Does Smoking Reduce Obesity? Reconciling the Conflicting Evidence

Rusty Tchernis* Keith Teltser† Arjun Teotia‡

November 24, 2020

Abstract

Smoking and obesity are the two leading causes of preventable deaths in the United States. Opposing trends in smoking and obesity in recent years raise the question of whether reduced smoking increases obesity. However, the literature addressing this question is mixed. One strand of research, which includes both observational and experimental evidence, suggests reduced smoking increases obesity. However, another strand of research using observational data finds reduced smoking reduces obesity. The experimental work has been criticized for using older and non-representative data, while the results of studies using observational data are sensitive to statistical methods and concerns about data integrity (e.g., misreporting). In this paper, we reconcile the two strands. To do this, we estimate the causal effect of smoking on obesity while accounting for potential misreporting. In particular, we use self-reported data from the Behavioral Risk Factor Surveillance System (BRFSS), cigarette taxes as a plausibly exogenous instrument for smoking behavior, and account for misreporting using the 2-step estimator developed by Nguimkeu et al. (2019). Using the standard two-stage least squares estimator, the results are consistent with earlier work suggesting that reduced smoking reduces obesity. However, we proceed to show that the results are quite sensitive to specification and functional form. Most notably, we find that using a non-linear first stage reverses the direction of the estimated effect. We then show that using a non-linear first stage *and* accounting for misreporting yields estimates consistent with the experimental literature in both magnitude and direction. Our findings suggest that specification choice and accounting for misreporting may be the missing links required to reconcile the two strands of research. In addition, our findings suggest that reduced smoking accounts for 6% of the concurrent rise in obesity, which adds to the evidence suggesting that increased obesity may be an unintended consequence of anti-smoking policy efforts.

*IZA, NBER, Georgia State University

†Georgia State University, corresponding author (kteltser@gsu.edu)

‡Georgia State University

1 Introduction

Smoking and obesity are the two leading causes of preventable deaths in the United States.[1] Almost 40% of adults in the United States are obese (Hales et al. (2017)), while roughly 14% of adults in the U.S. smoke according to the CDC.[2] In Figure 1, we show that smoking decreased from roughly 22% to 15% between 2002 and 2017 according to the Behavioral Risk Factor Surveillance System (BRFSS), while obesity rose from roughly 26.75% to 28.25% during the same time period. While these trends do not present causal evidence, they raise the question of whether the decline in smoking has contributed to rising obesity. If so, the health benefits of anti-smoking policies may be somewhat overstated.

Using Grossman's human capital model of health, we can consider health a capital stock that depreciates over time but can be replenished by investment (Grossman, 1972). Assuming an individual starts at her optimal body weight, both smoking and weight gain represent disinvestments. Cigarettes contain nicotine, which suppresses appetite and increases one's metabolic rate. Thus, smoking cigarettes tends to reduce body weight, all else equal. However, smoking also reduces lung capacity, which may restrict physical activity and thereby increase body weight. Because the causal effect of smoking on obesity is theoretically ambiguous, determining the true relationship is an empirical question.

There exists a large literature examining the causal relationship between smoking and obesity, though the findings are quite mixed. Using BRFSS data from 1984 to 1999, Chou et al. (2004) suggest that reduced smoking increases body mass index (BMI) by finding increasing cigarette prices have a large positive effect on BMI. Using BRFSS data from 1984 to 2002/2005, Gruber and Frakes (2006) and Courtemanche (2009) examine this relationship more closely using cigarette taxes instead of cigarette prices.[3] The authors find that an increase in cigarette taxes (i.e. reduced

[1] See https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm and https://www.commonwealthfund.org/blog/2018/rising-obesity-united-states-public-health-crisis

[2] See https://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/index.htm, last accessed February 19, 2020

[3] Note that Courtemanche (2009) actually uses BRFSS data to examine the robustness of his central estimates obtained using the National Longitudinal Survey of Youth. Indeed, the author finds that both data sources yield similar estimates.

smoking) *reduces* BMI. However, several other studies around the same time use different observational data or approaches to causal identification to suggest that reduced smoking increases BMI (Baum, 2009; Fang et al., 2009; Liu et al., 2010). Moreover, Courtemanche et al. (2018) also find that reduced smoking increases BMI using a randomized trial of smoking cessation treatments and clinically-measured carbon monoxide levels.

The mixed findings highlight the complexity of identifying the causal effect of smoking on BMI. In particular, two concerns stand out from the literature. First, estimates are sensitive to different approaches to identification and choice of specification. Second, estimates are sensitive to the source of data, which raises the concern that results obtained from self-reported data (as opposed to clinically-measured experimental data) are biased by systematic misreporting. For example, studies have found that accounting for misreporting can significantly impact estimates of marijuana usage, as well as the magnitude and direction of estimated effects of nutritional assistance programs on BMI and obesity (Greene et al., 2017; Almada et al., 2016; Nguimkeu et al., 2019). A large body of medical research has revealed widespread underreporting based on comparisons of self-reported smoking to biochemical indicators of tobacco intake (e.g., Gorber et al., 2009; Nesson, 2017). To our knowledge, no one has examined how misreporting affects estimates of the impact of smoking on BMI.

In this paper, we aim to reconcile the mixed findings in the existing literature. To do this, we use BRFSS data from 2002 to 2017 to examine how specification choice and accounting for misreporting affect estimates of the effect of smoking on BMI. In particular, we employ a variety of linear and non-linear instrumental variables approaches, using cigarette taxes as plausibly exogeneous shifters of smoking behavior. To account for misreporting, we use the 2-step estimation procedure proposed by Nguimkeu et al. (2019). First, we replicate the findings of Gruber and Frakes (2006) and Courtemanche (2009) using our more recent data, obtaining estimates suggesting that reduced smoking reduces BMI. Second, we show that using a non-linear first stage yields estimates of roughly the same magnitude but the sign is flipped, suggesting that reduced smoking *increases* BMI. Finally, using a non-linear first stage and accounting for misreporting, we obtain

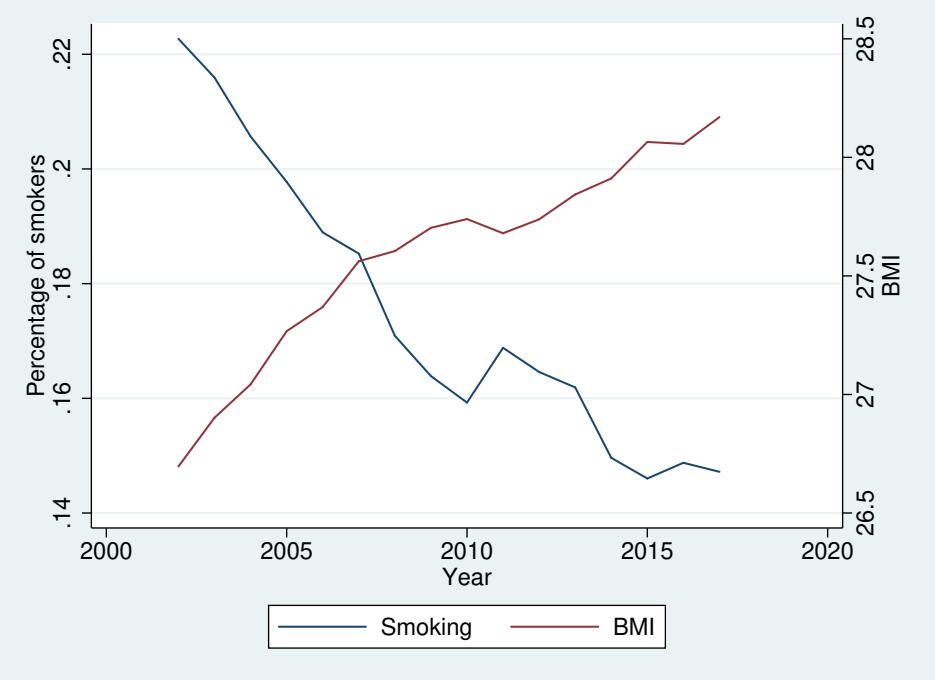

Figure 1: Trends in smoking and obesity (BRFSS 2002-2017)

estimates consistent with the experimental literature in both sign and magnitude. These findings suggest that both specification choice and accounting for misreporting may help reconcile the discrepancies among earlier studies. Our preferred estimate suggests that quitting smoking results in a 1.06 increase in BMI. From 2002-17, our findings suggest that the decline in smoking explains roughly 6% of the concurrent rise in obesity, reinforcing the notion that increased obesity may be an unintended consequence of anti-smoking policy efforts.

The rest of the paper is organized as follows. Section 2 provides a more detailed review of the literature. Section 3 describes our data. Section 4 develops our various approaches to estimation and presents our results. Section 5 concludes.

2 Literature Review

Instrumental variables (IV) approaches are common in the literature on the relationship between smoking and body weight. This is because smoking behavior may be endogenous, as individuals generally self-select into smoking. There are also concerns about reverse causality. That is,

body weight may affect an individual's decision of whether or not to smoke (Cawley et al., 2004; Rees and Sabia, 2010). The literature has primarily used cigarette prices and taxes as plausibly exogeneous instruments, as they shift the cost of smoking but do not seem to have a direct causal relationship with BMI.

In Table 3, we catalog the main estimates from earlier literature on the relationship between smoking and BMI. First, Chou et al. (2004) use BRFSS data from 1984 to 1999 to estimate the determinants of BMI and obesity, and find large positive effects of cigarette prices. Assuming cigarette prices affect individuals only through smoking, these results suggest that reductions in smoking contribute to increases in body weight. In contrast, using BRFSS data from 1984 to 2002 and cigarette taxes instead of prices, Gruber and Frakes (2006) examine this question more closely and find that reduced smoking reduces BMI and obesity rates. The authors propose two possible explanations for why their results differ from those of Chou et al. (2004). First, taxes may be better instruments than prices, as variation in prices may be driven by factors affecting both BMI and demand for cigarettes. Second, Gruber and Frakes (2006) use month-year fixed effects rather than quadratic time trends to more flexibly control for unobservable location-invariant shocks over time. However, the authors caution that their IV estimates may be implausibly large, as they suggest that individuals who quit smoking are 56% less likely to be obese. Courtemanche (2009) revisits Gruber and Frakes (2006) using both BRFSS and the National Longitudinal Survey of Youth (NLSY), and also finds that reduced smoking reduces BMI and obesity rates. Moreover, the author shows that this result is robust to a range of alternative linear specifications.

Other studies around the same time, however, challenge these findings. Baum (2009) use the NLSY for years 1979 through 2002 to estimate the body-weight effects of real cigarette prices and taxes. Using a standard differences-in-differences approach, they find that an increase in the cost of cigarettes (i.e. a reduction in smoking) increases BMI and obesity rates. Fang et al. (2009) use data from the China Health and Nutrition Survey (CHNS), leveraging number of cigarettes smoked and the local per-pack price of the most popular cigarette brand as IVs. Similarly, they find that reduced smoking increases BMI, though they find small and statistically insignificant effects on

obesity rates. Liu et al. (2010) use worksite smoking bans along with BRFSS data from 1998 to 2006 to find further evidence suggesting reduced smoking increases obesity rates.

Moreover, a more recent study by Courtemanche et al. (2018) uses clinical measurements from an experimental setting to provide additional consistent evidence. In particular, they use data from a randomized trial of smoking cessation treatments. Rather than relying on self-reported smoking, which seems to suffer from widespread underreporting, they observe clinically-measured carbon monoxide levels. They find that, in the short run, quitting smoking leads to a gain of 10-11 pounds at average height. In the long run, this effect increases to 11-12 pounds. However, this study focuses on individuals who are trying to quit smoking during the years 1990-1994, which some may argue casts doubt on the external validity of the findings.

If nothing else, the mixed evidence highlights the challenges inherent in estimating the causal relationship between smoking and body weight. In the following sections, we aim to reconcile these seemingly contradictory results. In particular, we argue that specification choice, functional form, and misreporting may explain the apparent discrepancies across previous studies.

3 Data

We start with repeated cross-sectional data from the 2002 through 2017 waves of the BRFSS, which is a health-focused telephone survey conducted by the CDC. The survey collects information on health-related risky behaviors among U.S. residents, as well as chronic health conditions and use of preventative services. The BRFSS was first established in 1984, with 12,258 participants across 14 states. In 2017, the data include approximately 430,000 participants from all 50 states and the District of Columbia. Starting in 2002, the BRFSS contains information on interviewer performance and whether a respondent fully or partially completed the interview. Because we use this information to account for possible misreporting, we restrict our main estimation sample to the years 2002-2017.

In particular, we use the core module that includes self-reported height, weight, demographics,

tobacco use, health status, and chronic health conditions. Following Gruber and Frakes (2006), we restrict our sample to respondents under 65 years of age. After also dropping observations with missing data for our key variables of interest, our sample contains roughly 3.6 million observations. We calculate BMI using respondents' self-reported height and weight, removing extreme outliers by dropping the bottom and top 1% of BMI values, which leaves us with a range of 17.75 to 46.39. Using CDC guidelines, we define a current smoker as one who reported smoking at least 100 cigarettes during their lifetime, and, at the time of participation, reported that they currently smoke either some days or every day. We also use several additional individual-level covariates, including marital status, race, education, and income. As in Gruber and Frakes (2006), we include age-by-gender group indicators.

We supplement our BRFSS data with information on state excise cigarette taxes from the CDC's Tax Burden on Tobacco (TBOT) dataset. The TBOT reports tax rates at the end of October every year, along with the exact date(s) of any rate changes, which we use to determine monthly tax rates. We convert nominal tax rates to 2017 dollars using the Consumer Price Index for All Urban Consumers (CPI-U). Because the BRFSS asks individuals about their behavior over the past 30 days, we follow the previous literature in assigning tax rates to individuals based on the month preceding their survey interview. Finally, we include a time-varying measure of state-level unemployment from the Bureau of Labor Studies to account for economic conditions that may be correlated with body weight, smoking, and/or cigarette taxes.

In Table 1, we report the summary statistics for the BRFSS 2002-2017, adjusted for the stratified sampling weights provided by the CDC. Using survey weights is important in this context, as BRFSS oversamples different population groups across different years depending on the guidance from the CDC. In our sample, 18% of respondents are identified as smokers. The average BMI among smokers is 27.11, which is lower than the average non-smoker BMI of 27.48. In addition, we see that 66% of individuals are white, 11% are black, and 6% are hispanic. In addition, 51% of individuals are female, 28-29% have a high school and/or technical degree, and 32% are college graduates.

In the sample, the mean (real) cigarette excise tax is $2.14. The highest observed nominal tax is $4.35 in New York and Connecticut in 2017, while the lowest in 2017 is $0.17 in Missouri (see Table 2). In addition to substantial across-state variation in tax magnitudes, taxes vary substantitally across time within states. Between 2002 and 2017, there were 132 nominal tax increases.

Finally, the variable we use to account for misreporting is called "Final Disposition" in the BRFSS data. This indicates whether the respondent fully completed their interview, and we use it to predict whether an individual is less likely to answer questions truthfully. We find that approximately 10% of respondents in our sample only partially complete their interview.

4 Estimation and Results

In this section, we start by obtaining naïve OLS estimates of the relationship between smoking and BMI via OLS, and then estimate a variety of IV specifications. Throughout, we follow the previous literature in specifying the following structural linear model of the relationship between BMI and smoking:

$$BMI_{isy} = \alpha S_{isy} + X_{isy}\beta + \tau_y + \mu_s + \epsilon_{isy} \tag{1}$$

where S indicates current smoking status for individual i in state s and month-year y. X is a vector of individual- and state-level covariates, τ_y are month-year fixed effects, μ_s are state fixed effects, and η_{isy} is an idiosyncratic error term. X includes individuals' income, race, marital status, education, age, gender, an age-by-gender interaction term, and state unemployment rate from the BLS.

In the first row, second panel of Table 4, we find that the naïve OLS estimate of the relationship between smoking and BMI is negative. For OLS, and all our other estimation approaches, we present three specifications. Column 1 presents the results when we only include state and month-year fixed effects. Column 2 presents the results when we only include the set of covariates (income, race, marital status, educational attainment, age-by-gender dummies, and state unemployment rate). Column 3 presents the results when including both the fixed effects and covariates.

In our preferred specification, column 3, we estimate that quitting smoking is associated with a BMI increase of 0.7 (2.5% relative to average BMI of 27.65). For someone who is 5' 7" tall and 180 pounds (BMI = 28.2), this corresponds to an increase of roughly 4.5 pounds.

Our first IV approach is a standard two-stage least squares (2SLS) approach, which follows Gruber and Frakes (2006) closely. Next, we obtain estimates from two alternative IV approaches where the first stage is non-linear. The first is a corrected version of the "forbidden regression." The second is a control function approach. Finally, we implement a 2-step estimator proposed by Nguimkeu et al. (2019) to account for the possibility of misreporting. Following the large economic literature on smoking, we use cigarette taxes to instrument for smoking status throughout these analyses. In addition, we weight our regressions throughout by the BRFSS-provided stratified sampling weights.[4]

4.1 Two-Stage Least Squares

In this approach, following Gruber and Frakes (2006), the first stage is specified as a linear probability model:

$$S_{isy} = \psi Tax_{sy} + X_{isy}\phi + \tau_y + \mu_s + \eta_{isy} \tag{2}$$

Tax is the identifying instrument, which reflects the real excise cigarette tax (in 2017 $). The corresponding second stage is given by:

$$BMI_{isy} = \alpha \hat{S}_{isy} + X_{isy}\beta + \tau_y + \mu_s + \epsilon_{isy} \tag{3}$$

where $\hat{S}$ is predicted smoking status from the first stage, and α is the causal parameter of interest.

Table 4 presents the results from the analysis using 2SLS. In the first stage (row 1, top panel, column 3), we estimate that a $1 increase in the real cigarette excise tax significantly reduces the probability of being a smoker by 0.6 percentage points. We also present the F-statistic, which is 16.4 for this specification, suggesting that cigarette taxes are indeed a strong predictor of smoking

[4] See appendix for non-weighted estimates.

behavior. Note that we do not find a statistically significant direct relationship between BMI and cigarette taxes, where the reduced form estimate presented in column 3 of the middle panel suggests that a $1 tax increase statistically insignificantly reduces BMI by 0.025. In row 2, column 3 of the bottom panel, our 2SLS estimate suggests that quitting smoking reduces BMI by roughly 4.2 (or a reduction of 27 pounds for someone 5' 7" tall and 180 pounds). While the standard error of the 2SLS estimate is roughly 50% larger than the coefficient, the magnitude is in line with the estimate of 5.7 from Gruber and Frakes (2006) using BRFSS data from 1984-2002 and our replication of that estimate of 5.34 (s.e. = 3.25). That said, as pointed out by the authors themselves, it still seems to be unreasonably large when compared to the rest of the literature.

In addition, it is also worth noting that the 2SLS estimates are relatively unstable across specifications. When including fixed effects but no covariates, the 2SLS estimate is similar but smaller (3.49). However, when including covariates and excluding fixed effects, the sign flips and the magnitude is much smaller (-0.79, more in line with the stable "naïve" OLS estimates of -0.4 to -0.7). In the following subsections, we show (1) the sign of the estimates is consistently negative when using a nonlinear first stage and (2) conditional on including covariates, the results are not nearly as sensitive to the inclusion of fixed effects. Moreover, we show that when using a nonlinear first stage that accounts for misreporting, the results are stable across all three specifications.

4.2 Nonlinear-Fits-as-Instruments

Next, we would like to estimate the relationship between smoking and BMI by using a probit model for the first stage to obtain predicted smoking status, and then substituting predicted smoking status for observed smoking status in a linear second stage. However, this was famously dubbed the "forbidden regression" by Hausman, since only a linear first stage estimated via OLS is "guaranteed to produce first-stage residuals that are uncorrelated with fitted values and covariates" (Angrist and Pischke, 2008).

To correct for this, we use the nonlinear-fits-as-instruments (hereafter, NFI) approach suggested by Newey (1990). The NFI approach has three stages, which we describe below as stages zero,

one, and two. Stage zero is given by:

$$S_{isy} = \psi Tax_{sy} + X_{isy}\phi + \tau_y + \mu_s + \eta_{isy}, \tag{4}$$

which we estimate as a probit model to obtain predicted smoking status $\hat{S}$. Then, in stage one, we estimate S as a function of $\hat{S}$ along with our set of controls:

$$S_{isy} = \theta\hat{S}_{isy} + X_{isy}\gamma + \tau_y + \mu_s + u_{isy}. \tag{5}$$

Finally, we estimate stage two of the NFI model, where BMI is specified as a function of predicted smoking from stage one, $\tilde{S}$, and the set of controls:

$$BMI_{isy} = \alpha\tilde{S}_{isy} + X_{isy}\beta + \tau_y + \mu_s + \epsilon_{isy}. \tag{6}$$

Note that this approach identifies α using both taxes and nonlinearities in stage zero. While using nonlinearities as identifying information may be somewhat undesirable, Newey (1990) shows that the resulting estimates are more efficient when the probit model is a better approximation of the first-stage conditional expectation function.

Turning to Table 4, row 2, column 3 of the top panel reveals a very similar first stage relationship as in 2SLS. Specifically, the estimate suggests that a $1 increase in the real cigarette excise tax reduces the probability of being a smoker by 0.006 percentage points. The first stage F-statistic is 11.38, again, suggesting that cigarette taxes are a strong predictor of smoking. However, turning to row 3, column 3 of the bottom panel, we find that the estimated effect is similar in magnitude to 2SLS but the direction of the relationship is flipped. That is, our NFI estimate suggests that quitting smoking increases BMI by 6.2 (roughly 40 pounds for someone 5' 7" tall and 180 pounds). In this specification, as opposed to 2SLS, the standard errors are much smaller and the coefficient is highly statistically significant. That said, the magnitude still appears to be unreasonably large relative to the rest of the literature. While the corresponding NFI estimate when we include fixed effects and

exclude covariates is particularly unreasonably large (-20.9) and imprecise (s.e. = 16.2), when we include covariates the estimates are not as sensitive as 2SLS to the inclusion of fixed effects (-4.5 without fixed effects vs. -6.2 with fixed effects).

4.3 Control Function

Next, we estimate the causal relationship of interest using a nonlinear (probit) first stage via the control function (hereafter, CF) approach. Again, the first stage is given by:

$$S_{isy} = \psi T\alpha x_{sy} + X_{isy}\phi + \tau_y + \mu_s + \eta_{isy}. \tag{7}$$

After estimating this probit model, we obtain the residuals (Res) and use them as a covariate in the second stage. The second stage is given by:

$$BMI_{isy} = \alpha S_{isy} + \delta Res_{isy} + X_{isy}\beta + \tau_y + \mu_s + \epsilon_{isy} \tag{8}$$

where S is self-reported smoking status, Res are the residuals from the first stage, along with the usual set of additional controls. The estimated α from the second stage is our causal parameter of interest.

The results presented in row 2, column 3 of the top panel of Table 4 reflect the first stage estimates for both the NFI and CF approaches, as they both use the exact same first stage specification. Turning to row 4, column 3 of the bottom panel, we see that the estimate is quite similar to the NFI estimate (negative relationship) but slightly larger in magnitude (-7.4 vs. the NFI estimate of -6.2). As in the NFI specification, the standard error is much smaller than that of the 2SLS estimate and the coefficient of interest is highly statistically significant. Again, while the corresponding CF estimate when we include fixed effects and exclude covariates is unreasonably large (-11.3) and imprecise (s.e. = 8.6), when we include covariates the estimates are not as sensitive as 2SLS to the inclusion of fixed effects (-5.5 without fixed effects vs. -7.4 with fixed effects).

Even though the sign flips when using a nonlinear first stage, which is in line with the exper-

imental literature, the estimates are still unreasonably large compared to the rest of the literature on the relationship between smoking and BMI. In the next subsection, we show that accounting for misreporting in smoking status has a substantial impact on the stability of the estimates while bringing them back within a reasonable range.

4.4 2-step estimator

Using the 2-step estimator allows us to account for misreporting of smoking in the BRFSS data, following Nguimkeu et al. (2019). Suppose the outcome variable of interest, BMI, is a function of correctly-measured exogeneous covariates X and the (true) smoking indicator, S^*. In particular, we specify:

$$BMI_{isy} = \alpha S^*_{isy} + X_{isy}\beta + \tau_y + \mu_s + \epsilon_{isy}. \tag{9}$$

where α is the key parameter of interest.

We model true smoking status (S^*) as a function of X and the exogenous instrument Tax:

$$S^*_{isy} = \mathbf{1}(\theta Tax_{sy} + X_{isy}\beta + \tau_y + \mu_s + \nu_{isy} \geqslant 0). \tag{10}$$

However, we only observe self-reported smoking status, S, which is a surrogate of one's true smoking status. Thus, we can consider $S_i = S^*_i d_i$, where d is an indicator of misreporting. If an individual truthfully reports their smoking status, $d = 1$ (0 otherwise). Note that $S = S^*$ for true non-smokers regardless of misreporting. That is, we assume a true non-smoker would not falsely report smoking. Next, we model d as a function of observable covariates X and disposition $Disp$:

$$d_{isy} = \mathbf{1}(\rho Disp_{isy} + X_{isy}\gamma + \tau_y + \mu_s + u_{isy} \geqslant 0) \tag{11}$$

such that:

$$\begin{aligned} S_{isy} &= S^{*}_{isy} d_{isy} \\ &= \mathbf{1}(\theta Tax_{sy} + X_{isy}\beta + \tau_y + \mu_s + \nu_{isy} \geqslant 0 \; ; \; \rho Disp_{isy} + X_{isy}\gamma + \tau_y + \mu_s + u_{isy} \geqslant 0). \end{aligned} \tag{12}$$

In the first step, we estimate equation 12 using a partial observability probit model, which allows us to predict each individual's true smoking status, $\hat{S}^{*}$. In the second stage, we estimate equation 9, substituting $\hat{S}^{*}$ in for S^{*}. The resulting α is the 2-step estimate of the true causal effect of smoking on obesity when accounting for endogenous misreporting.

The results of this approach, presented in row 3 of the top panel of Table 4, again yield very similar first-stage estimates of the effect of cigarette taxes on smoking. In our preferred specification (column 3), the first-stage F-statistic is 36.95, again suggesting that cigarette taxes are strong predictors of smoking. Turning to row 5 of the bottom panel, we find a negative relationship between smoking and obesity as in the NFI and CF approaches. However, here we find that quitting smoking leads to an increase in BMI of only 1.06 (about 7 pounds for someone 5' 7" tall and 180 pounds). Relative to our nonlinear specifications that do not consider misreporting, this estimate is much closer to the experimental 2SLS estimate obtained by Courtemanche et al. (2018), who find that quitting smoking increases BMI by 2.2 (about 14 pounds) when using clinically-measured carbon monoxide levels to determine smoking behavior among participants in a smoking cessation randomized trial. Moreover, we find that this 2-step approach yields a much more stable pattern of estimates across our three specifications, where the estimates only range between -0.50 and -1.06.

4.5 Heterogeneity

Finally, we use the 2-step estimator to examine heterogeneity in the effects of smoking on BMI across age, gender, and U.S. census region. While we will focus primarily on presenting the 2-step results, note that we also present the results of OLS, 2SLS, NFI, and CF estimation for comparison purposes.

Turning first to Table 5, our 2-step estimates suggest that cigarette taxes have a slightly larger

impact on smoking among respondents aged 41-65: a reduction of 1.1 percentage points in the probability of being a smoker for a $1 increase in taxes compared to 0.8 points for those aged 18-40. In terms of the effects of smoking status on BMI, however, we find that quitting smoking has a larger positive impact on BMI among the 18-40 age group: an increase of 4.2 (27 pounds for someone 5' 7" and 180 pounds) compared to only 0.8 (5 pounds) among those aged 41-65.

Next, looking at Table 6, our 2-step estimates suggest that cigarette taxes have a substantially larger impact on smoking among female respondents: a 1.2 percentage point smoking reduction compared to 0.6 for men. Interestingly, we find that quitting smoking results in a significantly larger weight gain for men than women: 6.9 BMI points for men compared to a statistically insignificant decrease of 1.3 BMI points for women.

Finally, in Table 7, we examine regional differences in the relationship between smoking and BMI. Looking in row 5 of the bottom panel, we find that quitting smoking increases BMI by 3.2 to 4.5 in the northeast and midwest, while we find no statistically discernable relationship among respondents in the southern and western regions of the United States.

5 Conclusion

In this paper, we show that carefully considering functional form and accounting for misreporting may help explain the conflicting evidence across previous experimental and observational studies of the relationship between smoking and obesity. In particular, we show that using a nonlinear first stage to estimate the relationship between cigarette taxes and smoking status yields a negative relationship between smoking and BMI, which is consistent with the experimental literature and some of the earlier literature using observational data.

Moreover, we show that accounting for misreporting and using a nonlinear first stage with observational data yields estimates of similar sign and magnitude to this body of work. For example, our preferred estimate suggests that quitting smoking increases BMI by 1.06 (about 7 pounds for someone 5' 7" and 180 pounds), which is quite similar to the experimental estimate of 2.2 (14

pounds) from Courtemanche et al. (2018), but quite different from the estimated *decrease* in BMI of 4.2 to 5.7 (27 to 36 pounds) when following the approach of Gruber and Frakes (2006) using 2SLS and observational data.

Using our estimation sample and incorporating survey weights, the proportion of smokers declined from 25.4% in 2002 to 18.8% in 2017 while BMI rose from 26.7 to 27.9. Multiplying the 6.6 percentage point decline in smoking by our preferred estimate of 1.06 yields a BMI increase of 0.07, which explains roughly 6% of the overall BMI increase of 1.2 across the same time period.

More broadly, our work shows how certain functional form choices and failing to account for misreporting might result in biased estimates when using survey data and instrumental variables methods. These considerations have receive little attention in the literature on the effects of smoking. As such, a key contribution of this paper is to demonstrate how researchers might effectively address them in future work.

Bibliography

Almada, L., McCarthy, I., and Tchernis, R. (2016). What can we learn about the effects of food stamps on obesity in the presence of misreporting? *American Journal of Agricultural Economics*, 98(4):997–1017.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baum, C. L. (2009). The effects of cigarette costs on bmi and obesity. *Health Economics*, 18(1):3–19.

Cawley, J., Markowitz, S., and Tauras, J. (2004). Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation. *Journal of health economics*, 23(2):293–311.

Chou, S.-Y., Grossman, M., and Saffer, H. (2004). An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. *Journal of health economics*, 23(3):565–587.

Courtemanche, C. (2009). Rising cigarette prices and rising obesity: Coincidence or unintended consequence? *Journal of Health Economics*, 28(4):781–798.

Courtemanche, C., Tchernis, R., and Ukert, B. (2018). The effect of smoking on obesity: Evidence from a randomized trial. *Journal of health economics*, 57:31–44.

Fang, H., Ali, M. M., and Rizzo, J. A. (2009). Does smoking affect body weight and obesity in china? *Economics & Human Biology*, 7(3):334–350.

Gorber, S. C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., and Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine & Tobacco Research*, 11(1):12–24.

Greene, W., Harris, M. N., Srivastava, P., and Zhao, X. (2017). Misreporting and econometric modelling of zeros in survey data on social bads: An application to cannabis consumption. *Health economics*.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255.

Gruber, J. and Frakes, M. (2006). Does falling smoking lead to rising obesity? *Journal of health economics*, 25(2):183–197.

Hales, C., Carroll, M., Fryar, C., and Ogden, C. (2017). Prevalence of obesity among adults and youth: United states, 2015-2016. *NCHS data brief*, (288):1–8.

Liu, F., Zhang, N., Cheng, K.-W., and Wang, H. (2010). Reduced smoking and rising obesity: Does smoking ban in the workplace matter? *Economics Letters*, 108(3):249–252.

Nesson, E. (2017). The impact of tobacco control policies on adolescent smoking: comparing self-reports and biomarkers. *American Journal of Health Economics*, 3(4):507–527.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, pages 809–837.

Nguimkeu, P., Denteh, A., and Tchernis, R. (2019). On the estimation of treatment effects with endogenous misreporting. *Journal of econometrics*, 208(2):487–506.

Rees, D. I. and Sabia, J. J. (2010). Body weight and smoking initiation: Evidence from add health. *Journal of Health Economics*, 29(5):774–777.

Table 1: Summary Statistics

	Non-Smoker		Smoker		Full Sample	
	Mean	SD	Mean	SD	Mean	SD
BMI	27.77	5.47	27.18	5.41	27.65	5.46
Current Cigarette Smoker	-	-	-	-	0.21	0.41
Cigarette Tax (2017 $)	2.25	1.06	2.10	1.03	2.22	1.06
Male	0.43	0.49	0.46	0.50	0.43	0.50
Black	0.08	0.28	0.09	0.29	0.08	0.28
White	0.78	0.42	0.77	0.42	0.78	0.42
Hispanic	0.08	0.27	0.06	0.24	0.07	0.26
Married	0.63	0.48	0.43	0.49	0.59	0.49
Divorced	0.13	0.34	0.23	0.42	0.15	0.36
Widowed	0.03	0.17	0.05	0.21	0.03	0.18
Grade 9-11	0.03	0.18	0.10	0.31	0.05	0.22
High School	0.24	0.43	0.38	0.49	0.27	0.44
Technical Degree	0.27	0.45	0.31	0.46	0.28	0.45
College Graduate	0.44	0.50	0.18	0.38	0.38	0.49
Disposition Rate	0.93	0.26	0.92	0.27	0.93	0.26
Income (2017 $)	69,497	38,164	48,445	35,449	65,088	38,575
State Unemployment Rate	5.96	2.09	5.97	2.04	5.96	2.08
Observations	2,843,312		753,331		3,596,643	

Table 2: Variation in Nominal Cigarette Taxes (2002-2017)

State	Tax in January 2002 ($)	Tax in Dec 2017 ($)	No. of increases
Alabama	0.165	0.675	2
Alaska	1	2	3
Arizona	0.60	2	2
Arkansas	0.34	1.15	2
California	0.87	2.87	1
Colorado	0.20	0.84	1
Connecticut	0.50	4.35	8
Delaware	0.24	2.10	4
District of Columbia	0.65	2.50	2
Florida	0.339	1.339	1
Georgia	0.12	0.37	1
Hawaii	1	3.2	9
Idaho	0.28	0.57	1
Illinois	0.58	1.98	2
Indiana	0.155	0.995	2
Iowa	0.36	1.36	1
Kansas	0.24	1.29	3
Kentucky	0.03	0.60	3
Louisiana	0.24	1.08	3
Maine	1	2	1
Maryland	0.66	2	2
Massachusetts	0.76	3.51	3
Michigan	0.75	2	2
Minnesota	0.48	3.04	5
Mississippi	0.18	0.68	1
Missouri	0.17	0.17	0
Montana	0.18	1.70	2
Nebraska	0.34	0.64	1
Nevada	0.35	1.80	2
New Hampshire	0.52	1.78	5
New Jersey	0.80	2.70	5
New Mexico	0.21	1.66	2
New York	1.11	4.35	3
North Carolina	0.05	0.45	3
North Dakota	0.44	0.44	0
Ohio	0.24	1.60	3
Oklahama	0.23	1.03	1
Oregon	0.68	1.32	4
Pennsylvania	0.31	2.60	4
Rhode Island	1	4.25	7
South Carolina	0.07	0.57	1
South Dakota	0.33	1.53	2
Tennessee	0.13	0.62	2
Texas	0.41	1.41	1
Utah	0.515	1.70	2
Vermont	0.44	3.08	7
Virginia	0.025	0.30	2
Washington	1.425	3.025	3
West Virginia	0.17	1.20	2
Wisconsin	0.77	2.52	2
Wyoming	0.12	0.60	1

Table 3: Estimates from Previous Literature where BMI is Dependent Variable

Study	Indep. Var.	Data	Method	Estimate	Interpretation
Smoking Increases BMI					
Gruber and Frakes (2006)	Smoking Status	BRFSS 1984-2002	2SLS	5.705	Effect of smoking using cigarette taxes as an instrument
Courtemanche (2009)	Cigarette Prices	BRFSS 1984-2005	OLS	-0.234	Effect of a $1 increase in real price of cigarettes
Courtemanche (2009)	Cigarette Taxes	NLSY 1979	OLS	-0.343	Effect of a $1 increase in real cigarette taxes
Smoking Decreases BMI					
Chou et al. (2004)	Cigarette Prices	BRFSS 1984-2002	OLS	0.486	Effect of a $1 increase in real price of cigarettes
Baum (2009)	Cigarette Taxes	NLSY 1979	OLS	0.328	Effect of a $1 increase in real cigarette taxes
Baum (2009)	Cigarette Prices	NLSY 1979	OLS	0.162	Effect of a $1 increase in real price of cigarettes
Fang et al. (2009)	Smoking Status	China Health and Nutrition Survey 2006	2SLS	-0.124	Effect of a one-cigarette increase in smoking (instruments: community smoking rate, cigarette prices)
Liu et al. (2010)	Smoking Status	BRFSS 1998-2006	2SLS	-2.054	Effect of workplace smoking bans
Courtemanche et al. (2018)	Smoking Status	Lung Health Study	RCT/2SLS	-2.202	Effect of smoking cessation treatment

Table 4: Effect of Smoking on BMI

	(1)	(2)	(3)
First Stage: Regress Smoking Status on Cigarette Taxes			
2SLS	-0.005***	-0.015***	-0.006***
	(0.002)	(0.003)	(0.001)
	[9.18]	[20.77]	[16.38]
NFI & CF	-0.006***	-0.016***	-0.006***
	(0.002)	(0.004)	(0.002)
	[6.74]	[20.48]	[11.38]
2-Step	-0.005***	-0.018***	-0.008**
	(0.002)	(0.000)	(0.000)
	[128.16]	[66.90]	[36.95]
Linear Reduced Form: Regress BMI on Cigarette Taxes			
Cigarette Taxes	-0.019	0.012	-0.025
	(0.050)	(0.048)	(0.042)
Second Stage: Regress BMI on Predicted Smoking Status			
OLS	-0.404***	-0.707***	-0.707***
	(0.085)	(0.089)	(0.080)
2SLS	3.489	-0.792	4.200
	(8.260)	(3.081)	(6.302)
NFI	-20.890	-4.526***	-6.210***
	(16.168)	(1.248)	(0.631)
CF	-11.270	-5.482***	-7.366***
	(8.615)	(1.510)	(0.760)
2-Step	-0.678	-0.495	-1.059**
	(1.484)	(0.937)	(0.499)
Observations	3,596,643	3,596,643	3,596,643
State FE	✓		✓
Quarter-Year FE	✓		✓
All covariates		✓	✓

Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. Note: the coefficient on the disposition variable in the first stage of the 2-step estimator is -0.021 (s.e. is 0.010). * p<0.10, ** p<0.05, *** p<0.01.

Table 5: Effect of Smoking on BMI by Age

	OLS	2SLS	NFI	CF	2-Step
Panel A: Age 18-40					
	First Stage (DV = Smoking Status)				
Cigarette Taxes	-	-0.006***	-0.006**	-0.006**	-0.008***
	-	(0.002)	(0.003)	(0.003)	(0.000)
	-	[8.62]	[5.17]	[5.17]	[75.27]
	Second Stage (DV = BMI)				
Smoking Status	-0.195**	4.039	-2.575***	-3.276***	-4.217***
	(0.082)	(9.465)	(0.539)	(0.676)	(0.542)
Observations	1,218,758	1,218,758	1,218,758	1,218,758	1,218,758
Panel B: Age 41-65					
	First Stage (DV = Smoking Status)				
Cigarette Taxes	-	-0.006***	-0.006***	-0.006***	-0.011***
	-	(0.001)	(0.001)	(0.001)	(0.000)
	-	[23.46]	[21.99]	[21.99]	[31.19]
	Second Stage (DV = BMI)				
Smoking Status	-1.422***	4.112	-6.058***	-7.184***	-0.784
	(0.079)	(4.690)	(0.485)	(0.595)	(0.791)
Observations	2,377,885	2,377,885	2,377,885	2,377,885	2,377,885

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: Effect of Smoking on BMI by Gender

	OLS	2SLS	NFI	CF	2-Step
Panel A: Male					
	First Stage (DV = Smoking Status)				
Cigarette Taxes	-	-0.005***	-0.005***	-0.005***	-0.006***
	-	(0.002)	(0.002)	(0.002)	(0.000)
	-	[7.57]	[6.77]	[6.77]	[76.07]
	Second Stage (DV = BMI)				
Smoking Status	-0.778***	3.871	-4.514***	-4.844***	-6.893***
	(0.091)	(5.632)	(0.532)	(0.579)	(0.499)
Observations	1,564,310	1,564,310	1,564,310	1,564,310	1,564,310
Panel B: Female					
	First Stage (DV = Smoking Status)				
Cigarette Taxes	-	-0.007***	-0.007***	-0.007***	-0.012***
	-	(0.001)	(0.002)	(0.002)	(0.000)
	-	[24.56]	[14.58]	[14.58]	[25.25]
	Second Stage (DV = BMI)				
Smoking Status	-0.521***	3.214	-0.876**	-1.002*	1.296
	(0.078)	(6.941)	(0.392)	(0.509)	(0.976)
Observations	2,032,333	2,032,333	2,032,333	2,032,333	2,032,333

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7: Effect of Smoking on BMI: By Census Region

	NE	MW	SO	WE
First Stage: Regress Smoking Status on Cigarette Taxes				
2SLS	-0.006***	-0.009**	-0.009	-0.000
	(0.002)	(0.003)	(0.006)	(0.002)
	[13.31]	[7.85]	[2.77]	[0.00]
NFI & CF	-0.007***	-0.010***	-0.011*	0.001
	(0.002)	(0.003)	(0.006)	(0.002)
	[10.63]	[9.17]	[3.50]	[0.39]
2-Step	-0.007***	-0.008***	-0.013*	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
	[10.15]	[24.43]	[4.99]	[10.92]
Second Stage: Regress BMI on Predicted Smoking Status				
OLS	-0.507***	-0.799***	-0.932***	-0.340*
	(0.092)	(0.064)	(0.086)	(0.160)
2SLS	15.345***	1.233	3.901	-3,394
	(1.914)	(5.931)	(4.456)	(344,919)
NFI	-5.741***	-6.904***	-5.825***	-6.330***
	(0.830)	(0.804)	(0.406)	(0.800)
CF	-6.521***	-7.680***	-7.111***	-7.180***
	(0.969)	(0.926)	(0.544)	(0.892)
2-Step	-3.198***	-4.492***	0.673	-0.900
	(0.654)	(0.389)	(0.577)	(0.742)
Observations	696,936	879,807	1,122,084	897,816

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.