methods, data, analyses | Vol. 18(1), 2024, pp. 33-58 DOI: 10.12758/mda.2022.14

Analyzing the Causal Effect of Obesity on Socioeconomic Status – the Case for Using Difference-in-Differences Estimates in Addition to Fixed Effects Models

Judith Lehmann
Otto-Friedrich-Universität Bamberg

Abstract

Recent studies use Fixed Effects (FE) models to estimate the causal effect of obesity on socioeconomic status, the so-called obesity penalty. In this paper, I will illustrate the advantages of using a Difference in Differences (DID) approach as an alternative method of causal analysis. Combining the German National Health Interview and Examination Survey 1998 (GNHIES98) and the German Health Interview and Examination Survey for Adults 2008 (DEGS1) allowed for a panel analysis of 3934 respondents. The dependent variable is a socioeconomic status score that integrates level of education, occupation and household income. The binary treatment variable is abdominal obesity. To estimate the causal effect of the treatment, FE and DID approaches were used.

Both the FE model and the DID estimate show no statistically significant causal effect of abdominal obesity on socioeconomic status for adults in Germany. However, both the respondents who became obese and those who stayed non-obese experience a rise in socioeconomic status over time. Nonetheless, the non-obese group had a more substantial increase in socioeconomic status than the obese group. Therefore, the obesity penalty does not necessarily have to be a decrease in socioeconomic status but could instead be a slowed growth or stagnation in status. The advantage of the DID approach is that the development in the control group is explicit. If obese individuals are more likely to have less favorable positive trends in socioeconomic status over time than other individuals, using DID estimates demonstrates the obesity penalty more effectively than using only FE models.

Keywords: Difference-in-Differences, Propensity Score Matching, Fixed-Effects, Obesity penalty, Germany

Fixed Effects (FE) models have become a popular and widely used method of panel analysis. Researchers apply FE models to identify causal effects through the within-comparison of cases (Brüderl, 2010). However, there are alternative methods for identifying causal effects with observational data that can add important insights into the topic under study (Gangl, 2010). Hence, in this paper I will illustrate the advantages of supplementing fixed effects analyses with Difference-in-Differences (DID) approaches.

To highlight the differences and advantages of both FE models and DID estimators, I chose the example of the obesity penalty (Averett & Korenman, 1996). The obesity penalty describes the finding that obese people earn lower wages and report more adverse labor market outcomes than non-obese people (Caliendo & Gehrsitz, 2016). To support these findings theoretically, different mechanism such as lower human capital, lower productivity and higher probability of health issues of the obese as well as discrimination and negative stereotyping are discussed (Bozoyan & Wolbring, 2018).

The obesity penalty is an interesting example to consider. On the one hand, FE models are frequently applied in research on the effect of body weight on socioeconomic status. Many previous studies focus on the question of what happens to an individual's socioeconomic status when they become obese. On the other hand, research on labor market outcomes has shown that it is important to observe an adequate control group (i.e. Angrist & Pischke, 2008). Knowledge concerning the development of socioeconomic status in the control group can change the interpretation of the development of status in the treatment group. Hence, DID estimators may contribute important new information on the obesity penalty. As a result, the main question of this study is: What are the advantages of using DID estimators in addition to using FE models in regards to the obesity penalty?

In this study, I will use FE models and DID estimators to identify the causal effect of abdominal obesity on socioeconomic status in a sample of adults in Germany. Propensity Score Matching will be applied to create an adequate control group for the DID estimator since treatment is not assigned randomly in observational data. I will use these two methods to the full sample and to female and male respondents separately. In the discussion, I will highlight the advantages of including DID approaches in this line of research and the additional information gained by introducing an adequate control group. I will also discuss the different perspectives offered by FE models and DID estimators to further evaluate their benefits.

Direct correspondence to

Judith Lehmann, M.A., Otto-Friedrich-Universität Bamberg, Lehrstuhl für Soziologie, insbesondere Soziale Ungleichheit, Feldkirchenstr. 21, 96052 Bamberg

E-mail: judith.lehmann@uni-bamberg.de

Previous Research

Studies on the causal relationship of obesity and socioeconomic status focus on both the social causation hypothesis and the health selection hypothesis. The social causation hypothesis states that socioeconomic status influences body weight and the probability of becoming obese. For example, Ball and Crawford (2005) show in their review that lower job position increases the probability of becoming obese compared to higher job position. Gebremariam et al. (2017) conclude that socio-economic position of the parents influences body weight of their children through mediators such as food consumption and TV usage. In a meta-analysis, Kim et al. (2017) show that both social causation and health selection exist in regards to education. However, the evidence for the health selection hypothesis is more consistent.

The health selection hypothesis states that obesity leads to lower socioeconomic status. Studies that focus on the health selection hypothesis overwhelmingly use FE models to identify the causal effect. For example, many studies used the National Longitudinal Survey of Youth, which is a panel study in the United States of America (US). Baum and Ford (2004) find negative effects of obesity on wages for men and women using this data. In this study, women experience stronger and more consistent negative effects than men. Cawley (2004) identifies a negative effect of Body Mass Index (BMI) on wages for white women. The effect for men is non-linear, with overweight men earning more than normal-weight or obese men. Han et al. (2009) report similar findings with overweight and obese white women and obese Black women earning less than their normal-weight peers. They report no causal effect for men. However, the authors can show that respondents in jobs with social interactions are especially affected. Harris (2019) builds on that and finds that high body weight leads to lower wages in jobs that are socially and mentally intensive and to higher wages in physically challenging jobs. He concludes that gender differences in the effect of body weight on wages can be explained through differing occupational positions.

Other recent studies use different data to analyze the obesity penalty. Bozoyan and Wolbring (2011) use fat free mass and body fat instead of BMI to model body weight. Using FE models, they cannot identify a significant effect of body weight on wages. Ahn et al. (2019) conclude that obese women and underweight men are disadvantaged on the labor market even if employment efforts are controlled for. When obese women find a job, Lee et al. (2019) show that their wages and other characteristics of the job (i.e. getting a bonus or having a job in a company with a labor union) are inferior to those of other women.

Another popular method of causal inference are instrument variables (IV) because exogenous instruments allow the identification of the treatment effect even in the presence of unobserved heterogeneity (Gangl, 2010). The IV method is applied in the context of the obesity penalty as well. For example, Cawley et al.

(2005) use this method and identify a negative effect of obesity on wages only for women in the US. Morris (2007) finds a negative effect of obesity on employment for men and women. Sari and Acan Osman (2018) show that obesity negatively influences labor market participation for women. Böckerman et al. (2019) identify negative effects of body weight on multiple dimensions of socioeconomic status, such as earnings, employment and social income transfers.

Some studies have combined FE models and IV methods to strengthen their findings. While IV methods control for unobserved time-varying heterogeneity in theory, it is challenging to find good instruments in practice. Therefore, results gained through IV methods are often viewed with caution and FE models are added as an alternative method of causal inference. Sabia and Rees (2012) find effects of body weight on wages only for white women using FE models. Their IV analyses confirm this finding; therefore, they conclude that this result is not influenced by unobserved time-varying heterogeneity. Katsaiti and Shamsuddin (2016) analyze a number of aspects of the socioeconomic position and find negative effects of higher body weight on wages, employment, promotions and a positive effect on duration of unemployment for women. There are no causal effects for men. Both FE models and IV method produce these findings. Wada and Tekin (2010) analyze the effects of body fat and fat free mass on wages. Their FE models find positive effects of fat free mass and negative effects of body fat on wages for white men and women. Using the IV method, only the effects for men can be confirmed. The authors do not interpret this further because of small sample sizes and restrictions of the IV. These studies mostly report similar findings for both methods, however, none compare the methods directly.

So far, DID approaches have not been applied to this field of research. While FE models focus on changes within the cases of the treatment group and use a control group implicitly when confounders are controlled for, DID estimators explicitly use the changes in the control group in addition to the ones in the treatment group to estimate the causal effect. For identifying causal effects, Angrist and Pischke (2008) have shown the importance of a comparable control group. They use examples from educational and labor market research to illustrate the advantages of DID estimators. Using an adequate control group can help identify the causal effect of abdominal obesity on socioeconomic status and provide important new insights. Therefore, I will address this research gap by using DID estimators in addition to FE models to show which additional information can be gained concerning the obesity penalty.

Research Question and Hypotheses

The aim of this study is to apply FE models and DID estimators to the same research question to illustrate the value of using both methods. To do this, I focus on the research question: Is there a causal effect of obesity on socioeconomic status? Previous research has reported mixed results on this question, especially for Germany. By using two different approaches to estimate the causal effect, I will strengthen the results and show the advantages of each method.

From previous research, I derived two hypotheses concerning the causal effect of obesity on socioeconomic status. First, I expect that obesity decreases socioeconomic status. This hypothesis is usually referred to as the obesity penalty. It is assumed that because of different mechanisms such as discrimination, differences in human capital and productivity or health problems obesity leads to a lower socioeconomic status (Bozoyan & Wolbring, 2018).

Second, I expect that the negative effect of obesity on socioeconomic status is stronger for women than for men. It is often assumed that women are judged more harshly for their appearance and body weight than men (Caliendo & Gehrsitz, 2016). Women have to comply with the norm for thinness more than men do (Magallares, 2016). Some research even indicates that overweight men are more privileged than other men (Cawley, 2004).

While the example of this study is the obesity penalty, the focus lies on the exploration of the benefits of combining FE models and DID estimators. Therefore, the main research question is: What are the advantages of using DID estimators in combination with Propensity Score Matching in addition to using FE models?

I argue that the DID approach can offer more information on the causal relationship between obesity and socioeconomic status. As will be shown in this study, the DID approach uses an explicit control group. Hence, it provides researchers with the chance to compare the development of the outcome variable in the treatment and control group. Furthermore, it requires a theoretical discussion of the comparability of treatment and control group.

Data & Methods

The next section will give an overview of the data and methods used. Since I will focus on discussing the use of FE models and DID estimators as methods of causal inference, I will present their advantages and disadvantages as well as their general logic and assumptions in more detail than usual.

Data

Both DID estimators and FE models usually require longitudinal data to estimate the causal effect of obesity on socioeconomic status. However, representative longitudinal data of the German adult population with a focus on health and the socioeconomic position of households or individuals is still sparse. Therefore, the German National Health Interview and Examination Survey 1998 (GNHIES98) and the German Health Interview and Examination Survey for Adults 2008 (DEGS1) conducted by the Robert Koch-Institute were combined to allow for panel analyses.

The German Health Interview and Examination Survey for Adults (DEGS) is the first representative longitudinal survey focusing on health and the socioeconomic position of adult respondents in Germany (Gößwald et al., 2012). The first wave of data was collected between 2008 and 2011. Respondents for DEGS1 were selected based on a previous study conducted by the Robert Koch-Institute: the German National Health Interview and Examination Survey 1998 (GNHIES98). Between 1997 and 1999, 7124 respondents were interviewed and examined for GNHIES98 (Thefeld et al., 1999). Those respondents who were still alive in 2008 were invited to also participate in DEGS1. Therefore, a longitudinal sample of 3959 respondents exists, covering two waves and a period of around ten years between the waves (Gößwald et al., 2012). Restricting the sample to cases with valid values for both the dependent and the central explanatory variable leads to 2835 cases that can be included in the analyses.

Both GNHIES98 and DEGS1 include medical interviews and medication history, health questionnaires and nutrition interviews, and laboratory and physical examinations. The data set contains anthropometric data such as height, body weight and waist circumference measured by health professionals as well as information on the socioeconomic situation of individuals and households (Scheidt-Nave et al., 2012, Gößwald et al., 2012). Hence, I chose this data set for the following analyses.

Variables

The dependent variable for the analyses is a socioeconomic status score that is provided by the Robert Koch-Institute and integrates information on the level of education, occupation and household income of the respondents. The socioeconomic status score is not a variable on the individual level, because it combines individual and household information. It creates a scale of socioeconomic status that integrates three different dimensions of social status (Lampert et al., 2013). For the subscale of education, schooling and vocational training of the respondents are combined and ranked from 1 (lowest education) to 7 (highest education). For the subscale of occupation, the jobs of the respondents and the main earners of their households are compared and the higher occupational position of the two is ranked from 1 (lowest occupational position) to 7 (highest occupational position) according to the average wages earned in that profession. For the subscale of household income, weighted net household income was ranked from 1 (lowest income) to 7 (highest income). The three subscales were summed up to form a socioeconomic status score ranging from 3 to 21 (Winkler & Stolzenberg, 1999; Lampert et al., 2013). The socioeconomic status score is considered a quasi-metric variable (Lampert et al., 2013). Since the socioeconomic status score is a relative measure of the social position, changes in the score over time can occur even if educational level, occupation and household income of the respondents did not change between waves. The socioeconomic status score is normally distributed. For the analyzed sample of this study, the mean of the socioeconomic status score was 11.5 in GNHIES98 and 11.6 in DEGS1.

The central explanatory variable is abdominal obesity defined by waist circumference. Health professionals measured waist circumference during both waves of data collection. The measurement was standardized as much as possible. For GNHIES98, waist circumference was measured at the midway point between the lowest rib and the pelvic crest while respondents wore a light layer of clothing (Bergmann, 1999). The same method was used for DEGS1; however, respondents were measured wearing only their underwear (Haftenberger et al., 2016). This change in measurement between the two waves might lead to small differences in waist circumference, even if the body weight of the respondents did not vary. The mean waist circumference of the analyzed sample is 89.8cm in GNHIES98 and 93.8cm in DEGS1, so in general, the respondents gained weight between the two waves.

Using waist circumference, I created a binary variable for abdominal obesity as the treatment variable. The binary variable allows for an easy separation of the sample into treatment and control group. The World Health Organization provides the following cut-offs to define abdominal obesity by waist circumference: 88cm for women and 102cm for men (WHO, 2011). These cut-offs were employed to generate the binary variable for abdominal obesity. According to this new variable, 31% of the sample were obese in GNHIES98 and 44% in DEGS1. Since the focus of this paper is on the causal effect of obesity, respondents who were not obese in GNHIES98, but were obese in DEGS1 constitute the treatment group. Approximately 24% of the non-obese respondents in the first wave became obese by the second wave. While respondents who were not obese in both waves constitute the control group, respondents who were already obese in the first wave were excluded from the analyses (888 cases).

The Counterfactual Framework

This study analyzes the causal effect of obesity on socioeconomic status using observational data. For this purpose, the counterfactual framework allows the integration of causal analyses and observational data (Gangl, 2010). This is necessary since most research questions in the social sciences cannot be analyzed using randomized experiments due to ethical and practical restrictions (Leszczensky & Wolbring, 2019). Randomized experiments are usually considered the golden standard of causal inference because respondents are randomly selected into the treatment or the control group and thus selection bias is eliminated (Gangl, 2010).

In contrast, in observational studies treatments are assigned in a socially structured way and therefore treatment assignment and the expected outcome might be correlated (Gangl, 2010). To estimate the causal effect, it is necessary to disrupt this correlation by conditioning on covariates. The aim is to achieve conditional independence, which states, "conditional on covariates, variation in [treatment variable] D is as good as randomly assigned" (Gangl, 2010, p. 27). If the conditional independence assumption (CIA) holds, conditioning on the covariates will lead to unbiased causal effects.

The identification of causal effects is complex because, in theory, the individual causal effect is calculated by subtracting the outcome of a person i who receives the treatment (Y_i^1) from the outcome of the same person i if they do not receive the treatment (Y_i^0) (Rosenbaum & Rubin, 1983). The only difference between the two states of person i is the treatment status so that any changes in the outcome can be attributed to the treatment. In practice, it is not possible to observe the outcome of person i in both treatment states at the same time – so it is not possible to calculate the individual causal effect (Holland, 1986; Dehejia & Wahba, 1999).

Therefore, Holland introduced the following statistical solution: replace "the impossible-to-observe causal effect of "X" on a specific unit with the possible-to-estimate average causal effect of "X" over a population of units" (Holland, 1986, p. 947). This is unproblematic since the research interests in the social sciences usually focus more on average causal effects in groups than individual causal effects. Still, with the methods of causal analyses of observational data we can only explore the causal effect indirectly and under the validity of certain assumptions (Brüderl, 2010; Gangl, 2010).

The counterfactual framework proposes the use of counterfactuals to identify the average causal effect. Counterfactuals are defined as the unobservable outcome of person i if their treatment status had been different (Gangl, 2010; Pearl, 2009). In place of the unobservable outcome, observational data can be used to estimate the outcome the treatment group would have had, if they had not received the treatment (Oakes & Johnson, 2006). The estimation strategy of this counterfactual outcome is a very crucial decision because different methods use different approaches. If plausible counterfactuals are estimated, they can be used to calculate the average causal effect, which is usually expressed as Average Treatment Effect on the Treated (ATT). To estimate the ATT for the treatment group, the counterfactual outcome Y_i^0 is subtracted from the observed outcome Y_i^1 (Dehejia & Wahba, 1999). The most important assumption is that no factors other than the treatment are responsible for the differences in the outcome of treatment and control group (Brüderl, 2010).

In this paper, I will highlight two different approaches of causal analysis: Fixed Effects (FE) models and Difference-in-Differences (DID) estimators. These methods use different approaches to estimate the counterfactuals and therefore underlie different assumptions. The aim of this paper is to show the advantages and disadvantages of both methods using a practical example from research on health and social inequalities.

Fixed Effects Models

Fixed Effects models have been employed widely in recent studies using panel data in the social sciences and are often used to evaluate the obesity penalty. FE models are appealing because they automatically condition on all time-constant unobserved heterogeneity (Gangl, 2010). Therefore, time-constant covariates cannot bias the causal effect. Thus, using FE models has clear advantages over traditional regressions (Brüderl, 2010).

Returning to the counterfactual framework of causality, the question is how FE models create the counterfactual to estimate the causal effect. In short, FE models estimate the causal effect within person i over time. The outcome of person i at time t1 before the treatment is used to construct the counterfactual. To estimate the causal effect this counterfactual is subtracted from the outcome of person i at time t2 after the treatment (Brüderl, 2010). Hence, the difference in the outcome between t2 and t1 is viewed as the causal effect.

Since FE models compare person i with itself, they automatically control for all unobserved heterogeneity that is time-constant (Brüderl, 2010). This is achieved through within transformation of the data. Within transformation removes the person-specific time-constant error by using only variation within individuals over time for the estimation of the treatment effect (Brüderl & Ludwig, 2015). Due to within transformation, the effect of characteristics of respondents that are stable over time is removed and thus changes in outcome are influenced only by treatment status, time-varying covariates and time-varying idiosyncratic error (Gangl, 2010). Therefore, the CIA is weaker than in traditional regression analysis; however, the assumption that time-varying unobserved characteristics do not bias the causal effect is still a strong one (Gangl, 2010).

Hence, the problem of unobserved heterogeneity that is not time-constant still remains (Hill et al., 2019). As long as information on the influencing factors that change over time is available in the data, conditioning on these variables will lead to unbiased estimates. Beyond that, the assumption of FE models is that there is no unobserved time-varying heterogeneity. Therefore, to strengthen the results of FE models, it is necessary to discuss explicitly which influencing factors might bias the causal effect and whether they can be controlled for in the model. Hence, substantive theoretical models must be the base of causal analysis (Gangl, 2010).

Further, within transformation of the data can also lead to higher risk of bias because only a selective group – the treated – contribute within information for the estimation (Gangl, 2010). This also enhances problems of measurement error due to misreporting or miscoding because small changes can lead to a big bias in the estimates (Angrist & Pischke, 2008). Additionally, FE models might also remove valuable information on the causal relationship of interest because of the within transformation of data.

Difference-in-Differences Estimator and Propensity Score Matching

Difference-in-Differences (DID) estimators use the same logic of comparing cases before and after treatment to estimate the causal effect (Gangl, 2010). However, DID estimators use the aggregate level, not the individual level (Angrist & Pischke, 2008). In contrast to FE models, DID estimators employ a control group to identify the causal effect of the treatment. Thus, the development of the outcome variable over time in the control group is used as the counterfactual for the changes in outcome the treatment group would have had, if they had not received the treatment (Halaby, 2004). DID estimates subtract the average change over time in the outcome variable of the control group from the average change over time in the outcome variable of the treatment group (Halaby, 2004; Stuart et al., 2014). Consequently, DID estimators condition on all group-specific time-constant unobserved heterogeneity (Gangl, 2010).

Additionally, DID estimators can reduce time-varying unobserved heterogeneity by using a control group. However, the central assumption of the DID approach is the parallel trends assumption: it is assumed that treatment and control group would have had the same development over time if the treatment had not happened in the treatment group (Caniglia & Murray, 2020; Cataife & Pagano, 2017). Therefore, the choice of control group is of utmost importance, as is shown by Angrist and Pischke (2008). If the parallel trends assumption does not hold, the DID estimate will be biased because the effect of time-varying unobserved heterogeneity is not statistically controlled for (Cataife & Pagano, 2017).

Similar to CIA in the case of FE models, the parallel trends assumption cannot be proven in a mathematical sense; however, it can be made plausible through theoretical arguments. One way to strengthen the assumption is to use Propensity Score Matching to weight the control group so it matches the treatment group in all relevant aspects (Godard-Sebillotte et al., 2019, Stuart et al., 2014; Heckman et al., 1997). Due to Propensity Score Matching, the treatment and control group are comparable to each other before the treatment. Therefore, it is more plausible that their further development would have been similar if the treatment had not happened (Caniglia & Murray, 2020; Cataife & Pagano, 2017). However, the selection of covariates chosen for Propensity Score Matching must be based on a strong theoretical model.

Since Propensity Score Matching was employed in the following analyses, a brief description of this method will be provided. The Propensity Score is "defined as the conditional probability of assignment to a particular treatment given a vector of observed covariates" (Rosenbaum & Rubin, 1983, p. 41). Conditioning on the Propensity Score, there should be no difference in the probability of receiving the treatment between the treatment and control group. Thus, it is used to reduce the bias that exists in observational data due to self-selection into the treatment (Austin, 2007).

The Propensity Score is usually estimated via logit models, using the treatment as the dependent variable (Gangl, 2010). The relevant covariates that influence the probability of receiving the treatment are used as independent variables in these models (Oaks & Johnson, 2006). The covariates are chosen based on theoretical considerations (Rosenbaum & Rubin, 1983), usually based on the idea of d-separation (Pearl, 2009, p. 106): all paths that could bias the effect of the treatment on the outcome are closed conditioning on the Propensity Score. Thus, the causal effect can be estimated (Pearl, 2009).

Once the Propensity Score has been estimated, treatment and control group can be matched accordingly. The aim is to pair a treated and a control case with very similar Propensity Score values and compare their outcome. In practice, Propensity Score Matching is a way of weighting the data so that treatment and control group are comparable (Dehejia & Wahba, 2002). After Propensity Score Matching, the DID estimate can be used to calculate the causal effect.

In conclusion, DID estimators condition on group-specific time-constant unobserved heterogeneity. Propensity Score Matching will provide an adequate control group for the estimation if all relevant covariates are available in the data and a good matching quality can be achieved. Therefore, the DID estimator after Propensity Score Matching will also reduce unobserved time-varying heterogeneity, as long as the parallel trends assumption holds. However, the assumption that time-varying unobserved characteristics influence treatment and control group in exactly the same way and therefore do not bias the DID estimator is still a strong one (Cataife & Pagano, 2017).

Analytical Strategy

After this brief overview of Fixed Effects models and Difference-in-Differences estimators, I will discuss the concrete analytical approach in this section.

First, I estimated several FE models. In these models, the dependent variable is the socioeconomic status score. Abdominal obesity constitutes the treatment variable. I excluded respondents that were pregnant during one of the interviews (4 cases) and disabled respondents (600 cases) from the analyses.

FE models control for time-invariant heterogeneity, however, time-varying heterogeneity might bias the effect. Therefore, the following time-variant control variables were chosen: marital status, number of adults and children in the household, years of education and age as well as age squared. Changes in marital status and household composition can directly affect socioeconomic status on the household level. At the same time, changes in marital status and household composition can influence body weight (Huyer-May, 2018). Changes in education directly affect socioeconomic status and can influence body weight indirectly through changes in health behavior (Brunello et al., 2013). Age affects both body weight and socioeconomic status positively but not necessarily linearly, thus it is a confounder of the causal relationship under study (Schienkiwitz et al., 2017; Krause & Schäfer, 2005). Respondents who had missing values on any of the control variables were excluded. I will present the results of the FE models with and without control variables. Separate models were estimated for men and women since the effect of obesity on socioeconomic status could vary by gender.

Second, I used Propensity Score Matching to prepare the data for the DID estimator. The choice of covariates to include in the estimation of the Propensity Score is of utmost importance. To allow for the interpretation of the DID estimator as a causal effect, all relevant variables need to be included as covariates in the estimation of the Propensity Score. Following a method introduced by Shrier and Platt (2008), I developed an explanatory model for the causal effect of obesity on socioeconomic status (Figure 1). Going through the six steps of the method led to the following list of covariates to include in the Propensity Score: gender, age, educational level, marital status and number of adults and children in the household, disability, diet and exercise. I estimated the Propensity Score using logit models including these covariates that were measured before the treatment. Respondents who had missing values on any of these variables were excluded.

After estimating the Propensity Score, I chose a matching algorithm. To identify the causal effect, a high matching-quality must be achieved. On the one hand, the overlap of the treatment and control group must be sufficient (Gangl, 2010).

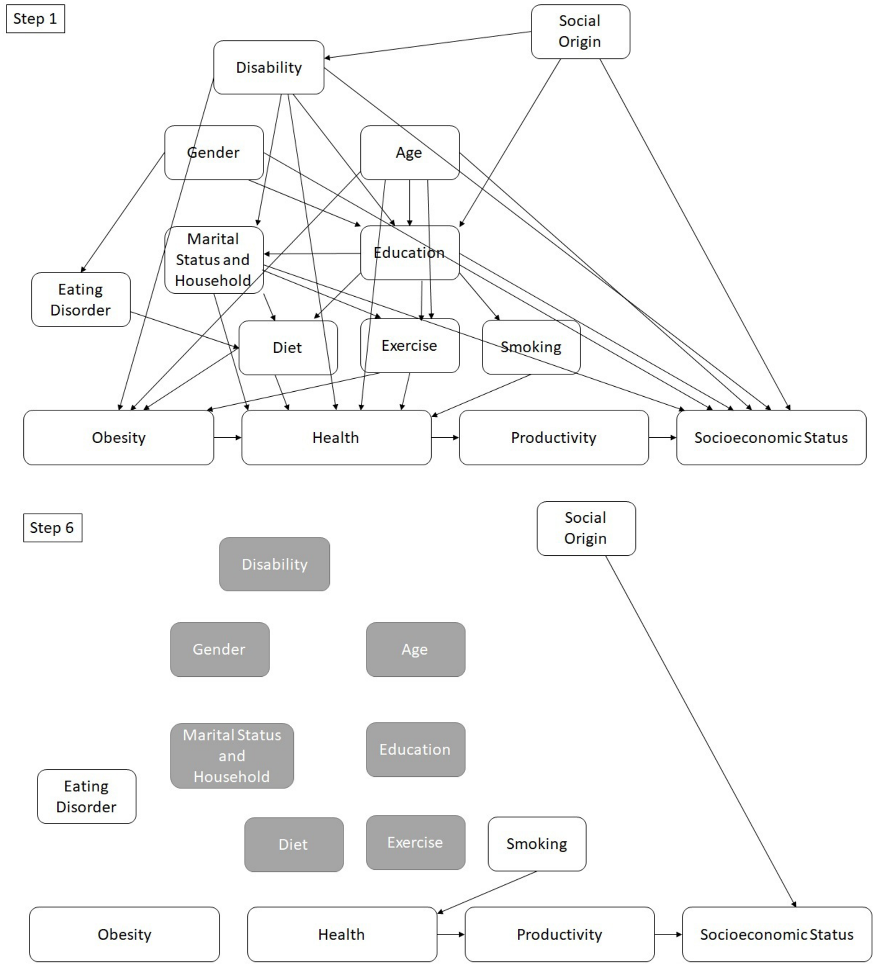

Figure 1 Explanatory model for the causal effect of obesity on socioeconomic status to choose covariates for Propensity Score Matching according to the method of Shrier and Platt (2008)

Therefore, there must be a reasonable number of respondents in each group that have a comparable Propensity Score (Figure 2, bottom). On the other hand, the matching algorithm that achieves the highest similarity in the chosen covariates between treatment and control group must be chosen. The best fit in this case was achieved using Radius Matching (Figure 2, top). Radius Matching is a variation of Caliper Matching where all possible matches with a certain maximum distance in the Propensity Score are used to create the counterfactual of each treated case (Caliendo & Kopeinig, 2008).

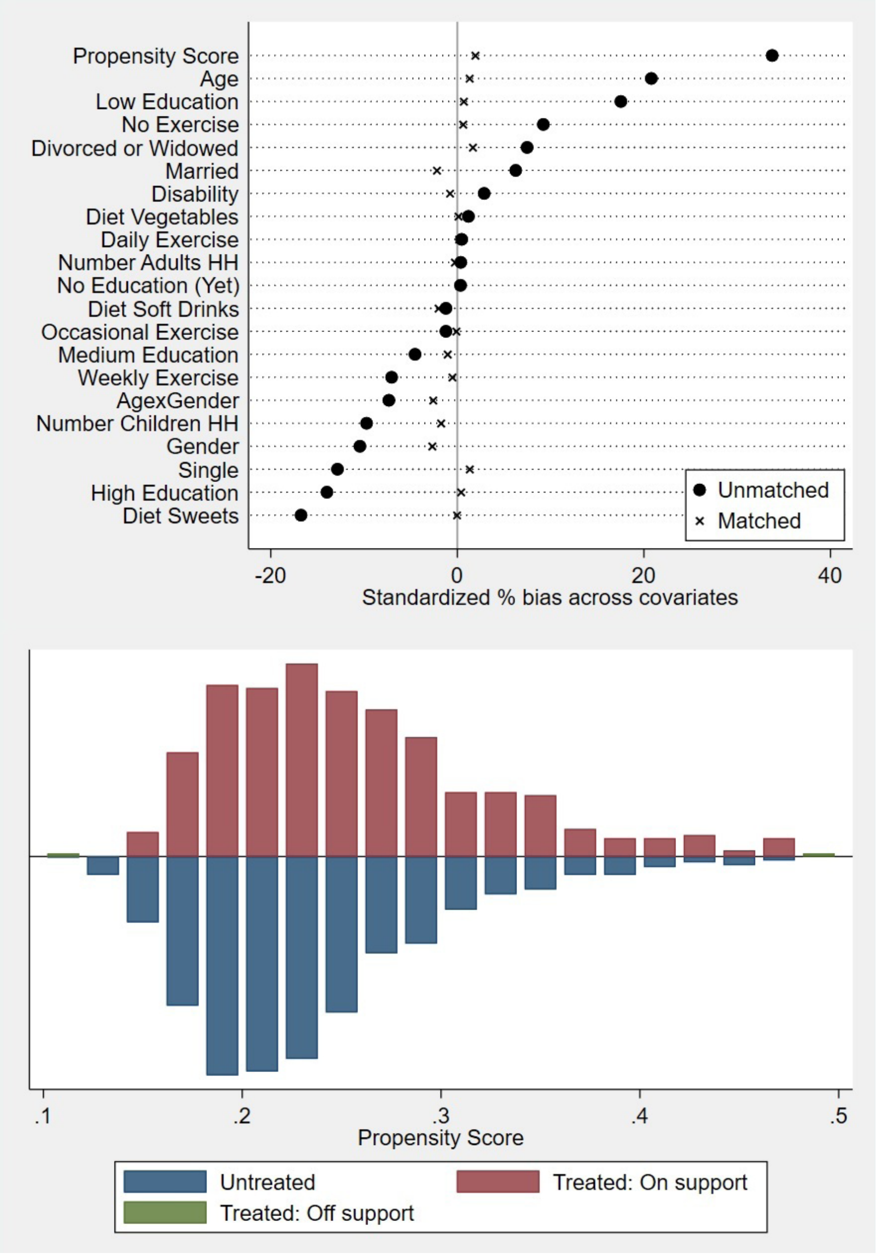

Figure 2 Bias reduction due to Propensity Score Matching (top) and overlap in the Propensity Score in treatment and control group (bottom) achieved through Radius Matching (Data: DEGS1 & GNHIES98)

Third, I calculated the DID estimator after Propensity Score Matching using the PSMATCH2 Stata module by Leuven and Sianesi (2003). Respondents who were already obese in the first wave of data collection and pregnant respondents were excluded. As the dependent variable, I used the socioeconomic status score and calculated the difference in the score between the first and second wave. This difference represents the change in socioeconomic status for each respondent during the observation period. The DID estimator then shows the difference in the changes over time between treatment and control group. I used bootstrapping to calculate standard errors (Gangl, 2010). Men and women were analyzed separately in case the causal effect varies by gender. The Propensity Score Matching process was repeated for each DID estimation.

Results

First, I will present the results of the FE models (Table 1). Model 1 represents the full sample and only includes the treatment variable without other covariates. Abdominal obesity has a non-significant positive effect on socioeconomic status according to Model 1 (b = .197, p = .138). We can see a non-significant .197 scale-points increase in socioeconomic status (on a scale ranging from 3-21) when respondents become obese. However, this result can be biased due to time-varying heterogeneity. Model 2 shows the results for the whole sample after conditioning on the time-varying control variables. The effect is still not statistically significant; however, it is now negative (b = -.05, p = .719). Controlling for changes in education, age, marital status, and composition of household, we see a slight decrease of socioeconomic status in respondents who become obese.

We observe the same pattern in the separate models for men and women. Model 3 shows a non-significant positive effect of abdominal obesity for female respondents (b = .154, p = .368). However, after conditioning on the control variables, in Model 4 the non-significant effect is negative (b = -.097, p = .607). For male respondents, the effect is not significant and positive in Model 5 (b = .256, p = .222) and not significant and negative in Model 6 after controlling for confounders (b = -.067, p = .747).

For both female and male respondents, FE models do not identify a significant causal effect of abdominal obesity on socioeconomic status. While the effect appears positive when confounders are not controlled for, it is negative after conditioning on the control variables. Respondents who become obese see a small decrease in socioeconomic status because of their obesity. However, this finding is not statistically significant and may therefore be due to chance.

Second, I will present the results of the DID estimator after Propensity Score Matching. Table 2 shows the findings for the full sample. The results are more illu-

Table 1 Fixed Effects models, dependent variable: Socioeconomic Status Score

	Model 1 Full Sample	Model 2 Full Sample	Model 3 Female Re-spondents	Model 4 Female Re-spondents	Model 5 Male Re-spondens	Model 6 Male Re-spondents
Obesity	.197 (.138)	-.050 (.719)	.154 (.368)	-.097 (.607)	.256 (.222)	-.067 (.747)
Conditioning on Controls		X		X		X
σ_u	3.324	2.665	3.121	2.592	3.523	2.754
σ_e	1.804	1.630	1.764	1.627	1.848	1.627
ρ	.773	.728	.758	.717	.784	.741
Within-R^2	.001	.192	.001	.163	.002	.240
Observations	3,761	3,761	1,972	1,972	1,789	1,789
Groups	2,209	2,209	1,157	1,157	1,052	1,052

Note. Data: DEGS1 & GNHIES98; Obesity: abdominal obesity (>88cm Waist Circumference for women, >102cm Waist Circumference for men); Control variables: years of education, marital status, number of adults and children in household, age, age²; p-values in parentheses; σ_u error due to differences between units, σ_e error due to differences within units, ρ proportion of variance due to unit effects

Table 2 Difference-in-Differences estimator of the full sample; dependent variable: Socioeconomic Status Score

Propensity Score Matching	Treatment Group	Control Group	DID Estimator	S.E.
Before	.107	.276	-.169	.135
After	.100	.085	.015	.128[1]
N (on support)	455	1,446		
N (off support)	2	0		

Note. Data: GNHIES98 & DEGS1, Treatment: abdominal obesity (>88cm Waist Circumference for women, >102cm Waist Circumference for men); [1] S.E. boot-strapped (1000 repetitions) minating than in the FE models. We can see the changes in socioeconomic status in the treatment and control group before and after Propensity Score Matching as well as the DID estimator.

Before Propensity Score Matching, both treatment and control group see an increase in socioeconomic status over time. However, the increase of .276 points for the control group is larger than the increase of .107 in the treatment group. Therefore, the DID estimator before Propensity Score Matching is negative with -.169 points on the socioeconomic status score. This effect cannot be interpreted as causal, though, because differences in the composition of treatment and control group bias the results.

The bias becomes apparent when we consider the findings after Propensity Score Matching. While the increase in socioeconomic status for the treated group is only marginally smaller with .1 points, the increase of the control group is reduced to .085 points. The DID estimator is now positive with .015; however, it is not statistically significant. The finding that respondents who become obese gain less socioeconomic status over time than people who stay non-obese is explained by differences in the composition of both groups.

Table 3 shows the results for the female respondents. In this subgroup, the DID estimator is negative both before and after Propensity Score Matching. We can see that both the treatment and the control group experience an increase in socioeconomic status over time; however, the increase is only .041 in the treated group and .291 in the untreated group before Propensity Score Matching. After Propensity Score Matching, the DID estimator is not statistically significant with -.042 points. Among the female respondents, the differences in the changes in socioeconomic status over time between treatment and control group can be mostly explained by the different composition of the groups.

Table 3 Difference-in-Differences estimator for female respondents; dependent variable: Socioeconomic Status Score

Propensity Score Matching	Treatment Group	Control Group	DID Estimator	S.E.
Before	.041	.291	-.249	.182
After	.033	.076	-.042	.173[1]
N (on support)	251	722		
N (off support)	1	0		

Note. Data: GNHIES98 & DEGS1, Treatment: abdominal obesity (>88cm Waist Circumference for women, >102cm Waist Circumference for men); [1] S.E. boot-strapped (1000 repetitions)

Considering the male respondents, the findings are very similar. Table 4 shows that both treatment and control group see an increase in socioeconomic status over time both before and after Propensity Score Matching. The DID estimator after Propensity Score Matching is not statistically significant with .029 points. The differences in growth of socioeconomic status between treatment and control group over time can be explained by the different composition of the groups.

In conclusion, both FE models and DID estimators after Propensity Score Matching do not identify a causal effect of obesity on socioeconomic status. This is surprising because most previous studies find a negative effect of obesity on different aspects of socioeconomic status for women (Cawley, 2004; Han et al., 2009; Sabia & Rees, 2012, Katsaiti & Shamsuddin, 2016; Ahn et al., 2019; Lee et al., 2019). Others confirmed the obesity penalty for men as well (Baum & Ford, 2004; Wada & Tekin, 2010; Harris, 2019). Studies that cannot find a significant effect of body weight on socioeconomic status are rare. Bozoyan and Wolbring (2011) also do not find a significant effect of body weight on socioeconomic status. They use data from Germany and wages as dependent variable. Similarly, Cawley et al. (2005) use the IV method and find no significant effect of obesity on wages with German data. Thus, the presented results of this study are consistent with some previous research.

Table 4 Difference-in-Differences estimator for male respondents; dependent variable: Socioeconomic Status Score

Propensity Score Matching	Treatment Group	Control Group	DID Estimator	S.E.
Before	.187	.262	-.075	.202
After	.187	.159	.029	.198[1]
N (on support)	205	724		
N (off support)	0	0		

Note. Data: GNHIES98 & DEGS1, Treatment: abdominal obesity (>88cm Waist Circumference for women, >102cm Waist Circumference for men); [1] S.E. boot-strapped (1000 repetitions)

Discussion

The findings do not lend support to the first two hypotheses. Neither method shows a significant negative effect of abdominal obesity on socioeconomic status. Considering men and women separately, there is no significant effect of abdominal obesity on socioeconomic status for either gender. In conclusion, there is no evidence for an obesity penalty for adults in Germany. While respondents who become obese in general have fewer points on the socioeconomic status score than respondents who are not obese, this difference does not change over time because of obesity.

However, the aim of this study was to discuss the potential of using DID estimators combined with Propensity Score Matching in addition to FE models. The results indicate that using DID estimators can lead to more information on the obesity penalty because it explicitly estimates the outcome changes of the control group in addition to the treatment group. I find an increase of socioeconomic status for both the treatment and control group over time and differences in this increase are due to the different composition of these groups. Further, the use of Propensity Score Matching strengthens the focus on the correct choice of covariates based on theoretical considerations to achieve an unbiased causal effect.

FE models and DID estimators mainly differ in the way they construct the counterfactual to estimate the causal effect. While FE models use comparisons within individuals before and after treatment and construct the counterfactual from the before-measurement of the outcome, DID estimates compare the development in the outcome over time between a treatment and a control group. This has important implications for the results.

Since FE models compare the same individual before and after treatment, all time-constant heterogeneity cannot bias the causal effect (Brüderl, 2010). Therefore, these characteristics cannot be included and furthermore they need not be measured or even known (Angrist & Pischke, 2008). However, time-variant heterogeneity must be controlled for or it will bias the results (Hill et al., 2019). In comparison, DID estimators automatically control for time-variant heterogeneity, assuming it is the same in the treatment and control group (Cataife & Pagano, 2017). As long as the parallel trends assumption holds, these characteristics need not be measured or known. Thus, it is of utmost importance for the DID estimator that an adequate control group is found or constructed (Angrist & Pischke, 2008). One way of achieving such a control group is Propensity Score Matching (Godard-Sebillotte et al., 2019; Stuart et al., 2014). A drawback of this approach is the amount of covariates necessary to estimate the Propensity Score.

To produce unbiased causal effects, both methods need covariates based on theoretical considerations (Gangl, 2010). Usually, the theoretical model behind the chosen covariates stays implicit in many studies. None of the previous studies on the obesity penalty presents theoretical considerations as a base for their control variables. For example, some studies use general health status as a control variable (Wada & Tekin, 2010; Bozoyan & Wolbring, 2011; Katsaiti & Shamsuddin, 2016; Lee et al., 2019; Ahn et al., 2020) even though it can be argued that general health is a causal link through which obesity influences socioeconomic status. This issue is not discussed in the studies. Some studies also include information on perceived discrimination without discussing the theoretical implications (Lee et al., 2019; Ahn et al., 2020). In general, none of the previous studies that use FE models discusses the explanatory model that their chosen covariates are based on.

While any causal analysis should make these decisions explicit, it is much more common in studies that use DID estimators because they have to discuss the parallel trends assumption. In addition, using Propensity Score Matching increases the need to describe the theoretical model and the method of choosing the covariates explicitly (Imbens, 2019; Gangl, 2010). Furthermore, with Propensity Score Matching there exist different methods to confirm matching quality. For example, figures showing the overlap in the Propensity Score of treatment and control group illustrate whether the groups are even similar enough to be compared (Dehejia & Wahba, 2002; Gangl, 2010). Usually there is no similar discussions about FE models and their quality in bias reduction.

Another way of looking at this is through considering the assumptions behind these two methods. The main assumption for FE models is that there would be no change in the outcome variable if there were no treatment (Brüderl, 2010). Meanwhile the main assumption if DID estimators is that the change in the treatment and control group would be the same if there were no treatment. Both are strong assumptions, even though some might argue that the one in FE models is stronger than the DID one because the counterfactual outcome at t2 itself has to be equal to the observed one, not only the counterfactual difference in outcome (Caniglia & Murray, 2020, p. 209).

However, it all comes down to theoretical considerations and well-chosen covariates. Using FE models, we must focus on the changes over time that occur simultaneously as respondents become obese. If important time-variant confounders cannot be controlled for, then the causal effect cannot be estimated. Employing DID estimators, we must concentrate on the differences of people who become obese and those who do not. If there is no adequate control group and none can be constructed using methods like Propensity Score Matching, then DID estimators will be biased. Thus, both methods have slightly different perspectives on causal effects and can therefore be considered complementary.

The main point of this study is to show the potential of adding DID approaches in combination with Propensity Score Matching in future research on the obesity penalty. Apart from the advantages already discussed, the explicit look at the control group within the DID approach is a great benefit.

Considering the results of the DID estimators again, we can derive some important new information. If we look at the DID estimator before Propensity Score Matching, we find a negative effect of abdominal obesity on socioeconomic status. We expect this finding according to the framework of the obesity penalty. However, after Propensity Score Matching this finding does not hold. Since there is no significant DID estimator after Propensity Score Matching, I conclude that the differences in the development of socioeconomic status between treatment and control group can be explained by their different composition. One important aspect is educational level: while higher educational level decreases the risk of becoming obese, it also leads to better job opportunities and higher income. If treatment and control group were comparable in their composition, becoming obese would not lead to differences in the growth of socioeconomic status. Future research into these characteristics could employ decomposition analysis to gain more knowledge about the relative importance of factors that influence the probability of becoming obese and the growth of socioeconomic status over time.

Additionally, we can also see from the results presented with the DID estimator, that both treatment and control group increased their socioeconomic status over time. The negative DID estimator shows us, that the increase in the treated group is smaller than the increase in the untreated group; however, both groups in general gain more socioeconomic status between the two waves. This is also valuable information concerning the obesity penalty. Potentially, the obesity penalty is not a decrease in socioeconomic status of the obese, but rather a slowed growth or stagnation in status. Looking closely at the DID results illustrates that well.

To sum up, the following advantages of the DID approach should be noted: First, the development in the outcome variable for treatment and control group is made explicit. Second, the DID estimator can be calculated before and after Propensity Score Matching to reduce bias due to the different composition of the groups. Third, the theoretical framework behind the choice of covariates for Propensity Score Matching and the parallel trend assumption must be made explicit and discussed.

The aim of this study is to show the advantages of combining FE models and DID estimates, and I have applied these methods in an example concerning obesity and socioeconomic status. I used data collected by the Robert Koch-Institute that have some clear limitations. First, so far there are only two waves of data available. Both FE models and DID estimators would benefit from a dataset with more waves included. Second, the two waves cover a period of about ten years. While this leads to a sufficient number of people who become obese between the two waves, it also leads to a lot of uncertainty about what happened within those ten years. For example, people could have become obese and then lost weight again before the second wave of data collection. We also have no information on when exactly respondents became obese within those ten years. This could influence whether and how their socioeconomic status changed. Third, socioeconomic status is a variable on the household level. Therefore, other members of the household might level out changes in wages, income or job position that occur because of weight gain, especially for female respondents. Unfortunately, this is the only variable for socioeconomic status that has been measured for both waves of data. Thus, the presented results concerning the causal effect of obesity on socioeconomic status in Germany should be interpreted with caution and further research and better data on this topic are necessary.

In conclusion, the DID approach offers a new perspective and new insights in the obesity penalty. Evidently, the obesity penalty can be understood as a slowed growth or stagnation instead of a decrease in socioeconomic status. If obese individuals are more likely to have less favorable positive trends in socioeconomic status over time than other individuals, using DID estimates demonstrates the obesity penalty more effectively than using only FE models. Therefore, future research should employ the DID approach in addition to FE models to gain more information on the complex relationship of obesity and socioeconomic status.

References

Ahn, R., Kim, T. H., & Han, E. (2019). The Moderation of Obesity Penalty on Job Market Outcomes by Employment Efforts. *International journal of environmental research and public health 16* (16). DOI: 10.3390/ijerph16162974.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly harmless econometrics. An empiricist's companion*. Princeton, NJ: Princeton Univ. Press.

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in medicine 27* (12), 2037–2049. DOI: 10.1002/sim.3150.

Averett, S. & Korenman, S. (1993). The Economic Reality of the Beauty Myth. National Bureau of Economic Research. Cambridge, MA (NBER Working Paper Series, Working Paper No. 4521).

Ball, K. & Crawford, D. (2005). Socioeconomic status and weight change in adults: a review. *Social science & medicine (1982) 60* (9), 1987–2010. DOI: 10.1016/j.socscimed.2004.08.056.

Baum, C. L. & Ford, W. F. (2004). The wage effects of obesity: a longitudinal study. *Health economics 13* (9), 885–899. DOI: 10.1002/hec.881.

Bergmann, K. E. & Mensink, G. B. M. (1999). Körpermaße und Übergewicht. *Gesundheitswesen 61* (Sonderheft 2), S115-S120.

Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., Pehkonen, J., & Raitakari, O. (2019). The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health economics 28* (1), 65–77. DOI: 10.1002/hec.3828.

Bozoyan, C. & Wolbring, T. (2011). Fat, muscles, and wages. *Economics and human biology 9* (4), 356–363. DOI: 10.1016/j.ehb.2011.07.001.

Bozoyan, C. & Wolbring, T. (2018). The Weight Wage Penalty: A Mechanism Approach to Discrimination. *European Sociological Review 34* (3), 254–267. DOI: 10.1093/esr/jcy009.

Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In C. Wolf & H. Best (Eds.). *Handbuch der sozialwissenschaftlichen Datenanalyse* (963–994). VS Verlag für Sozialwissenschaften.

Brüderl, J. & Ludwig, V. (2014). Fixed-effects panel regression. In H. Best und C. Wolf (Eds.). *The SAGE handbook of regression analysis and causal inference* (327–357). London: SAGE Publications.

Brunello, G., Fabbri, D., & Fort, M. (2013). The Causal Effect of Education on Body Mass: Evidence from Europe. *Journal of Labor Economics 31* (1), S. 195–223.

Caliendo, M. & Gehrsitz, M. (2016). Obesity and the labor market: A fresh look at the weight penalty. *Economics & Human Biology 23*, 209–225.

Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *J Economic Surveys 22* (1), 31–72. DOI: 10.1111/j.1467-6419.2007.00527.x.

Caniglia, E. C. & Murray, E. J. (2020). Difference-in-Difference in the Time of Cholera: a Gentle Introduction for Epidemiologists. *Current epidemiology reports 7* (4), 203–211. DOI: 10.1007/s40471-020-00245-2.

Cataife, G. & Pagano, M. B. (2017). Difference in difference: simple tool, accurate results, causal effects. *Transfusion 57* (5), 1113–1114. DOI: 10.1111/trf.14063.

Cawley, J. (2004). The Impact of Obesity on Wages. *J. Human Resources XXXIX* (2), 451–474. DOI: 10.3368/jhr.xxxix.2.451.

Cawley, J. H., Grabka, M. M. & Lillard, D. R. (2005). A Comparison of the Relationship between Obesity and Earnings in the U.S. and Germany. *Schmollers Jahrbuch 125* (1), 119–129.

Dehejia, R. H. & Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics 84* (1), 151–161. DOI: 10.1162/003465302317331982.

Dehejia, R. H. & Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association 94* (448), S. 1053–1062.

Gangl, M. (2010). Causal Inference in Sociological Research. *Annu. Rev. Sociol. 36* (1), 21–47. DOI: 10.1146/annurev.soc.012809.102702.

Gebremariam, M. K., Lien, N., Nianogo, R. A. & Arah, O. A. (2017). Mediators of socioeconomic differences in adiposity among youth: a systematic review. *Obesity Reviews 18* (8), 880–898. DOI: 10.1111/obr.12547.

Godard-Sebillotte, C., Karunananthan, S. & Vedel, I. (2019). Difference-in-differences analysis and the propensity score to estimate the impact of non-randomized primary care interventions. *Family practice 36* (2), 247–251. DOI: 10.1093/fampra/cmz003.

Gößwald, A., Lange, M., Kamtsiuris, P., & Kurth, B.-M. (2012). DEGS: Studie zur Gesundheit Erwachsener in Deutschland. Bundesweite Quer- und Längsschnittstudie im Rahmen des Gesundheitsmonitorings des Robert Koch-Instituts. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz 55* (6-7), 775–780. DOI: 10.1007/s00103-012-1498-z.

Haftenberger, M., Mensink, G. B. M., Vogt, S., Thorand, B., Peters, A., & Herzog, B. et al. (2016). Changes in Waist Circumference among German Adults over Time - Compiling Results of Seven Prospective Cohort Studies. *Obesity facts 9* (5), 332–343. DOI: 10.1159/000446964.

Halaby, C. N. (2004). Panel Models in Sociological Research: Theory into Practice. *Annu. Rev. Sociol. 30* (1), 507–544. DOI: 10.1146/annurev.soc.30.012703.110629.

Han, E., Norton, E. C., & Stearns, S. C. (2009). Weight and wages: fat versus lean paychecks. *Health economics 18* (5), 535–548. DOI: 10.1002/hec.1386.

Harris, M. C. (2019). The Impact of Bofy Weight on Occupational Mobility and Career Development. *International Economic Review 60* (2), 631–660. DOI: 10.1111/iere.12364.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Rev Econ Stud 64* (4), 605–654. DOI: 10.2307/2971733.

Hill, T. D., Davis, A. P., Roos, J. M., & French, M. T. (2020). Limitations of Fixed-Effects Models for Panel Data. *Sociological Perspectives 63* (3), 357–369. DOI: 10.1177/0731121419863785.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association 81* (396), 945–960.

Huyer-May, B. (2018). Do relationship transitions affect body weight? Evidence from German longitudinal data. *Journal of Family Research 30* (3), S. 316–338.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature 58* (4), 1129–1179. DOI: 10.1257/jel.20191597.

Katsaiti, M.-S. & Shamsuddin, M. (2016). Weight discrimination in the German labour market. *Applied Economics 48* (43), 4167–4182. DOI: 10.1080/00036846.2016.1153791.

Kim, T. J., Roesler, N. M., & von dem Knesebeck, O. (2017). Causation or selection - examining the relation between education and overweight/obesity in prospective observational studies: a meta-analysis. *Obesity Reviews 18* (6), 660–672. DOI: 10.1111/obr.12537.

Krause, P. & Schäfer, A. (2005). Verteilung von Vermögen und Einkommen in Deutschland: große Unterschiede nach Geschlecht und Alter. *DIW Wochenbericht 72* (11), S. 199–207.

Lampert, T., Kroll, L., Müters, S., & Stolzenberg, H. (2013). Messung des sozioökonomischen Status in der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz 56* (5-6), 631–636. DOI: 10.1007/s00103-012-1663-4.

Lee, H., Ahn, R., Kim, T. H., & Han, E. (2019). Impact of Obesity on Employment and Wages among Young Adults: Observational Study with Panel Data. *International journal of environmental research and public health 16* (1). DOI: 10.3390/ijerph16010139.

Leszczensky, L. & Wolbring, T. (2019). How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. *Sociological Methods & Research*. DOI: 10.1177/0049124119882473.

Leuven, E. & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics, revised 01 Feb 2018.

Magallares, A. (2016). Drive for thinness and pursuit of muscularity: the role of gender ideologies. *Universitas Psychologica, 15*(2), 353-360.

Morris, S. (2007). The impact of obesity on employment. *Labour Economics 14* (3), 413–433. DOI: 10.1016/j.labeco.2006.02.008.

Oakes, J. M. & Johnson, P. J. (2006). Propensity Score Matching for Social Epidemiology. In J. M. Oakes und J. S. Kaufman (Eds.). *Methods in social epidemiology* (pp. 370–392). San Franscisco: John Wiley & Sons.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statist. Surv. 3* (none), 96–146 DOI: 10.1214/09-SS057.

Robert Koch-Institut, Abteilung Für Epidemiologie Und Gesundheitsmonitoring (2000). Bundes-Gesundheitssurvey 1998 (BGS98 1997-1999).

Robert Koch-Institut, Abteilung Für Epidemiologie Und Gesundheitsmonitoring (2015). Gesundheit von Erwachsenen in Deutschland (DEGS1 2008-2012).

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70* (1), 41–55. DOI: 10.1093/biomet/70.1.41.

Sabia, J. J. & Rees, D. I. (2012). Body weight and wages: evidence from Add Health. *Economics and human biology 10* (1), 14–19. DOI: 10.1016/j.ehb.2011.09.004.

Sari, N. & Acan Osman, B. (2018). The effect of body weight on employment among Canadian women: evidence from Canadian data. *Canadian journal of public health = Revue canadienne de sante publique 109* (5-6), 873–881. DOI: 10.17269/s41997-018-0097-7.

Scheidt-Nave, C., Kamtsiuris, P., Gößwald, A., Hölling, H., Lange, M., & Busch, M. A. et al. (2012). German health interview and examination survey for adults (DEGS) - design, objectives and implementation of the first data collection wave. *BMC Public Health 12* (730). DOI: 10.1186/1471-2458-12-730.

Schienkiwitz, A., Mensink, G. B. M., Kuhnert, R., & Lange, C. (2017). Übergewicht und Adipositas bei Erwachsenen in Deutschland. *Journal of Health Monitoring 2* (2), S. 21–28.

Shrier, I. & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC medical research methodology 8* (70). DOI: 10.1186/1471-2288-8-70.

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health services & outcomes research methodology 14* (4), 166–182. DOI: 10.1007/s10742-014-0123-z.

Thefeld, W., Stolzenberg, H., & Bellach, B. M. (1999). Bundes-Gesundheitssurvey: Response, Zusammensetzung der Teilnehmer und Non-Responder-Analyse. *Gesundheitswesen 61* (Sonderheft 2), S57–S61.

Wada, R. & Tekin, E. (2010). Body composition and wages. *Economics and human biology 8* (2), 242–254. DOI: 10.1016/j.ehb.2010.02.001.

Winkler, J. & Stolzenberg, H. (1999). Der Sozialschichtindex im Bundes-Gesundheitssurvey. *Gesundheitswesen 61* (Sonderheft 2), S178-S183.

World Health Organization (2011). *Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008*. Online: https://apps.who.int/iris/bitstream/handle/10665/44583/9789241501491_eng.pdf