notes: Standard deviations omitted. Means do not include observations with missing values.