Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Application of Causal Inference Techniques to Examine the Determinants and Outcomes of Obesity

Isaac Ofori ASARE
ioasare@atu.edu.gh

Accra Technical University
Alice Constant Mensah
Accra Technical University
Jacob Apibilla Ayembilla
Accra Technical University

Research Article

Keywords: Obesity, Lifestyle, Public Health, Instrumental Variables, Propensity Score Matching

Posted Date: April 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9142567/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Obesity is a multifaceted and chronic health condition with major implications for global public health and socio-economic systems. Guided by the socio-ecological model, this study frames obesity as the outcome of a complex interaction between biological, behavioral, and environmental determinants.

Methods

To empirically examine these determinants, the study employs advanced causal inference techniques, including Ordinary Least Squares (OLS), Instrumental Variables (IV), and Propensity Score Matching (PSM). A dataset of 2,111 records from Colombia, Peru, and Mexico is analyzed to identify factors associated with obesity while addressing confounding and endogeneity concerns.

Results

The study identifies key variables linked to obesity, such as age, gender, family history, and high-calorie food consumption. The overall prevalence of obesity is found to be 46.1%. Significant associations also emerge between obesity and factors including family history, physical activity, and mode of transportation.

Conclusion

By applying robust causal inference methods, the study provides reliable evidence on the determinants of obesity and offers actionable insights to guide targeted interventions and public health policies. The findings highlight the importance of combining theoretical frameworks with causal techniques to better understand and address the complexity of obesity.

1. INTRODUCTION

Obesity is a chronic, complex disease defined by excessive fat deposits that can impair health [1]. It represents a major global concern with profound impacts on healthcare systems [2]. According to the National Health and Nutrition Examination Survey (NHANES), the prevalence of obesity is slightly higher in women (40%) than men (35%) [3]. The condition increases the risk of type 2 diabetes, cardiovascular diseases, and certain cancers, while significantly affecting quality of life. Globally, the WHO reports that in 2022, 43% of adults were overweight a significant rise from 25% in 1990 along with 37 million children under the age of five. Recent comprehensive studies highlight a dramatic surge in prevalence, affecting 128.9 million individuals, with heightened vulnerability observed among younger cohorts [4].

The drivers of obesity are multifaceted, a complexity best understood through a Socio-Ecological Model (SEM) lens [57, 58]. The SEM posits that health outcomes are not solely determined by individual choices but are shaped by a dynamic interplay of factors across multiple levels: individual (genetics, biology), interpersonal (family, social norms), community (built environment, transportation), and public policy [54–56]. While intertwined factors such as genetic susceptibilities and urban environments pose challenges for management [5], traditional data handling methods have often failed to capture these intricate causative networks [6].

Although recent advancements in Artificial Intelligence [7] and Machine Learning [8, 9] have improved the handling of extensive parameters, and techniques like Structural Equation Modeling (SEM) offer insights into relationships, these methods often lack the ability to robustly address endogeneity and causality. Consequently, this study integrates Ordinary Least Squares (OLS), Instrumental Variables (IV), and Propensity Score Matching (PSM) to bridge this methodological gap.

OLS is employed to estimate linear relationships between obesity and predictors, assuming linearity and no omitted variable bias [10, 11]. To address endogeneity where unobserved characteristics (e.g., lifestyle factors) simultaneously impact predictors and obesity IV analysis is applied. This technique isolates exogenous variation using instruments correlated with predictors but not directly with the outcome, providing reliable causal estimates [12, 13]. Furthermore, PSM is implemented to minimize selection bias by balancing treated and control groups, offering a robust framework for comparing outcomes without strong linearity assumptions [14, 15].

The relevance of this study is that obesity is a multifaceted issue influenced by behavioral, genetic, and environmental factors. The usage of accurate causal inference is essential for: Designing effective interventions and policies. Identifying modifiable risk factors and evaluating the long-term impact of treatments. Using these methods ensures that the study provides actionable insights with a solid causal basis, contributing to evidence-based policymaking and clinical practices. The gap this study addresses is the lack of robust causal inference in existing obesity analyses, which often rely on descriptive or predictive models. By integrating OLS, IV, and PSM, the study (1) Provides a comprehensive framework to estimate causal effects; (2) Addresses confounding and endogeneity, which are critical in understanding obesity determinants and outcomes and (3) bridges the methodological gap between traditional regression models and advanced machine learning techniques [16].

2 RESEARCH METHODOLOGY

2.1 Data Source and Collection

The data for this study was sourced from an open-access repository on Kaggle, originating from a web-based survey on eating habits and physical condition collected from individuals in Colombia, Peru, and Mexico [17]. The dataset is publicly available at: https://www.kaggle.com/.../exploring-and-modelling-obesity-dataset/.

2.2 Data Composition and Preprocessing

The dataset comprises 2,111 records. To address potential class imbalance and enhance the robustness of the analytical sample, 77% of the data were synthetically generated using the Weka tool and the SMOTE (Synthetic Minority Over-sampling Technique) filter, with the remaining 23% constituting the originally collected data. The primary outcome variable is Obesity_Status, a multi-class label with seven levels: Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. For specific analyses, this was dichotomized into "Obese" (encompassing Obesity Types I-III) and "Not Obese" (all other categories). The dataset was complete, with no missing values requiring imputation.

2.3 Variable Description

The analysis utilized 17 variables, which are comprehensively described in Table 1. These include demographic factors (e.g., Age, Gender), biological measures (Height, Weight), behavioral and dietary habits (e.g., High_Calorie_Freq, Physical_Activity, Smoking), and familial and environmental context (Family_History, Transport). Body Mass Index (BMI) was calculated from Height and Weight to define the Obesity_Status outcome, consistent with World Health Organization (WHO) classifications [18].

Table 1
Data descriptions

Variable	Description
ID	A unique identifier for each record.
Gender	The gender of the individual (Male, Female).
Age	The age of the individual.
Height	The height of the individual (likely in centimeters or meters).
Weight	The weight of the individual (likely in kilograms).
Family History	Indicates whether there is a family history of being overweight (Yes, No).
High_Calorie_Freq	Frequency of eating high-calorie food (Always, Frequently, Sometimes, Never).
Vegetable_Freq	Frequency of consuming vegetables (Always, Frequently, Sometimes, Never).
Main_Meals	Number of main meals per day.
Between_Meals	Consumption of food between meals (Always, Frequently, Sometimes, Never).
Smoking	Smoking habits (Yes, No).
Water_Intake	Daily water intake (Less than a liter, 1–2 liters, more than 2 liters).
Calorie_Monitor	Calorie consumption monitoring (Yes, No).
Physical_Activity	Physical activity frequency (None, 1–2 days, 2–4 days, 4–5 days).
Tech_Time	Time using technology devices (0–2 hours, 3–5 hours, more than 5 hours).
Alcohol	Alcohol consumption (Never, Occasionally, Frequently, Daily).
Transport	Mode of transportation (Walking, Bike, Motorbike, Car, Public transportation).
Obesity_Status	obesity status [Obesity / No Obese]

2.4 Causal Framework and Directed Acyclic Graph (DAG)

To provide a transparent foundation for our analysis, we developed a Directed Acyclic Graph (DAG) outlining our core causal hypotheses (Fig. 3). This framework posits two primary pathways to obesity: a dietary pathway (Family History → High-Calorie Food →Obesity) and a physical activity pathway (Transport → Physical Activity → Obesity). We explicitly control for the confounding influence of Age and Gender. A key assumption of our model is the exclusion restriction, which requires that our instrumental variables (Family History and Transport) influence obesity only through their specified exposures and not via any other direct or hidden pathways. The Other Factors node represents additional covariates (e.g.,Smoking, Alcohol, Water Intake) included to reduce bias. This causal structure directly motivates our use of IV analysis to address unmeasured confounding and justifies our multi-method analytical strategy. Variables were selected based on theoretical relevance from the socio-ecological model and prior literature. All available covariates were included in propensity score models to minimize omitted variable bias.

2.5 Verification of the Quality of the data

To check the quality of the dataset, statistical tests were conducted to explore the relationships between key variables in the dataset. The Kolmogorov-Smirnov test was used to compare the distributions of height and weight in the dataset. The results indicated a statistically significant difference between the two distributions ($D = 0.0829D = 0.0829$, $p < 0.001$). The result suggests that height and weight follow distinct patterns, which are consistent with real-

world observations where Height is relatively stable in adults, while Weight is more variable and influenced by factors such as diet, physical activity, and genetics. This difference underscores the importance of treating these variables separately in obesity analysis and highlights the need for normalization or feature engineering (e.g., calculating BMI) for further analysis in predictive models. Also, a Chi-Square test was performed to examine the association between Gender and obesity levels. The results indicated a significant association ($\chi^2 = 657.75, df = 6, p < 0.001$) demonstrating that obesity levels vary significantly by gender. To assess the relationships among features and validate the quality of input variables, a correlation heatmap was generated (see Fig. 1 in supplementary material). The heatmap provides a visual representation of the pairwise correlations between numerical variables in the dataset, where values range from − 1 to 1. Strong positive correlations are represented by dark blue, indicating a direct relationship, while strong negative correlations appear in dark red, signifying an inverse relationship. Lighter shades reflect weaker correlations, implying minimal linear association between variables. As evidenced in Fig. 1 *Weight* and *family_hisOWt* (family history of overweight) exhibits a moderate positive correlation, suggesting that individuals with a family history of overweight are more likely to have higher body weight. This finding aligns with prior studies that emphasize genetic predisposition as a contributing factor to weight-related conditions [19, 20].

2.6 Model Performance Evaluation

A random forest classification model was trained and evaluated on the test dataset. The performance was assessed using a confusion matrix and relevant classification metrics, including accuracy, Kappa statistic, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and balanced accuracy. The model demonstrated a high level of accuracy, with an overall accuracy of 0.934 (95% CI: 0.918 to 0.957), indicating that approximately 94% of the observations were correctly classified. The model's performance was statistically significant, as evidenced by a p-value of less than 2.2e-16. Crucially, the Kappa statistic was 0.928, which suggests an "almost perfect" level of agreement beyond what would be expected by chance alone. This confirms that the high accuracy is a result of the model's robust learning, rather than coincidental correct predictions. A detailed class-wise breakdown (Table 2) reveals that the sensitivity and specificity were high across all classes, with values ranging from 0.851 to 0.998 and 0.954 to 1.000, respectively. The consistently high sensitivity indicates the model is highly effective at correctly identifying true cases (minimizing false negatives), while the near-perfect specificity shows it is excellent at ruling out non-cases (minimizing false positives). This balance is essential for clinical reliability. The positive predictive value (PPV) and negative predictive value (NPV) were also high, with most classes having values above 0.95. These metrics provide insight into the practical utility of the model; a high PPV implies that when the model predicts a specific obesity outcome, it is highly likely to be correct, thereby instilling confidence in its diagnostic suggestions. The prevalence of each class varied from 0.1155 to 0.1661, and the detection rate was relatively consistent across classes, with values ranging from 0.1155 to 0.1598. Given this slight variation in prevalence, the Balanced Accuracy metric becomes critical. Balanced accuracy was high for all classes (0.924 to 0.995), confirming that the model's performance is not biased toward the majority classes but is discriminative across all categories. Finally, the area under the curve (AUC) for the one-vs-rest approach was 0.996. This exceptional value indicates that the model possesses excellent discriminative ability, effectively separating each specific class from the others with near-perfect precision. The overall multi-class AUC of 0.996 suggests that the model is highly effective in distinguishing between all classes, validating its potential as a robust tool for categorizing obesity determinants.

Table 2
Model evaluation statistics by class

Sensitivity	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**	**Class 7**
	0.901	**0.977**	**0.851**	**0.920**	**0.962**	**0.966**	**0.990**
Specificity	0.998	0.954	0.996	0.993	0.989	1.000	1.000
Pos Pred Value	0.987	0.771	0.974	0.952	0.944	1.000	1.000
Neg Pred Value	0.986	0.996	0.977	0.987	0.992	0.995	0.998
Prevalence	0.128	0.136	0.138	0.138	0.166	0.141	0.154
Detection Rate	0.116	0.133	0.117	0.127	0.160	0.136	0.152
Detection Prevalence	0.117	0.173	0.120	0.133	0.169	0.136	0.152
Balanced Accuracy	0.950	0.966	0.924	0.956	0.975	0.983	0.995
Area Under Curve (AUC)	0.999	0.989	0.992	0.992	0.999	1.000	1.000
Overall multi-class AUC: 0.996							

2.7 Statistical Analysis

All statistical analyses were conducted using R software (version 4.50). The dataset was complete, with no missing observations, thus requiring no data imputation. Our analytical strategy employed a multi-method causal inference approach to ensure robust and unbiased estimates. We began with preliminary bivariate analyses, including chi-square and Kolmogorov-Smirnov tests, to explore initial variable associations. To account for unobserved heterogeneity between Colombia, Peru, and Mexico, we incorporated country-level fixed effects into all subsequent multivariate models. Our causal analysis proceeded in three stages: First, we estimated baseline associations using Ordinary Least Squares (OLS). Second, to address endogeneity from unmeasured confounding, we implemented Instrumental Variable (IV) analysis, using Family History as an instrument for High-Calorie Food Frequency and Transport for Physical Activity; the validity of this approach was confirmed using the Wu-Hausman test. Third, to mitigate selection bias, we employed Propensity Score Matching (PSM) using nearest-neighbor matching with a caliper of 0.2 standard deviations. Model fit and validity were assessed through specific diagnostic criteria: for OLS, we examined R-squared, and root mean squared error (RMSE); for IV, we reported the first-stage F-statistic, with a value > 10 indicating a strong instrument; and for PSM, we evaluated covariate balance using standardized mean differences (SMD < 0.1). The robustness of our findings was verified through comprehensive sensitivity analyses, including testing alternative instrument specifications, varying PSM algorithms, and assessing residual

confounding. In line with best practices, we present 95% confidence intervals (including coefficients, odds ratios) alongside p-values to provide a comprehensive interpretation of the precision and magnitude of the estimated effects.

3. RESULTS

3.1 Prevalence of obesity status at seven levels and its associated factors

To better understand the association between the levels of obesity and its associated risk factors, the chi-square test (χ^2) of independence was performed to identify whether there is a statistically significant association between obesity and all covariates (p value ≤ 0.05). The test highlights relationships that might suggest which factors are prevalent among individuals with obesity, helping pinpoint which variables are worth investigating for further analysis [21]. Rao- Scott test of independence showed a significant association between disability and all covariates (p value ≤ 0.05). As evidenced in Table 3, the prevalence level for Insufficient Weight was 12.9%, normal weight was 13.6%, Overweight Level I was 13.7%, Overweight Level II was 13.7%, obesity type I was 16.6%, obesity II was 14.1% while obesity Type III was 15.3%. The results suggest that all factors are associated significantly with the level of obesity and overweight. For example, gender shown significantly association with obesity with most of the males been classified as obese, 55.6% of the males were classified as obesity Type I, 99.3% classified as obesity Type II and indication that men stood a higher chance of been overweight or obese than women. Table 4 provides further details of the association between the two categorical variables.

3.2 Prevalence of obesity status at two levels and its associated factors

After determining the association between the seven (7) levels of obesity, overweight against the covariates, the result showed statistically significant insight, it was necessary to re-group the outcome variable to take two levels (Obese and No obesity) for better understanding of the association among the variables. To do that Insufficient, Normal, Overweight, Level I, Overweight Level II were classified as No obesity while Obesity Type I, Obesity Type II and Obesity Type III were classified as obesity for better and easy understanding. As shown in Table 3, The prevalence of obesity was 46.1% and chi-square test of independence showed a significant association between obesity and all covariates (p value $\leq$ 0.05) except for Gender and Smoking habits (p value $>$ 0.05) The results suggest that men were more likely to have the condition than women with an estimated value of 50.6% and 49.4% respectively, although gender shown no significant association with obesity. Family history was considered a significant factor contributing to obesity as those with family history were more likely, 99.2%, to be obese than those without family history. Individuals who frequently eat high-calorie foods were more likely, 98%, to be obese than those who do not eat high-calorie food as shown in Table 4. Individuals who frequently eat high-calorie foods were more likely, 98% to be obese than those who do not eat high-calorie food. Alcohol consumption is significantly associated with obesity as those who consume stand a high chance, 74% of been obese than the other levels of alcohol consumption. The mode of transportation significantly influences obesity status. The results suggest that individuals who use public transportation stand a high chance,78.1% of whom have been obese than those who use other means of transportation such as Automobile, Bike, Motorbike and Walking. The results suggest that individuals within the age range of 21–25 years stand a higher chance of been obese than other age groups, as they formed 30.6% as shown in Table 4.

Table 3
Prevalence of obesity and its associated risk factors

Characteristics	Insufficient Weight	Normal Weight	Overweight Level I	Overweight Level II	Obesity Type I	Obesity Type II	Obesity Type III	Total	$\varvec{\chi}^2$	Sig
	12.9%	**13.6%**	**13.7%**	**13.7%**	**16.6%**	**14.1%**	**15.3%**			
Gender	Weighted %	Weighted %	Weighted %	Weighted %	Weighted %	Weighted %	Weighted %	%		
Male	36.4%	50.9%	50.0%	64.5%	55.6%	99.3%	0.3%	50.6%	657.746	0.000
Female	63.6%	49.1%	50.0%	35.5%	44.4%	0.7%	99.7%	49.4%		
Family history										
Yes	46.3%	54.0%	72.1%	93.8%	98.0%	99.7%	100.0%	81.8%	621.979	0.000
No	53.7%	46.0%	27.9%	6.2%	2.0%	0.3%	0.0%	18.2%		
Frequency of eating high-calorie food										
Yes	81.3%	72.5%	92.4%	74.5%	96.9%	97.6%	99.7%	88.4%	233.341	0.000
No	18.8%	27.5%	7.6%	25.5%	3.1%	2.4%	0.3%	11.6%		
Consumption of food between meals										
Always	0.7%	12.2%	1.7%	1.0%	1.7%	0.7%	0.0%	2.5%	802.977	0.000
Frequently	44.5%	28.9%	4.8%	5.5%	1.7%	0.3%	0.3%	11.5%		
Sometimes	53.7%	55.4%	81.4%	93.1%	96.3%	98.7%	99.7%	83.6%		
Never	1.1%	3.5%	12.1%	0.3%	0.3%	0.3%	0.0%	2.4%		
Smoking habits										
Yes	0.4%	4.5%	1.0%	1.7%	1.7%	5.1%	0.3%	2.1%	32.138	0.000
No	99.6%	95.5%	99.0%	98.3%	98.3%	94.9%	99.7%	97.9%		
Calorie consumption monitoring										
Yes	8.1%	10.5%	12.8%	1.4%	0.6%	0.3%	0.0%	4.5%	123.024	0.000
No	91.9%	89.5%	87.2%	98.6%	99.4%	99.7%	100.0%	95.5%		
Alcohol consumption										
Never	43.0%	37.3%	17.2%	44.1%	47.0%	23.9%	0.3%	30.3%	338.578	0.000
Occasionally	56.6%	56.1%	77.2%	49.3%	49.0%	75.4%	99.7%	66.4%		
Frequently	0.4%	6.3%	5.5%	6.6%	4.0%	0.7%	0.0%	3.3%		
Daily	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		
Mode of transportation										
Automobile	16.9%	15.7%	22.8%	32.4%	31.3%	32.0%	0.3%	21.6%	292.594	0.000
Bike	0.0%	1.4%	0.7%	0.0%	0.0%	0.3%	0.0%	0.3%		
Motorbike	0.0%	2.1%	0.3%	0.3%	0.9%	0.0%	0.0%	0.5%		
Public transportation	80.9%	69.7%	73.1%	65.2%	67.2%	67.3%	99.7%	74.8%		
Walking	2.2%	11.1%	3.1%	2.1%	0.6%	0.3%	0.0%	2.7%		
Age (Mean (± SD) = 24.31(6.35))	19.78(2.67)	21.74(5.10)	23.42(6.13)	27.00(8.06)	25.88(7.76)	28.23(4.87)	23.50(2.76)			
Age group										
14–20	66.2%	46.0%	32.4%	20.3%	21.7%	0.3%	13.3%	27.7%	864.913	0.000
21–25	25.0%	41.5%	41.0%	31.7%	39.3%	26.3%	25.0%	32.9%		

Note: The Correlation estimation is based on Fisher's r-to-z transformation with bias adjustment.

Characteristics	Insufficient Weight	Normal Weight	Overweight Level I	Overweight Level II	Obesity Type I	Obesity Type II	Obesity Type III	Total	$\varvec{\chi}^2$	Sig
	12.9%	**13.6%**	**13.7%**	**13.7%**	**16.6%**	**14.1%**	**15.3%**			
26–30	0.7%	6.3%	9.3%	8.3%	7.7%	19.5%	26.9%	11.5%		
31–35	0.7%	3.8%	6.2%	17.9%	3.4%	12.1%	0.0%	6.2%		
36–40	0.4%	1.7%	5.5%	6.6%	12.5%	5.7%	0.0%	4.8%		
> 40	7.0%	0.7%	5.5%	15.2%	15.4%	36.0%	34.9%	16.8%		
Obesity status -The age of the individual [M = 24.31, SD = 6.35; $\rho = 0.283, p < 0.001$]										
Obesity status - The height of the individual [M = 1.70, SD = 0.09; ρ =0.134, p < 0.001]										
Obesity status - The weight of the individual [M = 86.59, SD = 26.19; ρ =0.913, p < 0.001										
Obesity status - Frequency of consuming vegetables [M = 2.42, SD = 0.53; ρ =0.228, p < 0.001]										
Obesity status - Number of main meals per day. [M = 2.69, SD = 0.78; ρ =0.027, p = 0.220]										
Obesity status - Daily water intake [M = 2.01, SD = 0.61; ρ =0.133, p < 0.001]										
Obesity status - Physical activity frequency [M = 1.01, SD = 0.85; ρ =-0.200, p < 0.001										
Obesity status - Time using technology devices [M = 0.66, SD = 0.61; ρ =-0.108, p < 0.001]										
Obesity status – BMI [M = 29.70, SD = 8.01; ρ = 0.978, p < 0.001]										
Note: The Correlation estimation is based on Fisher's r-to-z transformation with bias adjustment.										

Table 4
Prevalence of obesity at two levels and its associated risk factors

Characteristics	No obesity	Obese	Total %	$\varvec{\chi^2}$	Sig
Prevalence	**53.9%**	**46.1%**			
Gender					
Male	50.7%	50.5%	50.6%	.004	0.947
Female	49.3%	49.5%	49.4%		
Family history of being overweight					
Yes	66.9%	99.2%	81.8%	366.387	0.000
No	33.1%	0.8%	18.2%		
Frequency of eating high-calorie food					
Yes	80.2%	98.0%	88.4%	163.564	0.000
No	19.8%	2.0%	11.6%		
Consumption of food between meals					
Always	4.0%	0.8%	2.5%	280.331	0.000
Frequently	20.5%	0.8%	11.5%		
Sometimes	71.2%	98.1%	83.6%		
Never	4.3%	0.2%	2.4%		
Smoking habits					
Yes	1.9%	2.3%	2.1%	.283	0.595
No	98.1%	97.7%	97.9%		
Calorie consumption monitoring					
Yes	8.2%	0.3%	4.5%	74.573	0.000
No	91.8%	99.7%	95.5%		
Alcohol consumption					
Never	35.3%	24.4%	30.3%	52.328	0.000
Occasionally	59.9%	74.0%	66.4%		
Frequently	4.7%	1.6%	3.3%		
Daily	0.1%	0.0%	0.0%		
Mode of transportation					
Automobile	22.0%	21.2%	21.6%	44.418	0.000
Bike	0.5%	0.1%	0.3%		
Motorbike	0.7%	0.3%	0.5%		
Public transportation	72.1%	78.1%	74.8%		
Walking	4.7%	0.3%	2.7%		
Age group					
14–20	40.8%	12.3%	27.7%	367.406	0.000
21–25	34.9%	30.6%	32.9%		
26–30	6.2%	17.7%	11.5%		
31–35	7.3%	4.9%	6.2%		
36–40	3.6%	6.3%	4.8%		
> 40	7.1%	28.2%	16.8%		

3.3 Model Justification or Selection

The analysis of the data used three main models. Using OLS, Instrumental Variables, and PSM together ensures robust analysis by addressing different biases: OLS provides baseline estimates, Instrumental Variables corrects for endogeneity, and PSM accounts for selection bias, leading to more credible and comprehensive results. The Wu-Hausman test for endogeneity was not significant (*p-value > 0.05*), suggesting that there is no strong evidence of endogeneity in our model. This implies that ordinary least squares (OLS) estimation may be appropriate and could provide unbiased estimates. We therefore present the results of both the OLS and IV models. Given the theoretical considerations around endogeneity in studies of obesity determinants, we include the IV model results for completeness but base our main interpretations on the OLS findings. As evidenced in Table 5, six variables were statistically significantly associated with obesity outcomes using the OLS technique : gender (*t=-2.342,p = 0.019*), height (*t=-17.608=,p < 0.001*); weight (*t = 54.040,p < 0.000*); high_calorie_freq (*t=-6.792,p < 0.000*); eating between meals (t=-2.862,p = 0.004) and alcohol intake (*t=-6.776,p < 0.000*) whiles in the case of the IV only height (*t=-4.539,p < 0.000*) and weights (*t = 7.012,p < 0.001*) were statistically significant. See Fig. 2 in the supplementary material for the relative importance of the variables.

3.4 Propensity Score Matching (PSM)

Propensity Score Matching (PSM) was employed to address selection bias and confounding by creating a balanced dataset, thereby approximating an experimental design for estimating the effect of obesity determinants. Variables identified in Table 5 were selected as covariates for the propensity score model based on their potential association with both the treatment assignment (obesity status) and the outcome. Table 6 presents the assessment of covariate balance before and after matching. The results indicate a general improvement in balance for the majority of covariates, as evidenced by the reduction in Standardized Mean Differences (SMD). The matching process successfully paired 972 treated individuals with 972 control individuals, resulting in a balanced analytical sample. However, it is important to note that residual imbalance persisted for specific covariates, notably family_history (SMD = -2.299) and distance (SMD = 1.746), which exceeded the recommended threshold of 0.1 [22]. To address this residual imbalance and ensure robust estimates, a doubly robust estimation approach was adopted. This involved adjusting for these specifically unbalanced covariates in the final logistic regression model to explicitly control for any remaining confounding bias.

Table 7 details the conditional logistic regression results on the matched cohort. The analysis identified significant causal effects of several key determinants on obesity. Age demonstrated a positive effect (OR = 1.107, 95% CI [Lower-Upper], p < 0.001), indicating that for each additional year, the odds of obesity increase by approximately 10.7%.

Regarding behavioral and biological factors, the results indicated significant protective effects for certain variables, interpreted as follows: Family History: Individuals with a family history of obesity showed significantly lower odds of obesity (OR = 0.023, p < 0.001), suggesting a 97.7% reduction in odds compared to those without, after adjustment; High-Calorie Food Consumption: Frequent consumption was associated with lower odds of obesity (OR = 0.089, p < 0.001), indicating a 91.1% decrease in likelihood holding other factors constant; Physical Activity: Engagement in physical activity was associated with a 22.7% reduction in the odds of obesity (OR = 0.773, p < 0.001). Conversely, other environmental and lifestyle factors showed positive associations: Water Intake: Higher water intake was associated with 32.8% higher odds of obesity (OR = 1.328, p = 0.002); Transportation: The mode of transportation was a strong predictor, with users of the specific transport mode exhibiting 48.1% higher odds of obesity (OR = 1.481, p < 0.001). Gender (OR = 1.127, p = 0.264) and smoking status (OR = 0.961, p = 0.916) did not demonstrate statistically significant associations with obesity in the matched sample.

Table 5
Summary of OLS and IV model results

Variable	OLS [MODEL I]					IV Inclusion [Model II]				
	Estimate	**SE**	***t value***	***p-value***	**95% CI**	**Estimate**	**SE**	***t value***	***p-value***	**95% CI**
Intercept	3.461	0.225	15.389	0.000	3.020 to 3.902	2.479	1.363	1.819	0.069	-0.192 to 5.150
weight	0.018	0.000	54.040	0.000	0.018 to 0.018	0.016	0.002	7.012	0.000	0.012 to 0.020
height	-1.777	0.101	-17.608	0.000	-1.975 to -1.579	-1.704	0.375	-4.539	0.000	-2.439 to -0.969
high_calorie_freq	-0.133	0.020	-6.792	0.000	-0.172 to -0.094	-0.082	0.042	-1.960	0.050	-0.164 to 0.000
vegetable_freq	-0.023	0.012	-1.914	0.056	-0.047 to 0.001	-0.049	0.098	-0.499	0.618	-0.241 to 0.143
main_meals	0.012	0.008	1.488	0.137	-0.004 to 0.028	0.036	0.103	0.352	0.725	-0.166 to 0.238
between_meals	-0.039	0.014	-2.862	0.004	-0.066 to -0.012	0.408	0.320	1.275	0.203	-0.219 to 1.035
smoking	-0.049	0.041	-1.186	0.236	-0.129 to 0.031	-0.280	0.608	-0.460	0.645	-1.472 to 0.912
gender	-0.038	0.016	-2.342	0.019	-0.069 to -0.007					
age	0.000	0.001	0.062	0.950	-0.002 to 0.002					
family_history	-0.010	0.018	-0.577	0.564	-0.045 to 0.025					
water_intake	-0.009	0.010	-0.886	0.376	-0.029 to 0.011					
calorie_monitor	0.053	0.029	1.793	0.073	-0.004 to 0.110					
physical_activity	-0.002	0.008	-0.320	0.749	-0.018 to 0.014					
tech_time	-0.009	0.010	-0.888	0.375	-0.029 to 0.011					
alcohol	-0.081	0.012	-6.776	0.000	-0.105 to -0.057					
transport	-0.005	0.006	-0.850	0.395	-0.017 to 0.007					
BMI	0.284	0.061	4.621	0.000	0.164 to 0.404					

Table 6
Propensity Score Matching before and after matching

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
Summary of Balance for All Data							
distance	0.619	0.325	2.147	0.242	0.301	0.506	
age	25.806	23.038	0.468	0.849	0.187	0.344	
gender	1.495	1.493	0.003	1.000	0.001	0.001	
family_history	1.008	1.331	-3.571	0.037	0.161	0.323	
high_calorie_freq	1.020	1.198	-1.291	0.121	0.089	0.179	
physical_activity	0.875	1.126	-0.326	0.734	0.091	0.202	
smoking	1.977	1.981	-0.022	1.168	0.002	0.003	
water_intake	2.073	1.953	0.196	1.006	0.076	0.243	
transport	3.362	3.368	-0.005	0.911	0.016	0.043	
Summary of Balance for Matched Data							
distance	0.619	0.380	1.746	0.267	0.248	0.436	1.753
age	25.806	23.441	0.400	0.782	0.168	0.321	1.042
gender	1.495	1.471	0.047	1.003	0.012	0.024	1.010
family_history	1.008	1.216	-2.299	0.048	0.104	0.208	2.299
high_calorie_freq	1.020	1.136	-0.839	0.163	0.058	0.116	1.062
physical_activity	0.875	1.056	-0.236	0.776	0.068	0.167	1.119
smoking	1.977	1.978	-0.007	1.047	0.001	0.001	0.284
water_intake	2.073	1.993	0.130	1.053	0.069	0.221	1.086
transport	3.362	3.348	0.012	0.905	0.018	0.037	0.909

Table 7
Summary of Fit (IV) an outcome model on the matched data

Variable	Estimate	Std. Error	z value	Pr(>|z|)	Odds Ratio	95% CI (OR)
Intercept	2.338	1.032	2.264	0.024	10.36	1.37 to 78.31
age	0.101	0.011	9.001	0.000	1.11	1.083 to 1.130
family_history	−3.757	0.378	-9.931	0.000	0.023	0.011 to 0.049
high_calorie_freq	−2.415	0.267	-9.052	0.000	0.089	0.053 to 0.151
physical_activity	−0.257	0.067	-3.862	0.000	0.773	0.678 to 0.882
water_intake	0.284	0.09	3.150	0.002	1.328	1.114 to 1.585
transport	0.393	0.054	7.219	0.000	1.481	1.333 to 1.647
gender	0.120	0.107	1.118	0.263	1.127	0.914 to 1.391
smoking	−0.040	0.375	-0.106	0.916	0.961	0.461 to 2.004

4. Discussions

Grounded in the Socio-Ecological Model (SEM) [57, 58], this study utilized advanced causal inference techniques to disentangle the multifaceted determinants of obesity. By examining factors across individual, interpersonal, and community levels, we provide robust evidence for policy formulation. To the best of our knowledge, this is one of the first studies to apply this specific suite of causal methods (OLS, IV, and PSM) to a nationally representative dataset comprising 2,111 records from Colombia, Peru, and Mexico. Notably, the associations discussed below remained robust even after controlling for inter-country heterogeneity, suggesting these are fundamental determinants within the region.

At the individual and biological determinants level, our findings reveal an obesity prevalence of 46.1%, with a higher prevalence observed among males (50.5%). This aligns with findings by [23, 59]. While global trends often show higher rates in women, the World Health Organization notes that obesity among men is accelerating rapidly due to shifting behavioral and environmental factors [24, 60]. Age also emerged as a critical individual-level factor; prevalence

was highest among young adults aged 21–25 (30.6%), followed by those over 40 (28.2%). This supports emerging research suggesting shifts in obesity onset [25]. While older age groups traditionally carry higher risk due to metabolic changes [26], the surge among adults in their 20s underscores the impact of modern sedentary lifestyles and socioeconomic transitions early in adulthood.

Consistent with the interpersonal level of the SEM, our findings confirm that family history significantly influences obesity status. This corroborates a study in Ghana which found that individuals with parents or grandparents living with obesity faced a more than two-fold increased risk compared to siblings without such history [27]. According to [28], family history creates a genetic and shared environmental predisposition, making individuals more susceptible to the condition, although they may ironically experience less stigma than those already diagnosed.

Our results indicate that men consume high-calorie foods more frequently than women, a behavior strongly linked to the observed male-dominant obesity prevalence. This gender disparity is driven by a convergence of biological and sociocultural factors [29, 30, 31]. Biologically, men have greater muscle mass and higher energy needs, driving a preference for calorie-dense foods [32]. Furthermore, higher testosterone levels are linked to increased appetite [33, 34]. Socio culturally, norms often encourage men to consume larger portions and energy-dense foods associated with "strength," such as meats [35, 36]. Unlike women, men may face less societal pressure to conform to thinness, leading to more frequent indulgence and emotional eating behaviors characterized by high-calorie consumption [37, 38]. The analysis identifies frequent snacking as a major contributor to obesity, a finding that highlights the "food environment" aspect of the SEM [39]. Frequent snacking, particularly on ultra-processed foods, leads to a significant caloric surplus. Evidence suggests that regular snackers consume more daily calories, which, when paired with inactivity, results in fat accumulation [40, 41]. Ultra-processed foods are particularly problematic; they are energy-dense, nutritionally poor, and fail to induce satiety [42]. Research by [40] indicates that consumption of these foods can increase daily intake by approximately 500 calories. Moreover, the timing of consumption matters; late-night snacking disrupts circadian rhythms and metabolic regulation, further exacerbating obesity risk [40, 43].

Our results indicate that physical activity remains a cornerstone of obesity prevention. Our causal estimates confirm a significant protective effect of physical activity. This aligns with [44], who found that 150 minutes of moderate exercise weekly reduces obesity risk by 30%. Conversely, sedentary behaviors are strongly linked to increased BMI [45]. Regarding water intake, Our causal analysis revealed a positive and significant effect on obesity. This finding contrasts with some literature, such as [46] and [47], which suggest that water consumption aids weight loss by displacing caloric beverages. Our counter-intuitive result might suggest reverse causality (where individuals with obesity increase water intake as a weight-control strategy) or unmeasured dietary confounders, highlighting the complexity of individual dietary behaviors.

The mode of transportation serves as a proxy for the built environment within the SEM. Our results demonstrate that transportation choices significantly impact obesity. Active transport (walking, cycling) is associated with lower body fat percentages [48], whereas car dependency correlates positively with obesity, particularly in urban areas with poor walkability [49]. This underscores the critical role of urban planning in shaping health outcomes.

While descriptive statistics showed higher obesity in males, our causal models suggested that gender itself did not have a significant causal effect after controlling for other covariates. This nuance reflects the literature: while women face barriers to physical activity [50], men often engage in riskier health behaviors [51]. Similarly, we found no significant association between smoking and obesity. The literature is mixed; while smoking cessation is associated with initial weight gain [52], the long-term cardiovascular benefits of quitting far outweigh these risks [53].

Implications for Future Studies

The results of this study carry significant implications for public health policy and future research, asserting that effective interventions must be multifactorial and framed within the Socio-Ecological Model (SEM), thereby targeting not only individual behaviors but also the environmental contexts that influence them. Policy actions should prioritize the Community and Environment by supporting urban planning initiatives for active transport, such as creating walkable neighborhoods and cycling infrastructure, to combat sedentary lifestyles; simultaneously, Public Health Campaigns must leverage findings on diet and water intake to promote healthier food environments and clear guidance on beverage choices, specifically advocating for replacing sugar-sweetened drinks. Furthermore, the findings underscore the need for Gender-Specific Strategies tailored to address unique risks, such as sociocultural barriers for women and unhealthy dietary norms among men, as well as Integrated Care strategies that incorporate weight management support directly into smoking cessation programs. Collectively, these insights strongly advocate for comprehensive, community-centered strategies to reduce healthcare burdens associated with obesity.

Limitations of the Study

Certain limitations must be acknowledged. First, the reliance on partially synthetic data, while validated for quality, may limit full generalizability. Second, despite the use of Instrumental Variables, unmeasured confounders such as psychological stress or specific genetic markers could still influence results. Third, while we controlled for country-level fixed effects to account for inter-country differences, the dataset's focus on Latin America may restrict applicability to other global regions. Future research should prioritize longitudinal data to validate these causal pathways and incorporate broader socio-economic variables.

Conclusion

This study demonstrates the utility of causal inference techniques specifically OLS, IV, and PSM in unraveling the complex determinants of obesity. Grounded in the Socio-Ecological Model, the findings underscore that obesity is not merely a result of individual choice but is shaped by a dynamic interplay of biological (age, genetics), behavioral (diet, physical activity), and environmental (transportation) factors. By addressing methodological gaps and robustly

controlling for confounding, this research provides a framework for evidence-based policy. To mitigate the obesity epidemic effectively, comprehensive strategies must move beyond siloed approaches and target the structural and social determinants of health.

Declarations

Acknowledgments

The authors are grateful to all the individuals who played a crucial role in getting this paper done.

Author Contributions

Isaac Ofori Asare**:** conceptualization, methodology, validation, formal analysis, investigation, data curation, writing – original draft, writing – reviewing and editing, visualization, and formal statistical analysis data. Alice Constant Mensah and Jacob Apibilla Ayembilla,conceptualization, reviewing and editing

Funding

No funding was received for this research.

Data availability

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/code/anetakovacheva/exploring-and-modelling-obesity-dataset/notebook.

Ethics approval and consent to participate

The study utilized publicly available secondary data, and therefore no ethical approval or participant consent was required. The study adhered to the principles of the Declaration of Helsinki.”

Consent for publication

Not applicable.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Młynarska E, Bojdo K, Bulicz A, Frankenstein H, Gąsior M, Kustosik N, Rysz J, Franczyk B. Obesity as a Multifactorial Chronic Disease: Molecular Mechanisms, Systemic Impact, and Emerging Digital Interventions. Curr Issues Mol Biol. 2025;47(10):787. https://doi.org/10.3390/cimb47100787.
2. Okunogbe A, Nugent R, Spencer G, Powis J, Ralston J, Wilding J. (2022). Economic impacts of overweight and obesity: current and future estimates for 161 countries. BMJ global health, 7(9), e009773.
3. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL. Trends in obesity among adults in the United States, 2005 to 2014. JAMA. 2016;315(21):2284–91.
4. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, Adams RJ, Aekplakorn W, Afsana K, Aguilar-Salinas CA, Agyemang C. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. lancet. 2017;390(10113):2627–42.
5. Valenzuela PL, Carrera-Bastos P, Castillo-García A, Lieberman DE, Santos-Lozano A, Lucia A. Obesity and the risk of cardiometabolic diseases. Nat reviews Cardiol. 2023;20(7):475–94.
6. Yi X, Heyang Z, Gao S, Li M. (2024). A review of the application of deep learning in obesity: From early prediction aid to advanced management assistance. Diabetes Metabolic Syndrome: Clin Res Reviews, 103000.
7. Colmenarejo G. Machine learning models to predict childhood and adolescent obesity: a review. Nutrients. 2020;12(8):2466. https://doi.org/10.3390/nu12082466.
8. Hernán MA, Robins J. (2020). Causal inference: What if. Boca Raton: Chapman & Hall/CRC.(2020). *Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations*.
9. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55.
10. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton University Press; 2009.
11. Rigdon J, Berkowitz SA, Seligman HK, Basu S. Re-evaluating associations between the Supplemental Nutrition Assistance Program participation and body mass index in the context of unmeasured confounders. Soc Sci Med. 2017;192:112–24.
12. Imbens GW, Wooldridge JM. Recent developments in the econometrics of program evaluation. J Econ Lit. 2009;47(1):5–86.
13. Averett SL, Smith JK. Financial hardship and obesity. Econ Hum Biology. 2014;15:201–12.
14. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66(5):688.

15. Zhao J, He J, Hong B, Hou Z, Wang Y, Luo Y, Li M, Wei D, Zhang N. Propensity score-matched analysis examining the role of obesity on outcomes in retroperitoneal laparoscopic adrenalectomy: a cohort study. Gland Surg. 2024;13(6):802.
16. Dubé L, Addy NA, Blouin C, Drager N. From policy coherence to 21st century convergence: a whole-of-society paradigm of human and economic development. Ann N Y Acad Sci. 2014;1331(1):201–15.
17. Palechor FM, De la Manotas H, A. Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data brief. 2019;25:104344.
18. Weir CB, Jan A. BMI Classification Percentile and Cut Off Points. In StatPearls. StatPearls Publishing; 2023.
19. Lowe MR, Shank LM, Mikorski R, Butryn ML. Personal history of dieting and family history of obesity are unrelated: Implications for understanding weight gain proneness. Eat Behav. 2015;17:144–8.
20. Eisenberg MH, Street RL, Persky S. It runs in my family … The association of perceived family history with body dissatisfaction and weight bias internalization among overweight women. Women Health. 2017;57(4):478–93. https://doi.org/10.1080/03630242.2016.1170095.
21. Tetteh J, Asare IO, Adomako I, Udofia EA, Seneadza NAH, Adjei-Mensah E, Calys-Tagoe BN, Swaray SM, Ekem-Ferguson G, Yawson A. Sex differences and factors associated with disability among Ghana's workforce: a nationally stratified cross-sectional study. BMJ open. 2021;11(3):e044246.
22. Zhang Z, Kim HJ, Lonjon G, Zhu Y. (2019). Balance diagnostics after propensity score matching. Annals translational Med, 7(1).
23. Muscogiuri G, Verde L, Vetrani C, Barrea L, Savastano S, Colao A. Obesity: a gender-view. J Endocrinol Investig. 2024;47(2):299–306. https://doi.org/10.1007/s40618-023-02196-z.
24. Kanter R, Caballero B. Global gender disparities in obesity: a review. Adv Nutr (Bethesda Md). 2012;3(4):491–8. https://doi.org/10.3945/an.112.002063.
25. Szczudlik E, Stępniewska A, Bik-Multanowski M, Brandt-Heunemann S, Flehmig B, Małecka-Tendera E, Mazur A, Petriczko E, Ranke MB, Wabitsch M, Zachurzok A, Wójcik M. The age of the obesity onset is a very important factor for the development of metabolic complications and cardiovascular risk in children and adolescents with severe obesity. Eur J Pediatrics. 2024;183(9):3833–41. https://doi.org/10.1007/s00431-024-05636-x.
26. Jura M, Kozak LP. Obesity and related consequences to ageing. Age (Dordrecht Netherlands). 2016;38(1):23. https://doi.org/10.1007/s11357-016-9884-3.
27. Amponsem-Boateng C, Oppong TB, Amertil NP, Boakye-Yiadom J, Kpongboe DG. Exploring the prevalence and predictors of overweight and obesity in second cycle schools in Ghana. Int J Community Med Public Health. 2024;11(6):2160.
28. Alammari SS, Almatrafi MA, Fallatah JM, Alahmadi KF, Aloufi YA, Alahmadi AS. The Effects of Sociodemographic Factors and BMI on Weight Stigma Among Adults With Obesity in Madinah Region, Saudi Arabia: A Cross-Sectional Analytical Study. Cureus. 2024;16(7):e63993. https://doi.org/10.7759/cureus.63993.
29. Gil M, Rudy M, Stanisławczyk R, Duma-Kocan P, Żurek J. Gender differences in eating habits of Polish young adults aged 20–26. Int J Environ Res Public Health. 2022;19(22):15280.
30. Westenhoefer J. Age and gender dependent profile of food choice. Forum Nutr. 2005;57:44–51. https://doi.org/10.1159/000083753.
31. Manippa V, Padulo C, van der Laan LN, Brancucci A. Gender Differences in Food Choice: Effects of Superior Temporal Sulcus Stimulation. Front Hum Neurosci. 2017;11:597. https://doi.org/10.3389/fnhum.2017.00597.
32. Feraco A, Armani A, Amoah I, Guseva E, Camajani E, Gorini S, Strollo R, Padua E, Caprio M, Lombardo M. Assessing gender differences in food preferences and physical activity: a population-based survey. Front Nutr. 2024;11:1348456. https://doi.org/10.3389/fnut.2024.1348456.
33. Karl JP, Berryman CE, Harris MN, Lieberman HR, Gadde KM, Rood JC, Pasiakos SM. Effects of Testosterone Supplementation on Ghrelin and Appetite During and After Severe Energy Deficit in Healthy Men. J Endocr Soc. 2020;4(4):bvaa024. https://doi.org/10.1210/jendso/bvaa024.
34. Feraco A, Armani A, Gorini S, Camajani E, Quattrini C, Filardi T, Karav S, Strollo R, Caprio M, Lombardo M. Gender Differences in Dietary Patterns and Eating Behaviours in Individuals with Obesity. Nutrients. 2024;16(23):4226.
35. Vandello JA, Bosson JK, Caswell TA, Cummings JR. Healthful eating as a manhood threat. J Mens Health. 2024;20:42–56.
36. Allan CA, McLachlan RI. Androgens and obesity. Curr Opin Endocrinol Diabetes Obes. 2010;17(3):224–32. https://doi.org/10.1097/MED.0b013e3283398ee2.
37. Yeung AWK. Sex differences in brain responses to food stimuli: a meta-analysis on neuroimaging studies. Obes Rev. 2018;19(8):1110–5.
38. Grzymisławska M, Puch EA, Zawada A, Grzymisławski M. (2020). Do nutritional behaviors depend on biological sex and cultural gender? Adv Clin Experimental Med, 29(1).
39. Aljefree NM, Shatwan IM, Almoraie NM. Impact of the Intake of Snacks and Lifestyle Behaviors on Obesity among University Students Living in Jeddah, Saudi Arabia. Healthc (Basel Switzerland). 2022;10(2):400. https://doi.org/10.3390/healthcare10020400.
40. Hall KD, Ayuketah A, Brychta R, Cai H, Cassimatis T, Chen KY, Chung ST, Costa E, Courville A, Darcey V, Fletcher LA, Forde CG, Gharib AM, Guo J, Howard R, Joseph PV, McGehee S, Ouwerkerk R, Raisingher K, Rozga I, Stagliano M, Walter M, Walter PJ, Yang S, Zhou M. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of ad libitum food intake. Cell Metabol. 2019;30(1):67–77.
41. Heitman K, Thomas SE, Kelly O, Fanelli SM, Krok-Schoen JL, Luo M, Taylor CA. Snacks contribute considerably to total dietary intakes among adults stratified by glycemia in the United States. PLOS Global Public Health. 2023;3(10):e0000802.
42. Skoczek-Rubińska A, Bajerska J. The consumption of energy dense snacks and some contextual factors of snacking may contribute to higher energy intake and body weight in adults. Nutr Res. 2021;96:20–36.
43. Peters B, Vahlhaus J, Pivovarova-Ramich O. Meal timing and its role in obesity and associated diseases. Front Endocrinol. 2024;15:1359772.

44. Smith AD, Crippa A, Woodcock J, Brage S. Physical activity and incident type 2 diabetes mellitus: A systematic review and dose-response meta-analysis of prospective cohort studies. Diabetologia. 2020;63(6):1071–84. https://doi.org/10.1007/s00125-020-05172-7.
45. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. Lancet. 2021;380(9838):219–29. https://doi.org/10.1016/S0140-6736(12)61031-9.
46. Stookey JD, Constant F, Popkin BM, Gardner CD. Drinking water is associated with weight loss in overweight dieting women independent of diet and activity. Obes (Silver Spring Md). 2008;16(11):2481–8. https://doi.org/10.1038/oby.2008.409.
47. Pan A, Malik VS, Hu FB. Water intake and risk of obesity: A systematic review and meta-analysis. Obes Rev. 2022;23(4):e13412. https://doi.org/10.1111/obr.13412.
48. Mueller N, Rojas-Rueda D, Cole-Hunter T, De Nazelle A, Dons E, Gerike R, Götschi T, Panis LI, Kahlmeier S, Nieuwenhuijsen M. Health impact assessment of active transportation: A systematic review. Prev Med. 2015;76:103–14. https://doi.org/10.1016/j.ypmed.2015.04.010.
49. Frank LD, Andresen MA, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars. Am J Prev Med. 2021;60(2):207–15. https://doi.org/10.1016/j.amepre.2020.08.015.
50. Wang Y, Beydoun MA, Min J, Xue H, Kaminsky LA, Cheskin LJ. Has the prevalence of overweight, obesity, and central obesity levelled off in the United States? Trends, patterns, disparities, and future projections for the obesity epidemic. Int J Epidemiol. 2020;49(3):810–23. https://doi.org/10.1093/ije/dyz273.
51. Jackson SE, Garnett C, Brown J. Gender differences in smoking cessation and obesity: A longitudinal study. Int J Obes. 2021;45(3):567–75. https://doi.org/10.1038/s41366-020-00723-z.
52. Clair C, Rigotti NA, Porneala B, Fox CS, D'Agostino RB, Pencina MJ, Meigs JB. Association of smoking cessation and weight change with cardiovascular disease among adults with and without diabetes. JAMA. 2019;321(5):458–67. https://doi.org/10.1001/jama.2018.20683.
53. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. The physical activity guidelines for Americans. JAMA. 2018;320(19):2020–8. https://doi.org/10.1001/jama.2018.14854.
54. Twersky SE, Scott T, Taracena L, Souza M. Applying the Social Ecological Model to Understand Support for Home Healthcare Workers and Covid-19 Vaccine Uptake in NJ. Home Health Care Manage Pract. 2024;36(3):217–29.
55. Ohri-Vachaspati P, DeLia D, DeWeese RS, Crespo NC, Todd M, Yedidia MJ. The relative contribution of layers of the Social Ecological Model to childhood obesity. Public Health Nutr. 2015;18(11):2055–66. https://doi.org/10.1017/S1368980014002365.
56. Naidoo D, Meyer-Weitz A, Govender K. The social ecological model: a framework for understanding COVID-19 vaccine uptake among healthcare workers—a scoping review. Vaccines. 2023;11(9):1491.
57. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q. 1988;15(4):351–77.
58. Bronfenbrenner U. The ecology of human development: Experiments by nature and design. Volume 352. Harvard University Press; 1979.
59. Solipaca A, Marino M. (2015). Italian Observatory on Healthcare Report 2015: Health Status and Quality of Care in the Italian Regions. *National Observatory on Health in the Italian Regions, Rome*.
60. Carrillo-Larco RM, Guzman-Vilca WC, Tarazona-Meza C, Xu X. & Bernabe-Ortiz, A. (2025). Prevalence of pre-clinical and clinical obesity in adults: Pooled analysis of 56 population-based national health surveys. PLOS Global Public Health, *5*(7), e0004838.

Figures

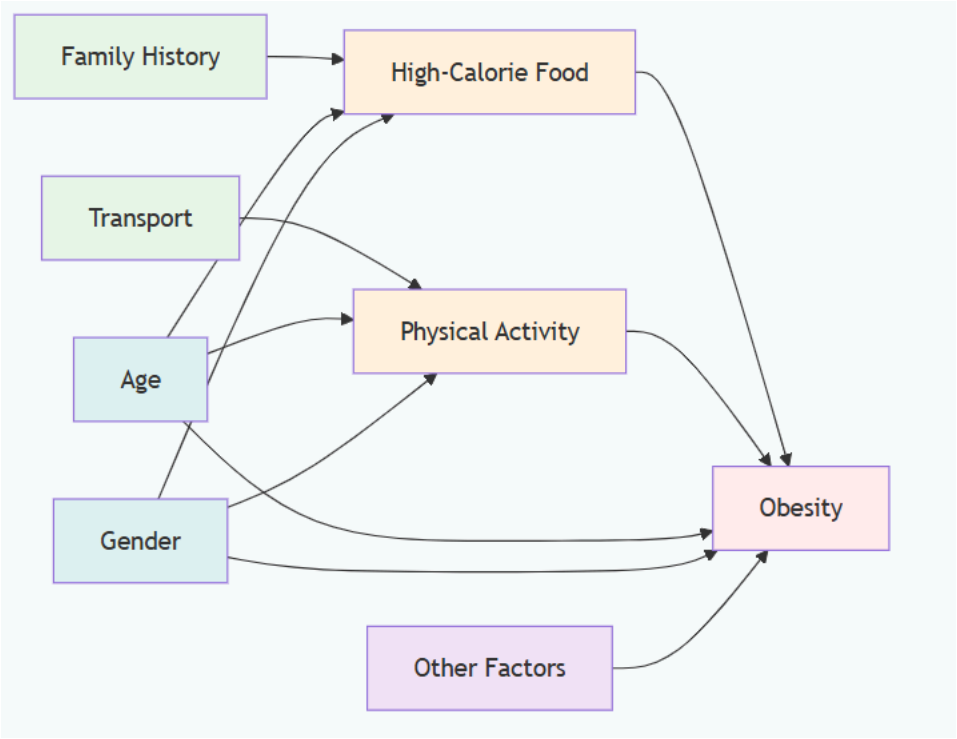


Figure 1

Figure 3. Causal Framework and Directed Acyclic Graph

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ObesitySupplementarymaterial.docx