REGION
The Journal of ERSA
Powered by WU

Volume 11, Number 2, 2024, 99–117
DOI: 10.18335/region.v11i2.533

journal homepage: region.ersa.org
ISSN: 2409-5370

The Impact of Education (Level of Knowledge) on the Prevalence of Obesity in Different Urban Environments

Yuval Arbel[1], Yifat Arbel[2], Amichai Kerner[2], Miryam Kerner[3][4]

[1] Western Galilee College, Acre, Israel
[2] Bar Ilan University, Ramat Gan, Israel
[3] Technion – Israel Institute of Technology, Haifa, Israel
[4] Emek Medical Center, Afula, Israel

Received: 14 March 2024/Accepted: 18 October 2024

Abstract. We investigate the impact of high-rise buildings on the prevalence of obesity in the US during 2011-2020 stratified by educated vs. non-educated populations. We use a quadratic specification that accounts for non-monotonic variation. Findings demonstrate that concentration of above 146 skyscrapers in a state is detrimental with regards to the projected prevalence of obesity. The main public policy repercussions of our study are: 1) the promotion of education for medical literacy due to the fact that for each number of skyscrapers the prevalence of obesity is lower among educated populations. 2) widening pavements, pathways and open spaces following urban development.

Key words: High-Rise Buildings, Obesity; Health Literacy

1 Background

Previous studies found fundamental differences in health behavioral patterns among different groups stratified by education levels. According to the allocated efficiency hypothesis, higher level of human capital, proxied by education level, is positively associated with accumulated health information and knowledge. This, in turn, permits more efficient health choices and decisions (Grossman 1972, Muurinen 1982, Wagstaff 1986). Recent empirical studies demonstrate that schooling is significantly associated with health knowledge levels, which explain up to 69% of the education effect on health lifestyle in the Philippines (Hoffmann, Lutz 2019). Higher education levels among women in Australia promotes non-monetary benefits, such as, hedonic well-being, and reduced psychological distress (Tran et al. 2021). Referring to acquired information from schools, Arbel et al. (2022) discuss children-parents knowledge spillover in the context of the COVID19 pandemic. This context may be extended to include physical activity and healthy nutrition so as to reduce obesity.

Another strand of the literature relates to the urban environment advantages and disadvantages in terms of physical activity, nutrition and obesity. On the one hand, such environments may promote walkability by including mixed uses of land, stairways, gyms, green spaces and parks, bicycle and running tracks. Several studies stress the claim that some built environment characteristics encourage people to increase their walking distances (Handy et al. 2006, Chaix et al. 2014, Sun et al. 2014). In this context and based on longitudinal study of Canadians, Wasfi et al. (2016) concluded that exposure to walkable neighborhoods in urban areas increases utilitarian walking. On the other hand, crowded cities, characterized by high population densities, may discourage physical activity and provide more opportunities for increased consumption of ill-nutritional food during the nights and concurrent sleep deprivation (Fatima et al. 2015, Winston 2015, Mayne et al. 2021). Given studies that found no correlation between specific built environments and walking, Feuillet et al. (2016) suggested that more subtle analysis is required. Chen, Zhou (2016) found that the densities of 4-way intersections and more than 5-way intersections and land use mixture are positively correlated with the pedestrian crash frequency and risk in Seattle, Washington. Recently, Nigg et al. (2021) demonstrated that for each additional ten citizens per square kilometer in Germany, less positive physical activity changes were observed during the first Covid-19 lockdown in April 2020 among children. Finally, referring to the older population in Japan, high walkability neighborhood (i.e., high population density, proximity to railway stations) adversely affected their step counts, whereas proximity to large parks had a positive effect during the COVID19 state-of-emergency period (Hino, Asami 2021).

The objectives of the current study are twofold. The first objective is to link between these two strands of the literature. This is done by investigation of the relationship between the prevalence of obesity (the endogenous variable); the time variable (the first exogenous variable) and the urban environment proxied by two variables: the number of skyscrapers (the second exogenous variable); population density (the third exogeneous variable). The former environmental variable is interacted with the education level (a dummy variable that equals one for educated and zero for non-educated populations; the fourth exogenous variables). Given the unclear relationship between built environments and walkability, the second objective is to propose a simple methodology to address this issue. We employ the quadratic model, which permits non-monotonic relationship between the number of skyscrapers and obesity prevalence[1]. With the exception of Sun, Yin (2018), we found no other paper that employs this methodology in this context. The underlined research hypothesis is that up to (above) a certain level of population density, the advantages (disadvantages) of agglomeration (congestion) effect dominates.

The validity of the number of skyscrapers as a measure of dense urban environments derives from the economic theory as well as from a long series of empirical studies, which show an association between high–rise construction and dense urban areas (for an extensive review on urban spatial structure see, for example, Anas et al. 1998). According to McDonald, McMillen (2011): “Residential buildings are typically built at a very high density – tall buildings with many units built atop small land parcels” (page 121).

Practitioners and urban planners utilize similar metrics to gauge the density of urban environments and the intensity of land use (McDonald, McMillen 2011, p. 128-131). One such metric is the floor-area ratio (FAR), which is defined as the ratio of a building’s total floor space to the area of the land it occupies. In Chicago, the R1 to R8 zoning categories demonstrate a range of FAR values: for example, R1 allows for a FAR of 0.5 for detached units on lots of 6,250 square feet, meaning the maximum floor area of the unit could be 3,125 square feet. In contrast, R8 zoning permits a FAR of up to 10 for apartment buildings on lots as small as 115 square feet, allowing a 20-story building to occupy 50% of the lot’s area.

Theoretically, there is a one-to-one match between population density and high-rise buildings. The implication of high (low) population density is residence in high rise buildings at the city centers where land is expensive and scarce (land detached single-family housing units in the suburbs, where land is cheap). This point is formally demonstrated, for instance, in Mills, Hamilton (1989, p. 425-434).

The validity of the relationships between dense urban environments and walkability comes from a long list of empirical studies (e.g., TRB 2005, Frank et al. 2010, Ewing et al. 2014, Mulalic, Rouwendal 2020). In TRB (2005), the US Committee on Physical

---

[1] When the relationship between two variables is quadratic and we run: 1) a simple linear model, and 2) a quadratic model, for the (former) latter model ($R^2 = 0$ indicating no correlation) $R^2 = 1$ implying perfect correlation. (e.g., Kmenta 1997, page 241 panel (b)). As a simple exercise, we artificially constructed a sample based on the following equation: $Y = \widehat{Y} = 6 + 5X^2$ for $X = -18, -17, -16, \ldots, 16, 17, 18$. While a simple linear regression between Y and X gives zero correlation, a quadratic regression between Y, X and X^2 yields a perfect fit (see Appendix C).

Activity, Transportation, and Land Use, the Transportation Research Board, and the Institute of Medicine identified the role of urban environments in physical activity levels as a relative new field of study (page 5). Physical inactivity has been identified as one of the leading risk factors for noncommunicable diseases, such as, obesity, mortality.

In sum, this study addresses the following research questions:

1a) What are the relationships between sparse/dense urban environments (proxied by the number of skyscrapers) and the prevalence of obesity?

1b) Are these relationships linear or quadratic?

2) Given that the urban environment is controlled, are the differences in obesity prevalences between educated and on-educated populations still preserved?

Based on a sample of 47 US States during 2011-2020, the outcomes show that, as anticipated, for each number of skyscrapers, compared to non-educated population, projected obesity is lower among educated population. For a model that includes standardized normal distribution transformations of each variable, education has the highest weight in explaining obesity prevalence. For both groups of educated and uneducated populations the model predicts a U-shaped curve. Yet, the incremental effect of each additional skyscraper is steeper among the educated group.

An important contribution of this study is the use of a quadratic model. Compared to the linear model, the quadratic model may describe more complex relationship. The model allows the fall and rise of the dependent variable with the number of skyscrapers thus demonstrating a crowding out effect in terms of obesity prevalence. Differently put, obesity prevalence reaches its basin when the urban environment is relatively sparse. This is represented by an amount of approximately 146 skyscrapers. This crowding out effect cannot be investigated by employing the linear model. Failure to test the possibility of quadratic relationships might cause a type two error namely non-rejection of the null hypothesis where it should be rejected. This possibility is demonstrated in Appendix C. In this context, Wilkins et al. (2019) argue that null association between obesity and food environment dominated across all measurement methods comprising 76.0% of 1937 associations in total.

The remainder of this manuscript is organized as follows. Section 2 gives a literature review on walkability, obesity and population density. Section 3 provides the description of data sources, the descriptive statistics of variables that are latter incorporated in the empirical model and describes the empirical model. Section 4 gives the results obtained from the empirical model. Section 5 provides robustness tests. Finally, Section 6 concludes and summarizes.

2 Literature Review (Walkability, Obesity and Population Density)

Walkability is a vital link between the number of skyscrapers and obesity. A few studies have tried to define objective and subjective measures for 'walkability' in a built environment. Mayne et al. (2013), for instance, stated that 'walkability' describes the ability of a built environment to support walking for multiple purposes. Increasing local opportunities for walking and sports activities through strategic development and efficient land use is a cornerstone of the city's correct policy. Walking for utilitarian purposes is related to the features of the built environment near certain destinations, mixed land use, connectivity between the streets and population density. The authors developed an objective measure of the ability to predict utilitarian walking in the city of Sydney.

In contrast, Rodrigue et al. (2022) developed a subjective index to measure the perceived friendliness of walking in a certain area or, in other words, what is the subjective 'walking experience' in that area. For example, the ability to predict walking for the purpose of a trip versus walking for the purpose of leisure or work at the street or neighborhood level.

In the context of walkability, Baobeid et al. (2021) noted that the need for rigorous assessment tools for policy evaluation and urban planning is important. Yet, there is no unified universal standard walking theory.

Other studies have indicated that the relationship between the built environment and excess weight lies, among other things, in the ability to walk and the physical activity that results from the infrastructures that exist in that environment. According to Mariela, one of the factors that lead to the desire to walk, and exercise is socioeconomic status, which prevents obesity and weight gain. Wei et al. (2016) also pointed out that one of the factors influencing the "willingness to go" is socioeconomic status, while the use of land has less influence on people's behavior.

Baobeid et al. (2021) recommended the creation of incentive structure associated with developing walking abilities for urban residents. This includes discouragement of private vehicles owning and transportation consumption, thus creating long-term health benefits from walking and physical activity. In addition, they examine objective and subjective indicators of the built environment that make walking possible and desirable. For example, connectivity, accessibility and proximity to destination points, the presence of greenery and parks, commercial retail and proximity to transit hubs and stations.

Lopez, Hynes (2006) state that factors such as low density, poor street connectivity and the absence of sidewalks will lead to a decrease in physical activity in the suburbs and an increased risk of being overweight, while in dense neighborhoods high density, excellent connectivity between streets and sidewalks will lead to obesity. In their opinion, the reasons for the consulting paradox lie in the complex interaction between land use, infrastructure and social factors that affect the city's population.

In addition to the number of skyscrapers, another proxy for the urban environment is population density. This variable is closely associated with but not identical to the number of skyscrapers (e.g. Mills, Hamilton 1989, McDonald, McMillen 2011, Pomponi et al. 2021). Net population density is defined as the ratio between population and land allocated for residence. This can be discussed from two perspectives:

Theoretically, there is a one-to-one match between population density and high-rise buildings. The implication of high (low) population density is residence in high rise buildings at the city centers where land is expensive and scarce (land detached single-family housing units in the suburbs, where land is cheap). This point is formally demonstrated, for instance, in Mills, Hamilton (1989, p. 425-434). Accordingly, the derivation procedure yields the following solutions: $R(u) = \bar{R}e^{tE(\bar{u}-u)}$ and $\frac{N(u)}{L(u)} = ER(u)$ where $R(u)$ is the land rent as a function of the distance from the city center (u), $\bar{R}$ is the lowest land rent at the city suburbs at the distance $\bar{u}$ from the city center, t is the commuting cost per two units of distance (either mile or km) and E is a parameter. Most importantly, $N(u)$ is population, $L(u)$ is land allocated for residence, where both are functions of the distance from the city center (u). Consequently, $\frac{N(u)}{L(u)}$ is population density, which is proportionate by a factor of E to the land rent. Differently put, as land price increases at the city center, construction will become higher and population density will rise.

Empirically, these two variables are closely associated, but not identical. McDonald, McMillen (2011, p. 120-122) distinguish between gross and net population density. As the authors demonstrate, gross population density is equal to $\frac{N(u)}{L(u)}$ where all land uses are included. Consequently, the tendency of population density is to rise until a certain distance from the city center, and then fall. Pomponi et al. (2021) argue that increasing population densities without construction of taller buildings will end up in reduction of green-house-gas (GHG) emission. A possible method to increase the population densities without heightening the structures is simply to populate larger families in each apartment.

Pomponi et al. (2021) stress further the difference between population density and tall buildings. According to the authors, part of the difference is the regulatory requirements to preserve public areas (staircases, elevators), preserve reasonable standards of daylight within the high-rise buildings, space between adjacent structures and green spaces and parks. Consequently, there is no maximum utilization of the area and therefore maximum urban density is not generated.

Table 1: Descriptive Statistics

Variable		Obs.	Mean	Std.Dev.	Min	Max
Obesity prevalence	Prevalence of population in the US state that suffers from obesity ($BMI \geq 30$ where) measured in percentage points	928	22.54	6.55	7.6	37.6
$(Year - 2011)$	The year in which the prevalence of obesity was measured in the state	928	4.53	2.86	0	9
Educated	1=Educated (College Education); 0=Non-Educated (Less than high-school education)	928	0.5	0.50	0	1
Skyscrapers	Number of skyscrapers in the state	928	15.86	42.87	0	267
Pop_Density	(in square km.)	928	192.68	629.95	0.5	4,361

Notes: The statistical test of Non-Educated vs. Educated obesity prevalence difference of means with unequal variances is given in Table 2.

3 Methods

3.1 Description of Data

Data for this study were obtained from several sources. Obesity prevalence data were provided by the CDC website. Prevalence of obesity is calculated as the ratio of obese persons (i.e., with BMI ≥ 30) and the group population. Data for the number of skyscrapers in each state were obtained from MapPorn. The population densities of the US states were obtained from the United States Census Bureau (2010, 2020).

Appendix A describes the data structure and stratification by states. Referring to obesity prevalence, the data includes up to ten years per state (Year$= 2011, 2012, \ldots, 2020$) for the most educated vs. most uneducated group. Consequently, the full sample per-state includes 20 observations (10 years×2 obs. for educated and non-educated in each state). There are 44 US states with the full sample of 20 observations ($44 \times 20 = 880$), and three US states with missing observations: (Idaho – 14 obs., North Dakota – 18 obs. and Utah – 16 obs.) The total sum is: $880 + 14 + 18 + 16 = 928$.

In addition, originally the CDC dataset includes four groups: 1) College graduates; 2) High School graduates; 3) Less than high school; and 4) Some college or technical schools. To simplify and avoid confusion, we chose to analyze only the two most extreme groups (the most uneducated – group 3; and the most educated – group 1). Appendix B demonstrate graphically the outcomes obtained for the four groups when the mid-range groups are not omitted.

3.2 Descriptive Statistics

Table 1 reports the descriptive statistics of the dependent variable and the predictors that are later incorporated in the empirical model. The variable (year-2011) is defined as the number of years (minus one) in which the prevalence of obesity was measured in the state starting from 2011. Note that following this transformation, the constant term in the empirical model displayed in the subsequent section becomes the baseline projected prevalence of obesity at states without skyscrapers in 2011 (Ramanathan 2002, Hoaglin 2016a,b)[2]. The sample mean of (year-2011) is 4.53, the standard deviation is 2.86, the minimum is 0 and the maximum is 9. The implication is that referring to the prevalence of obesity the sample covers 10 years.

Table 2 gives the outcomes of the difference of means statistical test. This is a crude measure for educated vs. non-educated differences of obesity prevalence without adjust-

---

[2] In this context, Hoaglin (2016a) states that: “Many presentations tend to use the same letters in models that involve different sets of other predictors, which makes it easy to overlook the role of the other predictors in the definition of the coefficient of each predictor. For example, if $2x + 5t$ is a good fit to the data on y, then $-3x + 5(t + x)$ is also a good fit to those data (it gives exactly the same predicted values).” (page 7)

Table 2: Obesity Prevalence: Test of Non-Educated vs. Educated

Group	Obs.	Mean	Std.Err.	Std.Dev.	95% Conf.Intervall
Educated	464	24.1321	0.213697	4.60317	23.71218 to 24.55205
Non-Educated	464	33.9748	0.192328	4.14286	33.59684 to 34.35273
combined	928	29.0535	0.21626	6.58796	28.62903 to 29.47787
diff		-9.8427	0.2875		-10.40691 to -9.278437

Notes: diff = mean(Educated) − mean(Non-Educated); Ho: diff = 0; Satterthwaite's degrees of freedom: 915.908. The calculated t-value with 915.908 degrees of freedom is -34.2354, compared to the 1% critical value of -2.5812.

Table 3: Pearson Correlation Matrix

	Obesity Prevalence	$(Year - 2011)$	Educated	Skyscrapers	Pop_Density
Obesity prevalence	1.0000				
$(Year - 2011)$	0.2365*** (<0.01)	1.0000			
Educated	-0.7474*** (<0.01)	0.0000 (1.000)	1.0000		
Skyscrapers	-0.1220*** (0.0002)	-0.0042 (0.8987)	-0.0000 (1.0000)	1.0000	
Pop_Density	-0.0886*** (0.0069)	0.0036 (0.9122)	-0.0000 (1.0000)	0.2323*** (<0.01)	1.0000

Notes: P-values for the rejection of zero correlation are given in parentheses. Number of observations: 928. ***: $p < 0.01$

ments to other predictors[3]. The test clearly demonstrates an average lower prevalence of obesity among educated population in the state by 9.8427% (95% confidence interval of [-10.40691, -9.278437]). The null hypothesis of zero difference of means is clearly rejected. The calculated t-value with 915.908 degrees of freedom is -34.2354, compared to the 1% critical value of -2.5812.

3.3 The Pearson Correlation Matrix

Prior to the estimation of the empirical model, Table 3 reports the Pearson correlation matrix. This is considered important information regarding the relationships between variables. High Pearson correlations between independent variables might distort the sign and significance of the coefficients.

It may be readily verified that the highest Pearson correlation is between the variables Education and Obesity Prevalence (-0.7474) and the null hypothesis of zero correlation is clearly rejected ($p < 0.01$). The implication is the decrease in obesity prevalence with higher education level. In addition, there is a positive correlation between obesity prevalence and the Year variable (+0.2365) and the null hypothesis of zero correlation is clearly rejected ($p < 0.01$). The indication here is the increase of obesity prevalence with the time variable. Finally, the Pearson correlations between obesity prevalence and the number of skyscrapers and density has the "correct" minus sign. This indicates a drop in obesity prevalence with densely populated environment. For both variables, the null hypothesis of zero correlation is clearly rejected ($p = 0.0002$ and $p = 0.0069$).

The Pearson correlation between the variables skyscrapers and population density measured in square kilometers is +0.2323. Despite the fact that: a) the null hypothesis of zero correlation is rejected, and b) the sign of the correlation is in the right direction; this measure is not considered to be a very high Pearson correlation. As discussed in the literature review section, these variables are similar but not identical.

---

[3] This is the correct terminology that should replace the terminology of "control of other explanatory variables". The correct interpretation actively keeps in view the adjustments for the contributions of the other predictors (Hoaglin 2016a,b).

3.4 The Empirical Model

Consider the following interaction model consisting of the structural equation:

$$\begin{aligned} \mathit{Obesity\_Prevalence} \;=\;& a_1^{'}(\mathit{Year}-2011)^2 + a_2^{'}(\mathit{Year}-2011) + a_1 \mathit{Skyscrapers}^2 + \\ & a_2 \mathit{Educated} \times \mathit{Skyscrapers}^2 + b_1 \mathit{Skyscrapers} + \\ & b_2 \mathit{Educated} \times \mathit{Skyscrapers} + c_1 \mathit{Educated} + d_1 \mathit{Pop\_Density} + \\ & c_2 + \mu_1 \end{aligned} \tag{1}$$

where $\mathit{Obesity\_Prevalence}$ is the dependent variable, $(\mathit{Year}-2011)^2$, $(\mathit{Year}-2011)$, $\mathit{Skyscrapers}^2$, $\mathit{Skyscraper}$, $\mathit{Educated}$ and $\mathit{Pop\_Density}$ are the independent variables, $a_1^{'}$, a_1, a_2, b_1, b_2, c_1, c_2, d_1 are the parameters, and μ_1 is the classical random disturbance term.

According to Chiang, Wainwright (2005, p. 229-231), the general form of the quadratic function is: $y = ax^2 + bx + c$ $(a \neq 0)$ with a second derivative equals to $2a$. Given that this derivative will always have the algebraic sign of the coefficient a, a U-shaped curve with a global minimum at $(\frac{-b}{2a}, \frac{-b^2+4ac}{4a})$ is obtained if $a > 0$, and an inverted a U-shaped curve with a global maximum at $(\frac{-b}{2a}, \frac{-b^2+4ac}{4a})$ is obtained if $a < 0$.

Compared to the linear model, the quadratic model may describe more complex relationships. The model allows the fall and rise of the dependent variable with the number of skyscrapers thus demonstrating a crowding out effect in terms of obesity prevalence. Differently put, and as demonstrated below, obesity prevalence reaches its basin when the urban environment is relatively sparse. This is represented by an amount of approximately 146 skyscrapers. This crowding out effect cannot be investigated by employing the linear model.

In their review, Wilkins et al. (2019) argue that null association between obesity and food environment dominated across all measurement methods comprising 76.0% of 1937 associations in total.

A possible interpretation for the domination of this null association is the linear restriction imposed on the empirical model. In his econometric textbook, Kmenta (1997) demonstrates a quadratic relationship between variables. However, the imposition of linear restriction yields a poor fit, namely, no association between Y and X (page 241).

To demonstrate this point, we now performed the following exercise. We constructed a tailored made quadratic function $\widehat{Y} = 6 + 5X^2$ and ran a linear and a quadratic regression. The outcomes are given in Appendix C. As can be seen, while the quadratic relationship exhibits a perfect fit, the linear model exhibits no fit at all.

To test whether the specification of the model is appropriate, namely, whether the model excludes important omitted variables, we employ the Ramsey's RESET procedure, where the RESET stands for Regression Specification Error Test (e.g., Ramanathan 2002, p. 270). This procedure is based on two steps. The first step of the procedure is the construction of vector of predictions $(\widehat{Y})$ from the model given in equation (1). The second step is the incorporation of $\widehat{Y}^2$, $\widehat{Y}^3$ and $\widehat{Y}^4$ in equation (1) as additional independent variables and testing the joint null hypothesis that their coefficients equal zero. If the null hypothesis is not rejected, one could argue that the model specification is appropriate.

Another concern the empirical model addresses is the possibility of spurious or nonsense correlation in time series analysis. This is done by incorporation of the time variable $(\mathit{Year}-2011)$. According to Johnston, DiNardo (1997, p. 9), series, responding to unrelated mechanisms, such as, death rates in England and Wales and the proportion of all marriages solemnized in the Church of England from 1866 to 1911 (Yule 1926), may display contemporaneous upward or downward movement[4]. This problem may be addressed by fitting trends to such series.

---

[4] Referring to Yule (1926), Johnston, DiNardo (1997) state that: “However, no British politicians proposed closing down the church of England to confer immortality on the electorate.” (page 10)

Table 4: Regression Analysis

Individual Effect Time Dummies (Full/Stepwise)	(1) No (Full)	(2) No (Stepwise)	(3) Yes (Stepwise)
Variables	Obesity Prevalence	Obesity Prevalence	Obesity Prevalence
$(Year - 2011)^2$	0.00768	-	-
	(0.655)	-	-
$(Year - 2011)$	0.472***	0.542***	-
	(0.00261)	(<0.01)	-
Skyscrapers^2	0.000139**	0.000145**	0.000146**
	(0.0172)	(0.0136)	(0.0139)
Educated× Skyscrapers^2	0.000323***	0.000323***	0.000323***
	(1.55×10^{-5})	(1.49×10^{-5})	(1.58×10^{-5})
Skyscrapers	-0.0407***	-0.0426***	-0.0426***
	(0.00532)	(0.00383)	(0.00406)
Educated×Skyscrapers	-0.0933***	-0.0933***	-0.0933***
	(1.13×10^{-6})	(1.06×10^{-6})	(1.20×10^{-6})
Educated	-9.036***	-9.036***	-9.036***
	(<0.01)	(<0.01)	(<0.01)
Pop_Density	-0.000115	-	-
	(0.703)	-	-
Constant	31.99***	31.89***	31.90***
	(<0.01)	(<0.01)	(<0.01)
Observations	928	928	928
R-squared	0.667	0.667	0.667
Minimum Obesity College Education			
Skyscrapers	145 [141, 149]	145 [142, 149]	145 [141, 149]
Projected Prevalence of Obesity	17 [16, 18]	15 [14, 17]	15 [14, 17]
Minimum Obesity less than High School Education			
Skyscrapers	146 [122, 169]	146 [109, 183]	146 [110, 183]
Projected Prevalence of Obesity	30 [28, 31]	31 [29, 33]	31 [29, 33]

Notes: The Educated variable receives 1 for college education and zero for less than high school education in the state. The Ramsey's RESET (Regression Specification Error Test – see Ramanathan 2002, p. 270) procedure is based on two steps. The first step of the procedure is the construction of vector of predictions ($\widehat{Y}$) from the model. The second step is the incorporation of $\widehat{Y}^2$, $\widehat{Y}^3$ and $\widehat{Y}^4$ as additional independent variables and testing the joint null hypothesis that their coefficients equal zero. If the null hypothesis is not rejected, one could argue that the model specification is appropriate. According to this procedure, the null hypothesis is not rejected ($F(3, 918) = 2.59$; $p = 0.0514$). Robust p-values are given in parentheses. 5% confidence intervals are given in square brackets. *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

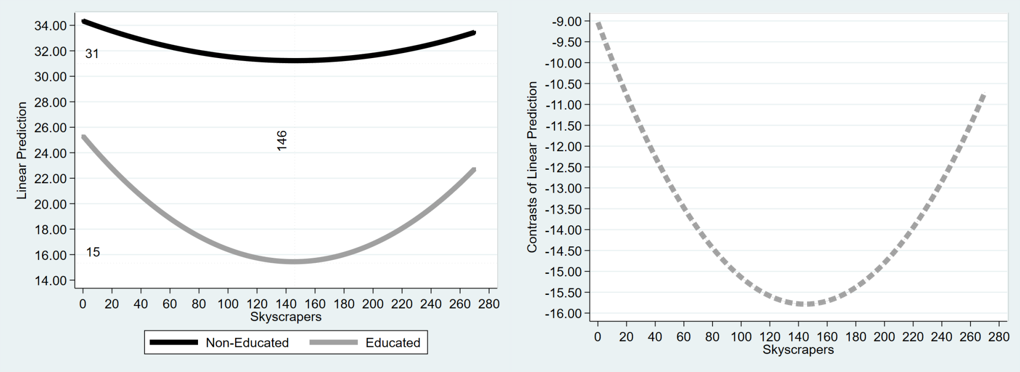
4 Results

Table 4 reports the regression outcomes. The table is divided to three columns. Column (1) gives the outcomes of the full model described by equation (1). Column (2) provides the outcomes obtained from the stepwise regression. This procedure gradually omits variables with the most insignificant coefficients (one variable for each step), until we are left with significant coefficients at a pre-determined level. Finally, column (3) gives the results obtained from the stepwise model, where the time variable is replaced with dummies for each year.

As can be seen from Table 4, 66.7% of the variance of the dependent variable (at the US States) is explained by the independent variables of education levels, the time variable and the number of skyscrapers. Application of the RESET procedure supports the conclusion that the incorporation of $\widehat{Y}^2$, $\widehat{Y}^3$, and $\widehat{Y}^4$, is unnecessary at the 5% and 1% levels ($F(3, 918) = 2.59$; $p = 0.0514$). Consequently, according to the statistical test, the specification of the empirical model described by equation (1) is appropriate.

The outcomes of column (1) in Table 4 demonstrate that the coefficient of the variable $(Year - 2011)^2$ is statistically insignificant ($p = 0.655$). In addition, the coefficient of the "control" variable population density, where land area is measured in square km., has the "correct" minus sign, but the variable is statistically insignificant ($p = 0.703$).

According to column (2) in Table 4, the 2011 baseline projected obesity prevalence



(a) Obesity Prevalence by Education

(b) Difference

Notes: Figure 1b refers to the difference between projected obesity prevalence of educated minus non-educated populations at the same US states obtained from column (2) in Table 4. All the projection differences are statistically different from zero at the 1% level and range between −16 and −9.

Figure 1: Impact of Skyscrapers on the Prevalence of Obesity: Educated vs. Non-Educated

for non-educated population in regions without skyscrapers is 31.89 percent ($p < 0.01$). Projected obesity prevalence *rises* by 0.542 percent per annum ($p < 0.01$) and *drops* by 9.036 percent ($p < 0.01$) with a shift from the non-educated to the educated population after the adjustment of contributions of other predictors[5]. When we add the variable population density measured in square kilometers to the regression analysis as a “control” variable, the parameter has the “correct” minus sign, but the variable is statistically insignificant ($p = 0.703$).

Figure 1 reports the impact of skyscrapers, as a proxy of denser urban environment, on the prevalence of obesity among educated vs. non-educated persons. Figure 1a is based on the outcomes obtained from column (2) in Table 4. Figure 1b refers to the difference between projected obesity prevalence of non-educated minus Educated populations at the same US states obtained from Table 1. All the projection differences are statistically different from zero at the 1% level and range between −16 and −9.

In line with previous literature, Figure 1b indicates lower projected obesity prevalence among educated population, where projections are adjusted for the number of skyscrapers. The gap ranges between 9 and 16 percent in favor of the educated population. A possible explanation to these outcomes is that as part of more efficient health literacy, compared to non-educated, educated people make more extensive use in urban infrastructure, such as, stairways, gyms, green spaces and parks, bicycle and running tracks.

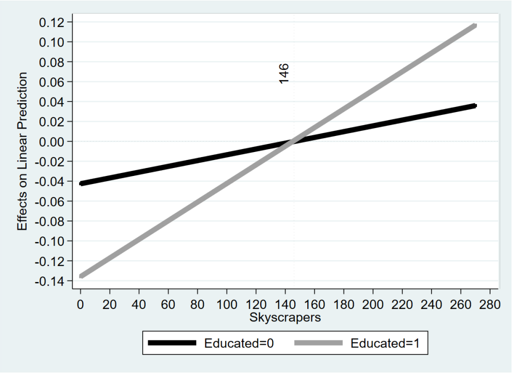
Figure 1a demonstrates a U-shaped curve for both groups. The implication is that within the range of 0 to 146 (146 to 267), skyscrapers projected obesity prevalence *drops* (*rises*) with the number of skyscrapers. Very few studies tested the possibility of non-monotonic change. Based on population density analysis in China, Sun, Yin (2018) demonstrated similar outcomes to our Study. The authors demonstrate that within the range of population density of 50 to 1255 (1,255 to 202,800) persons per square kilometer, projected obesity decreases (increases).

Figure 1 may be interpreted in the following manner:

1. The descending part of the graph: educated persons exploit the urban infrastructure more efficiently. Urban infrastructure includes bicycle and walking lanes; benches and gym installations in parks; the possibility to engage in utilitarian walk on the way to work and on the way from work. This is possible as long as congestion effect is weak and the extent of crime incidences are low.

---

[5] In this context see footnote 3. Hoaglin (2016b) states that: “If one wants to estimate the effect of making a unit change in x while holding w constant, one must have data in which values of x differ by one unit and w remains constant, so that one can actually observe that effect. Designed experiments in applied science often do this.” (page 32).



Notes: The graphs are based on the outcomes reported on column (2) in Table 4. The Educated variable receives 1 for college education and zero for less than high school education in the state.

Figure 2: Incremental impact of Skyscrapers on the Prevalence of Obesity among Educated vs. Non-Educated populations

2. The ascending part of the graph: As urban areas become more crowded, the urban infrastructure are exploited beyond their capacity (the congestion effect). This is particularly true for tourist attractions, which bring more people. The growth of cities avoids utilitarian walking in shaded parks at night due to increased crime rates. Consequently, educated persons gain weight and return to their initial weight.

Our interpretation to the educated vs. non educated difference is the following: the construction and study of the exploitation of the urban infrastructure evolves with the development of the urban areas. Walkable lanes are constructed at a later stage. Consequently, the mid-range of skyscrapers captures the full relative advantage of the educated population, where congestion effect is still weak.

Finally, based on the same regression analysis, Figure 2 displays the incremental impact of skyscrapers on the prevalence of obesity among Educated vs. Non-Educated populations. According to the figure, the incremental change of one additional skyscraper is steeper among the educated population. Up to 146 skyscrapers, the exclusion of individuals from the group of obese population with each additional skyscraper is faster among educated persons. While the first skyscraper reduces the expected proportion of obese population by 14% among educated persons, the equivalent figure is only 4% reduction among non-educated persons. But the picture reverses above 146 skyscrapers. For 260 skyscrapers, the other extreme, while the last skyscrapers increases the anticipated proportion of obese population by 12% among the educated population, the rise among non-educated becomes only 4%.

5 Robustness Tests

To further corroborate the relative contribution of education on the reduction of obesity prevalence we run a robustness test. We transformed the variables given in equation (1) to the standard normal distribution function. The analysis enables equals units of measurement of one standard deviation increase for each explanatory variable. This gives the possibility to rank the explanatory power of variables based on the absolute values of the coefficients. Results of this exercise are given in Table 5.

The outcomes show that education has the highest explanatory power (0.6862 in absolute value), followed by the number of skyscrapers (0.2771 in absolute value) and the time variable (0.2352). To offset the impact of the education *drop* on projected obesity prevalence, an *increase* of almost three standard deviations are required in the time variable.

As an additional robustness test, Table 6 includes a random effect regression for the educated vs. non-educated groups. This procedure corrects the serial correlation between the generic dummies for each state and the independent variables and thus improves the

Table 5: Regression Analysis: Beta Coefficients

Variables	(1) $\mathcal{Z}$(Obesity Prevalence)
$\mathcal{Z}(Year - 2011)$	0.2352*** (<0.01)
$\mathcal{Z}$(Educated)	-0.6862*** (<0.01)
$\mathcal{Z}$(Skyscrapers)	-0.2771*** (0.000623)
$\mathcal{Z}$(Educated×Skyscrapers)	-0.4440*** (1.29×10^{-7})
$\mathcal{Z}$(Skyscrapers×Skyscrapers)	0.2312*** (0.00428)
$\mathcal{Z}$(Educated×Skyscrapers×Skyscrapers)	0.3664*** (7.84×10^{-6})
Constant	0 (<0.01)
Observations	928
R-squared	0.667

Notes: The $\mathcal{Z}(.)$ is the standard normal distribution transformation with zero mean and standard deviation of 1. P-values are given in parentheses. *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

Table 6: Random effect Regressions

Variables	(1) Educated Obesity prevalence	(2) Non-Educated Obesity prevalence
$(Year - 2011)$	0.598*** (<0.01)	0.481*** (<0.01)
Skyscrapers×Skyscrapers	0.000467*** (0.000614)	0.000145 (0.266)
Skyscrapers	-0.135*** (0.000206)	-0.0424 (0.233)
Constant	22.59*** (<0.01)	32.16*** (<0.01)
Observations	464	464
Number of States	47	47

Notes: The random effect regression corrects for serial correlation between the 46 dummy variables included in the random disturbance term (one for each state with the exception of the base category) and the independent variables. Robust p-values are given in parentheses. 5% confidence intervals are given in square brackets. *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

efficiency of the model (e.g., Wooldridge 2009, 489-491). The outcomes remain robust with respect to previous procedures.

As can be seen from Table 6, obesity prevalence initially drops by 0.135 percent ($p = 2.0610^{-4}$) for the educated group and by only 0.0424 percent ($p = 0.233$) per additional high-rise building for the uneducated group. Moreover, for the uneducated population, the random effect regression makes the skyscrapers variables irrelevant.

Finally, one issue that warrants attention is the validity of the relationships among the variables. The validity of using the number of skyscrapers as a proxy for dense urban environments is supported by economic theory and a considerable body of empirical studies that demonstrate a correlation between high-rise construction and dense urban areas (for an extensive review on urban spatial structure, see, for example, Anas et al. 1998). According to McDonald, McMillen (2011), "Residential buildings are typically built at a very high density—tall buildings with many units built atop small land parcels" (p. 121). High-rise buildings and skyscrapers are distinctive forms of construction that have been infrequently explored in academic literature. The development of skyscrapers generates a blend of residential and commercial uses, job opportunities, and cultural amenities (Brueckner et al. 1999) and influences the behaviors of residents in these areas. This necessitates specialized planning characterized by pedestrian access between buildings and decreased reliance on automobiles.

The validity of the connection between dense urban environments and walkability is substantiated by a substantial number of empirical studies (e.g., TRB 2005, Frank et al. 2010, Ewing et al. 2014, Mulalic, Rouwendal 2020). In TRB (2005), the US Committee on Physical Activity, Transportation, and Land Use, and the Institute of Medicine recognized the influence of urban environments on physical activity levels as an emerging field (p. 5). Physical inactivity is recognized as one of the leading risk factors for non-communicable disease-related mortality (Ewing et al. 2014, WHO 2022). The probability of mortality is projected to increase by 20% to 30% due to insufficient physical activity (WHO 2022). It is expected that physical activity levels will decline over time due to the ongoing decentralization of urban areas, resulting in longer travel distances and making private vehicles the most convenient mode of transport. Indeed, in 2005, 55% of the US adult population reported not meeting the recommended standard of 30 minutes of daily brisk walking (TRB 2005, p. 2). In urban planning, compact development (i.e., dense construction) is linked to decreased automobile dependence (Ewing et al. 2014).

Other confounding factors include access to healthcare services. In this context, Hamidi et al. (2020) investigate the relationship between COVID-19 infection and dense urban environments in the United States. One the one hand, dense areas lead to more face-to-face interaction among residents. On the other hand, dense areas may have greater implementation of social distancing practices and policies and better access to health care facilities. Indeed, as the authors suggest, counties with higher densities have significantly lower virus-related mortality rates than do counties with lower densities, possibly due to superior health care systems.

Using a worldwide examination, Arbel et al. (2023b) found that a one STD increase in the population density and the number of beds is expected to decrease the STD of COVID-19 mortality rate by 0.127 and 0.0920, respectively.

In addition to the aforementioned justifications, we conducted a cross-validation procedure as illustrated in Table 7. This method stems from the foundational work of Milton Friedman (Friedman 1966, p. 9), who differentiated between "on sample," "off-sample," and "out-of-sample" groups, where the latter pertains to anticipated future events (forecasts) and the former relates to past events—whose outcomes remain uncertain due to insufficient information (predictions)[6]. The cross-validation procedure is applicable only to cross-sectional datasets because the off-sample subset is randomly selected from the entire sample pool. In time series analysis, the sequence is critical, preventing the procedure from sampling a random subset. The cross-validation process generates a prediction vector, denoted as $\widehat{P}$, where each of the five folds randomly assigns a portion of the off-sample group and performs an OLS regression on the training on-sample group. The vector $\widehat{P}$ comprises only predictions from the off-sample group. The table presents the Pearson correlations between Obesity Prevalence and $\widehat{P}$. The markers † ($p < 0.01$), indicate the rejection of the null hypothesis of zero correlation. The results of this procedure revealed a Pearson correlation coefficient ranging from 0.2926 to 0.7109. In all cases, the null hypothesis of zero correlation was rejected at the 1% level.

6 Discussion

A possible interpretation of the outcomes is the failure to utilize the urban infrastructures (e.g., bicycle and walking lanes; pedestrian pavements; benches and gym installations in parks) due to congestion problems at a densely populated urban environments (a possibility that cannot be considered in an empirical study that imposes a linear and a monotonic restriction). Up to 146 skyscrapers, the density is sufficiently low to facilitate

---

[6] According to Friedman (1966): "To avoid confusion, it should perhaps be noted explicitly that the 'predictions' by which the validity of a hypothesis is tested need not be about phenomena that have not yet occurred, that is, need not be forecasts of future events; they may be about phenomena that have occurred but observations on which have not yet been made or are not known to the person making the prediction. For example, a hypothesis may imply that such and such must have happened in 1906, given some other known circumstances. If a search of the records reveals that such and such did happen, the prediction is confirmed; if it reveals that such and such did not happen, the prediction is contradicted." (page 9)

Table 7: Cross Validate for Educated vs. Non-Educated

Rounds	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
First	$0.6222^{\dagger}$	$0.3821^{\dagger}$	$0.7109^{\dagger}$	$0.3847^{\dagger}$	$0.4650^{\dagger}$	$0.4957^{\dagger}$	$0.5938^{\dagger}$	$0.5012^{\dagger}$	$0.3563^{\dagger}$	$0.2926^{\dagger}$	
Obs.	90	90	92	94	94	94	94	94	92	94	928

Notes: The cross-validation procedure can only be performed on cross-sectional datasets. This procedure creates a vector of predictions denoted as $\widehat{P}$ where each of the five folds randomly assigns a subset of the off-sample group and runs an OLS regression on the training on-sample group. Vector $\widehat{P}$ contains only the predictions from the off-sample group. The table shows the Pearson correlations between Obesity Prevalence and $\widehat{P}$. †: $p < 0.01$, for the rejection of the null hypothesis of zero correlation.

the use of the urban infrastructures within and outside the high-rise buildings. In terms of density of people per pathways, it is more convenient to use staircases and gyms inside the high-rise buildings, where population densities (proxied by the number of skyscrapers) are low. Under such circumstances, it is still convenient to walk on pavements, use the bicycle and running tracks, swimming pools, basketball halls and engaging in recreational sport. As the number of skyscrapers (a proxy for population density) become higher, the type and objective of walking modifies. The pace of walking becomes slower due to elevated density, so that it can no longer be considered walking for the objectives of physical activity. The public transportation system also becomes crowded. The convenience in delivery services, along with the disincentive for walking, promotes the consumption of innutritious food and the reduction in physical activity. Combined with the slower pace of walking, the multitude of stimuli outside the perimeter (restaurants, food stands) and availability of fast-food increase food consumption. These reasons, in turn, promote the obesity prevalence of the population. The growth of cities avoids utilitarian walking in shaded parks at night due to increased crime rates. Consequently, educated persons gain weight and return to their initial weight with the evolution of the urban environment.

7 Summary and Conclusions

The objective of the current study is to examine the relationship between the prevalence of obesity and the urban environment proxied by the number of skyscrapers and stratified by education level (educated vs. non-educated populations). The study is based on a sample of 47 US States during 2011-2020. A unique feature of this study is the employment of a quadratic model, which permits non-monotonic change in obesity prevalence with the number of skyscrapers. With one exception (Sun, Yin 2018), we are unaware of any study in the field that employed a quadratic model.

The public policy repercussions of our study may be divided into the short and long run. The long run objectives should be the increase and improvement of the education attainment and level. This type of solution was proposed or indicated *inter alia* by Ross, Wu (1995), Devaux et al. (2011), and Tran et al. (2021). Using data from France, Devaux et al. (2011) find support to a causal relationship from education to obesity, and not vice versa. Their argument is based on the minimal effect on the strength of the association when reduced educational opportunities for those who are obese in young age are considered (page 140).

Given the difficulty to raise the education levels, particularly among the adult population, the short run objectives should be acquired health literacy via the media and schools. Another possibility is the use of pharmacological means and bariatric surgeries in extreme cases combined with proper nutrition and physical activity (Cannon, Kumar 2009).

In sum, the role of the government should be manifested in the following fields:

1. Improving urban planning in the areas of establishing schools and academic educational institutions, cultural institutions, educational youth movements. All these issues will fall into the realm of land allocation for urban educational development, including health institutions, sports infrastructure of all kinds, and programs for teaching medical literacy. The planning should include provisions on educational and health infrastructure.

2. Children-parents knowledge spillover via special nutritional and physical activity training programs in schools.

3. Popular training courses, particularly for uneducated populations.

4. Generating a new index, which accounts for optimal exploitation of urban infrastructure for health objectives with respect to congestion.

5. Constructing an incentive structure to encourage walkability instead of transportation.

Like any other research this study has its strength and limitations. A possible limitation is the grid at the US statewide level. Yet this grid has its advantage and disadvantage. Another limitation is the potential problem of omitted variable. Yet, the RESET test rejects the possibility of an additional independent variables.

Strengths

A comparison across US states permits a global perspective. This comparison controls for weather, economic and cultural differences and intensity of land use, particularly given the generic dummies for each state used in a random effect regression framework (see Table 6).

A cross comparison at a country level (a higher grid than our own research) is a very well-known and a conventional methodology. Three examples are Barro et al. (2020) and Arbel et al. (2023a,b). Barro et al. (2020) compared the mortality rate from the Spanish flue pandemic during the Great Influenza Pandemic, 1918-1920 and War Death Rates for Military in Combat during World War I, 1914-1918 at a country level. Arbel et al. (2023a,b) investigated a data source at a country level and demonstrated that from the examined independent variables the most influential on COVID19 morbidity and mortality is the age variable.

Moreover, this is the first article that investigates the relationship between education, the urban environment and obesity – using a non-linear model, which allows non-monotonic increase or decrease, and relaxes the linear restrictions. Results show that a quadratic model better fits the data particularly as far as the educated population is concerned. For the uneducated population. the random effect regression, given in Table 6, makes the skyscrapers variables irrelevant. The outcomes also demonstrate that: 1) the few independent variables explain 66.7% of the variance of the dependent variable and 2) the RESET (Regression Specification Error Test – see Ramanathan 2002, p. 270) procedure supports the absence of omitted variables.

Limitations

Given the data structure at a macro level, incorporating additional "control" variables is problematic – because there is a one-to-one match between the independent and dependent variable. Differently put, if we compare education with gender, for instance, the numerical prevalence of obesity (the dependent variable) for the same state will be different. Note, however, that the RESET procedure clearly supports the null hypothesis that all the empirical models employed in this article are correctly specified.

The grid of the data might be considered too high. Average data might be exposed to aggregation bias and information loss referring to the variance of variables at a lower grid of the data. Future research should employ datasets at a lower grid, at least at a city level.

Finally, future research regarding non-linearities in the data could be benefitted from the employment of non-parametric or semi-parametric methodologies instead of the parametric ones. These methodologies are beyond the scope of the current research.

Abbreviations

BMI = Body Mass Index
CDC = Center for Disease Control and Prevention
RESET = Regression Specification Error Test

References

Anas A, Arnott R, Small K (1998) Urban spatial structure. *Journal of Economic Literature* 36: 1426–1464

Arbel Y, Arbel Y, Kerner A, Kerner M (2022) To lockdown or not against COVID19: Can the power of education provide the answer? *Cities* 121. CrossRef

Arbel Y, Arbel Y, Kerner A, Kerner M (2023a) What is the optimal country for minimum COVID 19 morbidity and mortality rates. *Environmental Science and Pollution Research* 30: 59212–59232. CrossRef

Arbel Y, Fialkoff C, Kerner A, Kerner M (2023b) COVID-19 morbidity and mortality factors: An international comparison. *REGION* 10: 31–44. CrossRef

Baobeid A, Koç M, Al-Ghamdi S (2021) Walkability and its relationships with health, sustainability, and livability: Elements of physical environment and evaluation frameworks. *Frontiers in Built Environment* 7. CrossRef

Barro R, Ursúa J, Weng J (2020) The coronavirus and the great influenza pandemic: Lessons from the “spanish flu” for the coronavirus’s potential effects on mortality and economic activity. *NBER Working Papers* w26866. CrossRef

Brueckner J, Thisse JF, Zenou Y (1999) Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review* 43: 91–107. CrossRef

Cannon C, Kumar A (2009) Treatment of overweight and obesity: Lifestyle, pharmacologic, and surgical options. *Clinical Cornerstone* 9: 55–71. CrossRef

Chaix B, Simon C, Charreire H, Thomas F, Kestens Y, Karusisi N, Vallée J, Oppert JM, Weber C, Pannier B (2014) The environmental correlates of overall and neighborhood based recreational walking (a cross-sectional analysis of the RECORD study. *International Journal of Behavioral Nutrition & Physical Activity* 11: 1–28. CrossRef

Chen P, Zhou J (2016) Effects of the built environment on automobile-involved pedestrian crash frequency and risk. *Journal of Transport & Health* 3: 448–456. CrossRef

Chiang A, Wainwright K (2005) *Fundamental Methods of Mathematical Economics* (Fourth Edition ed.). McGraw Hills

Devaux M, Sassi F, Church J, Cecchini M, Borgonovi F (2011) Exploring the relationship between education and obesity. *OECD Journal: Economic Studies*: 121–159. CrossRef

Ewing R, Meakins G, Hamidi S, Nelson A (2014) Relationship between urban sprawl and physical activity, obesity, and morbidity – update and refinement. *Health & Place* 26: 118–126. CrossRef

Fatima Y, Doi S, Mamun A (2015) Longitudinal impact of sleep on overweight and obesity in children and adolescents: a systematic review and bias-adjusted meta-analysis. *Obesity Reviews* 16: 137–149. CrossRef

Feuillet T, Salze P, Charreire H, Menai M, Enaux C, Perchoux C, Hess F, Kesse-Guyot E, Hercberg S, Simon C, Weber C, Oppert JM (2016) Built environment in local relation with walking: Why here and not there? *Journal of Transport & Health* 3: 500–512. CrossRef

Frank L, Sallis J, Saelens B, Leary L, Cain K, Conway T, Hess P (2010) The development of a walkability index: Application to the neighborhood quality of life study. *British Journal of Sports Medicine* 44: 924–933. CrossRef

Friedman M (1966) The methodology of positive economics. In: Friedman M (ed), *Essays In Positive Economics*. University of Chicago Press, Chicago, 3–16, 30–43

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

Hamidi S, Sabouri S, Ewing R (2020) Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association* 86: 495–509. CrossRef

Handy S, Cao X, Mokhtarian P (2006) Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California. *Journal of the American Planning Association* 72: 55–74,. CrossRef

Hino K, Asami Y (2021) Change in walking steps and association with built environments during the COVID-19 state of emergency: A longitudinal comparison with the first half of 2019 in Yokohama, Japan. *Health and Place* 69. CrossRef

Hoaglin D (2016a) Regressions are commonly misinterpreted. *Stata Journal* 16: 5 22. CrossRef

Hoaglin D (2016b) Regressions are commonly misinterpreted: A rejoinder. *Stata Journal* 16: 30–36. CrossRef

Hoffmann R, Lutz S (2019) The health knowledge mechanism: Evidence on the link between education and health lifestyle in the Philippines. *European Journal of Health Economics* 20: 27–43. CrossRef

Johnston J, DiNardo J (1997) *Econometric Methods* (Fourth ed.). McGraw-Hill

Kmenta J (1997) *Elements of Econometrics* (Second Edition ed.). The University of Michigan Press, Ann Arbor. CrossRef

Lopez R, Hynes H (2006) Obesity, physical activity, and the urban environment: Public health research needs. *Environmental Health* 5: 25–10. CrossRef

Mayne D, Morgan G, Willmore A, Rose N, Jalaludin B, Bambrick H, Bauman A (2013) An objective index of walkability for research and planning in the Sydney Metropolitan Region of New South Wales, Australia: An ecological study. *International Journal of Health Geographics* 12: 1–22. CrossRef

Mayne S, Morales K, Williamson A, Grant S, Fiks A, Basner M, Dinges D, Zemel B, Mitchell J (2021) Associations of the residential built environment with adolescent sleep outcomes. *Sleep* 44. CrossRef

McDonald J, McMillen D (2011) *Urban Economics and Real Estate: Theory and Policy* (Second ed.). John Wiley & Sons Inc

Mills E, Hamilton B (1989) Appendix a: Simplified mathematical model of urban structure. In: Mills E, Hamilton B (eds), *Urban Economics* (Fourth ed.). 425–434

Mulalic I, Rouwendal J (2020) Does improving public transport decrease car ownership? Evidence from a residential sorting model for the Copenhagen metropolitan area. *Regional Science and Urban Economics*: 83. CrossRef

Muurinen JM (1982) Demand for health: A generalized Grossman model. *Journal of Health Economics* 1: 5–28. CrossRef

Nigg C, Oriwol D, Wunsch K, Burchartz A, Kolb S, Worth A, Woll A, Niessner C (2021) Population density predicts youth's physical activity changes during Covid-19 – Results from the MoMo study. *Health & Place* 70: 102619. CrossRef

Pomponi F, Saint R, Arehart J (2021) Decoupling densities from tallness in analysing the life cycle greenhouse gas emissions of cities. *npj Urban Sustainability* 1: 33. CrossRef

Ramanathan R (2002) *Introductory Econometrics with Applications* (Fifth ed.). South Western Thomson Learning

Rodrigue L, Daley J, Ravensbergen L, Manaugh K, Wasfi R, Butler G, El-Geneidy A (2022) Factors influencing subjective walkability: Results from built environment audit data. *Journal of Transport and Land Use* 15. CrossRef

Ross C, Wu CL (1995) The links between education and health. *American Sociological Review* 60: 719–745. CrossRef

Sun B, Yin C (2018) Relationship between multi-scale urban built environments and body mass index: A study of China. *Applied Geography* 94: 230–240. CrossRef

Sun G, Oreskovic N, Lin H (2014) How do changes to the built environment influence walking behaviors? A longitudinal study within a university campus in Hong Kong. *International Journal of Health Geographics* 13: 28–37. CrossRef

Tran D, Pham T, Nguyen T (2021) The influence of education on women's well-being: Evidence from Australia. *PLoS ONE* 16: 1–15. CrossRef

TRB – Transportation Research Board (2005) Does the built environment influence physical activity? Examining the evidence. Transportation research board special report 282. available at: https://www.trb.org/publications/sr/sr282.pdf

United States Census Bureau (2010) States area measurements and internal points coordinates. Available at: https://web.archive.org/web/20180316004512/https://www.census.gov/geo/reference/state-area.html

United States Census Bureau (2020) Change in resident population of the 50 States, the District of Columbia, and Puerto Rico: 1910 to 2020. Available at: https://www2.census.gov/programs-surveys/decennial/2020/data/apportionment/population-change-data-table.pdf

Wagstaff A (1986) The demand for health: A simplified Grossman model. *Bulletin of Economic Research* 38: 93–95. CrossRef

Wasfi R, Dasgupta K, Eluru N, Ross N (2016) Exposure to walkable neighbourhoods in urban areas increases utilitarian walking: Longitudinal study of Canadians. *Journal of Transport & Health* 3: 440–447. CrossRef

Wei Y, Xiao W, Wen M, Wei R (2016) Walkability, land use and physical activity. *Sustainability* 8. CrossRef

WHO – World Health Organization (2022) Physical activity. Available at: https://www.who.int/news-room/fact-sheets/detail/physical-activity

Wilkins E, Radley D, Morris M, Hobbs M, Christensen A, Marwa W, Morrin A, Griffiths C (2019) A systematic review employing the GeoFERN framework to examine methods, reporting quality and associations between the retail food environment and obesity. *Health & Place* 57: 186–199. CrossRef

Winston T (2015) *Handbook on Burnout and Sleep Deprivation: Risk Factors, Management Strategies and Impact on Performance and Behavior.* Nova Science Publishers, Inc

Wooldridge J (2009) *Introductory Economics: A Modern Approach* (Fourth ed.). Cengage Learning

Yule G (1926) Why do we sometimes get nonsense-correlations between time-series? – A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society* 89: 1–63. CrossRef

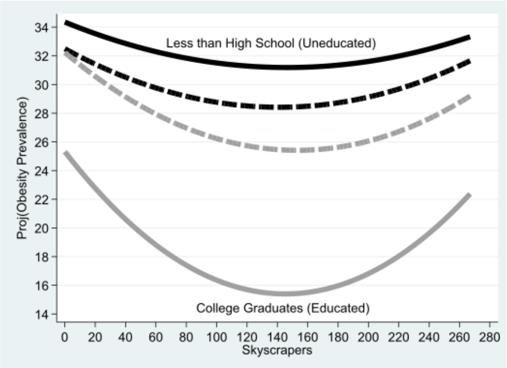
A Appendix: Data Structure

Table A.1: US states and observation numbers

Number	State	Observations	Number	State	Observations
1	Alabama	20	25	Missouri	20
2	Alaska	20	26	Montana	20
3	Arizona	20	27	Nebraska	20
4	Arkansas	20	28	Nevada	20
5	California	20	29	New Jersey	20
6	Colorado	20	30	New Mexico	20
7	Connecticut	20	31	New York	20
8	DC	20	32	North Dakota	18
9	Florida	20	33	Ohio	20
10	Georgia	20	34	Oklahoma	20
11	Hawaii	20	35	Oregon	20
12	Idaho	14	36	Pennsylvania	20
13	Illinois	20	37	Rhode Island	20
14	Indiana	20	38	South Dakota	20
15	Iowa	20	39	Tennessee	20
16	Kansas	20	40	Texas	20
17	Kentucky	20	41	Utah	16
18	Louisiana	20	42	Vermont	20
19	Maine	20	43	Virginia	20
20	Maryland	20	44	Washington	20
21	Massachusetts	20	45	West Virginia	20
22	Michigan	20	46	Wisconsin	20
23	Minnesota	20	47	Wyoming	20
24	Mississippi	20		Total	928

Notes: The full sample per-state includes 20 observations (10 years × 2 obs. for educated and non-educated in each state).

B Appendix: Four Educational Groups



Notes: the CDC dataset includes four groups: 1) College graduates (the most educated); 2) High School graduates; 3) Less than high school (the least educated); and 4) Some college or technical schools.

Figure B.1: Four Educational Groups

C Appendix: Simple Exercise of a Quadratic vs. Linear Relationship

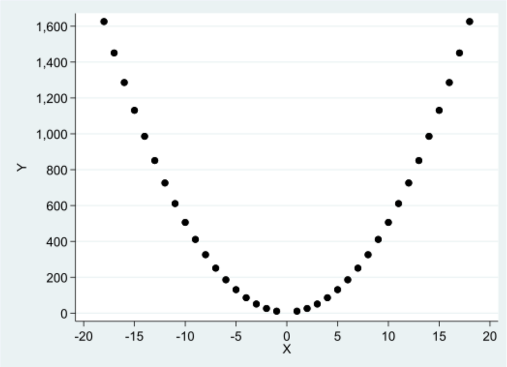


Figure C.1: A Scatter Diagram of the Quadratic Model $\widehat{Y} = 6 + 5X^2$

Table C.1: Regression outcomes

Coefficient of	Quadratic Y	Linear Y
X^2	5	-
		-
X	0	0
		(8.037)
Constant	6	591.83
		(86.99)
R-squared	1	0

Notes: Number in parentheses are standard errors.