



Does moderate weight loss affect subjective health perception in obese individuals? Evidence from field experimental data

Lucas Hafner¹ · Harald Tauchmann^{1,2,3} · Ansgar Wübker^{2,4,5}

Received: 20 June 2020 / Accepted: 19 October 2020 / Published online: 12 January 2021
© The Author(s) 2021

Abstract

This paper analyzes whether moderate weight reduction improves subjective health perception in obese individuals. Besides simple regression models, in a simultaneous equation framework we use randomized monetary weight loss incentives as instrument for weight change, to address possible endogeneity bias. In contrast to related earlier work that also employed instrumental variables estimation, identification does not rely on long-term, between-individuals weight variation, but on short-term, within-individual weight variation. Yet, our result does not suggest that the simple regressions suffer from much endogeneity bias, since instrumental variables estimation yields similar—though far noisily estimated and statistically insignificant—estimates. In qualitative terms, our results do not contradict previous findings pointing to weight loss in obese individuals resulting in improved subjective health. Our results suggest that a reduction of body weight by one BMI unit is associated with an increase in the probability of reporting self-rated health to be ‘satisfactory’ or better by 3 to 4 percentage points. This finding may encourage obese individuals in their weight loss attempts, since they are likely to be immediately rewarded for their efforts by subjective health improvements.

Keywords Self-rated health · BMI · Obesity · Randomized experiment · Short-term effect · Instrumental variable

✉ Harald Tauchmann
harald.tauchmann@fau.de

¹ Universität Erlangen-Nürnberg, Professur für Gesundheitsökonomie, Findelgasse 7/9, Erlangen, Germany

² RWI – Leibniz Institut für Wirtschaftsforschung, Essen, Germany

³ CINCH – Health Economics Research Center, Essen, Germany

⁴ Hochschule Harz, Wernigerode, Germany

⁵ Leibniz Science Campus Ruhr, Essen, Germany

JEL Classification I12 · C26 · C93

1 Introduction

It is well documented in the literature that excessive accumulation of body fat (obesity) is associated with many undesirable health outcomes such as heart disease (Hubert et al. 1983), type 2 diabetes (Mokdad et al. 2003), and several forms of cancer (Calle et al. 2003). A recent meta-analysis (Di Angelantonio 2016) even finds that obese individuals face a higher risk of all-cause mortality compared to their normal-weight counterparts. Although, at least in western societies, the general public seems meanwhile to be well aware of the health risks associated with obesity (Tompson et al. 2012), its prevalence is at an all-time high and further increasing worldwide (WHO 2000; Ng et al. 2014).

Even for moderate weight loss (5–10% percent of body weight) in obese individuals, substantial benefits for objectively measurable health outcomes, such as blood lipid profiles or cardiovascular risk factors, have been established (Blackburn 1995; Wing et al. 2011). However, despite likely health benefits from losing weight, many obese struggle with realizing even small, sustained reductions in body weight. This ubiquitous everyday experience is also well documented in the scientific empirical literature. In a systematic review of long-term weight management schemes, Loveman et al. (2011), for instance, find that short-run reductions of body weight are commonly offset by subsequent weight regain. A better understanding of the mechanisms that make weight loss sustainable and the factors that let weight loss efforts fail is, hence, crucial for battling the obesity ‘epidemic’.

One possible explanation is that moderate weight loss insufficiently induces short-term improvements in perceived health. Objective health measures, for which beneficial effects are well established, do not necessarily reflect patient’s subjective health perception. Yet the latter is likely to matter much for health- and obesity-related behavior. If one realizes some weight loss under great efforts without feeling better, it may be tough to keep up the discipline to maintain or further reduce one’s body weight.

In order to contribute to the discussion, we empirically address the question of whether moderate weight loss causally influences the subjective health perception of obese individuals. Several analyses have examined the relationship between self-rated health (SRH) and excess body weight. The vast majority of the existing literature find a significant negative association that is poor self-rated health accompanies obesity. Using a national survey with Americans, Ferraro and Yu (1995) find that—even after controlling for morbidity and functional limitations—obese individuals have a higher probability of bad self-rated health compared to normal weight individuals. Okosun et al. (2001) approve this finding also analyzing a sample of Americans. Phillips et al. (2005), Prosper et al. (2009), and Baruth et al. (2014) are further, more recent examples for analyses yielding similar results based on US data.

This general pattern is not confined to studies using data from the USA. Guallar-Castillón et al. (2002), for instance, analyze a sample of Spanish women and find that overweight and obese individuals are significantly more likely to report poor health

compared to normal weight women. Molarius et al. (2007) found that overweight (body mass index [BMI] $\geq 25\text{kg/m}^2$) and obese (BMI $\geq 30\text{kg/m}^2$) Swedes have a higher probability to rate their health as poor, compared to normal weight survey respondents. Using data from Finland, Johansson et al. (2009) establish a statistically significant and negative correlation between self-assessed good health and any measure for overweight they consider in their analysis, specifically raw weight, fat mass, waist circumference, and BMI. This result holds for both men and women. Examining health surveys from Portugal and Switzerland, Marques-Vidal et al. (2012) find that obese subjects rated their health significantly worse compared to their normal weight counterparts. This also holds for UK residents as shown in Ul-Haq et al. (2013b).

Only very few empirical studies yield mixed findings or do not find a significant association between self-rated health and obesity at all. Looking at American cross-sectional data over a time span of 30 years (1976–2006), Macmillan et al. (2011) confirm the above pattern for women. Yet, for men, the association between obesity and SRH is weaker and only significant in roughly half of the considered years. Imai et al. (2008) find that the association of BMI and SRH varies significantly across different ages and sexes. They generally confirm previous findings, stating that being underweight or severely obese is associated with bad SRH. However, they find no significant association for obese men older than 65. Darviri et al. (2012) find no significant association between SRH and BMI for a rural population in Greece, neither do Kepka et al. (2007) using a sample of Hispanic immigrants in the USA.

Although a close association of excess body weight and SRH is very well documented in the literature, the question remains unsettled whether excess body weight causally affects self-rated health. Such effect is crucial for subjectively perceived health improvements encouraging obese individuals in their weight loss efforts. However, the mere correlation may just capture the influence of confounding third factors such as certain lifestyles that affect both body weight and self-perceived health. An example for a confounding third factor is sleep duration. Studies find short sleep duration to be associated with poor self-rated health (Frange et al. 2014) as well as obesity (Patel and Hu 2008). Stress may serve as another example for such confounding factors. An increase in stress is likely to have detrimental effects on self-rated health. At the same time, stress may induce overeating (Zellner et al. 2006). Moreover, reverse causality may also be an issue. One may, for instance, think of individuals who feel well and healthy and are motivated by this to practice an active lifestyle that prevents them from becoming overweight.

The above mentioned studies analyze the relationship of inter-individual weight-variation and self-rated health in cross-sectional data sets. They spend little effort in establishing causality in the link between SRH and obesity. One notable exception in this literature is Cullinan and Gillespie (2016) who employ instrumental variables estimation to identify a causal link. Following several examples from the literature (Ali et al. 2014; Cawley and Meyerhoefer 2012; Sabia and Rees 2011; Kline and Tobias 2008; Lindeboom et al. 2010), they use body weight of biological relatives (children) as instrumental variable. This choice of instrument seems to be well justified by evidence from adoption (Vogler et al. 1995; Sacerdote 2007) and twin studies (see Elks et al. 2012; Maes et al. 1997, for survey soft his literature), which suggests that shared genetics explain intra-family correlation of BMI much better than the shared

social environment.¹ However, despite the major importance of genetic disposition, household level environmental conditions may still play some role for intra-family correlation of body weight.² They may, in turn, contaminate biological relatives' body weight as instrument, since such conditions may also matter for health and subjective health perception. More importantly, even if close relatives' body weight is a valid and strong instrument for the level of BMI or overweight status in a cross section of data, it can hardly be used as instrumental variable if the analysis is concerned with the effects of relatively small *changes* in body weight, which are observed over a relatively short period of time.

This is precisely the focus of the present analysis that aims at identifying subjective health effects of a moderate, short-term weight loss in obese individuals. Our contribution is to develop an empirical strategy that allows for identifying such intra-individual short-term effects. Following Cullinan and Gillespie (2016) and earlier work, we rely on instrumental variables estimation to establish a causal link. Yet, we do not adopt their instrument, which provides an exogenous source of variation in the long-term level of BMI. We rather make use of a randomized controlled experiment that exogenously induced short-term variation in body weight,³ and hence provides a basis for identifying short-term effects attributable to moderate weight loss.

To summarize our analysis in a nutshell, we use data of 695 obese patients of four rehabilitation clinics, who voluntarily participated in a field experiment. Upon discharge, all participants were set an individual weight-loss target which they were prompted to realize within 4 months. The participants were randomly assigned to one control group and two incentive groups. Only the latter could earn monetary rewards of up to € 150 or € 300, respectively, for successfully reducing body weight. The participants were asked about their subjective health both by the end of the rehab stay and by the end of the 4 months weight-loss phase. Weight loss over 4 months turns out to be significantly associated with self-assessed health reported by the end of the weight-loss phase. In an instrumental variables (IV) estimation approach, we only use that weight variation for identification that is externally induced by the monetary incentives. In the IV estimation, the point estimates do not change much compared to the simple 'naïve' estimation approach. Yet the estimates become by far noisier, not allowing for judging the IV estimates as statistically significant. Yet, since statistical tests do not point to endogeneity being a major issue and since the point estimates are similar, we still regard the IV results as in concordance with our earlier findings. In quantitative terms, our results suggest that reducing body weight by one BMI unit increases the probability of rating his or her health as 'satisfactory' or better by roughly three percentage points.

The remainder of this paper is structured as follows: In Sect. 2, we introduce our data. In Sect. 3, we describe our estimation procedure. In Sect. 4, we show the results

¹ A closely related identification strategy is to directly use genetic information as instrument for body weight. Few recent contributions (Norton and Han 2008; Fletcher and Lehrer 2011; von Hinke et al. 2016; Willage 2018), which consider different outcomes than subjective health, have adopted this strategy.

² Price and Swigert (2012) find substantial weight differences among siblings who reside in the same households. The authors mention differing parental behaviors across siblings as a possible explanation.

³ Reichert (2015) and Reichert et al. (2015) use the same source of exogenous weight-variation but consider different outcomes than health.

of our estimations. Finally, in Sect. 5, we summarize and discuss our main findings and present a conclusion.

2 Data

2.1 The field experiment

The data used in the present analysis originate from a field experiment that was conducted by RWI—Leibniz-Institut für Wirtschaftsforschung. Its prime objective was to test whether monetary incentives are an effective instrument for assisting obese individuals in losing body weight. Four medical rehabilitation clinics operated by the German Pension Insurance of the federal state of Baden–Württemberg and the association of pharmacists of Baden–Württemberg cooperated with RWI in this project. The *Pakt für Forschung und Innovation*, which is part of the excellence in research initiative of the German federal government, provided funding. The study protocol of the project was approved by the ethics commission of the Chamber of Medical Doctors of Baden–Württemberg. See Augurzky et al. (2018) and Augurzky et al. (2014) for a more detailed discussion of the project.

Upon admission to one of the four involved clinics, 695⁴ obese individuals were recruited for participation in the experiment between March 2011 and August 2012. The medical staff in charge was advised to approach any new patient whose BMI exceeded 30⁵ and to invite him or her to take part in the experiment. Yet, participation was entirely voluntary and had no consequence for any treatment or advice the patient received over their rehab stay, which usually takes 3 weeks. The prime objective of rehab stays in these clinics is to preserve, or to restore, patients' workableness. Our study population is hence biased toward the working population, which is, however, no challenge to the internal validity of our analysis. For the vast majority of participants, obesity was not the prime reason for being sent to rehabilitation. Yet, many suffered from health problems related to overweight such as chronic back pain. Hence, all obese patients, irrespective of participation in the experiment, were advised to reduce their body weight.

At rehab discharge, participants' body weight was measured again and participants were set an individual weight-loss target by the physician in charge, which they were prompted to realize within 4 months. Physicians were asked to choose a weight-loss target of about 6–8% of current body weight. Yet they were in principal free to deviate from this guideline. Near the end of the rehab, the participants received a questionnaire, which they were prompted to answer. The questionnaire covered a wide range of questions regarding socio-economic characteristics and weight-related

⁴ Originally 700 patients were recruited, yet five had to be excluded because of ex-post violation of the inclusion criteria (pregnancy, developing cancer) or missing documents.

⁵ In addition to BMI > 30, a detailed list of inclusion and exclusion criteria needed to be met. In detail the inclusion criteria were: age between 18 and 75 years and resident of the federal state of Baden–Württemberg; while the exclusion criteria were: pregnancy, psychiatric illness, eating disorder, carcinosis within the past 5 years, drug or alcohol abuse, a significant language barrier, and a severe generalized disease. Since the latter broadly defined criterion refers to a generalized disease, local diseases that affect only specific organs or bodily functions were in principal no criterion for being excluded.

behavior, such as exercising or eating habits. Most importantly, participants were also asked about their current health status. Two health questions addressed self-rated health and physical well-being, in a standard fashion. The questionnaire was collected (in a sealed envelope) at the appointment with the physician in which the weight-loss target was fixed.

Right after rehab discharge, the participants were randomly assigned to one control and two treatment/incentive groups, and subsequently informed about the result of the randomization by regular mail (intervention). While in this letter all participants were prompted to realize their weight-loss target, treatment group members were informed about the monetary reward they could earn by being successful in losing weight.⁶ For one treatment group, the maximum reward was € 150; for the other, it was € 300. If participants failed to realize at least 50% of the contractual weight loss, they did not receive any money. If they were partially successfully, i.e., they lost more than 50% but less than 100%, they were rewarded proportionally to the degree of target achievement.⁷

By the end of the 4 month weight loss period, all participants received another letter, by which they were prompted to visit a specified pharmacy in a specific week for a weigh-in. Body weight measured in the pharmacy served as basis for the cash out of rewards. Upon attending the weigh-in all participants, irrespective of the experimental group they were assigned to, received an expense allowance of € 25. Each letter was accompanied by a questionnaire, which included the same set of questions as the questionnaire the participants had answered by the end of the rehab stay. In particular, the questions addressing subjective health were exactly the same and made no reference to information the patients had provided earlier.

The experiment included two further phases: a 6-month weight-maintenance phase, which directly followed the weight reduction phase, and a subsequent 12-month follow-up phase. In the weight-maintenance phase, participants who were at least partially successful in meeting their weight-loss target were offered another monetary reward for not exceeding their target weight. In the follow-up phase, participants were not exposed to any monetary incentives for weight loss. In both phases, the weigh-in procedure was the same as for the weight-loss phase. The present analysis only uses information up to the end of the weight reduction phase. The reason for this is that in the weight reduction phase the exogenous source of weight variation, i.e., being

⁶ At recruitment, all participants were informed about the design of the experiment (randomization, monetary rewards). Control group members, hence, knew that they missed the chance of financially benefitting from losing weight. In consequence, the intervention may have had adverse motivational effects in the control group. Indeed, 55% of the members of this group reported (in the second survey) disappointment about the randomization outcome. Twelve individuals even reported to have eaten more in response to not being assigned to an incentive group. Nevertheless, the data does not reveal a significant correlation of the level of disappointment and the achieved weight loss in the control group. Moreover, possible adverse effects on the weight loss motivation is no challenge to using the group assignment as instrument. It still provides an exogenous source of weight variation and the monotonicity assumption is not violated, since possible adverse motivational effects operate in the same direction as the lack of financial incentives.

⁷ One may suspect that individuals who participate in the experiment are more likely to be motivated to lose weight than obese individuals in the general population. Due to the random treatment assignment this is however no threat for the internal validity of our analysis. However, we cannot rule out that the effect of weight loss on health for our sample considered differs from the general obese population, which is why external validity may be limited.

member of the control or the treatment arm of the experiment, is clearly random by the design of the experiment. This applies less to the subsequent weight-maintenance phase, since the second randomization was conditional on success in the previous phase.

The econometric analysis rests on information which was collected at rehab discharge and by the end of the weight-loss phase. While the information regarding body weight is complete for the first time of measurement, this does not hold for the second, since roughly one-fourth of the participants did not attend the weigh-in by the end of the weight-loss phase. In consequence, weight-change information is available for only 517 participants. Augurzyk et al. (2018) comprehensively discuss the issue of experiment drop-out and its possible implications. Using a battery of different econometric techniques, they find that the results are rather robust to correcting for selective drop-out. Unlike body weight, which was measured in the clinic or the pharmacy, the information regarding self-rated health and physical well-being was collected through a written questionnaire. This renders item non-response an issue, which further reduced the size of the estimation sample to 485 individuals in the self-rated health estimation and 468 in the physical well-being estimation, for which weight and health information is available for either time of measurement.

2.2 Variables used in the empirical analysis

We employ two variables to measure the outcome subjective health perception: (i) self-rated health (*SRH*) and (ii) physical well-being (*PWB*). Self-rated health is measured by asking the respondents “how would you describe your current health status?” and allowing for five possible answers: “excellent”, “good”, “satisfactory”, “poor” and “bad”.⁸ Physical well-being is measured by asking the respondents “how would you describe your current physical well-being?”, allowing for the same five possible answers.

While either variable measures subjective health perception, they potentially capture different aspects of it. *PWB* emphasizes subjectiveness in health perception even stronger, while *SRH* leaves more room for objectifying the reported health status. For instance, an obese individual without any health impairments might rate her physical well-being as excellent. At the same time, she is probably aware that her excess weight is a risk for her health. Although feeling healthy she might therefore report a relatively poor *SRH*, to account for potential health risks.

While any questionnaire the participants were asked to fill in included questions about *SRH* and *PWB*, the present empirical analysis focusses on *SRH* and *PWB* that was reported by the end of the 4-month weight reduction phase. These variables,

⁸ A wide variety of methods to assess subjective health perception have been suggested in the literature. These methods include multi-item measures as well as single item-measures. An example of a multi-item measure is the often used Medical Outcomes Study Short Form 36 (SF-36) (Ware et al. 1993). Most studies using the SF-36 find obesity to be associated with poor subjective health perception (see Kroes et al. 2016; Ul-Haq et al. 2013a; Kolotkin et al. 2001; Fontaine and Barofsky 2001, for reviews).

Table 1 Joint and marginal distribution of SRH_0 and SRH_1

		SRH_1					Marginal distribution
		Bad	Poor	Satisfactory	Good	Excellent	
SRH_0	Bad	7	5	6	0	0	18
	Poor	14	37	50	18	2	121
	Satisfactory	5	28	106	67	4	210
	Good	2	8	38	69	8	125
	Excellent	0	2	3	4	2	11
Marginal distribution		28	80	203	158	16	485

denoted as SRH_1 and PWB_1 enter the econometric model at the left-hand side.⁹ The analysis also makes use of self-rated health and physical well-being reported at rehab discharge, i.e., at the outset of the weight reduction phase. As single item measures that do not refer to any objective health indicator but are purely subjective in nature, SRH and PWB are well suited for analyzing self-perceived rather than objectively measured health effects.

Table 1 displays the (joint and marginal) sample distribution of SRH for both considered times of measurement. Not surprisingly—all respondents underwent medical rehabilitation for some reason—the share of individuals who regarded themselves in excellent or good health is smaller than in general population surveys such as the German Socioeconomic Panel (SOEP). Nevertheless, SRH exhibits substantial heterogeneity between individuals. From Table 1, it also becomes obvious that self-rated health considerably varies at the individual level over the observation period.¹⁰ For 54% of the participants, we observe a change in SRH (off-diagonal elements in Table 1), while 46% report the same category of SRH at the beginning and by the end of the weight reduction phase (values highlighted bold in Table 1). 60% of all changes are improvements in SRH (cells above the principal diagonal). Among the participants who reported SRH changes, 81% report a change to an adjacent category. Yet, some rather drastic shifts in SRH , e.g., from ‘excellent’ to ‘poor’ or the other way round, are observed.

The corresponding (joint and marginal) sample distribution of PWB at rehab discharge (PWB_0) and at the end of the weight-loss phase (PWB_1) is displayed in Table 2.¹¹ Comparable to SRH , physical well-being exhibits substantial heterogeneity between individuals and varies at the individual level over time. For 60% of the

⁹ The subscript 1 is a time index that refers to the information gathered by the end of the weight-loss phase (period 1). The subscript 0 indicates pre-intervention values that is SRH_0 (PWB_0) denotes self-rated health (physical well-being) at rehab (period 0) discharge. This notation analogously applies to all variables that are measured at different points in time such as the body mass index BMI_1 and BMI_0 .

¹⁰ If no within-individual variation of SRH was observed, linking changes in SRH to weight change would arguably make little sense.

¹¹ The number of observations for the variables measuring subjective health perception differ - individuals reported their self-rated health status slightly more often.

Table 2 Joint and marginal distribution of PWB_0 and PWB_1

		PWB_1					Marginal distribution
		Bad	Poor	Satisfactory	Good	Excellent	
PWB_0	Bad	9	9	5	5	1	29
	Poor	22	48	58	20	3	151
	Satisfactory	10	37	82	53	5	187
	Good	0	6	29	48	8	91
	Excellent	0	1	2	5	2	10
Marginal distribution		41	101	176	131	19	468

Table 3 Joint and marginal distribution of PWB_1 and SRH_1

		SRH_1					Marginal distribution
		Bad	Poor	Satisfactory	Good	Excellent	
PWB_1	Bad	19	16	7	1	0	43
	Poor	9	40	43	9	0	101
	Satisfactory	1	18	122	40	3	184
	Good	0	2	29	102	2	135
	Excellent	1	0	2	6	10	19
Marginal distribution		30	76	203	158	15	482

participants, we observe a change in PWB (off-diagonal elements in Table 2), while 40% report the same category of PWB at the beginning and by the end of the weight reduction phase (values highlighted bold in Table 2). 60% of all changes are improvements in PWB (cells above the principal diagonal). Among the participants for which reported PWB changes, 79% report a change to an adjacent category.

Self-rated health and physical well-being are obviously closely related measures and are strongly correlated in the sample. However, as their correlation is far from perfect, the two variables seem to capture different aspects of subjective health perception. Table 3 displays the (joint and marginal) sample distribution of SRH and PWB at the end of the weight-loss phase. Most respondents report the identical answer category for both variables (61%). However, 25% of the respondents reported better SRH , while 14% of the respondents reported better PWB .¹² Only 1% of the respondents deviated by more than two answer categories (bad SRH and excellent PWB).¹³

¹² This pattern is similar when we look at the relationship of self-rated health and physical well-being at the end of the rehab-phase (correlation coefficient of 0.64). Here 58% of respondents reported the same answer category for both variables, while 28% of respondents reported better SRH and 14% reported better PWB . See Table 7 in Appendix.

¹³ Excluding these individuals from the analysis does not change our results in qualitative terms.

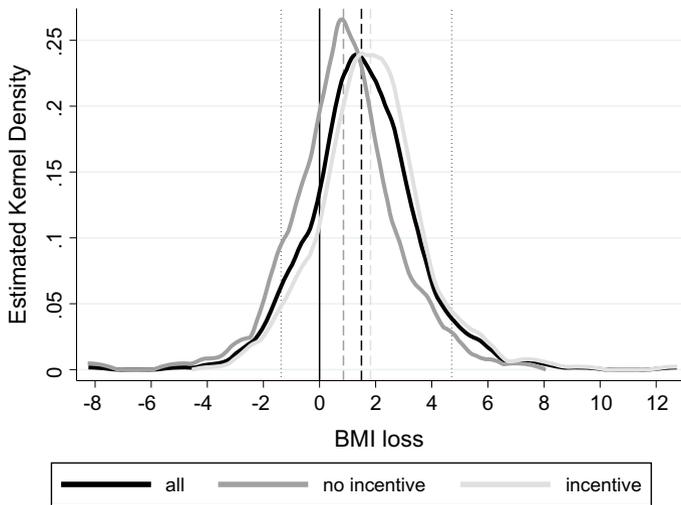


Fig. 1 Distribution of *BMI loss* in the full sample and the experimental groups. *Notes:* Estimated kernel densities; dashed lines mark the medians (full sample 1.49, no incentive 0.85, incentive 1.82); dotted lines mark 5th (−1.37) and the 95th (4.71) the percentile in the full sample.

Source: Own calculations and Augurzyk et al. (2018)

Body weight, which is the key explanatory variable in the present analysis, is measured in terms of the body mass index.¹⁴ Rather than its level, we consider the absolute change ($BMI\ loss \equiv BMI_0 - BMI_1$) between rehab discharge and the end of the weight-reduction phase as regressor. By this choice, we emphasize that the focus of the analysis is on the effects of within-individual weight loss rather than between-individual heterogeneity in the level of *BMI*.¹⁵

The variation of weight change in the sample is quite substantial. 81% of the participants lost weight. Mean weight change is 1.56 BMI units. The median of the weight loss distribution (1.49) is close to the mean. The 95% quantile is 4.71, indicating that a substantial share of participants managed to materially reduce body weight over the 4 month weight-loss phase. Yet, the 5% quantile is −1.37, pointing to substantial weight gain being not a rare phenomenon in the sample; see Fig. 1 for sample distribution of *BMI loss*. Figure 1 also illustrates that members of the incentive groups were on average clearly more successful in reducing body weight (cf. Augurzyk et al. 2018).

¹⁴ For decades, BMI and its commonly used threshold value of 30kg/m² (WHO 2016) have been criticized as an, at least in certain circumstances, inappropriate measure of clinical obesity (Garn et al. 1986). We nevertheless stick to this frequently used measure. Since we consider changes of BMI over a relatively short period of time, rather than comparing the level of BMI between individuals, several shortcomings (age dependence, indifference regarding lean and fat tissue, etc.) of the BMI are arguably of little importance. Using percentage change in body weight instead of absolute change in *BMI* as weight change measure yields largely equivalent results in our empirical analysis. Moreover, the problem of misreported height and weight (cf. Gorber et al. 2007) is of little relevance to our study since body weight is not self-reported but measured by clinic staff or pharmacy staff.

¹⁵ We include the pre-intervention level BMI_0 as control. Hence technically, our preferred specification is equivalent to including both the pre- and post-intervention level BMI_0 and BMI_1 at the right-hand-side of the regression model.

Table 4 Self-rated health and physical well-being by BMI loss

	Bad (%)	Poor (%)	Satisfactory (%)	Good (%)	Excellent (%)	Marginal distribution*
<i>SRH₁</i>						
No BMI loss	16.48	21.98	42.86	15.38	3.30	91
BMI loss	3.30	15.23	41.62	36.55	3.30	394
<i>PWB₁</i>						
No BMI loss	13.95	32.56	38.37	11.63	3.49	86
BMI loss	7.59	19.11	37.43	31.68	4.19	382
<i>SRH₀</i>						
No BMI loss	8.79	30.77	34.07	25.27	1.10	91
BMI loss	2.54	23.60	45.43	25.89	2.54	394
<i>PWB₀</i>						
No BMI loss	6.98	38.37	36.05	18.60	0.00	86
BMI loss	6.02	30.89	40.84	19.63	2.62	382

Shares and *absolute numbers of individuals

While the median weight loss is 1.82 BMI units for the former, it is only 0.85 BMI units for the latter. Yet, it also becomes visible that the weight loss variation in the respective group is substantial and exceeds the variation between the groups.

From the first panel of Table 4, one can see that—in a descriptive sense—participants who lost weight are more likely to report good health. Among this group of the participants around 40% reported good or excellent health, while this only holds for around 19% of the participants who gained weight. 38% of the latter reported poor or bad health. In contrast, the corresponding share of participants who lost weight is only 19%. According to a Wilcoxon rank-sum test, the distribution of *SRH₁* clearly differs (p -value 0.000) between individuals who lost weight and individuals who did not. The estimated probability for an individual from the former group to be in better health than an individual from the latter is 0.65. These descriptive findings line up with the general pattern of results found in the literature that less body weight is associated with better self-ratings of health. Considering physical well-being instead of self-rated health yields a very similar picture. Again, according to a Wilcoxon rank-sum test, the distribution of *PWB₁* clearly differs (p -value 0.000) between individuals who lost weight and individuals who did not.

If the same descriptive analysis is applied to self-rated health measured at the beginning of the weight-loss phase, i.e., to *SRH₀* instead of *SRH₁*, we still find a significant (p -value 0.044), though less distinct, deviation in the distribution of self-rated health. At the one hand, this suggests that weight loss might be endogenous and, in turn, calls for an empirical approach that does not interpret the mere correlation as causal effect. On the other hand, this pattern suggests analyzing the effect of *BMI loss* on *SRH* conditionally on its initial level *SRH₀* in order to account for persistent unobserved heterogeneity and to eliminate variation in the dependent variable that cannot be explained by a change in *BMI*. For this reason, *SRH₀* enters the econometric analysis as control variable.

If the same analysis is applied to physical well-being measured at the beginning of the weight-loss phase (PWB_0), we do not find a clearly significant difference (p -value 0.159) in the distribution of physical well-being. Yet, the share of respondents who rate their physical well-being at least satisfactory is still higher for respondents who lost weight. Hence, analogously to the regression explaining self-rated health, we control for PWB_0 when our outcome variable is PWB_1 in order to account for persistent unobserved heterogeneity.¹⁶

As another approach to account for unobserved heterogeneity, we also control for initial body mass index BMI_0 . Though all participants were obese at the time of recruitment, BMI_0 exhibits pronounced heterogeneity ranging from 28 up to 60.¹⁷ The average of the initial BMI is 37.26, while the median value of 36.03 is somewhat smaller, indicating that distribution of initial BMI is skewed to the right.

Due to the relatively small estimation sample, we abstain from specifying a rich regression model with a large number of controls. As basic socioeconomic characteristics we only control for age and gender.¹⁸

As discussed in Sect. 1, we use exposition to monetary weight loss incentives as instrument for weight change. Though random assignment to the experimental groups is a very strong argument for the instrument being exogenous, direct effects of the group assignment on subjective health might still be a challenge for exogeneity. One such channel is anxiety of not earning the reward because of insufficient weight loss, which may negatively impact on health.¹⁹ We cannot rule out that this channel plays some role. However, this effect should downward bias the estimated effect since only the members of the incentive groups are subject to such adverse effects of the incentives. This possible direct effect should hence not generate a spurious regression result in the IV estimation. Moreover, to dig deeper into this issue, we stratified the analysis of weight loss effects with respect to: (i) the weight-loss target [kg] (sample split at the median) and (ii) the size of the reward (€ 150 and € 300). One may hypothesize that a more ambitious target and a higher amount of money at stake are more prone to elicit anxiety. Yet, regarding the effect of weight loss, we see no significant differences between the respective groups.²⁰ We take this as indication that possible anxiety of not earning the reward does not generate a major endogeneity problem. Another possible channel is that not weight loss itself, but the measures taken to achieve the reduction in body weight, affect subjective health. Though it is almost impossible to disentangle these two channels, the results of Augurzky et al. (2018), who find a much stronger effect of weight loss incentives on weight loss than on weight reducing activities such as doing sports and health eating, argue in favor of weight loss being the prime channel through which the incentives operate.

¹⁶ Both SRH_0 and PWB_0 enter the model in a linear way. Estimating the model with dummy variables for the different categories of SRH and PWB does not alter our results.

¹⁷ Many individuals already lost weight over the rehab stay. This is the reason for some participants entering the weight-loss phase with a BMI smaller than 30.

¹⁸ We also estimated models with more explanatory variables (controlling for education, income and employment), however, the results of those models (reported in Tables 12 and 13 in the Appendix) are similar to the results of our preferred specification, where the number of observations is higher.

¹⁹ We would like to thank the reviewer for pointing us to this issue.

²⁰ Results are available upon request.

Table 5 Descriptive statistics for estimation sample

	Obs.	Mean	S.D.	Min.	Max.	Median
Dependent variables						
<i>SRH</i> ₁	485	3.11	0.92	1	5	3
<i>PWB</i> ₁ [†]	468	2.97	1.00	1	5	3
Explanatory variables						
<i>BMI loss</i>	497	1.56	1.98	-8.28	12.78	1.51
<i>BMI</i> ₀	497	37.26	6.16	28.03	60.22	36.03
<i>SRH</i> ₀	497	2.97	0.87	1	5	3
<i>female</i>	497	0.33	0.47	0	1	0
<i>age</i>	497	49.16	8.56	20	68	50
Instrumental variable						
<i>incentive</i> [‡]	497	0.69	0.46	0	1	1

The number of observations differs between the main (485, 468) and the auxiliary (497, 487) equation of the econometric model. Descriptive statistics for the explanatory variables are for the auxiliary equation in the model explaining *SRH*₁. [†]See Appendix, Table 8 for comprehensive descriptive statistics of the estimation sample with *PWB*₁ serving as dependent variable. [‡]162 individuals belong to the treatment group, where the maximum reward is € 150 while 183 individuals belong to the treatment group with a maximum reward of € 300

Though the experiment involved two treatment groups which were offered incentives of different size, in the regression analysis we use a simple dummy that indicates random assignment to one of the treatment groups. Pooling the treatment groups is in line with the finding of Augurzky et al. (2018) that the size of offered monetary reward proved to be immaterial for realized weight loss. Descriptive statistics for all variables that enter the preferred regression model are provided in Table 5.

3 Estimation procedure

In order to take the ordered categorical nature of our dependent variables *SRH*₁ and *PWB*₁ into account, the econometric analysis rests on ordered probit models. We start with estimating a conventional specification of this model that regards all regressors as exogenous. Besides the key explanatory variable *BMI loss*, pre-intervention body weight *BMI*₀, age and gender enter the models at the right-hand-side. Additionally, we control for pre-intervention self-rated health (*SRH*₀) or pre-intervention physical well-being (*PWB*₀), depending on the dependent variable that is used. This basic model specification serves as reference.

Yet, as discussed above, results from conventional ordered probit estimation are most likely biased, due to unobserved confounders affecting both subjective health perception and *BMI loss*, as well as reverse causality. To tackle possible endogeneity bias, and to allow for identifying a causal effect of *BMI loss* on subjective health perception, in our preferred empirical model we do not only rely on naïve ordered probit estimation, but tap an exogenous source of variation in body weight for identifying the effect under scrutiny. Random assignment to either the control or the treatment

arm of the experiment generates weight variation, which by the experimental design is exogenous. Moreover, as shown elsewhere (Augurzky et al. 2018), the incentive treatment was clearly effective and hence induced exogenous variation in weight loss. Technically, the binary indicator *incentive*, which indicates assignment to one of the two incentive groups, serves as instrument for *BMI loss*.²¹

If health was measured on a continuous scale, two-stage least squares would be an obvious choice for the estimation procedure. However, this choice would conflict with the ordered categorical nature of *SRH*₁ and *PWB*₁.²² We, hence, opt for a more parametric approach to instrumental variables estimation. That is, we augment the equation of prime interest by a second equation that specifies the endogenous regressor *BMI loss* as a function of the instrument *incentive* and the covariates that enter the main equation, and assume joint normality of the two error terms. The cross-equation error-correlation, hence, captures possible endogeneity of *BMI loss*. Joint estimation by full-information maximum likelihood (ML) is straightforward for this model.²³

4 Estimation results

4.1 Results for the basic model

In this section, we present and discuss results for the regression models we introduced in the previous section. Columns one and two in Table 6 display the estimation results of the naïve model that does not take possible endogeneity into account.

The results are in line with those of the majority of the related literature. In both specifications we find a statistically significant association between weight change and subjective health perception, where weight loss is positively associated with the inclination to report a better status of subjective health. In terms of magnitude, the estimated coefficient is similar for both measures of subjective health perception.

In quantitative terms the point estimate of 0.15 (self-rated health) translates into an average increase in the probability of rating one's health 'satisfactory' or better of 3.7 percentage points if one reduces her or his body weight by one BMI unit. For physical well-being the average marginal effect is of similar magnitude. A reduction of body weight by one BMI unit is associated with an increase in the probability of rating one's physical well-being 'satisfactory' or better by 4.7 percentage points.

Turning to the coefficients of the control variables, *BMI*₀ is not significantly associated with subjective health perception. Yet, not surprisingly, the coefficients estimated for initial subjective health perception (*SRH*₀ and *PWB*₀) are positive and highly sig-

²¹ As an alternative model specification we used two indicators, each indicating membership in one of the two incentive groups, as instrument. This affects the results just marginally.

²² We also estimated the model by two stage least squares. In this robustness check we reduced the number of categories to just two. In qualitative terms, the results from this less parametric model are largely equivalent to our preferred specification; see Table 11 in Appendix for detailed results.

²³ We used the user-written Stata (R) command *cmp* (Roodman 2009) for estimation. It generalizes the familiar full information ML approach to estimating binary probit models with endogenous explanatory variables (cf. Wooldridge 2002, 472–477). Although referring to joint ML estimation with exclusion restrictions as 'instrumental variables estimation' is a questionable choice of terminology, we stick to this nomenclature, which is common in the applied empirical literature.

Table 6 Coefficient estimates of Naïve ordered probit and IV-ordered probit models (full sample)

Dependent variable	Naïve ordered probit		IV-ordered probit	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.150*** (0.026)	0.160*** (0.026)	0.126 (0.111)	0.072 (0.110)
<i>BMI</i> ₀	-0.005 (0.008)	-0.004 (0.008)	-0.004 (0.010)	0.001 (0.011)
<i>SRH</i> ₀	0.598*** (0.062)	-	0.601*** (0.062)	-
<i>PWB</i> ₀	-	0.542*** (0.059)	-	0.550*** (0.059)
<i>female</i>	-0.362*** (0.107)	-0.339*** (0.107)	-0.367*** (0.109)	-0.357*** (0.108)
<i>age</i>	-0.009 (0.006)	-0.005 (0.006)	-0.009 (0.006)	-0.007 (0.006)
Marginal effect of <i>BMI loss</i> [†]	0.037*** (0.006)	0.047*** (0.007)	0.032 (0.027)	0.022 (0.033)
Auxiliary equation (dependent variable: <i>BMI loss</i>)				
<i>incentive</i>	-	-	0.978*** (0.185)	1.004*** (0.187)
<i>BMI</i> ₀	-	-	0.051*** (0.014)	0.056*** (0.014)
<i>SRH</i> ₀	-	-	0.167* (0.099)	-

Table 6 continued

Dependent variable	Naïve ordered probit		IV-ordered probit	
	SRH_1	PWB_1	SRH_1	PWB_1
PWB_0	-	-	-	0.196** (0.097)
<i>female</i>	-	-	-0.299 (0.182)	-0.318* (0.184)
<i>age</i>	-	-	-0.014 (0.010)	-0.017* (0.010)
<i>Constant</i>	-	-	-0.721 (0.871)	-0.820 (0.854)
<i>Cross-eq. error corr.</i>	-	-	0.048 (0.214)	0.173 (0.208)
Instrument relevance (F -statistic)	-	-	27.72	28.54
No. of observations (over all)	485	468	497	487
No. of observations (main equation)	-	-	485	468

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

† Sample average of individual marginal effects of BMI loss on the probability of rating one's health or physical well-being 'satisfactory' or better. Number of observations differ in the IV-Ordered Probit model: Due to joint estimation, participants for which SRH_1 and PWB_1 , respectively, is not observed but the endogenous regressor BMI loss is observed contribute to the log-likelihood function and enter the estimation sample. Cut points of the ordered probit estimations are displayed in Table 14 in Appendix

nificant, revealing pronounced persistence in subjective health perception, which has already been observed in Tables 1 and 2. The simple ordered probit regressions indicate a gender differential in subjective health perception, with women exhibiting a less favorable subjective health rating both in terms of *SRH* and *PWB*. The regression analysis does not yield a significant association between age and subjective health perception²⁴. We also find no significant influence of age on physical well-being.

4.2 Results from IV estimation

As discussed above, the results presented in columns one and two of Table 6 might suffer from endogeneity bias regarding the coefficient attached to *BMI loss*. In this subsection, we discuss estimates that address this issue by the use of an instrumental variable.²⁵ Columns three and four of Table 6 display coefficients for the model that relies on weight variation induced by randomly assigned weight loss incentives for identifying the coefficient of prime interest. Besides the coefficients of the equation of prime interest (upper panel), average marginal effects of *BMI loss* as well as estimates for the auxiliary equation explaining *BMI loss* (lower panel) are also displayed.

Starting with the instrumental equation, estimation results indicate that cash incentives have a substantial effect on achieved weight loss, see second panel of Table 6 columns three and four. This result, which has already been established in the literature (e.g., Augurzyk et al. 2018; Volpp et al. 2008; John et al. 2011; Cawley and Price 2013; Paloyo et al. 2014), is important for the present analysis, as it points to the experiment generating exogenous variation in *BMI* that can be used for identification. Indeed, the indicator *incentive* proved to be a rather strong instrument for *BMI loss*. The relevant *F*-statistic²⁶ is 27.72 and 28.54, respectively,²⁷ which clearly exceed the conventional threshold value of 10 (Stock et al. 2002). Besides this key result regarding instrument relevance, the estimates for the auxiliary equation indicate that those who start with a high initial *BMI* are more likely to lose weight. Yet it is worth mentioning that including controls in the instrumental variable equation is of minor importance since the randomization balances the covariates between the groups.²⁸ Dropping the controls from the auxiliary equation thus has very little effect on the results.

Turning to the equation of main interest, we see almost no change in the coefficients of the control variables as compared to simple ordered probit. Yet, with respect to the effect of weight change on subjective health perception, we find a pattern of results that in some respect deviates from its counterpart from naïve estimation. The point estimates are smaller and the coefficients of *BMI loss* turn statistically insignif-

²⁴ Including age^2 as additional regressor does not point to a non-linear relationship between *SRH* and age.

²⁵ From a purely technical perspective, one could argue that instrumental variables are not required for identification that could solely rest on the non-linearity of the model. Yet, this will rarely work in practice. Indeed, in the present application the optimization procedure runs into serious convergence problems if *incentive* is not included as instrument.

²⁶ It is calculated from estimating the auxiliary equation separately by least squares.

²⁷ The first stage regressions differ slightly between the models explaining SRH_1 and PWB_1 , since either SRH_0 or PWB_0 enters the model at the right-hand-side.

²⁸ A joint balancing test (Pei et al. 2019, p. 212) yields an *F*-statistic as small as 1.06 (*p*-value 0.375), i.e., the test is far from rejecting the null of no association of any covariate with the group assignment.

icant. However, the estimated *BMI loss* coefficients are still positive. For the model explaining self-rated health, the deviation in the point estimates and the corresponding marginal effects is very small. Yet, due to the large standard errors the 95-% confidence interval around the IV-estimates is wide, more precisely $[-0.092, 0.345]$. For the model explaining physical well-being the deviation from the IV estimate and its counterpart from simple ordered probit is more pronounced. But still, the associated confidence interval, which is $[-0.144, 0.287]$, includes the coefficient from the naïve approach. The same line of argument applies to the corresponding marginal effects. The 95-% confidence intervals around the estimated mean effects are $[-0.022, 0.085]$ and $[-0.042, 0.086]$, respectively. They clearly include the relatively precisely estimated effects from simple ordered probit estimation. Yet, they are that wide that one cannot reject effects which are much smaller (even zero effects and effects to the opposite direction) or much bigger than those one obtains from the naïve estimation approach.

The lack of statistical significance of the estimated effects in the IV approach, thus, seems first of all to be a standard error issue, and can hardly be interpreted as evidence for the absence of a weight loss effect on subjective health. Evidently—although the instrument is not weak—augmenting the naïve model by the instrumental equation inflates the noisiness of the estimates substantially. The reason for this is that the instrument, despite the large *F*-statistic, still explains a relatively small fraction of the variation in the endogenous variable *BMI loss*. The partial *R*-squared is, indeed, just 0.051 despite the relatively large *F*-value of 27.72; compare Fig. 1.²⁹

The finding that instrumenting *BMI loss* does not fundamentally change the estimated key coefficients in our main specifications is mirrored by the estimates of the cross-equation error correlation. The estimate is positive but of moderate or even negligible magnitude and statistically insignificant. Though the positive sign argues in favor of unobserved confounders may play some role in the correlation of subjective health perception and *BMI*, the estimates still provide little evidence for endogeneity of *BMI loss* being a major issue. Moreover, based on linear specifications (see next subsection) that employ OLS and 2SLS rather than ordered probit and IV ordered probit, Hausman–Wu tests do not yield evidence for any systematic deviation between instrumental variables and naïve estimation.³⁰ One possible reason for this pattern is that relying on short-term, within-individual variation cuts or at least weakens already several channels—for instance, health-conscious attitudes, educational and family background, and certain genetic endowments—that are likely to be major sources of endogeneity in cross-sectional data based analyses.

²⁹ The issue of why a small share of variation in the endogenous regressor that can be attributed to variation in the instrument inflates the variance becomes clearer if one considers a two-step control function approach (e.g., Wooldridge 2015), which is a close alternative to joint ML estimation. In the present analysis the two-step estimator yields almost the same coefficient estimates as those reported in Table 6, columns 3 and 4. In the control function approach the first stage residual enters the main equation as regressor in addition to *BMI loss*. The less variation is explained by the first-stage regression the stronger the residual is correlated with the endogenous regressor. Hence, in technical terms, additionally including the first stage residual means that another explanatory variable enters the model that is substantially correlated with the regressor of primary interest. This necessarily inflates the variance. In this context it is important to note that the argument of a ‘sufficiently’ strong instrument primarily addresses the issue of the IV finite-sample bias, but not the issue of instrumental variables estimation inflating the variance; see e.g., Angrist and Pischke (2009, p. 208).

³⁰ For the majority of specifications the *p*-value exceeds 0.9.

To sum up this discussion, though the instrumental variables estimation does not yield a clear cut result regarding the effect of weight change on subjective health perception, the pattern of results is still telling. While, the point estimates argue for the naïve empirical approach suffering from some upward bias, the rather noisy IV estimates can hardly put the general finding of weight-gain in obese individuals affecting subjective health perception detrimentally into question. This in particular holds since the IV approach reveals little evidence for the naïve estimates suffering from a severe endogeneity bias. In qualitative terms our results, hence, do not conflict with the bulk of the literature. They also do not contradict those of Cullinan and Gillespie (2016)³¹, who carefully designed their analysis to allow for a causal interpretation of the link between body weight and self-rated health. This appears to be an interesting finding, given that the present analysis exploits a source of variation for identification that is very different from what is used in Cullinan and Gillespie (2016) and that, in consequence, the nature of the estimated effect also differs. While Cullinan and Gillespie (2016) rest on genetics as a persistent and long-term determinant of body weight, the present analysis uses short-term extrinsic incentives. Thus, the result of the former can be interpreted such that permanently reducing the BMI of an obese individual to a normal level will improve her SRH substantially. In contrast, our results—at least in terms of the point estimates—suggest that also a small reduction in body weight will make an obese individual instantaneously feel healthier. This distinction is important for the question of how to motivate obese individuals to lose weight. Even if substantial weight loss is known to pay-off in the long-run for sure, obese individuals may still need some instantaneous improvement in subjectively perceived health in order to keep the discipline to continue their weight loss efforts.

Comparing our results in quantitative terms to those of Cullinan and Gillespie (2016) is not straight forward as their analysis relies on a very different source of weight variation and since they consider a categorical measure of the weight status (healthy weight, overweight, obese grade I, obese grade II) as key explanatory variable, rather than a continuous measure of weight change as we do in our analysis. Nevertheless we use a simple simulation to translate our results into figures that allow for being compared to the results of Cullinan and Gillespie (2016). More specifically, based on the point estimates from the IV model (Table 6, column 3), for each participant we calculate the change in the probability of reporting to be in excellent health—this is the probability Cullinan and Gillespie (2016) focus on when discussing their results—that would occur if she reduced her body weight to a BMI of 25. We hence consider a shift from obesity to a healthy body weight as Cullinan and Gillespie (2016) do. Moreover, following the line of how they present their results, we examine the mean change in this probability separately for men and women and for grade II obese (BMI > 35) and grade I obese individuals.³² At least for grade I obese individuals our results are

³¹ They identify a strong link between *BMI* and *SRH* in obese individuals, while their instrumental variables estimation results are inconclusive for moderately overweight individuals.

³² Because some participants lost some weight during their rehab stay, this group includes few individuals who are not obese in the sense that BMI_0 does not exceed 30; see footnote 17. Yet, because of the small number of individuals to whom this applies and because of the fact that even the least overweight participant left the rehab clinic with a BMI that exceeded 28, we do not distinguish between being grade I obese and being just overweight.

surprisingly similar to those of Cullinan and Gillespie (2016). For women we calculate a mean change of 7.0 percentage points while the corresponding value in Cullinan and Gillespie (2016, IV with controls) is 9.1. For grade I obese men our approach yields a mean change of 9.7 percentage points while the corresponding value in Cullinan and Gillespie (2016) is 11.6. For the grade II obese the results are less well aligned, but the overall pattern is still similar. Cullinan and Gillespie (2016) report effects of 17.1 (women) and 20.1 (men) percentage points on the probability to be in excellent health, while we calculate mean effects of 31.7 (women) and 40.0 (men) percentage points, respectively.³³

4.3 Robustness checks

We ran several robustness checks in order to test how sensitive the results are to changes of the model specification. (i) We estimated all discussed model specifications with additional controls for education, employment, and income measured at rehab discharge; see Tables 12 and 13 in Appendix. This does not change the overall pattern of results. The coefficients of the naïve model are hardly affected. Their counterparts from instrumental variables estimation remain inconclusive, as they do not change consistently in the same direction. For *SRH* the point estimate of the *BMI loss* coefficient gets smaller, while it gets bigger for *PWB*. (ii) We reduced the number of health categories to just three merging ‘good’ with ‘excellent’ and ‘bad’ with ‘poor’. This just marginally affects the coefficient estimates; see Table 11 first panel in Appendix. As another robustness check, (iii) we excluded individuals with extreme changes in *BMI*, in order to check for few extraordinary cases possibly driving the empirical results. We considered two definitions of extreme, ($BMI\ loss > 5$) and ($BMI\ loss < -2 \mid BMI\ loss > 5$). For both, the pattern of results remains largely unchanged; see Table 17 and 18 in Appendix.

In order not to rely exclusively on fully parametric model, (iv) we also ran two-stage least squares (2SLS) regression. To avoid interpreting *SRH* as being measured on an interval scale, we recoded the dependent variable to have just two categories and, in consequence, we estimated linear probability models by 2SLS and as reference also by OLS. Since transforming five-category *SRH* to a binary health indicator involves the somewhat arbitrary choice of a cutoff category, we tried all four possible variants; see Table 11 second to fifth panel and Figs. 2–7 in Appendix. If the left-hand side variable is specified to indicate one of the two extreme categories (‘better than good’ or ‘bad’) the effect of *BMI loss* gets very small or even vanishes. Interestingly, this holds for both OLS and 2SLS. One possible explanation is that these categories are rather rare in the sample, hampering the identification of effects on these extreme categories; see Table 1. An alternative, less technical, explanation

³³ Note that heterogeneity across the grades of obesity is an empirical result in the analysis of Cullinan and Gillespie (2016). In our exercise the much stronger effect among the grade II obese occurs by construction, since we have to assume a much bigger weight loss in order to make the grade II obese reach the 25kg/m² threshold. The results from the above exercise have generally to be interpreted with much caution. For many (# 196) participants we assume a hypothetical reduction of body weight—35.2 BMI units at the extreme—that exceeds any weight loss that is actually observed in our data; cf. Table 5.

is that relatively small changes in body weight will rarely be the reason for a shift in subjective health perception to an extreme. If an interior category is chosen as cutoff, the linear model insofar mirrors the results of ordered probit estimation as the naïve estimator yields a significant and favorable effect of weight loss on *SRH*, while IV does not. For the variant with the left-hand side variable indicating poor or bad *SRH*, the 2SLS coefficient even turns negative. This does not apply to the variant with an indicator for ‘*SRH* neither good nor excellent’ serving as dependent variable. There, the 2SLS coefficients are similar or larger than their OLS counterparts.

Since due to drop-out the estimation sample is considerably smaller than the initial 695 recruited participants, non-random sample attrition might bias our results. (v) To address the concern, we estimated model specifications that include a third equation explaining experiment attrition that is jointly estimated with the remaining two. As an ‘instrument’ for attrition this additional equation includes a dummy variable ‘pharmacy in town’ which indicates whether the assigned pharmacy for the weigh-in lies in the same zip-code as the respondent’s place of residence. The results from these specifications suggest, that non-random sample attrition is likely to be no issue; see Table 15 and 16 in Appendix.³⁴

As displayed in Table 5, initial body weight varies substantially in our sample. To check whether this heterogeneity is mirrored by heterogeneity in the effect of weight loss on self-perceived health, (vi) we stratified the analysis by initial BMI. Figure 8 in Appendix displays the respective estimated coefficients³⁵ of *BMI loss*, which do not reveal a striking pattern of heterogeneity with respect to BMI_0 .

Since earlier studies found gender differences in the relationship between BMI and SRH (Imai et al. 2008), besides the pooled model, we also conduct separate regression analyses for males and females in the specification of reference as well as in the robustness checks. The naïve regression models yield results that are rather similar to those from the pooled model. However, for the stratified instrumental variable analysis the results deviate more from the naïve estimation compared to the pooled model: In the male sample, the estimated effect of *BMI loss* on *PWB* gets smaller, which is similar in the female sample for the effect of *BMI loss* on *SRH*. However, neither the cross-equation error correlation in the male nor the cross-equation error correlation in the female sample is significant, which is in line with our main results, stating that endogeneity of *BMI loss* does not seem to be a major issue. Due to the rather small sample sizes and the instrument becoming relatively weak in these sub-samples the instrumental variable analysis seems not to be overly informative and has to be taken with a grain of salt.

³⁴ Augurzyk et al. (2018) who use the same data as the present analysis also found, that selection-bias does not drive their results.

³⁵ We use our preferred set of right-hand-side variables and simple OLS as estimation method.

5 Discussion and conclusion

This paper analyzes the relationship between moderate weight loss and subjective health perception in obese individuals. We confirm the results of the related literature, which find a significant association of body weight and subjective health perception. Unlike the bulk of the existing literature, the present analysis is not only concerned with the association of body weight and self-rated health, but employs instrumental variables estimation to establish a causal link. In doing this, it follows Cullinan and Gillespie (2016) who also use instrumental variables estimation to identify a causal effect of BMI on SRH. Our analysis differs from this key reference, by tapping a completely different source of exogenous weight variation. While Cullinan and Gillespie (2016) use BMI of biological relatives as instrument and rest on exogenous genetically determined, long-term, between-individuals weight variation for identification, we use cash incentives of a weight loss intervention as instrument and, hence, rely on short-term, within-individual variation. Though, our instrumental variable approach does not establish statistical significance of the effect under scrutiny, the pattern of results suggests that the positive association of subjective health perception and weight loss is not primarily due to unobserved confounders. Our results hence appear not to conflict with the finding of Cullinan and Gillespie (2016). It, nevertheless, adds a relevant aspect to the insights into how weight loss affects subjective health perception in obese individuals. While Cullinan and Gillespie (2016) establish that obese individuals' health perception will improve if they manage to become normal-weight, our analysis yields some evidence for subjective health improvements accompanying even small initial weight reduction. This finding may encourage obese individuals in their weight loss attempts, since they can expect to be immediately rewarded for their efforts by subjective health improvements.

With respect to external validity our results have, however, to be interpreted with some caution. Our analysis uses a very specific sample of obese individuals. In particular—besides being admitted to a rehab clinic and meeting the inclusion criteria discussed in section 2—all participants actively selected themselves into the sample by agreeing to participate in the field experiment. It is, hence, likely that our results are based on a sample that is selective with respect to the motivation for losing body weight and probably with respect to the subjective likelihood of being successful in reducing overweight. Our findings may hence not one-to-one apply to the obese population in general. Nevertheless, we regard our results as relevant as our discussion focusses on obese individuals who try to lose body weight that is a subpopulation which in this respect is similarly selected as the study population. Moreover, as discussed above, our results are astonishingly similar to earlier findings that are based on very different samples of data. We take this as evidence that our conclusions are not completely specific to our study population.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included

in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Appendix

See Tables 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18.

Table 7 Joint and marginal distribution of PWB_0 and SRH_0

		SRH_0					Marginal distribution
		Bad	Poor	Satisfactory	Good	Excellent	
PWB_0	Bad	12	14	3	1	0	30
	Poor	6	73	55	14	1	149
	Satisfactory	0	27	127	42	1	197
	Good	0	3	25	63	3	94
	Excellent	0	0	1	3	6	10
Marginal distribution		18	117	211	123	11	480

Table 8 Descriptive statistics for estimation sample (PWB_1 as dep. var.)

	Obs.	Mean	S.D.	Min.	Max.	Median
Dependent variable						
PWB_1	468	2.97	1.00	1	5	3
Explanatory variables						
BMI_{loss}	487	1.57	1.99	-8.28	12.78	1.52
BMI_0	487	37.26	6.15	28.03	60.22	36.05
PWB_0	487	2.79	0.89	1	5	3
<i>female</i>	487	0.32	0.47	0	1	0
<i>age</i>	487	49.05	8.50	20	68	50
Instrumental variable						
<i>incentive</i>	487	0.70	0.46	0	1	1

The number of observations differs between the main (468) and the auxiliary (487) equation of the econometric model. Descriptive statistics for the explanatory variables are for the auxiliary equation

Table 9 Coefficient estimates of Naïve ordered probit and IV-ordered probit models for the male sample

Dependent variable	Naïve ordered probit		IV-ordered probit	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.135*** (0.031)	0.174*** (0.032)	0.146 (0.155)	0.018 (0.151)
<i>BMI</i> ₀	-0.016 (0.011)	-0.011 (0.011)	-0.017 (0.014)	-0.001 (0.015)
<i>SRH</i> ₀	0.611*** (0.076)	-	0.608*** (0.085)	-
<i>PWB</i> ₀	-	0.574*** (0.073)	-	0.580*** (0.074)
<i>age</i>	-0.020*** (0.007)	-0.009 (0.007)	-0.020*** (0.008)	-0.011 (0.007)
Marginal effect of <i>BMI loss</i> [†]	0.030*** (0.007)	0.048*** (0.008)	0.032 (0.034)	0.005 (0.043)
Auxiliary equation (dependent variable: <i>BMI loss</i>)				
<i>incentive</i>	-	-	0.821*** (0.226)	0.857*** (0.226)
<i>BMI</i> ₀	-	-	0.061*** (0.018)	0.066*** (0.019)
<i>SRH</i> ₀	-	-	0.232* (0.126)	-
<i>PWB</i> ₀	-	-	-	0.221* (0.121)
<i>age</i>	-	-	-0.010 (0.013)	-0.013 (0.013)
<i>Constant</i>	-	-	-1.347 (1.152)	-1.360 (1.136)
<i>Cross-eq. error corr.</i>	-	-	-0.023 (0.309)	0.306 (0.292)
Instrument relevance (<i>F</i> -statistic)	-	-	12.96	14.16
No. of observations (over all)	329	319	335	329
No. of observations (main equation)	-	-	329	319

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; [†]Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better. Number of observations differ in the IV-Ordered Probit model: Due to joint estimation, participants for which *SRH*₁ and *PWB*₁, respectively, is not observed but the endogenous regressor *BMI loss* is observed contribute to the log-likelihood function and enter the estimation sample. Cut points of the ordered probit estimations are displayed in Table 14 in the Appendix

Table 10 Coefficient estimates of Naïve ordered probit and IV-ordered probit models for the female sample

Dependent variable	Naïve ordered probit		IV-ordered probit	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.192*** (0.047)	0.131*** (0.047)	0.038 (0.147)	0.130 (0.144)
<i>BMI</i> ₀	0.009 (0.013)	0.008 (0.014)	0.015 (0.014)	0.008 (0.016)
<i>SRH</i> ₀	0.586*** (0.106)	–	0.569*** (0.111)	–
<i>PWB</i> ₀	–	0.467*** (0.103)	–	0.467*** (0.105)
<i>age</i>	0.010 (0.010)	0.001 (0.010)	0.007 (0.010)	0.001 (0.010)
Marginal effect of <i>BMI loss</i> [†]	0.057*** (0.013)	0.044*** (0.015)	0.012 (0.046)	0.044 (0.046)
Auxiliary equation (dependent variable: <i>BMI loss</i>)				
<i>incentive</i>	–	–	1.402*** (0.318)	1.412*** (0.332)
<i>BMI</i> ₀	–	–	0.033 (0.021)	0.040* (0.022)
<i>SRH</i> ₀	–	–	0.047 (0.154)	–
<i>PWB</i> ₀	–	–	–	0.164 (0.160)
<i>age</i>	–	–	–0.023 (0.016)	–0.027 (0.017)
<i>Constant</i>	–	–	0.114 (1.308)	–0.257 (1.278)
<i>Cross-eq. error corr.</i>	–	–	0.295 (0.267)	0.002 (0.272)
Instrument relevance (<i>F</i> -statistic)	–	–	18.80	17.47
No. of observations (over all)	156	149	162	158
No. of observations (main equation)	–	–	156	149

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

[†] Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better. Number of observations differ in the IV-Ordered Probit model: Due to joint estimation, participants for which *SRH*₁ and *PWB*₁, respectively, is not observed but the endogenous regressor *BMI loss* is observed contribute to the log-likelihood function and enter the estimation sample. Cut points of the ordered probit estimations are displayed in Table 14 in the Appendix

Table 11 Estimated coefficients of *BMI loss* for different specifications of the dependent variable

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
Trinary						
<i>Ordered Probit</i>	0.172*** (0.028)	0.174*** (0.029)	0.162*** (0.034)	0.187*** (0.035)	0.200*** (0.050)	0.151*** (0.050)
<i>IV-ordered probit</i>	0.110 (0.120)	0.047 (0.119)	0.166 (0.167)	0.007 (0.162)	-0.024 (0.148)	0.086 (0.156)
Marginal effect ordered Probit [†]	0.042*** (0.007)	0.050*** (0.008)	0.035*** (0.008)	0.051*** (0.009)	0.058*** (0.013)	0.050*** (0.015)
Marginal effect IV-ordered Probit [†]	0.028 (0.029)	0.014 (0.035)	0.036 (0.036)	0.002 (0.046)	-0.007 (0.047)	0.030 (0.052)
Instrument relevance	27.38	28.05	12.80	13.84	18.69	17.24
No. of observations (over all)	497	487	335	329	162	158
No. of observations (main equation)	485	468	329	319	156	149
Binary_{bad}						
<i>OLS</i>	0.014*** (0.005)	0.015** (0.007)	0.011** (0.005)	0.015** (0.006)	0.022* (0.013)	0.013 (0.017)
<i>2 SLS</i>	0.030 (0.024)	0.014 (0.027)	0.015 (0.028)	-0.001 (0.034)	0.045 (0.038)	0.032 (0.045)
Instrument relevance	28.7	28.82	12.96	14.25	20.05	17.23
Binary_{poor}						
<i>OLS</i>	0.040*** (0.009)	0.045*** (0.010)	0.038*** (0.009)	0.047*** (0.010)	0.050** (0.020)	0.042* (0.023)
<i>2 SLS</i>	-0.026 (0.039)	-0.015 (0.043)	-0.012 (0.050)	-0.007 (0.059)	-0.059 (0.056)	-0.031 (0.062)
Instrument relevance	27.42	28.18	13.09	13.81	18.73	17.24
Binary_{satisfactory}						
<i>OLS</i>	0.055*** (0.010)	0.059*** (0.010)	0.048*** (0.012)	0.061*** (0.013)	0.071*** (0.017)	0.055*** (0.016)
<i>2 SLS</i>	0.083* (0.044)	0.042 (0.043)	0.084 (0.064)	0.002 (0.064)	0.054 (0.052)	0.091* (0.049)
Instrument relevance	27.10	27.65	12.38	13.55	18.80	17.19

Table 11 continued

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
Binary_{good}						
<i>OLS</i>	0.004 (0.004)	0.014** (0.007)	0.004 (0.006)	0.020** (0.009)	0.003 (0.003)	0.002 (0.002)
<i>2 SLS</i>	-0.008 (0.019)	0.001 (0.020)	-0.008 (0.029)	0.003 (0.031)	-0.012 (0.019)	-0.002 (0.021)
Instrument relevance	26.91	27.92	12.32	13.49	18.59	17.40
No. of observations	485	468	329	319	156	149

Trinary specification of *SRH* and *PWB*: standard errors in parentheses

† Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better. *SRH/PWB* (trinary) equals 1, for poor and bad *SRH/PWB*, 2 for satisfactory *SRH/PWB* and 3 for good and excellent *SRH/PWB*; Binary specifications of *SRH/PWB*: Robust standard errors in parentheses; *binary_{bad}* equals 0, for bad *SRH/PWB*, 1 for poor, satisfactory, good and excellent *SRH/PWB*. *Binary_{poor}* equals 0, for bad and poor *SRH/PWB*, 1 for satisfactory, good and excellent *SRH/PWB*. *binary_{satisfactory}* equals 0, for bad, poor and satisfactory *SRH/PWB*, 1 for good and excellent *SRH/PWB*. *binary_{good}* equals 0, for bad, poor, satisfactory and good *SRH/PWB*, 1 for excellent *SRH/PWB*. OLS stands for ordinary least squares, 2 SLS stands for two-stage least squares. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table 12 Estimated coefficients of ordered probit model with additional controls

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.166*** (0.030)	0.163*** (0.030)	0.155*** (0.036)	0.178*** (0.036)	0.197*** (0.055)	0.118** (0.054)
<i>BMI</i> ₀	-0.010 (0.009)	-0.009 (0.010)	-0.024** (0.012)	-0.014 (0.012)	0.014 (0.016)	0.003 (0.017)
<i>SRH</i> ₀	0.636*** (0.069)	-	0.670*** (0.087)	-	0.602*** (0.119)	-
<i>PWB</i> ₀	-	0.526*** (0.065)	-	0.562*** (0.080)	-	0.483*** (0.116)
<i>female</i>	-0.345*** (0.120)	-0.304** (0.120)	-	-	-	-
<i>age</i>	-0.010 (0.007)	-0.003 (0.007)	-0.018** (0.008)	-0.005 (0.008)	0.006 (0.012)	0.002 (0.012)
<i>No secondary</i>	-1.044** (0.485)	-0.494 (0.462)	-1.366** (0.590)	-0.491 (0.577)	-0.660 (0.929)	-0.532 (0.824)
<i>Lower secondary</i>	-0.573* (0.285)	-0.368 (0.368)	-0.901** (0.368)	-0.392 (0.392)	0.137 (0.392)	-0.222 (0.392)

Table 12 continued

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>Education</i>	(0.300)	(0.292)	(0.379)	(0.364)	(0.510)	(0.500)
<i>Medium secondary</i>	-0.556*	-0.287	-0.832**	-0.314	0.065	-0.211
<i>Education</i>	(0.311)	(0.304)	(0.392)	(0.378)	(0.539)	(0.536)
<i>Higher secondary</i>	-0.779**	-0.247	-1.039**	-0.327	-0.407	0.040
<i>Education</i>	(0.351)	(0.344)	(0.432)	(0.417)	(0.627)	(0.641)
<i>Employment</i> ₀	0.071	-0.001	0.126	-0.135	-0.014	0.269
	(0.160)	(0.162)	(0.210)	(0.206)	(0.251)	(0.270)
<i>Low income</i> ₀	-0.151	-0.299*	-0.101	-0.171	-0.239	-0.678**
	(0.172)	(0.171)	(0.209)	(0.207)	(0.345)	(0.342)
<i>Medium income</i> ₀	-0.037	-0.113	-0.020	-0.103	0.075	-0.305
	(0.143)	(0.141)	(0.163)	(0.161)	(0.322)	(0.318)
Marginal effect of <i>BMI loss</i> [†]	0.040***	0.048***	0.033***	0.050***	0.056***	0.038**
	(0.007)	(0.008)	(0.008)	(0.010)	(0.014)	(0.017)
No. of observations	412	397	281	273	131	124

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; for education/income, individuals with university degree/high income are the reference category, respectively; [†]Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better

Table 13 Estimated coefficients of instrumental variable ordered probit model with additional controls

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.082 (0.124)	0.114 (0.119)	0.035 (0.177)	0.017 (0.167)	0.078 (0.174)	0.250* (0.150)
<i>BMI</i> ₀	-0.006 (0.011)	-0.006 (0.011)	-0.018 (0.015)	-0.006 (0.014)	0.020 (0.017)	-0.005 (0.019)
<i>SRH</i> ₀	0.640*** (0.069)	-	0.671*** (0.090)	-	0.603*** (0.120)	-
<i>PWB</i> ₀	-	0.530*** (0.065)	-	0.559*** (0.084)	-	0.443*** (0.132)
<i>female</i>	-0.350*** (0.120)	-0.308** (0.120)	-	-	-	-
<i>age</i>	-0.012 (0.007)	-0.005 (0.007)	-0.020** (0.009)	-0.008 (0.009)	0.003 (0.013)	0.007 (0.013)
<i>No secondary School education</i>	-1.118** (0.491)	-0.536 (0.471)	-1.641** (0.676)	-0.877 (0.673)	-0.337 (1.031)	-0.837 (0.877)
<i>Lower secondary Education</i>	-0.551* (0.302)	-0.357 (0.293)	-0.917** (0.376)	-0.424 (0.361)	0.247 (0.527)	-0.368 (0.521)
<i>Medium secondary Education</i>	-0.535* (0.312)	-0.277 (0.305)	-0.872** (0.390)	-0.379 (0.378)	0.201 (0.565)	-0.398 (0.570)
<i>Higher secondary Education</i>	-0.808** (0.351)	-0.267 (0.347)	-1.149*** (0.445)	-0.489 (0.439)	-0.292 (0.646)	-0.122 (0.663)
<i>Employment</i> ₀	0.074 (0.159)	0.004 (0.163)	0.105 (0.212)	-0.123 (0.206)	0.018 (0.254)	0.253 (0.270)
<i>Low income</i> ₀	-0.146 (0.171)	-0.297* (0.171)	-0.076 (0.211)	-0.117 (0.214)	-0.250 (0.344)	-0.604* (0.357)
<i>Medium income</i> ₀	-0.026 (0.143)	-0.105 (0.142)	0.010 (0.168)	-0.058 (0.167)	0.065 (0.321)	-0.289 (0.318)
Marginal effect of <i>BMI loss</i> [†]	0.020 (0.029)	0.034 (0.035)	0.008 (0.038)	0.005 (0.048)	0.023 (0.050)	0.079* (0.043)

Table 13 continued

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
Auxiliary equation (dependent variable: <i>BMI loss</i>)						
<i>incentive</i>	0.982*** (0.193)	1.017*** (0.197)	0.808*** (0.232)	0.848*** (0.234)	1.358*** (0.343)	1.368*** (0.365)
<i>BMI</i> ₀	0.047*** (0.015)	0.051*** (0.016)	0.047** (0.019)	0.050** (0.020)	0.043* (0.023)	0.052** (0.025)
<i>SRH</i> ₀	0.132 (0.106)	–	0.139 (0.133)	–	0.101 (0.168)	–
<i>PWB</i> ₀	–	0.149 (0.103)	–	0.148 (0.126)	–	0.203 (0.174)
<i>female</i>	–0.167 (0.195)	–0.152 (0.198)	–	–	–	–
<i>age</i>	–0.019* (0.011)	–0.021* (0.011)	–0.019 (0.014)	–0.018 (0.014)	–0.025 (0.018)	–0.030 (0.019)
<i>No secondary School education</i>	–0.860 (0.787)	–0.854 (0.771)	–2.405** (0.955)	–2.424** (0.949)	2.633* (1.399)	1.850 (1.295)
<i>Lower secondary Education</i>	0.300 (0.470)	0.226 (0.489)	–0.150 (0.602)	–0.122 (0.603)	0.794 (0.723)	0.615 (0.817)
<i>Medium secondary Education</i>	0.257 (0.489)	0.209 (0.510)	–0.318 (0.625)	–0.293 (0.627)	0.828 (0.778)	0.737 (0.879)
<i>Higher secondary Education</i>	–0.355 (0.559)	–0.397 (0.576)	–0.957 (0.695)	–0.925 (0.693)	0.701 (0.923)	0.641 (1.020)
<i>Employment</i> ₀	0.132 (0.263)	0.194 (0.270)	–0.046 (0.347)	0.121 (0.347)	0.350 (0.381)	0.222 (0.417)
<i>Low income</i> ₀	–0.113 (0.283)	–0.128 (0.287)	0.094 (0.344)	0.182 (0.346)	–0.525 (0.535)	–0.761 (0.551)
<i>Medium income</i> ₀	0.054 (0.235)	0.084 (0.237)	0.223 (0.269)	0.235 (0.269)	–0.428 (0.500)	–0.398 (0.518)
<i>Constant</i>	–0.586 (1.053)	–0.703 (1.056)	0.035 (1.324)	–0.331 (1.343)	–0.816 (1.741)	–0.871 (1.732)
<i>Cross-eq. error corr.</i>	0.159 (0.225)	0.093 (0.218)	0.226 (0.324)	0.299 (0.306)	0.218 (0.300)	–0.270 (0.310)
Instrument relevance (<i>F</i> -statistic)	24.96	25.72	11.59	12.6	14.24	12.74
No. of observations (over all)	421	411	286	281	135	130

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; for education/income, individuals with university degree/high income are the reference category, respectively; †Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better

Table 14 Cut points of Naïve ordered probit and instrumental variable ordered probit estimation (coefficient estimates)

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>Naive ordered probit estimation</i>						
Cut 1: bad to poor	-0.626 (0.501)	-0.363 (0.491)	-1.517** (0.647)	-0.719 (0.635)	1.136 (0.818)	0.513 (0.782)
Cut 2: poor to satisfactory	0.360 (0.500)	0.625 (0.491)	-0.619 (0.644)	0.299 (0.634)	2.253*** (0.824)	1.466* (0.784)
Cut 3: satisfactory to good	1.715*** (0.504)	1.797*** (0.494)	0.806 (0.644)	1.469** (0.636)	3.515*** (0.844)	2.654*** (0.799)
Cut 4: good to excellent	3.435*** (0.518)	3.304*** (0.508)	2.505*** (0.653)	2.988*** (0.649)	5.420*** (0.911)	4.161*** (0.852)
No. of observations	485	468	329	319	156	149
<i>Instrumental variable ordered probit estimation</i>						
Cut 1: bad to poor	-0.621 (0.501)	-0.335 (0.491)	-1.525** (0.653)	-0.588 (0.648)	1.041 (0.822)	0.513 (0.782)
Cut 2: poor to satisfactory	0.365 (0.500)	0.640 (0.490)	-0.627 (0.652)	0.386 (0.633)	2.116** (0.844)	1.466* (0.785)
Cut 3: satisfactory to good	1.718*** (0.504)	1.796*** (0.494)	0.797 (0.653)	1.506** (0.632)	3.333*** (0.886)	2.654*** (0.799)
Cut 4: good to excellent	3.437*** (0.517)	3.282*** (0.514)	2.496*** (0.664)	2.960*** (0.661)	5.171*** (0.992)	4.161*** (0.852)
No. of observations (over all)	497	487	335	329	162	158
No. of observations (main equation)	485	468	329	319	156	149

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table 15 Naïve ordered probit estimation, controlling for selection (coefficient estimates)

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.145*** (0.029)	0.142*** (0.027)	0.122*** (0.033)	0.172*** (0.032)	0.192*** (0.047)	0.122*** (0.046)
<i>BMI</i> ₀	-0.002 (0.009)	-0.010 (0.008)	-0.009 (0.011)	-0.010 (0.011)	0.009 (0.015)	0.002 (0.014)
<i>SRH</i> ₀	0.563*** (0.095)	-	0.543*** (0.104)	-	0.587*** (0.119)	-
<i>PWB</i> ₀	-	0.513*** (0.065)	-	0.561*** (0.079)	-	0.475*** (0.101)
<i>female</i>	-0.359*** (0.107)	-0.253** (0.111)	-	-	-	-
<i>age</i>	-0.012* (0.007)	0.002 (0.006)	-0.024*** (0.007)	-0.011 (0.008)	0.010 (0.014)	0.007 (0.010)
Marginal effect of <i>BMI loss</i> [†]	0.033*** (0.008)	0.048*** (0.008)	0.023*** (0.007)	0.045*** (0.009)	0.057*** (0.015)	0.043*** (0.015)
No. of observations	665	650	453	447	212	203

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; [†]Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better

Table 16 Instrumental variable ordered probit estimation, controlling for selection (coefficient estimates)

Dependent variable	All		Males		Females	
	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁	<i>SRH</i> ₁	<i>PWB</i> ₁
<i>BMI loss</i>	0.061 (0.126)	0.070 (0.131)	0.043 (0.139)	-0.049 (0.134)	0.018 (0.198)	0.209 (0.131)
<i>BMI</i> ₀	0.002 (0.012)	-0.001 (0.015)	-0.003 (0.014)	0.005 (0.014)	0.017 (0.016)	-0.003 (0.016)
<i>SRH</i> ₀	0.575*** (0.088)	-	0.527*** (0.094)	-	0.556*** (0.151)	-
<i>PWB</i> ₀	-	0.552*** (0.061)	-	0.557*** (0.085)	-	0.456*** (0.109)
<i>female</i>	-0.377*** (0.106)	-0.331** (0.142)	-	-	-	-
<i>age</i>	-0.013* (0.007)	-0.004 (0.010)	-0.025*** (0.007)	-0.013 (0.008)	0.006 (0.014)	0.009 (0.011)
Marginal effect of <i>BMI loss</i> [†]	0.014 (0.029)	0.022 (0.043)	0.008 (0.026)	-0.014 (0.038)	0.006 (0.062)	0.072* (0.041)
No. of observations	665	650	453	447	212	203

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; [†]Sample average of individual marginal effects of *BMI loss* on the probability of rating one's health or physical well-being 'satisfactory' or better

Table 17 Naïve ordered probit estimation, excluding extreme cases (coefficient estimates)

Dependent variable	Whole sample		$BMI\ loss \leq 5$		$-2 \leq BMI\ loss \leq 5$	
	SRH_1	PWB_1	SRH_1	PWB_1	SRH_1	PWB_1
$BMI\ loss$	0.150*** (0.026)	0.160*** (0.026)	0.179*** (0.031)	0.166*** (0.031)	0.204*** (0.036)	0.215*** (0.037)
BMI_0	-0.005 (0.008)	-0.004 (0.008)	-0.001 (0.009)	-0.003 (0.009)	-0.003 (0.009)	-0.005 (0.009)
SRH_0	0.598*** (0.062)	-	0.594*** (0.063)	-	0.590*** (0.064)	-
PWB_0	-	0.542*** (0.059)	-	0.558*** (0.060)	-	0.554*** (0.062)
<i>female</i>	-0.362*** (0.107)	-0.339*** (0.107)	-0.380*** (0.109)	-0.317*** (0.109)	-0.375*** (0.110)	-0.327*** (0.111)
<i>age</i>	-0.009 (0.006)	-0.005 (0.006)	-0.008 (0.006)	-0.005 (0.006)	-0.006 (0.006)	-0.003 (0.006)
Marginal effect of $BMI\ loss^\dagger$	0.037*** (0.006)	0.047*** (0.007)	0.045*** (0.007)	0.050*** (0.009)	0.051*** (0.009)	0.064*** (0.010)
No. of observations	485	468	464	448	449	433

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; † Sample average of individual marginal effects of $BMI\ loss$ on the probability of rating one's health or physical well-being 'satisfactory' or better

Table 18 Instrumental variable ordered probit estimation, excluding extreme cases (coefficient estimates)

Dependent variable	Whole Sample		$BMI\ loss \leq 5$		$-2 \leq BMI\ loss \leq 5$	
	SRH_1	PWB_1	SRH_1	PWB_1	SRH_1	PWB_1
$BMI\ loss$	0.126 (0.111)	0.072 (0.110)	0.117 (0.132)	0.037 (0.125)	0.143 (0.169)	0.073 (0.161)
BMI_0	-0.004 (0.010)	0.001 (0.011)	0.000 (0.009)	0.001 (0.009)	-0.002 (0.010)	-0.001 (0.010)
SRH_0	0.601*** (0.062)	-	0.600*** (0.063)	-	0.597*** (0.065)	-
PWB_0	-	0.550*** (0.059)	-	0.569*** (0.060)	-	0.570*** (0.062)
<i>female</i>	-0.367*** (0.109)	-0.357*** (0.108)	-0.376*** (0.109)	-0.314*** (0.109)	-0.372*** (0.111)	-0.318*** (0.112)
<i>age</i>	-0.009 (0.006)	-0.007 (0.006)	-0.008 (0.006)	-0.006 (0.006)	-0.007 (0.006)	-0.005 (0.006)
Marginal effect of $BMI\ loss^\dagger$	0.032 (0.027)	0.022 (0.033)	0.030 (0.033)	0.011 (0.038)	0.036 (0.042)	0.023 (0.049)
No. of obs. (over all)	497	487	476	467	461	452

Standard errors in parentheses; * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$; † Sample average of individual marginal effects of $BMI\ loss$ on the probability of rating one's health or physical well-being 'satisfactory' or better

See Figs. 2, 3, 4, 5, 6, 7 and 8.

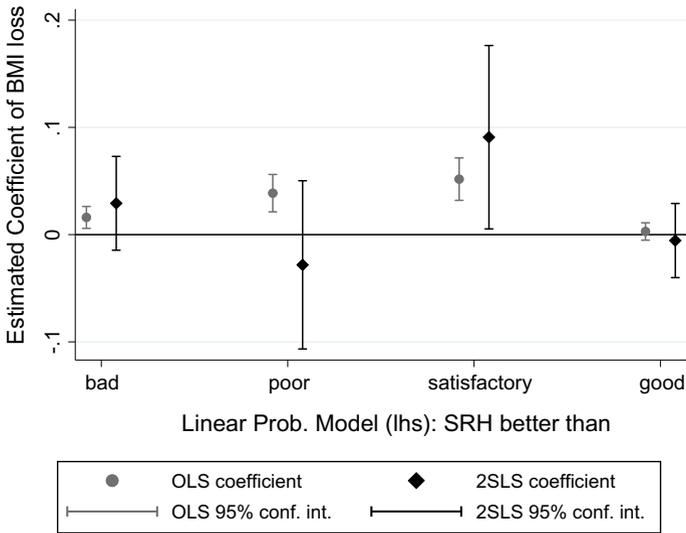


Fig. 2 Est. coef. for linear prob. model (*self-rated health, full sample*)

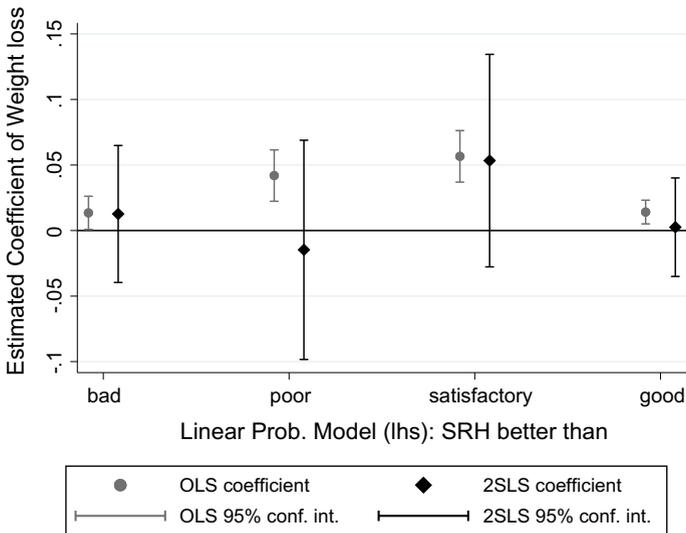


Fig. 3 Est. coef. for linear prob. model (*physical well-being, full sample*)

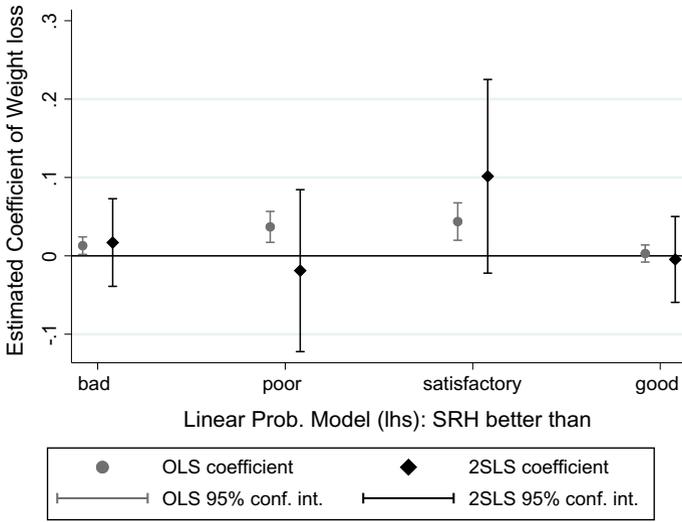


Fig. 4 Est. coef. for linear prob. model (*self-rated health, men*)

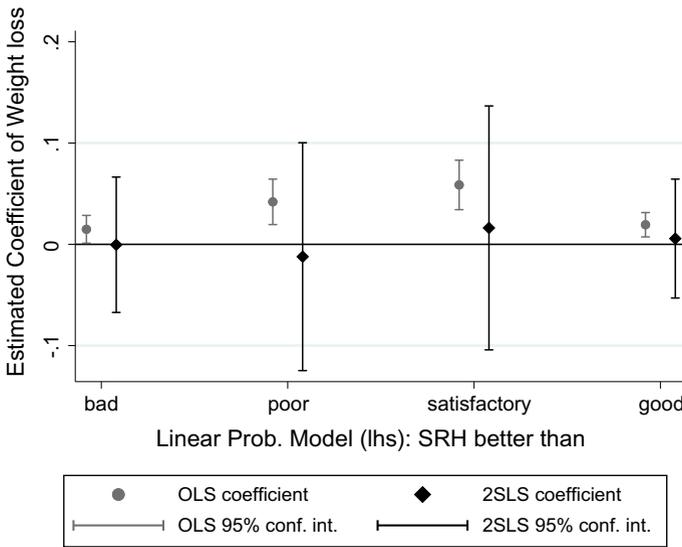


Fig. 5 Est. coef. for linear prob. model (*physical well-being, men*)

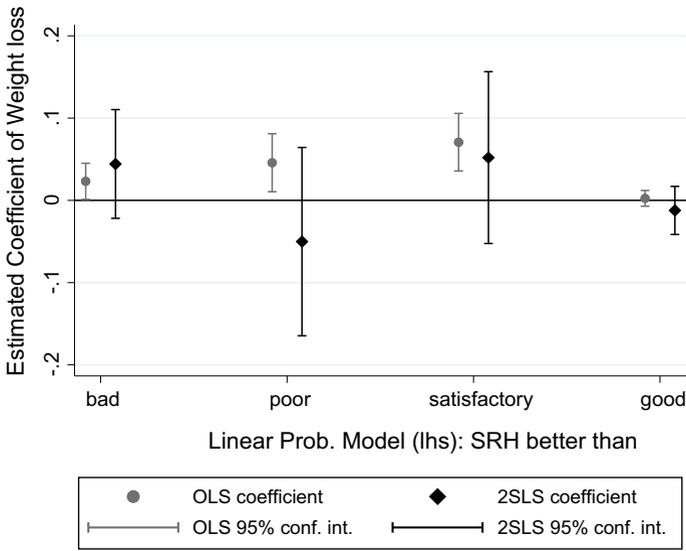


Fig. 6 Est. coef. for linear prob. model (*self-rated health, women*)

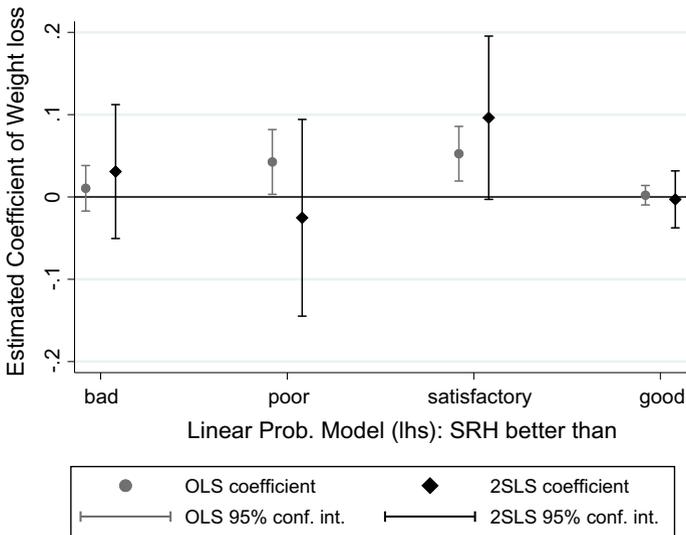


Fig. 7 Est. coef. for linear prob. model (*physical well-being, women*)

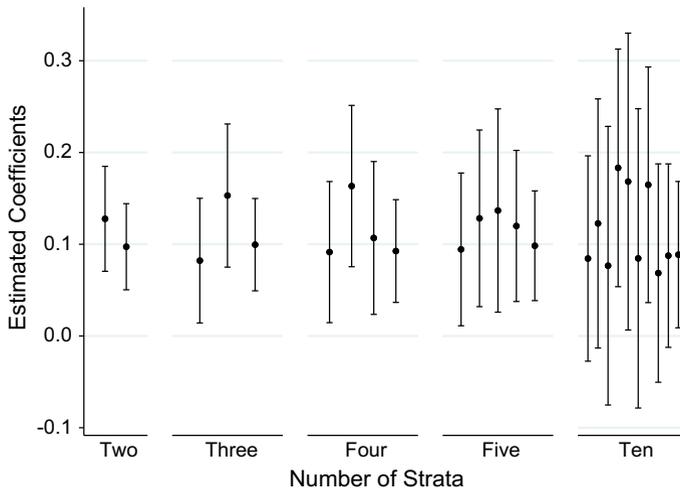


Fig. 8 Est. coef. (with 95% conf. int.) from OLS estimation stratified by different quantiles of the initial weight distribution (dependent variable: *self-rated health*)

References

- Ali MM, Rizzo JA, Amialchuk A, Heiland F (2014) Racial differences in the influence of female adolescents' body size on dating and sex. *Econ Hum Biol* 12:140–152
- Angrist JD, Pischke J-S (2009) *Mostly harmless econometrics: an empiricist's companion*, 1st edn. Princeton University Press, Princeton
- Augurzky B, Bauer TK, Reichert AR, Schmidt CM, Tauchmann H (2014) Small cash rewards for big losers—experimental insights into the fight against the obesity epidemic. *Ruhr economic papers*, 530
- Augurzky B, Bauer TK, Reichert AR, Schmidt CM, Tauchmann H (2018) Habit formation, obesity, and cash rewards. *Ruhr economic papers*, 750
- Baruth M, Becofsky K, Wilcox S, Goodrich K (2014) Health characteristics and health behaviors of African American adults according to self-rated health status. *Ethn Dis* 24(1):97–103
- Blackburn G (1995) Effect of degree of weight loss on health benefits. *Obes Res* 3(S2):211s–216s
- Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ (2003) Overweight, obesity, and mortality from cancer in a prospectively studied cohort of us adults. *N Engl J Med* 348(17):1625–1638
- Cawley J, Meyerhoefer C (2012) The medical care costs of obesity: an instrumental variables approach. *J Health Econ* 31(1):219–230
- Cawley J, Price JA (2013) A case study of a workplace wellness program that offers financial incentives for weight loss. *J Health Econ* 32(5):794–803
- Cullinan J, Gillespie P (2016) Does overweight and obesity impact on self-rated health? Evidence using instrumental variables ordered probit models. *Health Econ* 25(10):1341–1348
- Darviri C, Fouka G, Gnardellis C, Artemiadis AK, Tigani X, Alexopoulos EC (2012) Determinants of self-rated health in a representative sample of a rural population: a cross-sectional study in Greece. *Int J Environ Res Public Health* 9(3):943–954
- Di Angelantonio E (2016) Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet* 388:776–786
- Elks C, Den Hoed M, Zhao JH, Sharp S, Wareham N, Loos R, Ong K (2012) Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol* 3:29
- Ferraro KF, Yu Y (1995) Body weight and self-ratings of health. *J Health Soc Behav* 36(3):274–284
- Fletcher JM, Lehrer SF (2011) Genetic lotteries within families. *J Health Econ* 30(4):647–659
- Fontaine K, Barofsky I (2001) Obesity and health-related quality of life. *Obes Rev* 2(3):173–182
- Frangé C, de Queiroz SS, da Silva Prado JM, Tufik S, de Mello MT (2014) The impact of sleep duration on self-rated health. *Sleep Sci* 7(2):107–113

- Garn S, Leonard W, Hawthorne V (1986) Three limitations of the body mass index. *Am J Clin Nutr* 44(6):996–997
- Gorber SC, Tremblay M, Moher D, Gorber B (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev* 8(4):307–326
- Guallar-Castillón P, López GE, Lozano PL, Gutiérrez-Fisac J, Banegas BJ, Lafuente UP, Rodríguez AF (2002) The relationship of overweight and obesity with subjective health and use of health-care services among Spanish women. *Int J Obes Relat Metab Disord J Int Assoc Study Obes* 26(2):247–252
- Hubert HB, Feinleib M, McNamara PM, Castelli WP (1983) Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the framingham heart study. *Circulation* 67(5):968–977
- Imai K, Gregg EW, Chen YJ, Zhang P, Rekeneire N, Williamson DF (2008) The association of BMI with functional status and self-rated health in US adults. *Obesity* 16(2):402–408
- Johansson E, Böckerman P, Kiiskinen U, Heliövaara M (2009) Obesity and labour market success in Finland: the difference between having a high BMI and being fat. *Econ Hum Biol* 7(1):36–45
- John L, Loewenstein G, Troxel A, Norton L, Fassbender J, Volpp K (2011) Financial incentives for extended weight loss: a randomized, controlled trial. *J Gen Intern Med* 26:621–626
- Kepka D, Ayala GX, Cherrington A (2007) Do Latino immigrants link self-rated health with BMI and health behaviors? *Am J Health Behav* 31(5):535–544
- Kline B, Tobias JL (2008) The wages of BMI: Bayesian analysis of a skewed treatment-response model with nonparametric endogeneity. *J Appl Econ* 23(6):767–793
- Kolotkin R, Meter K, Williams G (2001) Quality of life and obesity. *Obes Rev* 2(4):219–229
- Kroes M, Osei-Assibey G, Baker-Searle R, Huang J (2016) Impact of weight change on quality of life in adults with overweight/obesity in the United States: a systematic review. *Curr Med Res Opin* 32(3):485–508
- Lindeboom M, Lundborg P, van der Klaauw B (2010) Assessing the impact of obesity on labor market outcomes. *Econ Human Biol* 8(3):309–319
- Loveman E, Frampton G, Shepherd J, Picot J, Cooper K, Bryant J, Welch K, Clegg A (2011) The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. *Health Technol Assess* 15(2):
- Macmillan R, Duke N, Oakes JM, Liao W (2011) Trends in the association of obesity and self-reported overall health in 30 years of the integrated health interview series. *Obesity* 19(5):1103–1105
- Maes HHM, Neale MC, Eaves LJ (1997) Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 27(4):325–351
- Marques-Vidal P, Ravasco P, Paccaud F (2012) Differing trends in the association between obesity and self-reported health in Portugal and Switzerland. Data from national health surveys 1992–2007. *BMC Public Health* 12(588):1–8
- Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, Marks JS (2003) Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA* 289(1):76–79
- Molarius A, Berglund K, Eriksson C, Lambe M, Nordström E, Eriksson HG, Feldman I (2007) Socio-economic conditions, lifestyle factors, and self-rated health among men and women in Sweden. *Eur J Public Health* 17(2):125–133
- Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, Mullany EC, Biryukov S, Abbafati C, Abera SF et al (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the global burden of disease study 2013. *Lancet* 384(9945):766–781
- Norton EC, Han E (2008) Genetic information, obesity, and labor market outcomes. *Health Econ* 17(9):1089–1104
- Okusun IS, Choi S, Matamoros T, Dever GA (2001) Obesity is associated with reduced self-rated general health status: evidence from a representative sample of white, black, and Hispanic Americans. *Prev Med* 32(5):429–436
- Paloyo A, Reichert AR, Reinermann H, Tauchmann H (2014) The causal link between financial incentives and weight loss: an evidence-based survey of the literature. *J Econ Surv* 28(3):401–420
- Patel SR, Hu FB (2008) Short sleep duration and weight gain: a systematic review. *Obesity* 16(3):643–653
- Pei Z, Pischke J-S, Schwandt H (2019) Poorly measured confounders are more useful on the left than on the right. *J Bus Econ Stat* 37(2):205–216
- Phillips LJ, Hammock RL, Blanton JM (2005) Predictors of self-rated health status among Texas residents. *Prev Chronic Dis* 2(4):

- Price J, Swigert J (2012) Within-family variation in obesity. *Econ Hum Biol* 10(4):333–339
- Prosper M-H, Moczulski VL, Qureshi A (2009) Obesity as a predictor of self-rated health. *Am J Health Behav* 33(3):319–329
- Reichert AR (2015) Obesity, weight loss, and employment prospects—evidence from a randomized trial. *J Hum Resour* 50(3):759–810
- Reichert AR, Tauchmann H, Wübker A (2015) Weight loss and sexual activity in adult obese individuals: establishing a causal link. *Ruhr Econ Pap* 561
- Roodman D (2009) Estimating fully observed recursive mixed-process models with CMP. *Stata J* 11:159–206
- Sabia JJ, Rees DI (2011) The effect of body weight on adolescent sexual activity. *Health Econ* 20(11):1330–1348
- Sacerdote B (2007) How large are the effects from changes in family environment? A study of Korean American adoptees. *Q J Econ* 122(1):119–157
- Stock JH, Wright JH, Yogo M (2002) A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 20(4):518–529
- Tompson T, Benz J, Agiesta J, Brewer K, Bye L, Reimer R, Junius D (2012) Obesity in the United States: public perceptions. *Food Ind* 53(26):21
- Ul-Haq Z, Mackay DF, Fenwick E, Pell JP (2013a) Meta-analysis of the association between body mass index and health-related quality of life among adults, assessed by the SF-36. *Obesity* 21(3):E322–E327
- Ul-Haq Z, Mackay DF, Martin D, Smith DJ, Gill JM, Nicholl BI, Cullen B, Evans J, Roberts B, Deary IJ et al (2013b) Heaviness, health and happiness: a cross-sectional study of 163 066 UK biobank participants. *J Epidemiol Commun Health* 68:340–348
- Vogler G, Sørensen T, Stunkard A, Srinivasan M, Rao D (1995) Influences of genes and shared family environment on adult body mass index assessed in an adoption study by a comprehensive path model. *Int J Obes Relat Metab Dis* 19(1):193–197
- Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G (2008) Financial incentive-based approaches for weight loss. *J Am Med Assoc (JAMA)* 300(22):2631–2637
- von Hinke S, Smith GD, Lawlor DA, Propper C, Windmeijer F (2016) Genetic markers as instrumental variables. *J Health Econ* 45:131–148
- Ware JE, Snow KK, Kosinski M, Gandek B (1993) SF-36 health survey manual and interpretation guide. The Health Institute, New England Medical Center, Boston
- WHO (2000) Obesity: preventing and managing the global epidemic. Technical report 894
- WHO (2016) Obesity and overweight, fact sheet 311. <http://www.who.int/mediacentre/factsheets/fs311/en/>
- Willage B (2018) The effect of weight on mental health: new evidence using genetic IVS. *J Health Econ* 57:113–130
- Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, Hill JO, Brancati FL, Peters A, Wagenknecht L et al (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 JM (2002) *Econometric analysis of cross section and panel data*. MIT Press, Cambridge
- Wooldridge JM (2015) Control function methods in applied econometrics. *J Hum Resour* 50(2):420–445
- Zellner DA, Loaiza S, Gonzalez Z, Pita J, Morales J, Pecora D, Wolf A (2006) Food selection changes under stress. *Physiol Behav* 87(4):789–793