The Distributional Effect of Education on Body Mass

Young-Joo Kim and Vince Daly

We investigate the effect of education on mid-life obesity, with particular attention to potential heterogeneity across the Body Mass Index (BMI) distribution. Applying quantile regression methods to British men and women, we first find that childhood and parental BMI are critical determinants of obesity in middle age. We then establish that even when controlling for various weight-related factors in childhood and a potential endogeneity bias, a higher education level reduces the probability of being obese in middle age. We show that this education effect is obtained by a compression of the distribution of BMI (kg/m2) and a shifting of its center leftward toward a more healthy BMI range. We further show that income and physical activity are important channels of the education effect, and the significant effect of education at the upper quantile of the BMI distribution is neither a disguised income effect nor a healthy behavior effect.

Keywords: Obesity, Body Mass Index, Education, Quantile regression

JEL Classification: I12, I14, I26

Young-Joo Kim, Associate Professor, Department of Economics, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul, 04066 Korea. (Tel): 82-2-320-1759, (E-mail): y.j.kim@hongik.ac.kr; Vince Daly, Professor Emeritus, Department of Economics, Kingston University, 55-59 Penrhyn Rd, Kingston upon Thames KT1 2EE, U.K. (Tel): 44-208-417-9000, (E-mail): vdaly@kingston.ac.uk

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A8019707 & NRF-2020S1A5A2A03046422).

[**Seoul Journal of Economics** 2021, Vol. 34, No. 2]
DOI: 10.22904/sje.2021.34.2.001

I. Introduction

In recent decades, the proportion of overweight and obese people has rapidly increased globally to the point of being recognized as a public health issue in many countries and by international institutions, including the World Health Organization (WHO). Data from the WHO in 2020 revealed that the UK has one of the largest proportions of obese adults in Europe, with an estimated adult obesity rate of 27.8 percent in 2016 (see Figure 1). Developed countries, particularly, exhibit increased interest in the possibility of social gradients in the incidence of obesity, an education gradient being a leading example. Given the evidence of increased health risks associated with obesity (Gatineau *et al.* 2014; Guh *et al.* 2009; Jung and Choi 2014; Renehan *et al.* 2008; Peeters *et al.* 2003), such a gradient can imply differences in morbidity and mortality rates across social groups. This presents a challenging issue to policymakers and researchers concerned about advancing public health equity.

The existing literature provides some evidence of an education gradient in obesity. Cutler and Lleras-Muney (2008) show how education attainment is negatively associated with the mean value of the Body Mass Index (BMI) or the probability of being obese (BMI≥30). Subsequent studies have investigated the potential causal effect of education using instrumental variables or a regression discontinuity approach, but the findings are mixed. Kemptner *et al.* (2011), Brunello *et al.* (2013), and Bockerman *et al.* (2017) find significant effects of education on BMI, whereas Reinhold *et al.* (2010), Jurges *et al.* (2011), and Clark and Royer (2013) find no such effect. The empirical findings from these investigations are based on least squares regressions, which focus on modeling the mean of the BMI distribution. Importantly, the question of how education is associated with aspects of BMI distribution other than in its mean remains largely unanswered even though the incidence of obesity is more directly evidenced by the upper quantiles of the BMI distribution rather than its mean.

In this study, we assess the extent to which educational attainment affects the BMI distribution by using quantile regressions. Quantile regression is a useful tool for identifying the impact of education on the whole of the BMI distribution. As the left and right tails of the BMI distribution both contain regions of unhealthy body size, we expect that an ideal education effect would compress the BMI distribution by reducing the probability in both tails, but conventional regression methods may not detect this feature. Binary dependent variable models might offer an adequate approach to modeling tail behavior in the BMI distribution. For example, Cutler and Lleras-Muney (2008) and Webbink *et al.* (2010) estimate the schooling effect on the probability of being overweight. Yatchew and Chriliches (1985), however, establish that the usual numerical maximum likelihood estimators for binary dependent variable models are inconsistent in the presence of heteroskedastic errors, but this situation is not the case for quantile regression.

The next question to be explored is to ascertain the channels through which education might affect obesity. Some studies shed light on the mechanism of the education effect. For example, Brunello *et al.* (2013) show that part of the education effect for women is explained by income, employment, and physical activity. A higher level of education is more likely to induce higher income and better access to health care. At the same time, education attainment may motivate an individual to maintain regular physical activity regardless of income level, and this characteristic can play an essential role in the formation of body shape. Further examination of factors that might explain the link between education and BMI is warranted as useful for the development of effective strategies and policies to mitigate the obesity pandemic.

We extend the literature in three important ways. First, our use of quantile regression is novel in this area of study and can reveal the heterogeneous effects of education across the BMI distribution. Second, drawing on several studies (Vogler *et al.* 1995; Sacerdote 2007; Elks *et al.* 2012; Classen and Thompson 2016) that document strong transmission effects in weight from parents to children, we attempt to control for childhood factors that are highly correlated with BMI in adulthood, such as birthweight; personal and parental BMIs; and health status at ages 5, 10, and 16. Third, we exploit a long panel dataset from the British Cohort Study (BCS) that has tracked its respondents for more than 40 years (from birth in 1970 to middle age). Using this rich information on individual life histories, we further examine whether and how physical activity and income in early and later stages of life play any role in the formation of education impact on the health status captured by BMI in adulthood while recognizing and addressing other potential determinants of obesity.

II. Methods

A. Data

We use data from the BCS, a survey of approximately 17,000 people born in England, Scotland, and Wales in a single week of April in 1970. The respondents have been followed since birth and have now reached middle age. We extract data from the 2012 wave of this survey in which the BMI and health status of the respondents were recorded. Our sample consists of 1,295 individuals with income at age 42 observed for 959 of those respondents. Table 1 provides the summary statistics of the key variables used for this work.

B. Covariates

The BCS is exceptionally rich in providing details of the respondents' family backgrounds and childhood characteristics. For the intergenerational transmission effect and childhood environment effect, we control for the childhood histories of BMI and health, parental BMI, parental income, socioeconomic status, education, and other characteristics collected at birth and at ages 5, 10, and 16. For other measures that are likely to be associated with adult obesity, we use an individual's regular physical activity indicators at age 16 and 42 and their parents' regular physical activity indicators when the respondent was 16.

For the measures of educational attainment, we employ four categories defined by the highest qualification obtained: no academic qualification (dropout); General Certificate of Secondary Education (GCSE); an undergraduate degree (Bachelor); higher degree (MA or PhD). We also use an indicator for tertiary education that combines undergraduate and graduate qualifications. For the individuals in this dataset, compulsory schooling typically finished at age 16 and was assessed by the GCSE, which we select as the reference category of educational attainment for modeling purposes.

C. Statistical Methods

We use quantile regression methods developed by Koenker and Bassett (1978) and Koenker (2005) to estimate the schooling effect on BMI distribution. We focus on the 25th, 50th, and 75th percentiles of the BMI distribution. We address endogeneity issues that may arise from a potential correlation between schooling and the unexplained part of BMI with alternative approaches. First, we adopt the instrumental variable (IV) quantile regressions proposed by Abadie *et al.* (2002). We implement this approach using the procedure developed by Frölich and Melly (2010). As this approach is based on one endogenous variable matched with a binary instrument, we exclude individuals with no qualification and use a single indicator for tertiary education (undergraduate and graduate). For a binary instrumental variable, we employ an indicator of whether family income at age 16 was above the median of the income distribution. Four assumptions are given in Abadie *et al.* (2002) for the instrument validity in the IV quantile regression. The first is that potential outcomes and educational attainment are jointly independent of an instrument given observed covariates. The second assumption is that the assignment of an instrument is nontrivial. For the first assumption, the instrument-error independence assumption is not easily testable, but Abadie *et al.* (2002) suggest the testing of the exclusion restriction by showing that the potential outcome is not directly affected by the instrument. In our context, we check whether BMI in adulthood is not directly affected by family income in childhood and have confirmed that this is the case. We further test if our instrument is not associated with BMI at 42 after controlling for parental and childhood BMI and other covariates. The *p*-value of the test indicates that our instrument satisfies the exclusion restriction. The remaining two assumptions are satisfied if respondents are more likely to obtain university education with parental income above the median, an outcome which is plausible in our case.

As an alternative approach to the sensitivity testing of the distributional effect of education on BMI with respect to endogeneity bias, we test the equivalence of the estimated coefficients of education across quantiles. By testing whether a significant difference occurs in the estimated effects across quantiles, we infer whether the estimated education effect at a specific quantile is driven by endogeneity bias. We discuss the test results in the next section.

III. Results

The quantile regression results in Panel A of Table 2 clearly indicate that university education has a significant impact on BMI distribution even after controlling for the various childhood and environment factors listed in Table 1. The most notable finding is that a higher level of education has a greater impact at the upper quantile of the distribution, whereas the effect is negligible at the lower quantile.

Building on these findings, we next examine the role of physical activity and income as potential channels for an education effect on the BMI distribution. We first consider the possibility of an indirect effect of education via income by including income at age 42 in addition to parental income at respondents' ages 10 and 16. We also explore the possible transmission of education effects via health-promoting physical activity while controlling for family physical activity in childhood by including indicators of own and parental regular physical activity when respondents were aged 16 and an indicator of own regular physical activity at 42. Since opportunities for physical activity in childhood may have been influenced by local environmental factors (such as whether the respondent lives in a well-provided and safe neighborhood), we control for parental socioeconomic status at respondents' age 16.

In this extended specification, the sample size is reduced from 1,295 to 959, and the estimation results are presented in Panel B of Table 2. Consistent with our prior results, the broad pattern of a substantial education effect that becomes large at the upper quantile is maintained. Among the activity measures and income, regular exercise at 42 is estimated to be the most important and significant factor associated with the BMI distribution. Specifically, regular exercise at 42 is associated with a 0.34 decrease of BMI at the median and 0.47 decrease of BMI at the 0.75 quantile, thereby shifting the BMI distribution leftward into a healthier region. Income at age 42 is also estimated to significantly impact the median of the BMI distribution and may thus explain part of the education effect observed at the median. Nevertheless, the education effect remains significant regardless of any income effect at the upper quantile. Tertiary educational qualifications (undergraduate or graduate) are associated with a decrease of BMI at the upper quantile, thereby confirming a heterogeneous effect of education across the BMI distribution.

We next examine whether our findings from quantile regressions are robust to a potential selection bias driven by the correlation between education attainment and the error term. In the empirical framework of IV quantile regression, we reduce the four education levels into the two categories of secondary education with GCSE and tertiary education with a university degree. Then we compare how the latter has an additional effect relative to the former.

To facilitate comparison for models with a single combined indicator for tertiary education, we first present the standard quantile regression results and the IV quantile regression results in Panels A and B, respectively, of Table 3. We have similar findings of the largest proportional impact at upper quantiles. The magnitude of the point estimate becomes large with this approach. However, a substantial and significant effect of education is still observed at the median and 75th percentile of the distribution, thereby confirming the leftward shift plus compression of the BMI distribution in response to university education. The consistent patterns of the education effects on the conditional distribution of BMI across specifications suggest that the endogeneity bias in the quantile regression estimates, if any, is not large enough to negate the education effect observed earlier.

The equivalence test in Table 4 also supports the limited impact of unobserved confounding factors. Table 4 presents the estimated difference in education effects between the 75th and 25th percentiles, with p-values for the F test of the constraint that the education effect is the same at both quartiles. We find no evidence that heterogeneous education effects along the distribution are led by the bias, provided that the endogeneity bias is of similar magnitude across quantiles.

For a graphical summary of the key findings of this study, we provide three different estimates of the effect of having a university degree in Figure 2. The dashed line of "Quantile regression" is the estimate from the quantile regression as shown in Panel A of Table 3. The solid line labeled "observed differences" reflects the differences in raw quantiles between university degree holders and GCSE holders. The dotted line is the IV quantile regression estimate as shown in Panel B of Table 3. The standard quantile regression presents the monotonic increase of the education effect from the lower to the upper quantile of the distribution. The IV quantile regression suggests that the impact is largest at the upper quantile. Overall, most of the estimates show that education has significant effects on the BMI distribution, with the most considerable effects at the upper quantiles which characterize overweight and obese status.

IV. Discussion

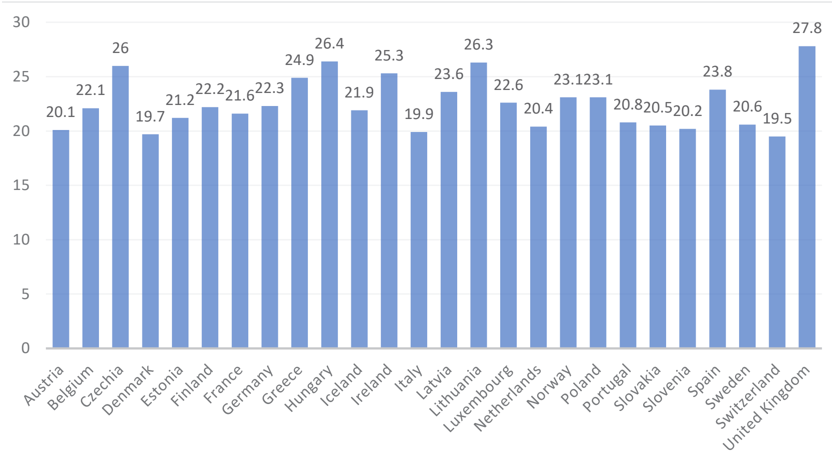
We provide a more comprehensive picture of the education effect on obesity by establishing that university education has a significant effect on the shapes of the conditional distribution of BMI. A higher level of education not only compresses the BMI distribution but also shifts its center leftward toward a more healthy BMI range, thereby reducing obesity risk. The findings of this study corroborate previous ones (Kemptner *et al*. 2010; Brunello *et al*. 2013; Kim 2016; Bockerman *et al*. 2017) on the educational effects on the mean level of BMI or the probability of being obese from developed countries by showing that education has a protective effect on BMI distribution.

This study has some limitations. Although we carefully address the endogeneity issue in estimating the education effect, we cannot exclude the possibility of other confounding effects such as reverse causality. BMI in childhood may affect academic achievement and ultimately education attainment in later years, an issue that is an important concern in child obesity literature. In addition, the sample used here, being extracted from long-term panel data, is small relative to registration data, as is increasingly used in recent literature.

Despite these shortcomings, the current study does add value to the literature. First, by using quantile regressions, we provide new evidence on the heterogeneous effect of education by showing how it evolves differently across the BMI distribution. Second, we confirm that the heterogeneous effect of education on BMI is impervious to various weight-related factors in childhood, including childhood histories of BMI, health, and also parental BMI. Third, we explore potential mechanisms of the education effect by examining income and physical activity at different stages of life and reveal that these features are important channels of the education effect and consequently expand the evidence for the physical activity impact documented in Brunello *et al.* (2013). Our findings further indicate that these factors are insufficient to explain the education effect observed at the upper quantile of the BMI distribution.

V. Conclusion

This study presents new evidence of the significant protective effects of education on the BMI distribution, especially at the university education level. From the perspective of public health and policy intervention, it is noteworthy that the BMI-reducing effect of university education becomes larger as we move from the left tail to the right tail



Notes: Age-standardized estimates of the percent of obese adults (BMI≥30) in 2016
Source: WHO, Global health observatory, World health data platform

Figure 1
Prevalence of obesity among adults in Europe

of the distribution. Therefore, the most substantial impact occurs for individuals who are most likely to be overweight or obese and thus face a higher health risk. Given that income and physical activity alone do not explain the education effect, further study of other channels will be useful for a better understanding of the mechanisms that drive the education effect at the higher education level.

Table 1
Summary Statistics

Variables	Mean	Standard Deviation
Male	0.402	0.490
Asian	0.014	0.117
Higher degree	0.069	0.254
Bachelor degree	0.306	0.461
Dropout	0.185	0.389
Mother's age at birth < 20	0.067	0.250
Birth weight	3.339	0.517
Father's school leaving age	15.620	1.247
Being ill at 5	0.011	0.107
Family Income at 10	137.158	54.084
BMI of Mother at 10	23.178	3.573
BMI of Father at 10	24.311	2.815
BMI at 10	16.826	2.072
Being ill at 10	0.059	0.235
Family Income at 16	234.905	120.278
BMI at 16	20.689	2.570
Being ill at 16	0.240	0.427
BMI at 42	26.528	5.162
Income at 42	48,028.52	28,660.51
Sample Size	1,295	

Notes: BMI is measured in kg/m2. Income at 42 is annual income, while family income at 10 and 16 are weekly income. The sample size for income at 42 is 959.

Table 2
OLS and Quantile Regressions of BMI on Educational Levels with 4 Groups of Education Level

	(1)		(2)		(3)		(4)	
Explanatory Variables	OLS		Quantile Regression					
			0.25 Quantile		0.50 Quantile		0.75 Quantile	
	Panel A: Basic Models							
Higher degree	–1.338*	(0.518)	–0.064	(0.445)	–1.041*	(0.484)	–1.082†	(0.626)
Bachelor	–0.720*	(0.302)	–0.186	(0.299)	–0.495	(0.314)	–0.909*	(0.388)
Dropout	–0.414	(0.346)	0.227	(0.346)	–0.052	(0.379)	0.012	(0.460)
Sample Size	1,295		1,295		1,295		1,295	
	Panel B: Extended Models with Activity Measures and Income							
Higher degree	–1.077†	(0.565)	–0.053	(0.522)	–0.832	(0.562)	–1.358†	(0.764)
Bachelor	–0.882*	(0.340)	–0.258	(0.317)	–0.458	(0.369)	–1.715*	(0.509)
Dropout	–0.581	(0.399)	0.228	(0.391)	–0.099	(0.475)	–0.609	(0.616)
High SES at 16	–0.331	(0.364)	0.338	(0.316)	0.032	(0.383)	–0.469	(0.575)
Exercise at 16	–0.533	(0.390)	–0.647†	(0.372)	–0.688	(0.472)	–0.527	(0.553)
Mother exercised at 16	–0.072	(0.418)	0.233	(0.408)	0.021	(0.552)	0.209	(0.576)
Father exercised at 16	0.797*	(0.394)	0.485	(0.367)	0.616	(0.480)	0.652	(0.545)
log income at 42	–0.294	(0.242)	–0.298	(0.220)	–0.481†	(0.249)	–0.094	(0.362)
Exercise at 42	–0.367*	(0.065)	–0.277*	(0.059)	–0.341*	(0.084)	–0.470*	(0.095)
Sample Size	959		959		959		959	

Notes: * indicates statistical significance at the 5% level and † for the 10% level. For quantile regressions, heteroskedasticity robust standard errors are in parentheses. The reference education level is the GCSE. Models in Panel A include covariates listed in Table 1 except for income at 42, and models for Panel B use all variables.

TABLE 3

QUANTILE AND IV QUANTILE REGRESSIONS OF BMI ON EDUCATIONAL LEVELS WITH 2 GROUPS OF EDUCATION LEVEL

	(1)		(2)		(3)	
Explanatory variables	0.25 Quantile		0.50 Quantile		0.75 Quantile	
	Panel A: Quantile Regression					
Bachelor+	–0.221	(0.288)	–0.537†	(0.304)	–0.909*	(0.379)
Sample Size	1,055		1,055		1,055	
	Panel B: IV Quantile Regression					
Bachelor+	–2.424*	(1.466)	–2.813	(1.749)	–3.605*	(1.622)
Sample Size	1,055		1,055		1,055	

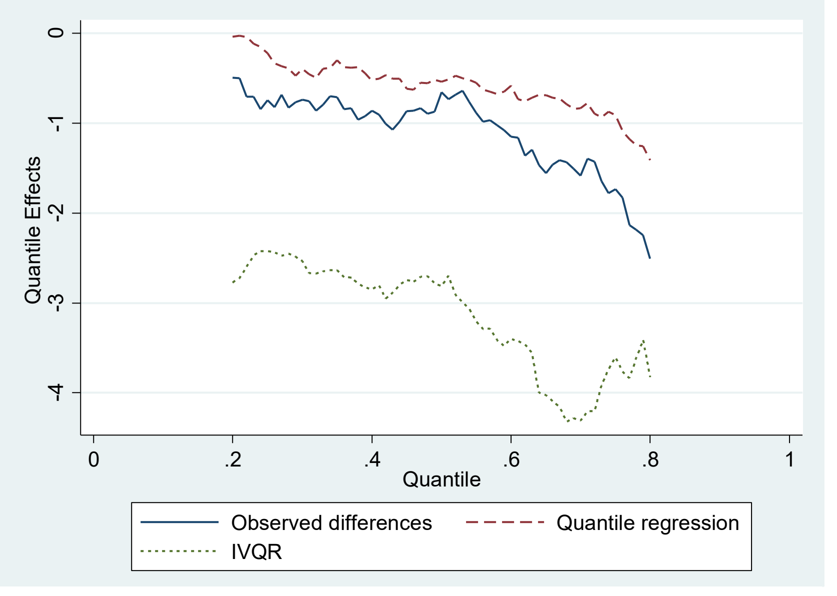
Notes: * indicates statistical significance at the 5% level and † for the 10% level. For quantile regressions, heteroskedasticity robust standard errors are in parentheses. The dropout sample is excluded, and the reference education level is the GCSE. All models include covariates listed in Table 1 except for income at 42. The instrument is an indicator of having family income above the median at age 16.

TABLE 4

TESTING EQUALITY OF QUARTILE REGRESSION COEFFICIENTS

	$\delta_{j,\,0.75}-\delta_{j,\,0.25}$ with [p-values]			
Explanatory Variables	(1)		(2)	
Higher degree	–1.018	[0.060]	–0.845	[0.156]
Bachelor	–0.723	[0.058]	–0.800	[0.076]
Log income at 42	No		Yes	
Sample Size	1,295		959	

Notes: Each cell reports the estimated difference, $\delta_{j,\,0.75}-\delta_{j,\,0.25}$, between the effects of education level j on the upper and lower quartiles of the BMI distribution. The figures in parentheses are p-values for testing the null hypothesis of zero difference. All explanatory variables listed in Table 1 are included in the quantile regressions.



The estimated effects of having a university degree on BMI quantiles are illustrated. *Observed differences* are estimated from quantile regressions without covariates. Quantile regression represents the estimates in Panel A of Table 3, and *IVQR* represent the IV quantile regression estimates in Panel B of Table 3.

Figure 2
Comparison of Quantile Regression Estimates

(Received March 3 2020; Revised April 30 2020; Accepted August 20 2020)

References

Abadie, A., J. Angrist, and G. Imbens. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings" *Econometrica* 70 (No. 1 2002): 91-117.

Böckerman, P., J. Viinikainen, L. Pulkki-Råback, C. Hakulinen, N. Pitkänen, T. Lehtimäki, J. Pehkonen, O. T. Raitakari. "Does Higher Education Protect Against Obesity? Evidence using Mendelian Randomization." *Preventive Medicine* 101 (2017): 195-98.

Brunello, G., D. Fabbri, and M. Fort. "The Causal Effect of Education on

Body Mass: Evidence from Europe." *Journal of Labor Economics* 31 (No. 1 2013): 195-223.

Classen, T. J., and O. Thompson. "Genes and the Intergenerational Transmission of BMI and Obesity." *Economics and Human Biology* 23 (2016): 121-33.

Cutler, D.M. and A. Lleras-Muney. "Education and Health: Evaluating Theories and Evidence." House, J., R. Schoeni, G. Kaplan, and H. Pollack (eds), *Making Americans Healthier: Social and Economic Policy as Health Policy*. New York: Russell Sage Foundation. 2008.

Elks, C.E., M. Den Hoed, J. H. Zhao, S. J. Sharp, N. J. Wareham, R. J. Loos, and K. K. Ong. "Variability in the heritability of body mass index: a systematic review and meta-regression." *Frontiers in endocrinology* 3 (No. 29 2012): 1-16.

Frölich, M. and B. Melly. "Estimation of Quantile Treatment Effects with Stata." *The Stata Journal* 10 (No. 3 2010): 423-57.

Gatineau, M., Hancock, C., Holman, N., Outhwaite, H., Oldridge, L., Christie, A., & Ells, L. Adult obesity and type 2 diabetes. *Public Health England* (2014): 5-14.

Guh, D. P., W. Zhang, N. Bansback, Z. Amarsi, C. L. Birmingham, and A. H. Anis. "The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis." *BMC Public Health* 9 (No. 1 2009): 88.

Jung, U. J., and M. S. Choi. "Obesity and its metabolic complications: the role of adipokines and the relationship between obesity, inflammation, insulin resistance, dyslipidemia and nonalcoholic fatty liver disease." *International Journal of Molecular Science* 15 (No. 4 2014): 6184-6223.

Kemptner, D., H. Jurges, and S. Reinhold. "Changes in Compulsory Schooling and the Causal Effect of Education on Health: Evidence from Germany." *Journal of Health Economics* 30 (No. 2 2011): 340-354.

Kim, Y-J. "The Long-Run Effect of Education on Obesity in the US." *Economics and Human Biology* 21 (2016): 100-109.

Koenker, Roger and G. Bassett, Jr. "Regression Quantiles." *Econometrica* 46 (No. 1 1978): 33-50.

Koenker, R. "Quantile Regression." Econometric Society Monograph Series, Cambridge University Press. 2005.

Peeters, A., J. J. Barendregt, F. Willekens, F., J. P. Mackenbach, A.

A. Mamun, and L. Bonneux. "Obesity in adulthood and its consequences for life expectancy: a life-table analysis." *Annals of internal medicine* 138 (No. 1 2003): 24-32.

Renehan, A. G., M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen. "Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies." *Lancet* 371 (No. 9612 2008): 569-578.

Sacerdote, B. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *The Quarterly Journal of Economics* 122 (No. 1 2007): 119-57.

Vogler, G. P., T. I. Sorensen, A. J. Stunkard, M. R. Srinivasan, and D. C. Rao. "Influences of Genes and Shared Family Environment on Adult Body Mass Index Assessed in an Adoption Study by a Comprehensive Path Model." *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* 19 (No. 1 1995): 40-45.

World Health Organization. Prevalence of Obesity among Adults, age-standardized. 2020. (https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/prevalence-of-obesity-age-standardized, accessed July 1, 2020.)

Yatchew, Adonis and Z. Griliches. "Specification Error in Probit Models." *Review of Economics and Statistics* 67 (No. 1 1985): 134-139.