ECONSTOR Make Your Publications Visible.

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Suyunov, Alisher; Akramov, Kamiljon

Article

Economic AI on the move: The relationship between sugar-sweetened beverages consumption and weight gain in Kyrgyzstan

Journal of Applied Economics

Provided in Cooperation with: University of CEMA, Buenos Aires

Suggested Citation: Suyunov, Alisher; Akramov, Kamiljon (2024) : Economic AI on the move: The relationship between sugar-sweetened beverages consumption and weight gain in Kyrgyzstan, Journal of Applied Economics, ISSN 1667-6726, Taylor & Francis, Abingdon, Vol. 27, Iss. 1, pp. 1-27, https://doi.org/10.1080/15140326.2024.2399457

This Version is available at: https://hdl.handle.net/10419/314293

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



WWW.ECONSTOR.EU

https://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.







Journal of Applied Economics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/recs20

Economic AI on the move: the relationship between sugar-sweetened beverages consumption and weight gain in Kyrgyzstan

Alisher Suyunov & Kamiljon Akramov

To cite this article: Alisher Suyunov & Kamiljon Akramov (2024) Economic AI on the move: the relationship between sugar-sweetened beverages consumption and weight gain in Kyrgyzstan, Journal of Applied Economics, 27:1, 2399457, DOI: <u>10.1080/15140326.2024.2399457</u>

To link to this article: https://doi.org/10.1080/15140326.2024.2399457

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 06 Sep 2024.

Ø

Submit your article to this journal \square

Article views: 490



View related articles 🗹

🌔 View Crossmark data 🗹

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Routledge

Taylor & Francis Group

Economic AI on the move: the relationship between sugar-sweetened beverages consumption and weight gain in **Kyrgyzstan**

Alisher Suyunov (D^a and Kamiljon Akramov (D^b

^aGraduate School, Westminster International University in Tashkent, Tashkent, Uzbekistan; ^bDevelopment Strategy and Governance, International Food Policy Research Institute, Washington DC, USA

ABSTRACT

The prevalence of overweight and obesity has been increasing in Central Asia. Among numerous potential dietary determinants of overweight and obesity, sugar-sweetened beverages (SSBs) have received a great deal of attention because they are a significant source of calories and added sugars in both children and adults. In this study, we examine the relationship between SSB consumption and overweight and obesity in Kyrgyzstan using household-level data from the Kyrgyzstan Integrated Household Survey 2011 using propensity score matching in tandem with machine-learning models. The findings of our study indicate SSBs consumption is associated with 1.6 percentage point higher risk of obesity among both men and women in the context of Kyrgyzstan. Besides, we have observed the positive association between age, income, calorie intake, and share of calorie intake from staples and the likelihood of being obese.

ARTICLE HISTORY

Received 6 September 2023 Accepted 22 August 2024

KEYWORDS

sugar-sweetened beverages; obesity; machine-learning; covariate balance propensity scores

1. Introduction

Four widespread forms of malnutrition - undernourishment, overweight & obesity, and micronutrient deficiencies - coexist in Central Asia (FAO, 2018). While overall malnutrition in the region has improved, overweight among children and obesity among adults continue to rise and now constitute a significant issue (FAO, 2018).

Obesity is widespread around the globe. WHO (2024) estimated 43% and 16% of adult population in the world suffered overweight and obesity in 2022. According to World Bank (Figure 1), in Central Asian countries – Kazakhstan (53%), Turkmenistan (56.9%), Kyrgyzstan (57.8%), Uzbekistan (62.3%), and Tajikistan (58.8%) - overweight and obesity are considerably greater than the world average.

Obesity is associated with several health problems - hypertension, cardiovascular diseases, musculoskeletal disorders, diabetes, depression, and cancer - contributing to death rates (FAO, 2018; Malik & Hu, 2015; Malik et al., 2006, 2010; WHO 2024). As body mass index increases, the risk of those diseases also rises (FAO, 2018).

CONTACT Alisher Suyunov 🖾 alishersuyunov.as@gmail.com 🖃 Graduate School, Westminster International University in Tashkent, Tashkent, Uzbekistan

^{© 2024} The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/ licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



Figure 1. Prevalence of overweight (% of adults) in 2022. Source: WHO (2024), "Prevalence of overweight among adults, BMI \geq 25 (age-standardized estimate) (%)", https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-overweight-among-adults-bmi-=-25-(age-standardized-estimate)-(-).

In addition to adverse effects of obesity on health, research studies emphasized economic consequences of obesity. Obesity worsens labor market outcomes, leading to lower wages and employment probability, and greater healthcare expenditure (Cawley, 2015; Cawley & Meyerhoefer, 2012). Cawley and Meyerhoefer (2012) argued obesity is associated with \$2,741 (in 2005 dollars) higher annual medical care costs in the US, while overweight and obesity cost 4.1% of China's GNP annually (cited in Ecker, 2019).

One of factors contributing to greater than normal body mass index (BMI) is sugarsweetened beverage (SSB) consumption (Burgermaster et al., 2017; Hu, 2013; Malik & Hu, 2015; Malik et al., 2006, 2010; WHO 2024). Hu (2013) and O'Connor et al. (2016) emphasized positive association between SSB intake and high risk of Type 2 Diabetes. Guerrero-López and Arantxa Colchero (2018) found 92.1% of premature death-related loss in productivity attributed to diabetes caused by SSB consumption. Approximately 180,000 deaths annually are ascribed to SSBs consumption—133,000 deaths from diabetes 44,000 from cardiovascular diseases, and 6,000 from cancer (cited in The Heart and Stroke Foundation, 2014). As roughly 80% of individuals suffering from Type 2 Diabetes reside in low- and middle-income countries (Hu, 2013; IFPRI, 2016), steadily increasing overweight and obesity rates in Central Asia are an indicator of emerging public health issue.

Changes in global and national food systems, combined with local socio-cultural and economic environments, seem to drive rising overweight and obesity (Swinburn et al., 2011). Due to heavy marketing campaigns of SSBs throughout the world, consumption of sugar sweetened beverages has increased worldwide, especially in developing economies (Basu et al., 2013; Lobstein, 2014). As a result, SSB sales are growing (Basu et al., 2014; Du et al., 2018; Luger et al., 2017; Sinclair, 2016; WHO, 2014). According to Statista Market Insights (2024), in the US, approximately 195 liters per capita or 63.14bn liters of soft drinks were produced in 2022. Meanwhile, per capita consumption of sparkling drinks in

Central Asia is 40 liters in Kazakhstan, 42 liters in Turkmenistan, 23 liters in Kyrgyzstan, and 9 liters in Tajikistan (Coca-Cola Içecek, 2016).

In this study, we examine the relationship between SSB consumption and overweight and obesity using propensity score matching in tandem with machine learning techniques in the case of Kyrgyzstan. Kyrgyzstan has been experiencing a nutritional transition, characterized by a shift from traditional diets to more Westernized eating patterns, which includes an increased consumption of SSBs. This transition provides a relevant context to study the impact of SSBs on weight gain and obesity in a setting that differs from those of high-income countries where most similar studies have been conducted. The country's diverse socioeconomic landscape, with significant urban-rural variations in income, education, and lifestyle, allows for a comprehensive analysis of these interactions with dietary habits and health outcomes. Findings from Kyrgyzstan can provide insights applicable to other Central Asian and low- to middle-income countries undergoing similar dietary transitions.

In the subsequent sections, we provide a review of the theoretical background on the relationship between sugar-sweetened beverages (SSBs) and weight gain (Section 2). Section 3 outlines the methodology and data sources used in this study, including the Kyrgyzstan Integrated Household Survey and the propensity score matching approach combined with machine learning techniques. Section 4 presents the results of our analysis, followed by a discussion of these findings, their implications for policy and future research, and a discussion of the limitations of the current study.

2. Background

Being sweetened with different sugar forms, SSBs are one of the contributors of added sugars, enhancing normal diets with extra calories (Malik & Hu, 2015; Stacey et al., 2017; U.S. Department of Health and Human Services, and U.S. Department of Agriculture, 2015). Since free sugars are a source of energy without additional nutrients, when they are consumed at high rates, food nutritional quality suffers (U.S. Department of Health and Human Services, and U.S. Department of Health and Human Services, 2015; WHO, 2015).

Food constituents determine the effect of calorie intake on satiation and satiety; this indicates excessive eating may not lead to obesity (Woodward-Lopez et al., 2011). Usually, individuals fail to adjust their physical activities to liquid energy intake from sugar-sweetened beverages as they are less satiating and consumed more frequently (The Heart and Stroke Foundation, 2014; Malik & Hu, 2015; Poppitt, 2015; Te Morenga et al., 2012; Woodward-Lopez et al., 2011). Therefore, roughly 90% of energy remains within a human organism; however, they burn 64% of calorie intake from solid foods through physical activities (Woodward-Lopez et al., 2011). In this case, weight gain from fluids can be greater than from solid foods.

Evidence suggests a positive correlation between sugar-sweetened beverage consumption and weight gain. Malik et al. (2006, 2015), Della Torre et al. (2016), and Luger et al. (2017)'s review showed a positive association between SSB consumption and weight gain among children and adults. Stacey et al. (2017) and Woodward-Lopez et al. (2011), similarly, attributed higher risk of obesity to SSB consumption.

A meta-analysis of regression studies on the nutritional outcomes of SSBs showed that limiting SSB consumption decreases the probability of obesity and weight gain (Hu, 2013). This effect was even stronger for individuals consuming large amounts. Based on a comprehensive analysis of observational and experimental studies, Malik and Hu (2015) concluded that SSBs lead to a dose-dependent increase in weight gain and obesity.

Experimental studies argued that the substitution of SSBs with non-caloric alternatives slows down the increase in BMI. Ebbeling et al. (2012) found that the treatment group that substituted SSBs with non-caloric drinks had a slower BMI increase than the control group after a year. Using quasiexperimental approach, Burgermaster et al. (2017), similarly reported that high SSB intake contributes to an increase in BMI. However, they argued that reducing SSB consumption was crucial but insufficient to combat obesity without adjusting overall consumption patterns.

Other researchers outlined the ambiguity of the relationship. Although in treatment group with higher SSB intake body weight increased, Te Morenga et al. (2012) and Kaiser et al. (2013) pointed up borderline statistical significance of the effect. Kaiser et al. (2013)¹ additionally emphasized there was still equivocal evidence on how consuming SSB affects obesity, thereby, implying a research gap. Similarly, the contribution of SSBs consumption on obesity in the context of low- and middle- income countries is underresearched (Basu et al., 2013).

To investigate how SSB intake affects BMI, Shang et al. (2012), Ebbeling et al. (2012), Laverty et al. (2015), and Burgermaster et al. (2017) controlled for socioeconomic factors, diet, fruit consumption, and exercise habits. To estimate the probability of consuming SSBs, Burgermaster et al. (2017) chose fifteen covariates: gender, age, race, marital status, education level, household income, number of children, household size, a dummy variable indicating if the participant was born in the US, the quantity of fruits and vegetables eaten yesterday, physical exercises apart from regular job during past 30 days, general health level, borough of residence, if the respondent lived in a designated chronic disease neighborhood with a district public health office, and if the respondent discussed body weight with primary care practitioner over the past year. Along with covariates mentioned above, other studies controlled for smoking and dietary factors (Mullie et al., 2012; Hu, 2013; Lee et al., 2017). Mullie et al. (2012) additionally emphasised age, body mass index, income, and dietary patterns to be strong determinants of SSB consumption.

3. Methodology and data

3.1. Data: Kyrgyzstan integrated household survey

Our study uses data from the nationally representative Kyrgyzstan Integrated Household Survey (KIHS). This survey provides detailed information on household consumption and spending, including household structure, education, health status, labor force participation, and expenditures on food and non-food items. Conducted quarterly by the National Statistical Committee of Kyrgyzstan since 2003, the KIHS is based on a stratified two-stage random sampling method derived from the 1999 population census. The sample is divided into 15 strata aligned with the country's administrative divisions (7 regions and Bishkek city) and area types (urban and rural). The dependent variable, a body mass index, is calculated using the respondent's height and weight. We classified individuals' BMI based on the World Health Organization (WHO)'s international classification of underweight, overweight, and obesity: underweight (BMI < 18.50), normal weight (BMI 18.50–24.99), overweight (BMI \ge 25.00, with pre-obesity defined as 25.00–29.99), or obese (BMI \ge 30.00). Based on the classification, we create two binary variables "Overweight" and "Obese" to indicate if the person is overweight or obese.

Our key independent variable is the consumption of sugar-sweetened beverages (SSBs), defined as mineral and soda water with added sugars. It indicates if a respondent consumed a sugar-sweetened beverage over the last 14 days. We do not include fruit juice in this category because, from the dataset, we cannot distinguish between fruit juice with added sugars and 100% organic juice, which is a healthy alternative to SSBs (Malik & Hu, 2015). Since subsequent rounds of the KIHS in 2016 and 2017 did not capture SSB intake, we use the 2011 survey data, which provides the latest information on SSB consumption.

Using 14-day recall data, we calculate total calorie intake per capita and share of food item groups – alcohol, fruit & vegetables, and staples (wheat and wheat products) – in total calorie intake. Calorie intake for each food item is estimated based on the USDA Food Composition Databases, which serve as a relatively accurate proxy for Central Asian food composition tables due to similar nutrition profiles.

3.2. Estimation strategy

In observational studies with non-random treatment assignment, the two groups are not comparable since observed and unobserved variables determine treatment status (Imai & Ratkovic, 2014; Johnson et al., 2018; Rosenbaum & Rubin, 1983; Ross et al., 2015). The validity of results may suffer due to selection bias, leading to an overestimated or underestimated effect (Linden & Yarnold, 2016). A solution to the problem is using Propensity Score Matching (PSM). PSM enables to estimate a treatment effect by balancing on observed covariates to achieve similar baseline characteristics to randomized experiments (Austin, 2009; Imai & Ratkovic, 2014; Johnson et al., 2018; Ramachandra, 2018; Rosenbaum & Rubin, 1983).

According to Imai and Ratkovic (2014), in classic propensity score (PS) model, we assume a parametric PS model to be as follows:

$$\Pr(T_i|X_i) = \pi_\beta(X_i) \tag{1}$$

where β is a dimensional column vector of unknown parameters.

Our outcome of interest is:

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$
(2)

Conditional probabilities of treatment assignment are primarily estimated using the logistic regression model (3) (Andersen & Kurth, 2018; Denkowska, 2017; Lee et al., 2009; Westreich et al., 2010; Xie et al., 2019).

$$\pi_{\beta}(X_i) = \frac{exp(X_i'\beta)}{1 + exp(X_i'\beta)}$$
(3)

To predict treatment assignment accurately, we maximise empirical model fit:

$$\hat{\beta}_{MLE} = \arg \max_{\beta \in \theta} \sum_{i=1}^{N} T_i \log \{ \pi_{\beta}(X_i) \} + (1 - T_i) \log \{ 1 - \pi_{\beta}(X_i) \}$$
(4)

given that the first-order conditions balancing a set of variables hold:

$$\frac{1}{N}\sum_{i=1}^{N}s_{\beta}(T_{i}, X_{i}) = 0s_{\beta}(T_{i}, X_{i}) = \frac{T_{i}\pi'_{\beta}(X_{i})}{\pi_{\beta}(X_{i})} - \frac{(1 - T_{i})\pi'_{\beta}(X_{i})}{1 - \pi_{\beta}(X_{i})}$$
(5)

When conditional independence assumption (6) holds, by conditioning propensity score only, as opposed to high-dimensional vector of variables (Abadie & Imbens, 2016; Guo et al., 2006; Loux, 2015), an unbiased estimate of the average treatment effect on the treated can be obtained (Austin, 2009; Burgermaster et al., 2017; Imai & Ratkovic, 2014; Merz, 2016; Ramachandra, 2018; Rosenbaum & Rubin, 1983).

$$\{Y_i(1), Y_i(0)\} \perp \perp T_i | \pi_\beta(X_i) \tag{6}$$

Since both PSM and regression models assume there are no unobserved relevant variables and control for observed characteristics, theoretically, propensity scores neither underperform or outperform conventional regression models under similar assumptions (Andersen & Kurth, 2018).

Nevertheless, PSM is advantageous over regression models. According to Andersen and Kurth (2018), matching enables immediate observation of analysis details and baseline characteristics between treated and control groups compared to regression models, where result interpretation varies based on the model. Matching disentangles the design and analysis steps of the research, allowing for the estimation of propensity scores, enhancing the PS model match, and improving covariate balance without jumping straight to results (Andersen & Kurth, 2018).

In randomised studies, propensity score is constant across individuals, and treatment assignment is independent of other variables (Loux, 2015). However, in non-randomized studies, true propensity scores are unknown and must be estimated based on data (Abadie & Imbens, 2016; Imai & Ratkovic, 2014; Rosenbaum & Rubin, 1983). How propensity scores are predicted and covariate balance are crucial. To predict treatment status in traditional PS models, the likelihood function is maximized. Due to a trade-off between prediction accuracy and covariate balance (Pirracchio et al., 2015), overlooking covariate balance results in poor balance of covariates (Friedman, 2012; Imai & Ratkovic, 2014).

As it is infeasible to control for all confounders, no theoretical framework specifies variables to be included in the PS model (Imai & Ratkovic, 2014). Since multiple sources contribute to obesity, there is a high risk of PS model misspecification. When the logistic regression model is misspecified, for example, if interactions between variables and non-linear relationships are not considered, the model's bias reduction abilities decline, thereby leading to a biased estimate of treatment effect (Imai & Ratkovic, 2014; Lee et al., 2009; Pirracchio et al., 2015). To overcome these

limitations, Imai and Ratkovic (2014) suggested Covariate Balance Propensity Scores (CBPS), which are more robust to propensity score model misspecification than traditional models. This robustness is achieved by focusing on a covariate balancing score and the conditional probability of treatment assignment (Imai & Ratkovic, 2014). Unlike conventional methods of estimating propensity scores, such as logistic regression, CBPS incorporates the covariate balancing property into propensity score estimation to provide robust and efficient parametric propensity score estimation (Imai & Ratkovic, 2014).

To estimate the average treatment effect on the treated, we control using inverse propensity score weighting. This approach ensures that the weighted covariate distribution of the treatment group is followed. We obtain the following moment condition (7):

$$\mathbb{E}\left\{\mathbb{T}_{\exists}\widetilde{\mathbb{X}_{\exists}} - \frac{\pi_{\beta}(\mathbb{X}_{\exists})(\mathscr{V} - \mathbb{T}_{\exists})\widetilde{\mathbb{X}_{\exists}}}{\mathscr{V} - \pi_{\beta}(\mathbb{X}_{\exists})}\right\} = \mathscr{V}$$
(7)

Imai and Ratkovic (2014) suggested using the following GMM estimator (8):

$$\hat{\beta}_{GMM} = \arg\min_{\beta \in \theta} \bar{g}_{\beta}(T, X)' \Sigma_{\beta}(T, X)^{-1} \bar{g}_{\beta}(T, X)$$
(8)

where $\bar{g}_{\beta}(T, X)$ is a sample mean of moment conditions and $g_{\beta}(T_i, X_i)$ is a combination of all moment conditions.

$$\bar{g}_{\beta}(T,X) = \frac{1}{N} \sum_{i=1}^{N} g_{\beta}(T_i, X_i)$$
(9)

As covariate balancing property is derived from propensity score definition, it does not require conditional independence assumption (6) to hold.

Conventional approaches, such as ordinary least squares and logistic regression, focus on unbiasedness but are weaker in solving prediction problems (Kleinberg et al., 2015). Typical propensity score (PS) models do not consider interactions and higher-order polynomials (Westreich et al., 2010). This implies that non-linearities were assumed not to contribute to the overall model fit, which is a strong assumption (Westreich et al., 2010). Therefore, they suggested finding better alternatives to the logit model.

One option is to use machine learning (ML) models. Since they are robust to sample size, non-additivities, and non-linearities, ML models are relatively better than traditional logistic regression in achieving good covariate balance and estimates (Westreich et al., 2010). Using training data, ML models maximize prediction accuracy while considering the bias-variance trade-off. Unlike conventional models, they can make predictions even if the number of variables exceeds the number of observations (Kleinberg et al., 2015; Westreich et al., 2010). In this case, ML models make fewer assumptions and better capture interactions and non-linearities compared to logistic regression, which relies on strong assumptions (Westreich et al., 2010).

To take advantage of these models, Linden and Yarnold (2016) suggested using a ML model in combination with matching techniques to evaluate treatment effect in observational studies and check consistency of estimates from conventional approaches. Despite difficulties in interpreting ML models (Westreich et al., 2010), they can be used to

estimate propensity scores, as interpretability is not an essential component at the propensity score estimation stage (Lee et al., 2009).

Existing studies have suggested several ML models to estimate the conditional probability of treatment assignment: classification and regression trees (CART), naïve Bayes, neural networks, and ensemble methods. Being insensitive to monotonic transformations and outliers, CART works with categorical, continuous, ordinal, and missing observations and captures interactions and non-linearities (Lee et al., 2009).

Zigler and Dominici (2014) suggested using Bayesian methods to construct propensity score models to estimate the conditional probability of treatment assignment. Unlike regression models that require domain-specific knowledge – such as odds ratios, statistical significance, interaction terms, and confounders – researchers can build neural networks without a complete understanding of the comprehensive network structure (Ayer et al., 2010). Although overweight and obesity are multifactorial issues, neural networks are suitable as a PS model because they detect and model relationships between variables without prior knowledge of the data. In contrast, a logistic regression model clearly (Ayer et al., 2010). Because neural networks provide less biased estimates of propensity scores (Lee et al., 2009; Westreich et al., 2010), previous studies have suggested using neural networks to generate propensity scores (Cannas & Arpino, 2018; Ramachandra, 2018; Setoguchi et al., 2008; Westreich et al., 2010).

Neural networks are composed of input and output layers; if they are enhanced with hidden layers between input and output layers, they are referred to as deep neural networks (DNN) (Cichy & Kaiser, 2019). DNNs learn by identifying the strength of connections between units when trained with data (Chollet & Allaire, 2017; Cichy & Kaiser, 2019). In DNNs, data flows unidirectionally from input to output, and intermediate hidden layers capture features from previous layers (Ramachandra, 2018). However, there are no explicit rules on the number of hidden layers in neural networks (Westreich et al., 2010). Unlike the logistic regression model, DNNs capture non-linearities and non-additivities automatically while estimating propensity scores, thereby automating the covariate balancing process (Cannas & Arpino, 2018).

Unlike other ML models, ensemble methods run multiple samples and trees to increase prediction performance and avoid model overfitting (Lee et al., 2009). In CART decision trees and bagged CART, multiple bootstrap samples are drawn from the data, and for each sample, a separate CART tree is built. Following this, predictions obtained from all samples are averaged to estimate the propensity score (de Vries et al., 2018). In this case, bagged CART decreases variance. While logistic regression models may overestimate the effect in the presence of non-additivities and non-linearities, bagged CART and random forest provide lower bias (Lee et al., 2009). Similarly, Westreich et al. (2010) suggested using random forests to predict the conditional probability of treatment assignment. Although the boosted CART model has relatively poor covariate balance, it provides less biased propensity scores compared to logit (Lee et al., 2009; cited in; Westreich et al., 2010).

Based on existing studies on building ML models to estimate propensity scores, we implemented a supervised ML framework with 10-fold cross-validation to achieve high

accuracy (ROC). We built a DNN with three hidden layers and a 0.2 dropout rate to avoid model overfitting.

Considering the advantages and disadvantages of various approaches, we use three methods to benefit from the strengths of these models and check if our estimate of the average treatment effect on the treated is consistent across different PS models, see Figures A1–A4 in Appendix for the balance of covariates:

- (i) Traditional PSM with a logistic regression model to predict the conditional probability of treatment assignment.
- (ii) PSM with covariate balance propensity scores as suggested by Imai and Ratkovic (2014).
- (iii) PSM in tandem with machine learning methods.

4. Results

Descriptive statistics (Table 1) show that average body mass index (BMI) of respondents is 24.28 kg/m². While 27.5% of respondents are overweight ($25 \le BMI < 30$), 8.91% of respondents are classified as obese ($BMI \ge 30$).

In the sample, 10.8% of respondents confirmed SSB consumption (Table 1), with individuals having secondary and secondary-special education constituting the largest share (Table 2). Despite minority of SSB consumers, ML models account for class imbalance while training the PS model.

Most SSB consumers are individuals aged 15–19 and 30–49 (Table 3), while those within age group 25–29 have the lowest consumption.

Variables	Ν	Mean	Std.Dev	Min – Max
Dependent variables				
Body Mass Index	13,243	24.28	4.069	13.62-55.47
Overweight (Binary)	13,243	0.275	0.447	0–1
Obese (Binary)	13,243	0.0891	0.285	0–1
Independent variables				
SSB (Binary)	13,243	0.108	0.311	0–1
Age	13,243	40.44	17.91	15–105
Urban (Binary)	13,243	0.574	0.495	0–1
Male (Binary)	13,243	0.451	0.498	0–1
Marital Status (Binary)	13,243	0.591	0.492	0–1
Level of education (Tertiary $=$ 4)	13,243	2.463	0.899	0–4
Visited doctor this year (Binary)	13,243	0.4	0.49	0–1
General health status (Very Good = 5)	13,243	3.867	0.672	1–5
Exercise habit (at least once a week) (Binary)	13,243	0.239	0.426	0–1
lsSmoker (Binary)	13,234	0.159	0.366	0–1
#adults in household (≥18)	13,243	2.955	1.233	1–9
Presence of children (<18) (Binary)	13,243	0.332	0.471	0–1
IsRetired (Binary)	13,243	0.212	0.409	0–1
Share of alcohol	13,243	0.00203	0.00655	0-0.117
fruit/vegetables	13,243	0.0775	0.0272	0-0.243
staples in total calorie intake	13,243	0.402	0.109	0-0.785
Total calorie intake (kcal)	13,243	2,338	878.4	478.4 –11,672
Total household expenditure (KGS)	13,188	8,864	13,147	455 –268,271

Table 1. Descriptive statistics of the KIHS survey 2011 Q1.

Source: Authors' estimates based on the dataset.

The share of men and women with overweight status was roughly the same in 2011, at 28% and 27%, respectively (Table 3). However, the obesity rate is higher among females. Top three regions with SSB consumers are Jalal-Abad (26%), Issykkul (24%), and Chui (14%), while the lowest are Talas (5%) and Osh (7%) (Table 4).

The highest proportion of people with overweight status was observed in Bishkek (15%), while Batken and Talas had relatively lower share of overweight (10%) in 2011 (Figure 2). Meanwhile, relatively high rate of obesity is captured in Chui (19%) and Issykkul (16%), with the lowest rate in Talas (6%).

	SS	SB	
	No	Yes	Total
Education level	Col%	Col%	Col%
No education	1.4%	0.9%	1.3%
Primary education	7.1%	6.3%	7.1%
Secondary education	52.4%	55.6%	52.7%
Secondary special education	21.9%	21.9%	21.9%
Tertiary education	17.2%	15.3%	17.0%
Total	100.0%	100.0%	100.0%

Table 2. Education and SSB consumption.

Source: Authors' estimates based on the dataset.

|--|

		SSB – No			SSB – Yes			Total		
Age	Male	Female	Total	Male	Female	Total	Female	Male	Total	
15–19	6.9	6.9	13.8	1.0	1.0	2.0	8.0	7.9	15.8	
20-24	4.6	4.5	9.1	0.6	0.5	1.1	5.0	5.2	10.2	
25-29	3.4	3.1	6.5	0.4	0.4	0.8	3.5	3.9	7.4	
30–49	13.5	17.6	31.1	1.9	2.4	4.2	20.0	15.4	35.4	
50-59	6.3	8.4	14.7	0.7	0.8	1.5	9.2	7.0	16.2	
60+	5.2	8.7	13.9	0.5	0.6	1.2	9.3	5.7	15.0	
Total	40.0	49.2	89.2	5.1	5.7	10.8	54.9	45.1	100.0	

Source: Authors' estimates based on the dataset.

	S	SB	Total
	No	Yes	
Region	Col %	Col %	Col %
Bishkek	14%	8%	13%
Chui	12%	14%	12%
lssyk-kul	12%	24%	13%
Jalal-Abad	13%	26%	14%
Naryn	12%	8%	11%
Batken	11%	8%	11%
Osh	14%	7%	14%
Talas	12%	5%	11%
Total	100%	100%	100%

 Table 4. Sugar-sweetened beverages consumption across regions.



Figure 2. Overweight and obese status across regions of Kyrgyzstan. Source: Authors' estimates based on the dataset (see Table A1 (A and B) in appendix).

4.1. SSB consumption and BMI

A two-sample mean comparison shows no statistically significant difference in BMI between group 1 and the control group (group 2) before matching (Table 5).

Similarly, all PSM results (Table 6) suggest no significant link between SSB consumption and BMI in Kyrgyzstan. The difference in BMI (ATT) between group 1 and group 2 is around 0.1 kg/m^2 , though it is not statistically different from zero at any level. Meanwhile, the ATT estimates from ML models are inconsistent, though deep neural networks and bagged CART models provide closer estimates to those from the logit and CBPS models.

4.2. SSB consumption and overweight

Likewise, in the relationship between SSB intake and BMI, a two-sample mean comparison shows (Table 7) no statistically significant difference in overweight.

PSM results similarly suggest no significant link between SSB consumption and overweight in Kyrgyzstan (Table 8).

	1 1		,	
Control group – Group 2 (SSB = 0)	Group 1 (SSB = 1)	Difference	t-statistics	<i>p</i> -value
24.282	24.267	-0.016	-0.134	0.893

Table 5. Two sample mean comparison (SSB consumption & BMI).

	Group 2	Group 1				
Model	(SSB = 0)	(SSB = 1)	ATT	t-statistics	<i>p</i> -value	Accuracy
Logit	24.167	24.267	0.100	0.749	0.454	-
CBPS	24.160	24.267	0.106	0.797	0.426	-
Machine-learning models						
Deep Neural Network	24.082	24.267	0.184	1.383	0.167	0.809
CART	24.378	24.267	-0.112	-0.874	0.382	0.766
Bagged CART	24.155	24.267	0.111	0.870	0.385	0.934
Naive Bayes	23.996	24.267	0.271**	2.147	0.032	0.789
Random Forest	24.228	24.267	0.038	0.297	0.766	0.941

Tal	ble	6.	PSM	results	(SSB	consum	ption	&	BMI)
-----	-----	----	-----	---------	------	--------	-------	---	-----	---

All models control for individual, household, location characteristics and regional fixed effects. *p < 0.1; **p < 0.05; ***p < 0.05; **p <0.01.

Table 7. Two-sample mean comparison (SSB consumption & overweight).

Group 2 (SSB = 0)	Group 1 (SSB = 1)	Difference	t-statistics	<i>p</i> -value
0.277	0.263	-0.014	-1.143	0.253

Table 8. PSM results (SSB consumption & overweight).

Model	Group 2 (SSB = 0)	Group 1 (SSB = 1)	ATT	t-statistics	<i>p</i> -value	Accuracy
Logit	0.275	0.263	-0.012	-0.870	0.385	-
CBPS	0.264	0.263	-0.002	-0.109	0.913	-
Machine-learning models						
Deep Neural Network	0.264	0.263	-0.001	-0.085	0.933	0.809
CART	0.286	0.263	-0.023*	-1.736	0.083	0.766
Bagged CART	0.267	0.263	-0.005	-0.363	0.717	0.934
Naive Bayes	0.267	0.263	-0.005	-0.346	0.730	0.789
Random Forest	0.275	0.263	-0.012	-0.897	0.370	0.941

All models control for individual, household, location characteristics and regional fixed effects. *p < 0.1; **p < 0.05; ***p < 0.01.

4.3. SSB consumption and obesity

A two-sample mean comparison of unmatched data shows no statistically significant difference (1.3 pp) in obesity between group 1 and the control group (group 2) (Table 9).

In contrast to results we observed so far, the estimates suggest a statistically significant link between SSB consumption and obesity in Kyrgyzstan. As presented in Tables 10 and 11, SSB consumption is associated with 1.5–3.1 pp. increase in obesity rates among individuals.

As mentioned earlier, logistic regression model (Table 10) is prone to model misspecification. Therefore, we compare estimates across models to check for consistency. The results (Table 11) show that the magnitude of the association in ML models converges to 1.6 percentage points, which is similar to the CBPS model estimate. In

Table 9. Two-sample mean comparison (SSB consumption & obesity).							
Group 2 (SSB = 0)	Group 1 (SSB = 1)	Difference	t-statistics	<i>p</i> -value			
0.088	0.101	0.013	1.591	0.112			

Table O Two complements of comparison (CCD comparation 8 abovity)

Variable	Obese
SSB	0.0212***
	(0.0076)
Share of alcohol in total calorie intake	0.3705
	(0.3472)
fruit/vegetables in total calorie intake	-0.1264
	(0.0948)
staples in total calorie intake	-0.0711***
	(0.0263)
Log of Total calorie intake	0.0059
A	(0.008)
Age	(0.0024"""
Marital status	(0.0002)
	(0.0056)
Presence of children	0.0030)
	(0.004
Education level (Primary, reference cat=No education)	0.0367*
Education level (Initialy, reference car no education)	(0.0202)
Education level (Secondary)	0.0285
	(0.0176)
Education level (Secondary special)	0.0314*
	(0.0179)
Education level (Tertiary)	0.0306*
	(0.0182)
Exercise habit (at least once a week)	-0.0093
	(0.0083)
IsMale	-0.046***
	(0.0064)
General health status (Bad, reference cat=Very bad)	0.0193
Consult has the status (Augus as)	(0.0469)
General health status (Average)	-0.042
Conoral health status (Good)	(0.0455)
General fieditif status (GOOG)	-0.0331
General health status (Very good)	-0.0507
General nearth status (very good)	(0.0463)
#adults in household	0.0003
	(0.0024)
IsRetired	-0.0027
	(0.0074)
IsSmoker	-0.0285***
	(0.0083)
Log of household expenditure	0.0065**
	(0.0029)
Visited doctor this year	0.0113**
	(0.0055)
IsUrban	0.0014
	(0.0054)
Regional fixed effects	Yes
Observations	13,243
Log-likelihood	-3,329.4
Akaike Inf. Crit.	6,724.7

Table 10. Average marginal effects of logistic regression model.

Standard errors are in parentheses. *indicates p < 0.1; **p < 0.05; ***p < 0.01.

addition to being statistically significant at least at the 10% level, all models provide consistent estimates in terms of both sign and magnitude.

As we extracted the contribution of covariates from ML models, the main predictors of the conditional probability of treatment assignment in the bagged CART model are the log of household expenditure, the log of total calorie intake, the share of fruit/vegetables

Model	Group 2 (SSB = 0)	Group 1 (SSB = 1)	ATT	t-statistics	<i>p</i> -value	Accuracy
Logit	0.079	0.101	0.022**	2.382	0.017	-
CBPS	0.085	0.101	0.016*	1.700	0.089	-
Machine-learning models						
Deep Neural Network	0.082	0.101	0.019**	2.054	0.040	0.809
CART	0.085	0.101	0.016*	1.776	0.076	0.766
Bagged CART	0.084	0.101	0.016*	1.829	0.068	0.934
Naive Bayes	0.070	0.101	0.031***	3.541	0.000	0.789
Random Forest	0.086	0.101	0.015*	1.698	0.090	0.941

Table 11. PSM results (SSB consumption & obesity).

All models control for individual, household, location characteristics and regional fixed effects. *p < 0.1; **p < 0.05; ***p < 0.01.

in total calorie intake, and the share of staples in total calorie intake (Figure 3). Surprisingly, the results suggest (Figures 3 and 4) that the share of alcohol in total calorie intake is one of the most important variables for predicting propensity scores.

As ML models indicate an association between independent variables and SSB consumption but do not imply the existence of a causal relationship, ML outputs should be interpreted cautiously. The deep neural network output (Figure 4) is more informative in terms of the direction of the relationship. While the share of fruit/vegetables, staples, and smoking are negatively associated with a higher probability of consuming SSBs,



Figure 3. Feature importance in a machine-learning model – bagged CART.



Influencing covariates

Figure 4. Influencing covariates. The assigned weights for SSB predictors in deep neural networks.

household expenditure, the presence of children in the household, and alcohol consumption are positively associated with SSB intake.

While education level plays a negligible role in the ML-based PS model, the CBPS model suggests no statistically significant association between education and SSB consumption, except for secondary special education (see Table A3 in appendix). Similarly, we found no statistically significant evidence at any level to support the association between gender and the probability of SSB consumption, which contradicts the findings of Rao et al. (2015).

5. Discussion

Being associated with adverse health and socioeconomic outcomes, the prevalence of overweight and obesity is increasing in Central Asia. Our study indicates that SSB intake is related to a higher risk of obesity among adults in Kyrgyzstan. In general, the findings suggest a 1.6 pp. increase in obesity rates associated with SSB consumption in Kyrgyzstan.

The key results corroborate the findings of other studies in different contexts, particularly those of Shin et al. (2018) among adult South Koreans and Burgermaster et al. (2017) among New Yorkers. However, we could not find strong evidence supporting the relationship between SSB consumption and BMI or overweight.

The results of the PS models (logit and CBPS) suggest that household expenditure, as a proxy for income, is positively associated with the probability of consuming SSBs. This finding, however, contrasts with the studies by Han and Powell (2013) and Burgermaster et al. (2017). The discrepancy in Kyrgyzstan's context implies that SSBs are more popular among people with higher household income, as evidenced by the higher percentage of SSB consumption observed in the upper expenditure quantiles (see Table A3 in the appendix). Additionally, the presence of children in the household and alcohol consumption are also positively associated with SSB intake.

Healthy eating patterns, including the consumption of fruits, vegetables, and staples, are negatively associated with SSB intake, which aligns with the findings of Mullie et al. (2012), Hu (2013), and Rao et al. (2015). Similarly, total calorie intake is negatively related to the probability of consuming SSBs. In contrast to the results of Mullie et al. (2012), Rao et al. (2015) in the United States, and Lee et al. (2017) in China, we found that smoking is negatively associated with SSB consumption in Kyrgyzstan.

The study suggests a significant relationship between SSB intake and obesity in Kyrgyzstan. Existing studies similarly imply that empty calories from SSB consumption should be constrained by limiting SSB intake to combat increasing obesity rates (Burgermaster et al., 2017; Hartman et al., 2017; Hsiao & Claire Wang, 2013; Hu, 2013). Although the findings of our study are insufficient to identify a strong causal relationship between SSB intake and obesity in Kyrgyzstan, they align with other relevant research, suggesting that reducing SSB consumption may be a valuable strategy in addressing the growing public health challenges.

A broad range of studies have suggested policy implications to control SSB consumption, such as implementing SSB warning labels to reshape people's perception and reduce SSB intake (The Heart and Stroke Foundation, 2014; Gollust et al., 2014; VanEpps & Roberto, 2016; Popova, 2016; Lee et al., 2018); increasing public awareness of the negative effects of SSBs (Du et al., 2018; Kansagra et al., 2015; Rao et al., 2015); regulating marketing campaigns promoting SSBs (Popkin & Hawkes, 2016; WHO, 2019); and imposing a consumption tax on SSBs (Battakova et al., 2017; Cawley, 2015; Hsiao & Claire Wang, 2013; Jou & Techakehakij, 2012; Park & Yu, 2019; Rao et al., 2015). The latter measure, being adopted in several countries including the United States, Colombia, Mexico, and EU countries (The Heart and Stroke Foundation, 2014), has been an effective way to decrease obesity prevalence at the population level, especially among low-income groups in Colombia (Vecino-Ortiz & Arroyo-Ariza, 2018), the United States (Andreyeva et al., 2011; Lin et al., 2011; Long et al., 2015), South Africa (Blecher, 2015), and Mexico (cited in Blecher, 2015; Popkin & Hawkes, 2016). However, SSB taxation may not be effective when misused (Hsiao & Claire Wang, 2013) or in countries with high tax rates (The Heart and Stroke Foundation, 2014; Jou & Techakehakij, 2012). Due to ongoing fiscal reforms and current tax rates in Central Asia, taxing SSB consumption can be a last resort

to decrease obesity rates. Since the effectiveness of one policy option over another is beyond the scope of this paper, we will leave the cost-benefit analysis of policy options to be discussed in further studies.

Although in the study we addressed different sources of bias and enhanced insensitivity to PS model misspecification, a few unresolved issues remain. We used food items captured at the household level and extrapolated for all household members, which may induce a measurement error in the variables' magnitude. Similarly, the study considers SSB consumption at home, overlooking instances when SSBs are consumed outside, indicating another source of bias. We would obtain more accurate estimates if outside SSB consumption were considered. However, neither individual-level data on food nor out-of-home consumption is available for Kyrgyzstan. Our study would also benefit in terms of policy implications if, in addition to statistical significance, the economic significance of the issue in Kyrgyzstan was considered. This aspect, however, was beyond the objectives of our research.

Acknowledgement

This research was supported by the CGIAR Research Initiative on Fragility to Resilience in Central and West Asia and North Africa (CWANA).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Alisher Suyunov correspondent author, is an applied economist and data scientist. He is currently a PhD candidate in Econometrics and Statistics at Westminster International University in Tashkent and works as a Senior Consultant at PwC Uzbekistan. His research interests include econometrics and artificial intelligence, health and nutrition economics, labor economics, and development policy. He is also actively engaged in sustainability and climate change topics and climate economic modeling.

Kamiljon Akramov PhD, Senior Research Fellow in International Food Policy Research Institute's Development Strategies and Governance Unit. He conducts research on agriculture, food security, and development policy issues using applied econometric analysis. His primary areas of interest are institutional change, agricultural development, and food security in Central Asia. His research also focuses on governance, aid effectiveness, decentralization, and efficiency of service delivery in transitional and developing countries.

ORCID

Alisher Suyunov () http://orcid.org/0000-0002-2782-2691 Kamiljon Akramov () http://orcid.org/0000-0003-4357-0508

References

- Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84 (2), 781–807. https://doi.org/10.3982/ECTA11293
- Andersen, L. W., & Kurth, T. (2018). Propensity scores a brief introduction for resuscitation researchers. *Resuscitation*, 125(January), 66–69. https://doi.org/10.1016/j.resuscitation.2018.02.003
- Andreyeva, T., Chaloupka, F. J., & Brownell, K. D. (2011). Estimating the potential of taxes on sugar-sweetened beverages to reduce consumption and generate revenue. *Preventive Medicine*, 52(6), 413–416. https://doi.org/10.1016/j.ypmed.2011.03.013
- Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and monte carlo simulations. *Biometrical Journal*, 51(1), 171–184. https://doi.org/10.1002/bimj.200810488
- Ayer, T., Chhatwal, J., Alagoz, O., Kahn, C. E., Woods, R. W., & Burnside, E. S. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *RadioGraphics*, 30(1), 13–22. https://doi.org/10.1148/rg.301095057
- Basu, S., McKee, M., Galea, G., & Stuckler, D. (2013). Relationship of soft drink consumption to global overweight, obesity, and diabetes: A cross-national analysis of 75 countries. *American Journal of Public Health*, 103(11), 2071–2077. https://doi.org/10.2105/AJPH.2012.300974
- Basu, S., Sukumar, V., Sutapa, A., David, S., Barry, P., Shah, E., & Blakely, T. (2014). Averting obesity and type 2 diabetes in India through sugar-sweetened beverage taxation: An economic-epidemiologic modeling study." *PLOS Medicine*, 11(1), e1001582. https://doi.org/ 10.1371/journal.pmed.1001582
- Battakova, J. E., Mukasheva, S. B., Slazhneva, T. I., Abdrakhmanov, S. Z., Buonkristiano, M., Adaeva, A. A., & Akimbayeva, A. A. (2017). Эпидемиологический мониторинг детского ожирения и факторов, его формирующих, в республике казахстан, 2015-2016 гг. *Almaty.* https://hls.kz/uploads/scientific-facts/rus/cosi-nac-otchet-detskoe-ojirenie.pdf
- Blecher, E. (2015). Taxes on tobacco, alcohol and sugar sweetened beverages: Linkages and lessons learned. *Social Science & Medicine*, 136–137(July), 175–179. https://doi.org/10.1016/j.socs cimed.2015.05.022
- Burgermaster, M., Bhana, H., Dot Fullwood, M., Luna Bazaldua, D. A., & Tipton, E. (2017). Exploring the role of sugar-sweetened beverage consumption in obesity among new yorkers using propensity score matching. *Journal of the Academy of Nutrition and Dietetics*, 117(5), 753–762. https://doi.org/10.1016/j.jand.2017.01.022
- Cannas, M., & Arpino, B. (2018). Machine learning for propensity score matching and weighting: Comparing different estimation techniques and assessing different balance diagnostics. Working Paper of Research and Expertise Centre for Survey Methodology Centre of Pompeu Fabra University. https://doi.org/10.13140/RG.2.2.14365.00485
- Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43(September), 244–268. https://doi.org/ 10.1016/j.jhealeco.2015.03.001
- Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics*, *31*(1), 219–230. https://doi.org/10.1016/j.jhealeco.2011. 10.003
- Chollet, F., & Allaire, J. J. (2017). Deep learning with R online. Manning Publications Company.
- Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, 23(4), 305–317. https://doi.org/10.1016/j.tics.2019.01.009
- Coca-Cola Içecek. (2016). CCI sustainability report 2016. http://www.cci.com.tr/Portals/3/images/ CCI_2016_CSR.pdf
- Della Torre, B., Sophie, A. K., Laure Depeyre, J., & Kruseman, M. (2016). Sugar-sweetened beverages and obesity risk in children and adolescents: A systematic analysis on how methodological quality may influence conclusions. *Journal of the Academy of Nutrition and Dietetics*, *116*(4), 638–659. https://doi.org/10.1016/j.jand.2015.05.020

- Denkowska, S. (2017). Assessing the robustness to an unobserved confounder of the average treatment effect on the treated estimated by propensity score matching. *Argumenta Oeconomica Cracoviensia*, (15), 57–74. https://doi.org/10.15678/aoc.2016.1504
- de Vries, P., Bas, B. L., & van Smeden, M. (2018). Propensity score estimation using classification and regression trees in the presence of missing covariate data. *Epidemiologic Methods*, 7(1), 1–18. https://doi.org/10.1515/em-2017-0020
- Du, M., Tugendhaft, A., Erzse, A., & Hofman, K. J. (2018). Sugar-sweetened beverage taxes: Industry response and tactics. *The Yale Journal of Biology and Medicine*, *91*(2), 185–190.
- Ebbeling, C. B., Feldman, H. A., Chomitz, V. R., Antonelli, T. A., Gortmaker, S. L., Osganian, S. K., & Ludwig, D. S. (2012). A randomized trial of sugar-sweetened beverages and adolescent body weight. *New England Journal of Medicine*, 367(15), 1407–1416. https://doi.org/10.1056/ NEJMoa1203388
- Ecker, O. (2019). Reshaping agriculture to reduce obesity. In S. Fan, S. Yosef, & R. Pandya-Lorch (Eds.), *Agriculture for improved nutrition: Seizing the momentum* (pp. 81–92). IFPRI & CAB International. http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/133070/filename/ 133281.pdf
- Esenaliev, D., Kröger, A., & Steiner, S. (2011). *The Kyrgyz Integrated Household Survey (KIHS):* A primer. Deutsches Institut für Wirtschaftsforschung (DIW). http://hdl.handle.net/10419/129271
- FAO. (2018). Regional overview of food security and nutrition Europe and Central Asia the role of migration, rural women and youth in sustainable development. http://www.fao.org/3/CA2703EN/CA2703EN.pdf
- Friedman, J. (2012). Tools of the trade: The covariate balanced propensity score. *World Bank Blogs: Impact Evaluation, 2012.* https://blogs.worldbank.org/impactevaluations/tools-of-the-trade-the -covariate-balanced-propensity-score
- Gollust, S. E., Barry, C. L., & Niederdeppe, J. (2014). Americans' opinions about policies to reduce consumption of sugar-sweetened beverages. *Preventive Medicine*, 63(6), 52–57. https://doi.org/ 10.1016/j.ypmed.2014.03.002
- Guerrero-López, C. M., & Arantxa Colchero, M. (2018). Productivity loss associated with the consumption of sugar-sweetened beverages in Mexico. *Preventive Medicine*, *115*(7), 140–144. https://doi.org/10.1016/j.ypmed.2018.08.014
- Guo, S., Barth, R. P., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children & Youth Services Review*, 28(4), 357–383. https://doi.org/10.1016/j.childyouth.2005.04.012
- Han, E., & Powell, L. M. (2013). Consumption patterns of sugar-sweetened beverages in the United States. *Journal of the Academy of Nutrition and Dietetics*, *113*(1), 43–53. https://doi.org/10.1016/j.jand.2012.09.016
- Hartman, T., Haardörfer, R., Greene, B., Parulekar, S., & Kegler, M. (2017). Beverage consumption patterns among overweight and obese African American women. *Nutrients*, 9(12), 1344. https://doi.org/10.3390/nu9121344
- The Heart and Stroke Foundation. (2014). Sugar, heart disease and stroke. *The Heart and Stroke Foundation of Canada*. http://www.heartandstroke.com/atf/cf/%7B99452D8B-E7F1-4BD6-A57D-B136CE6C95BF%7D/Sugar-Eng.pdf
- Hsiao, A., & Claire Wang, Y. (2013). Reducing sugar-sweetened beverage consumption: Evidence, policies, and economics. *Current Obesity Reports*, 2(3), 191–199. https://doi.org/10.1007/s13679-013-0065-8
- Hu, F. B. (2013). Resolved: There is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. *Obesity Reviews*, 14(8), 606–619. https://doi.org/10.1111/obr.12040
- Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 243–263. https://doi.org/10.1111/ rssb.12027
- Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., & Feldman, B. M. (2018). Propensity score methods for bias reduction in observational studies of treatment effect.

Rheumatic Disease Clinics of North America, 44(2), 203–213. https://doi.org/10.1016/j.rdc.2018. 01.002

- Jou, J., & Techakehakij, W. (2012). International application of sugar-sweetened beverage (SSB) taxation in obesity reduction: Factors that may influence policy effectiveness in country-specific contexts. *Health Policy*, 107(1), 83–90. https://doi.org/10.1016/j.healthpol.2012.05.011
- Kaiser, K. A., Shikany, J. M., Keating, K. D., & Allison, D. B. (2013). Will reducing sugar-sweetened beverage consumption reduce obesity? Evidence supporting conjecture is strong, but evidence when testing effect is weak. *Obesity Reviews*, 14(8), 620–633. https://doi.org/10.1111/obr.12048
- Kansagra, S. M., Kennelly, M. O., Nonas, C. A., Curtis, C. J., Van Wye, G., Goodman, A., & Farley, T. A. (2015). Reducing sugary drink consumption: New York City's approach. *American Journal of Public Health*, 105(4), e61–64. https://doi.org/10.2105/ajph.2014.302497
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *The American Economic Review*, *105*(5), 491–495. https://doi.org/10.1257/aer.p20151023
- Laverty, A. A., Magee, L., Monteiro, C. A., Saxena, S., & Millett, C. (2015). Sugar and artificially sweetened beverage consumption and adiposity changes: National longitudinal study. *The International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 137. https://doi.org/ 10.1186/s12966-015-0297-y
- Lee, B. K., Lessler, J., & Stuart, E. A. (2009). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3), 337–346. https://doi.org/10.1002/sim.3782
- Lee, B. Y., Ferguson, M. C., Hertenstein, D. L., Adam, A., Zenkov, E., Wang, P. I., Wong, M. S., Gittelsohn, J., Mui, Y., & Brown, S. T. (2018). Simulating the impact of sugar-sweetened beverage warning labels in three cities. *American Journal of Preventive Medicine*, 54(2), 197–204. https://doi.org/10.1016/j.amepre.2017.11.003
- Lee, Y. H., Wang, Z., Chiang, T. C., & Ti Liu, C. (2017). Beverage intake, smoking behavior, and alcohol consumption in contemporary China—A cross-sectional analysis from the 2011 China health and nutrition survey. *International Journal of Environmental Research & Public Health*, 14(5), 1–11. https://doi.org/10.3390/ijerph14050493
- Lin, B. H., Smith, T. A., Lee, J. Y., & Hall, K. D. (2011). Measuring weight outcomes for obesity intervention strategies: The case of a sugar-sweetened beverage tax. *Economics & Human Biology*, 9(4), 329–341. https://doi.org/10.1016/j.ehb.2011.08.007
- Linden, A., & Yarnold, P. R. (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 22 (6), 868–874. https://doi.org/10.1111/jep.12592
- Lobstein, T. (2014). Reducing consumption of sugar-sweetened beverages to reduce the risk of childhood overweight and obesity. *World Health Organisation*, 2014. https://www.who.int/elena/bbc/ssbs_childhood_obesity/en/
- Long, M. W., Gortmaker, S. L., Ward, Z. J., Resch, S. C., Moodie, M. L., Sacks, G., Swinburn, B. A., Carter, R. C., & Claire Wang, Y. (2015). Cost effectiveness of a sugar-sweetened beverage excise tax in the U.S. American Journal of Preventive Medicine, 49(1), 112–123. https://doi.org/10. 1016/j.amepre.2015.03.004
- Loux, T. M. (2015). Randomization, matching, and propensity scores in the design and analysis of experimental studies with measured baseline covariates. *Statistics in Medicine*, *34*(4), 558–570. https://doi.org/10.1002/sim.6361
- Luger, M., Lafontan, M., Bes-Rastrollo, M., Winzer, E., Yumuk, V., & Farpour-Lambert, N. (2017). Sugar-sweetened beverages and weight gain in children and adults: A systematic review from 2013 to 2015 and a comparison with previous studies. *Obesity Facts*, *10*(6), 674–693. https://doi. org/10.1159/000484566
- Malik, V. S., & Hu, F. B. (2015). Fructose and cardiometabolic health. *Journal of the American College of Cardiology*, 66(14), 1615–1624. https://doi.org/10.1016/j.jacc.2015.08.025
- Malik, V. S., Popkin, B. M., Bray, G. A., Despres, J. P., Willett, W. C., & Hu, F. B. (2010). Sugarsweetened beverages and risk of metabolic syndrome and type 2 diabetes: A meta-analysis. *Diabetes Care*, 33(11), 2477–2483. https://doi.org/10.2337/dc10-1079

- Malik, V. S., Schulze, M. B., & Hu, F. B. (2006). Intake of sugar-sweetened beverages and weight gain: A systematic Review 1–3. *The American Journal of Clinical Nutrition*, 84(2), 274–288. https://doi.org/10.1093/ajcn/84.2.274
- Merz, J. (2016). Policy evaluation in a nutshell. *Leuphana University of Lüneburg Working Paper*, 1–41. https://www.leuphana.de/fileadmin/user_upload/PERSONALPAGES/_mno/merz_joa chim/files/Merz_Policy_evaluation_in_a_nutshell.pdf
- Mullie, P., Aerenhouts, D., & Clarys, P. (2012). Demographic, socioeconomic and nutritional determinants of daily versus non-daily sugar-sweetened and artificially sweetened beverage consumption. *European Journal of Clinical Nutrition*, 66(2), 150–155. https://doi.org/10.1038/ejcn.2011.138
- O'Connor, L., Ye, Z., Imamura, F., Hayashino, Y., Bhupathiraju, S. N., Forouhi, N. G., & Mursu, J. (2016). Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: Systematic review, meta-analysis, and estimation of population attributable fraction. *British Journal of Sports Medicine*, 50(8), 496–504. https://doi. org/10.1136/bjsports-2016-h3576rep
- Olivier, E., Perrihan, A.-R., Clemens, B., & Rawia, E.-B. (2016). Nutrition and economic development: Exploring Egypt's exceptionalism and the role of food subsidies. Washington, DC: International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/9780896292383
- Park, H., & Yu, S. (2019). Policy review: Implication of tax on sugar-sweetened beverages for reducing obesity and improving heart health. *Health Policy and Technology*, 8(1), 92–95. https:// doi.org/10.1016/j.hlpt.2018.12.002
- Pirracchio, R., Petersen, M. L., & van der Laan, M. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. *American Journal of Epidemiology*, 181(2), 108–119. https://doi.org/10.1093/aje/kwu253
- Popkin, B. M., & Hawkes, C. (2016). Sweetening of the global diet, particularly beverages: Patterns, trends, and policy responses. *The Lancet Diabetes & Endocrinology*, 4(2), 174–186. https://doi.org/10.1016/S2213-8587(15)00419-2
- Popova, L. (2016). Sugar-sweetened beverage warning labels: Lessons learned from the tobacco industry. *Journal of the California Dental Association*, 44(10), 633–640. https://doi.org/10.1080/19424396.2016.12221070
- Poppitt, S. (2015). Beverage consumption: Are alcoholic and sugary drinks tipping the balance towards overweight and obesity? *Nutrients*, 7(8), 6700–6718. https://doi.org/10.3390/ nu7085304
- Ramachandra, V. (2018, February). Deep learning for causal inference. http://arxiv.org/abs/1803. 00149
- Rao, G., Kirley, K., Weiss-Coleman, R., Jeffrey Inman, J., Bauer, V., Zhou, Y., & Hledin, V. (2015). Consumption patterns of sugar-sweetened carbonated beverages among children and adolescents. *Current Cardiovascular Risk Reports*, 9(4), 17. https://doi.org/10.1007/s12170-015-0445-6
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41. https://doi.org/10.1093/biomet/70.1.41
- Ross, M. E., Kreider, A. R., Huang, Y. S., Matone, M., Rubin, D. M., & Russell Localio, A. (2015). Propensity score methods for analyzing observational data like randomized experiments: Challenges and solutions for rare outcomes and exposures. *American Journal of Epidemiology*, 181(12), 989–995. https://doi.org/10.1093/aje/kwu469
- Setoguchi, S., Schneeweiss, S., Alan Brookhart, M., Robert, J. G., & Francis Cook, E. (2008). Evaluating uses of data mining techniques in propensity score estimation: A simulation study. *Pharmacoepidemiology and Drug Safety*, *17*(6), 546–555. https://doi.org/10.1002/pds.1555
- Shang, X. W., Ling Liu, A., Zhang, Q., Qi Hu, X., Ming Du, S., Ma, J., & Fa Xu, G. (2012). Report on childhood obesity in China (9): Sugar-sweetened beverages consumption and obesity. *Biomedical and Environmental Sciences*, 25(2), 125–132. https://doi.org/10.3967/0895-3988. 2012.02.001
- Shin, S., Kim, S. A., Ha, J., & Lim, K. (2018). Sugar-sweetened beverage consumption in relation to obesity and metabolic syndrome among Korean adults: A cross-sectional study from the 2012–

2016 Korean national health and nutrition examination survey (KNHANES). *Nutrients*, *10*(10), 1467. https://doi.org/10.3390/nu10101467

- Sinclair, B. (2016). Sugary drink companies target low & middle-income countries. *World Cancer Research Fund International*, 2016. https://www.wcrf.org/int/blog/articles/2016/02/sugary-drink-companies-target-low-middle-income-countries
- Stacey, N., Tugendhaft, A., & Hofman, K. (2017). Sugary beverage taxation in South Africa: Household expenditure, demand system elasticities, and policy implications. *Preventive Medicine*, 105(December), S26-31. https://doi.org/10.1016/j.ypmed.2017.05.026
- Statista Market Insights. (2024). Soft Drinks United States, Volume per Capita, Combined. Statista. https://www.statista.com/outlook/cmo/non-alcoholic-drinks/soft-drinks/united-states#volume
- Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L., & Gortmaker, S. L. (2011). The global obesity pandemic: Shaped by global drivers and local environments. *Lancet*, 378(9793), 804–814. https://doi.org/10.1016/S0140-6736(11)60813-1
- Te Morenga, L., Mallard, S., & Mann, J. (2012). Dietary sugars and body weight: Systematic review and meta-analyses of randomised controlled trials and cohort studies. *BMJ*, 346(3), e7492–e7492. https://doi.org/10.1136/bmj.e7492
- U.S. Department of Health and Human Services, and U.S. Department of Agriculture. (2015). 2015–2020 dietary guidelines for Americans. http://health.gov/dietaryguidelines/2015/guidelines/
- VanEpps, E. M., & Roberto, C. A. (2016). The influence of sugar-sweetened beverage warnings: A randomized trial of adolescents' choices and beliefs. *American Journal of Preventive Medicine*, 51(5), 664–672. https://doi.org/10.1016/j.amepre.2016.07.010
- Vecino-Ortiz, A. I., & Arroyo-Ariza, D. (2018). A tax on sugar sweetened beverages in Colombia: Estimating the impact on overweight and obesity prevalence across socio economic levels. *Social Science & Medicine*, 209(7), 111–116. https://doi.org/10.1016/j.socscimed.2018.05.043
- Westreich, D., Lessler, J., & Jonsson Funk, M. (2010). Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, 63(8), 826–833. https://doi.org/10.1016/j. jclinepi.2009.11.020
- WHO. (2014). Reducing consumption of sugar-sweetened beverages to reduce the risk of unhealthy weight gain in adults. *World Health Organization*. https://www.who.int/elena/bbc/ssbs_adult_weight/en/
- WHO. (2015). Guideline: Sugars intake for adults and children. 26(4), 34-36.
- WHO. (2019). *Monitoring food and beverage marketing to children via television in the Republic of Kazakhstan*. http://www.euro.who.int/__data/assets/pdf_file/0011/396389/WHO-Food-marketing-kazakhstan-report_v4.pdf
- WHO. (2024). Obesity and overweight. http://www.who.int/news-room/fact-sheets/detail/obesityand-overweight
- Woodward-Lopez, G., Kao, J., & Ritchie, L. (2011). To what extent have sweetened beverages contributed to the obesity epidemic? *Public Health Nutrition*, 14(3), 499–509. https://doi.org/10. 1017/S1368980010002375
- Xie, Y., Zhu, Y., Cotton, C. A., & Wu, P. (2019). A model averaging approach for estimating propensity scores by optimizing balance. *Statistical Methods in Medical Research*, *28*(1), 84–101. https://doi.org/10.1177/0962280217715487
- Zigler, C. M., & Dominici, F. (2014). Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects. *Journal of the American Statistical Association*, 109(505), 95–107. https://doi.org/10.1080/01621459.2013.869498

Appendix A

As the Figures A2–A4 presents, the ASAM differences of covariates in CBPS are considerably lower than those we obtained from both logit and deep neural network.



Distributional Balance for Covariate Balance Propensity Scores

Figure A1. Balance of covariates in covariate balance propensity scores (CBPS).



Covariate Balance Propensity Scores

Figure A2. The average standardized absolute mean differences (ASAM) in CBPS.

24



Balance of covariates in logit

Figure A3. The average standardized absolute mean differences (ASAM) in logit.



Covariate Balance

Figure A4. The average standardized absolute mean differences (ASAM) in deep neural network.

	Region							
	Bishkek	Chui	lssyk-kul	Jalal-Abad	Naryn	Batken	Osh	Talas
Overweight (A)								
No $(n = 9,600)$	13%	12%	13%	14%	11%	11%	13%	12%
Yes (n = 3,643)	15%	12%	14%	14%	11%	10%	14%	10%
Total (<i>n</i> = 13,243)	13%	12%	13%	14%	11%	11%	14%	11%
Obese (B)								
No (<i>n</i> = 12,063)	13%	11%	13%	15%	11%	11%	13%	12%
Yes (n = 1,180)	14%	19%	16%	8%	13%	8%	16%	6%
Total (n = 13,243)	13%	12%	13%	14%	11%	11%	14%	11%

Table A1. Overweight and obesity status across regions of Kyrgyzstan.

26

	Gen	der	
	Female	Male	Total
Overweight ^(25≤BMI<30)			
No (<i>n</i> = 9,600)	73%	72%	72%
Yes $(n = 3,643)$	27%	28%	28%
Total ($n = 13,243$) Obese ^(BMI \ge 30)	100%	100%	100%
No (<i>n</i> = 12,063)	88%	95%	91%
Yes (<i>n</i> = 1,180)	12%	5%	9%
Total ($n = 13,243$)	100%	100%	100%

 Table A2. Gender and the status overweight and obesity.

 Table A3.
 SSB consumption across total household expenditure quantiles.

	Summar	Summary of SSB (D)		
5 quantiles of household expenditure	Mean	Std. dev.		
1	0.070	0.255		
2	0.091	0.288		
3	0.091	0.288		
4	0.141	0.348		
5	0.151	0.358		
Total	0.108	0.311		