



Can we estimate the causal effects of diet and sedentary behavior on schoolchildren's overweight/obesity from observational studies?

Emil Kupek* 

Department of Public Health, Federal University of Santa Catarina, Florianópolis, Santa Catarina 88040-370, Brazil

***Correspondence:** Emil Kupek, Department of Public Health, Federal University of Santa Catarina, Florianópolis, Santa Catarina 88040-370, Brazil. emil.kupek@ufsc.br

Academic Editor: Pietro Vajro, University of Salerno, Italy

Received: November 2, 2022 **Accepted:** January 14, 2023 **Published:** April 27, 2023

Cite this article: Kupek E. Can we estimate the causal effects of diet and sedentary behavior on schoolchildren's overweight/obesity from observational studies? *Explor Med.* 2023;4:272–85. <https://doi.org/10.37349/emed.2023.00139>

Abstract

Aim: To investigate the causal impact of diet and sedentary behavior on Brazilian schoolchildren's overweight/obesity using the data from observational studies.

Methods: Annual cross-sectional nutritional surveys over the 2013–2015 period, with 26,712 children old 7–12 years in Florianópolis, Brazil, provided the data for this analysis. The surveys applied an online previous-day recall questionnaire on food intake and physical/sedentary activities. Outcome measures were overweight/obesity, whereas exposure variables were daily frequencies of consuming sugary drinks and ultra-processed foods, the total number of dietary items consumed and the total number of sedentary activities per day, and consuming breakfast, mid-morning snacks, lunch, afternoon snack, dinner, and evening snack. Control variables included child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, metabolic equivalents (METs) of physical activities (PAs), and the quality of dietary and PA reports. Causal effects were estimated by augmented inverse probability weighting.

Results: Daily consumption of sugary drinks, eating ten or more foods, and engaging in three or more sedentary behaviors per day significantly increased the odds ratios (ORs) of being overweight/obese in the range of 3–24% compared to the reference, with 95% confidence intervals in the range of 1–32%. Among 19 ORs with P -value ≤ 0.05 , only 3 exceeded 10%.

Conclusions: Under certain conditions, not uncommon in large-scale monitoring and surveillance studies, it is possible to evaluate the causal effects of diet and sedentary activities on overweight/obesity. Daily consumption of sugar-sweetened beverages, eating ten or more foods, skipping breakfast, and engaging in three or more sedentary behaviors per day significantly increased the odds of being overweight/obese.

Keywords

Statistics and numerical data, diet, food consumption, child, surveys and questionnaires, causality



Introduction

Although the relationship between diet, physical activity (PA), and sedentary behavior on the one hand, and nutritional status on the other hand, is based on strong scientific grounds in both theoretical and empirical terms, quantifying it in a causal framework remains a challenge, especially in observational studies. While randomized controlled trials (RCTs) continue as a gold standard for causal inference, observational studies are more suitable for monitoring and tracking, both of which allow rapid adjustments in policy recommendations, regulations, and other interventions to improve nutritional health. With the advent of computerized evaluations of diet and PA by questionnaires, the cost-effectiveness of these methods has made them popular worldwide [1]. Reproducibility, internal, and (more rarely) external validity for some of these instruments have been reported, but to the best of my knowledge, their validity for causal inference on nutritional status has not been evaluated. The reason for that probably has a foothold in a firm belief that only RCTs, preferably case-control longitudinal studies, are entitled to causal interpretation [2].

The absence of random allocation of treatment (intervention, exposure, or any putative causal agent) in observational studies makes it difficult—but not impossible—to attribute the differences between the groups under comparison to the treatment. The effect of confounding and other omitted variables cancels out between the groups under random assignment but observational studies cannot control this aspect of the study design. Even so, under certain conditions, it is possible to balance the differences between the groups by estimating the likely values of the units (e.g., subjects) had they received the treatment applied to another group. Therefore, for each group both observed and hypothetical (“potential”) outcomes are presented, in addition to the observed treatments and baseline covariates, so from a statistical perspective it becomes a missing data issue. As long as the missing outcomes can be consistently estimated based on their association with observed data—a scenario known as “missing at random” [3]—unbiased differences between treatment groups can be calculated in observational studies.

To call these differences causal effects, a substantial theoretical basis must be present. For example, there are sufficient endocrinology grounds to claim that the pathogenesis of metabolic diseases is influenced by the quantity of sugar intake due to its contribution to advanced glycation end products, which in turn cause liver and muscle tissue damage [4]. Another point worth mentioning is the comparison of the differences between the end-of-study *versus* baseline performance of cases and controls in longitudinal studies. This design is known as difference-in-differences [5, 6] and has been widely applied to draw causal inferences in longitudinal studies with a control cohort.

The aforementioned reasoning rests on a key concept known as a potential outcome [7, 8], which can be translated as the unobserved outcome value under a particular scenario, such as receiving another treatment. In the context of cohort studies with intervention at a certain point in time, hypothetical values that would have been observed had there been no intervention can be estimated using so-called synthetic cohorts. For example, time trends in food consumption before and after relevant policy regulations may use this method to evaluate the policy impact.

A propensity score is another key concept for balancing the probability of receiving a particular treatment between groups under comparison in observational studies. The probability is estimated by binary logistic or multinomial regression for two or more treatment groups, respectively, using independent predictor variables such as baseline characteristics, and applied to matching units with similar probabilities, thus mimicking random allocation of treatment in RCTs. Various matching methods with software implementation, and their advantages and disadvantages, have been the subject of intense discussions [6].

Causal analysis is technically an application of marginal structural models [6]. An accessible introduction to causal methods in nutritional observational studies can be found from Mazzocchi et al. [5]. A special type of instrumental variable method to avoid bias in these studies, known as Mendelian randomization, is nicely explained in Bennet and Du [9], whereas the statistical basis for causal effect estimation is didactically provided by Cunningham [6].

The study aims to estimate the causal impact of diet and sedentary behavior on overweight in schoolchildren based on observational data. The emphasis is on the implementation and interpretation of causal analysis rather than on substantive theory.

Materials and methods

Data and sampling

Three repeated cross-sectional nutritional surveys with 7–12-year-old children (2nd to 5th grades) from public schools in Florianópolis, the Santa Catarina state capital in southern Brazil, provided the data for this analysis. The surveys were performed annually over the 2013–2015 period and included about 95% of all municipal schools, thus providing a virtually complete population coverage.

Eligible classrooms were the primary and the schools were the secondary sampling units, with intraclass correlation accounted for in all analyses. The classrooms were randomly selected within each school, and all the children within the selected classrooms took part in the surveys, except those with a mental handicap or visual impairment. Also, the survey reports from the children who did not bring in informed consent from parents or legal guardians were excluded from the analysis. Among 9,100 children invited to participate in the survey over the 2013–2015 period, the response rate was 91%. The survey was applied on different days of the week to reflect dietary intake on both school days and Sundays.

Anthropometric measurements

Trained researchers used standardized protocols [10] to measure body weight and height at school. The former was measured with a portable digital scale Marte (model PP, 50 g precision), and the latter with a portable stadiometer Altuxata (1 mm precision). The body mass index (BMI) was calculated as weight (kg) divided by height (m) squared. World Health Organization criteria for children and adolescents aged 5–19 years [11] were applied to calculate age-and-sex-specific BMI z-scores. Nutritional status was categorized into non-overweight (BMI z-score for age < 1) or overweight including obesity (BMI z-score for age ≥ 1) or obesity (BMI z-score for age ≥ 2).

Food section of the questionnaire

An online questionnaire on schoolchildren's self-reported food intake and PA, known under the "Web sistema de monitoramento de Consumo Alimentar e Atividade Física em Escolares"—Web-CAAFE (Portuguese; Web-based system for monitoring food consumption and PA in schoolchildren), provided all the information analyzed in the present studies. The food intake section is a previous-day recall of the intake of 32 food items, divided into 6 eating events presented in chronological order. For each eating event, 32 images of foods, beverages, or food groups are presented on the computer screen. The Web-CAAFE uses foods and beverage drawings for each of 6 typical daily eating events, ordered chronologically and presented sequentially on the screen (breakfast, mid-morning snack, lunch, mid-afternoon snack, dinner, and evening snack). It was assumed that only one serving was consumed at each meal or snack. No energy intake was calculated for this 24 h-recall food frequency questionnaire.

PA section of the questionnaire

Thirty-two physical and sedentary activities were presented as drawings (icons) on a computer screen for each of the three parts of the day: morning, afternoon, and evening. The children were instructed to click on all icons that represented their activities on the previous day. For the present study, these activities were summed up separately for physical and for sedentary activities over the whole day. This section also contained questions on PA classes at school and the means of transport to and from school.

The questionnaire has been extensively tested for usability [12], reproducibility [13], internal [14], and external validity [15] for its use in the school setting among the children from 2nd to the 5th grades. The questionnaire details have been presented in the aforementioned publications, and its screen slides are available at <http://caafe.ufsc.br/portal/10/detalhes>.

Validity of the questionnaire

Test-retest reliability was shown in an earlier version of the questionnaire [16]. Usability tests showed the child's capacity to understand and respond to Web-CAAFE [12], while Perazi et al. [13] showed moderate-to-high reproducibility. The external validity of food consumption showed the Web-CAAFE's

accuracy [15] in the range of other similar instruments [17, 18]. For example, the percentage of matches between the reported and directly observed food intake at school was 39% for the morning snack, 44% for lunch, and 44% for the afternoon snack [15]. Of the 13 food groups analyzed, 10 showed a moderate reporting bias of $\pm 30\%$, with the lowest value of 4.3% found for sugar-sweetened beverages (SSB) [19].

The overestimation of metabolic equivalents (METs) for PA was found to be lower than the level observed for sedentary behavior and not statistically significant from directly observed PA [20].

Statistical methods

The Brazilian Institute of Geography and Statistics reported the average income of the census sector of each school [21]. As the family residential address determined the school a child was assigned to attend, the income was used as a proxy for family income and categorized into quintiles.

Outcome measures were overweight including obesity and obesity alone, representing less and more specific targets for the impact of diet and sedentary activities on nutritional status. The choice of these outcomes is a simplified sensitivity analysis for the causal impact range. Exposure variables were daily frequencies of consuming SSB, such as sodas, fruit juice, chocolate milk, and ultra-processed foods (nuggets, instant pasta), the total number of dietary items consumed per day, the total number of sedentary activities per day, consuming breakfast, mid-morning snack, lunch, afternoon snack, dinner, and evening snack. Two or more items per day were considered frequent consumption, following Giacomelli et al. [22].

The exposure variables were chosen for being intensively debated in recent literature on markers of a healthy and unhealthy diet, whereas other independent predictors of schoolchildren's nutritional status (control variables) were selected from a previous publication on this topic. The markers are rough-and-ready risk indicators, based on their association with the outcome of interest, and typically used for nutritional monitoring and surveillance, as opposed to quantitative evaluation of a suitable nutrient intake according to the healthy diet recommendations. For example, fruits and vegetables are considered parts of a healthy diet whereas canned food is a marker of an unhealthy diet, although it is the quantities of these foods that largely determine their impact on health.

The quality of dietary records was evaluated using Goldberg's categorization into adequate, underestimate, and overestimate [23], whereas the adequacy of PA reporting was judged by fitting within the mean ± 3 standard deviations (SDs) of reported PA frequency, based on the Poisson distribution. By taking into account the intensity of each PA (vigorous, moderate, light), these were converted into METs [24] and summed up for each individual.

Two methods of accounting for the quality of dietary reporting were applied. First, it was adjusted for in regression by fitting the three-categorical variable (underestimate, adequate, overestimate) for the Goldberg criterion [23] of adequacy as a covariate, thus preserving the sample size and maximizing statistical power. Second, inadequate records were excluded, thus reducing the sample size and power but providing a better assurance against low record quality. The adequacy of PA reports was also taken into account by adjusting for it in regression for both aforementioned methods. The exclusion method is more conservative and was used to probe the sensitivity of causal estimates under the most balanced covariate space possible with the available data. Taken together, both methods provide a range of effect size variations due to reporting quality.

Control variables included child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, METs of PA, and quality of dietary and PA records. All these variables were found predictive of overweight/obesity in previous analysis [25]. With earlier analyses of these data pointing to almost two-thirds of the children reporting a low MET level (≤ 50 per day), the number of sedentary activities gained importance in predicting overweight because it was a more variable feature also and theoretically important.

A 95% confidence interval (CI) was used to express uncertainty around the weighted non-linear least squares estimates of the average treatment/exposure effects. Augmented inverse probability weights [26, 27] were applied to calculate these effects. This algorithm combines inverse probability weight with data augmentation. The former is a well-known method used for the calculation of sampling weights,

among other applications, only here it was used to improve the balance in the probability of belonging to a specific exposure group given the control variables by assigning the weights inversely proportional to this probability for each subject. Data augmentation is a statistical algorithm that improves the representativeness of the sample available, in this case by oversampling a small number of subjects with very low probabilities of falling into an exposure category as predicted by multinomial logistic regression based on control variables. By combining data augmentation and inverse probability weighting with weighted nonlinear least squares estimation, robust causal effects may be obtained [27].

Stata software [28] was used for all statistical procedures.

Results

The describes the relationship between sociodemographic characteristics, diet, and PA on the one hand, and two measures of excess weight on the other hand are described in the Table 1. The latter showed a clear-cut decline since the age of 11 years. Extreme categories of METs showed a difference in overweight including obesity, but it was reduced when only obesity was focused on. Higher totals of SSB, food items, and sedentary activities per day corresponded to higher proportions of overweight/obesity. The children who reported consuming lunch or dinner were more often overweight than those who did not report consumption of these meals.

Table 1. Sociodemographic characteristics, eating behavior, physical and sedentary activities of Brazilian schoolchildren in Florianópolis

Characteristic	Category	n	Overweight (%)	
			Including obesity	Only obesity
Survey year	2013	8,316	29.8	10.8
	2014	8,412	33.7	11.1
	2015	9,984	29.2	10.6
Sex	Male	13,854	30.5	9.0
	Female	12,858	31.1	12.8
Age (years)	7	2,880	32.9	12.1
	8	5,964	31.0	10.1
	9	6,846	32.5	12.5
	10	6,930	31.2	11.0
	11	3,894	25.6	7.9
Family income quintiles	12	174	24.1	13.8
	1st	5,700	28.8	8.7
	2nd	5,274	30.3	9.9
	3rd	6,024	29.1	10.5
	4th	4,572	34.8	14.4
Days of the week of the survey	5th	5,142	32.0	11.4
	Weekend	5,868	31.0	12.3
	Week day	20,844	30.7	10.4
School shift	Morning	12,480	31.6	11.1
	Afternoon	13,698	30.2	10.6
	Integral	534	28.1	12.4
MET of all physical activities per day	0–9.99	4,536	27.4	9.0
	10–29.99	14,496	31.2	10.8
	30–49.00	5,574	30.7	12.3
	50+	2,106	35.6	11.1
Total of SSB per day	0	8,868	28.6	9.4
	1	8,400	29.1	10.3
	2–6	9,444	34.3	12.7

Table 1. Sociodemographic characteristics, eating behavior, physical and sedentary activities of Brazilian schoolchildren in Florianópolis (*continued*)

Characteristic	Category	n	Overweight (%)	
			Including obesity	Only obesity
Total of sedentary activities per day	< 3	10,008	28.2	9.0
	3–5	9,846	31.0	10.8
	6+	6,858	34.3	13.6
Total food items consumed per day	< 10	6,528	21.8	6.5
	10–20	18,864	33.4	12.1
	20+	1,320	38.6	14.1
Total of ultraprocessed foods per day	0	21,690	30.3	10.9
	1	4,140	33.2	10.6
	2–4	882	31.3	10.9
Breakfast	Not consumed	2,490	32.0	29.9
	Consumed	24,222	30.7	31.2
Mid-morning snack	Not consumed	8,868	30.6	10.0
	Consumed	17,844	30.9	11.3
Lunch	Not consumed	756	26.2	12.7
	Consumed	25,956	30.9	10.8
Afternoon snack	Not consumed	3,936	29.0	10.1
	Consumed	22,776	31.1	11.0
Dinner	Not consumed	1,860	26.1	9.0
	Consumed	24,852	31.1	11.0
Evening snack	Not consumed	8,820	29.9	11.0
	Consumed	17,892	31.2	10.8

N = 26,712 (the total number of samples counted); n: the number of statistical samples eligible for classification

No statistical significance is reported for the between-category differences concerning the outcomes as the data practically reached population coverage where the significance is assigned by definition, i.e. axiomatically.

Compared to not consuming sugary drinks, two or more of these per day significantly increased the odds of obesity by 3% (95% CI 0–6%), and overweight including obesity by 6% (2–10%) when unreliable diet records were excluded (Table 2). With regression adjustment instead of exclusion of these records, corresponding impacts were 6% (3–10%) and 9% (5–13%), in the same order.

Table 2. Causal effect estimates of diet on schoolchildren nutrition status, adjusted for child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, METs of PA, quality of dietary and PA reporting

Exposure variables	Quality reporting adjustment	Category level	Obesity				Overweight including obesity			
			OR	Lower	Upper	P-value	OR	Lower	Upper	P-value
Total number of dietary items consumed	CV	< 10	1.00*	-	-	-	1.00*	-	-	-
	EXCL									
	CV	10–19	1.18	1.12	1.24	< 0.001	1.24	1.16	1.32	< 0.001
	EXCL		1.06	1.04	1.08	< 0.001	1.13	1.09	1.17	< 0.001
Number of sugary drinks	CV	20+	1.04	0.69	1.59	0.844	0.99	0.53	1.83	0.971
	EXCL		1.09	1.02	1.16	0.009	1.17	1.06	1.29	0.002
	CV	0	1.00*	-	-	-	1.00*	-	-	-
	EXCL									
	CV	1	1.02	1.00	1.04	0.065	1.03	1.01	1.05	0.005
	EXCL		1.00	0.99	1.02	0.647	1.00	0.97	1.04	0.773
	CV	2–6	1.06	1.03	1.10	< 0.001	1.09	1.05	1.13	< 0.001
	EXCL		1.03	1.00	1.06	0.023	1.06	1.02	1.10	0.002

Table 2. Causal effect estimates of diet on schoolchildren nutrition status, adjusted for child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, METs of PA, quality of dietary and PA reporting (*continued*)

Exposure variables	Quality reporting adjustment	Category level	Obesity				Overweight including obesity			
			OR	Lower	Upper	P-value	OR	Lower	Upper	P-value
Number of ultraprocessed foods	CV	0	1.00*	-			1.00*	-		
	EXCL									
	CV	1	0.99	0.96	1.01	0.312	1.00	0.98	1.03	0.831
	EXCL		1.00	0.98	1.02	1	1.03	1.00	1.06	0.053
	CV	2–4	1.01	0.95	1.08	0.733	1.02	0.92	1.12	0.734
Number of sedentary activities	EXCL		1.00	0.96	1.03	0.869	1.01	0.93	1.10	0.788
	CV	< 2	1.00*	-			1.00*	-		
	EXCL									
	CV	3–5	1.08	1.04	1.12	0	1.12	1.08	1.15	< 0.001
	EXCL		1.01	0.98	1.05	0.37	1.07	1.02	1.12	0.002
	CV	6+	1.10	1.06	1.15	0	1.12	1.07	1.16	< 0.001
	EXCL		1.07	1.03	1.10	0	1.09	1.05	1.13	< 0.001

N = 26,712 (the total number of samples counted); OR: odds ratios (95% CI limits); Lower: lower OR; Upper: upper OR; *: reference category; CV: covariate adjustment; EXCL: exclusion from regression

Consuming 10–19 foods per day increased the odds of obesity by 6% (4–8%), and overweight including obesity by 13% (9–17%), compared to the children who consumed up to 9 food items per day, with the exclusion of unreliable dietary records. With the latter included and adjusted for as regression covariates, the corresponding figures climbed to 11% (8–15%) and 24% (16–32%), respectively. Compared to the same baseline, consuming 20 or more food items significantly increased the odds of overweight/obesity only when unreliable dietary records were excluded, by 17% (6–29%) and 9% (2–16%), respectively.

Using up to two sedentary activities as a baseline and the covariate adjustment, 3–5 such activities increased the odds of overweight including obesity by 12% (8–15%) and of obesity alone by 8% (4–12%). Only the former effect was statistically significant with the exclusion adjustment, increasing the odds by 7% (2–12%). Six or more sedentary activities significantly increased average odds in the range of 7–12% for adjustment methods of dietary records quality.

Consuming one or none of the ultra-processed foods per day increased the odds of being overweight/obese by 3% (0–6%) with the exclusion, but not with the covariate, adjustment. Eating 2 or more ultra-processed foods showed no significant effects on weight status.

Eating breakfast reduced the odds of overweight/obesity by 3% (0–7%) under covariate adjustment, whereas eating dinner increased these odds by 5% (1–10%) but only with the covariate adjustment method (Table 3).

Table 3. Causal effect estimates of consuming daily meals on schoolchildren nutrition status, adjusted for child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, METs of PA, quality of dietary and PA reporting

Meals/snacks	Meal consumed	Quality reporting adjustment	Obesity				Overweight including obesity			
			OR	Lower	Upper	P-value	OR	Lower	Upper	P-value
Breakfast*	No	CV	1.00**	-			1.00**	-		
		EXCL								
	Yes	CV	1.01	0.98	1.04	0.470	0.97	0.93	1.00	0.042
		EXCL	1.02	0.98	1.05	0.344	0.97	0.94	1.01	0.165
Mid-morning snack*	No	CV	1.14	-			1.00**	-		
		EXCL	1.11							
	Yes	CV	1.02	1.00	1.04	0.075	1.00	0.98	1.03	0.878
		EXCL	1.01	0.99	1.03	0.251	1.00	0.98	1.02	0.934

Table 3. Causal effect estimates of consuming daily meals on schoolchildren nutrition status, adjusted for child age, sex, family income, school shift, survey year, day of the week the questionnaire refers to, METs of PA, quality of dietary and PA reporting (*continued*)

Meals/snacks	Meal consumed	Quality reporting adjustment	Obesity				Overweight including obesity			
			OR	Lower	Upper	P-value	OR	Lower	Upper	P-value
Lunch*	No	CV	1.18	-			1.00**	-		
		EXCL	1.20							
	Yes	CV	0.97	0.91	1.05	0.485	1.04	0.96	1.13	0.342
		EXCL	0.93	0.86	1.01	0.08	1.02	0.91	1.15	0.706
Afternoon snack*	No	CV	1.14	-			1.00**	-		
		EXCL	1.11							
	Yes	CV	1.02	1.00	1.04	0.057	1.02	0.99	1.05	0.132
		EXCL	1.01	0.98	1.04	0.604	1.00	0.96	1.04	0.959
Dinner*	No	CV	1.12	-			1.00**	-		
		EXCL	1.09							
	Yes	CV	1.03	0.99	1.06	0.117	1.05	1.01	1.10	0.016
		EXCL	1.02	0.98	1.07	0.345	1.05	0.99	1.12	0.114
Evening snack*	No	CV	1.15	-			1.00**	-		
		EXCL	1.12							
	Yes	CV	0.99	0.98	1.01	0.454	1.00	0.97	1.03	0.948
		EXCL	0.99	0.97	1.01	0.577	1.01	0.98	1.05	0.544

N = 26,712 (the total number of samples counted); OR: odds ratios (95% CI limits); Lower: lower OR; Upper: upper OR; *: adjusted for consuming all other meals; **: reference category

No other effects of consuming specific meals/snacks were statistically significant.

To facilitate the understanding of the snacks, their top five food/beverage items are provided. In the morning, these consisted of bread/biscuits (15%), fruits (12%), yogurt (10%), cream biscuits (9%), and fruit juice (6%); in the afternoon, the preferred items were bread/biscuits (28%), cream biscuits (13%), fruits (12%), yogurt (12%), chocolate drinks (10%), and coffee with milk (10%); in the evening, the preference was for fruits (9%), sweets (8%), cream biscuits (8%), sodas (7%), fruit juice (7%), and bread/biscuits (7%).

Discussion

To the best of the author’s knowledge, the present study is the first to estimate the causal effects of diet and sedentary behavior in nutritional epidemiology based on observational data. Daily consumption of sugary drinks, eating ten or more foods, and engaging in three or more sedentary behaviors per day significantly increased the odds of being overweight/obese. Most of these odds followed a dose-response relationship, i.e. they increased with higher levels of exposure.

Causal analysis with observational study data has a strong foothold in econometrics and public health, especially for evaluating the impact of policy regulations on an area level (county, district, state), and in epidemiological analyses of the impact of public health interventions such as vaccination, using synthetic cohort method [6, 29–32]. It usually requires a large sample size to guarantee a sufficient number of comparable case and control units, particularly if the effect size is small, such as in the present study. However, even a small risk increase may translate into a large disease burden if the cause is frequent in the population. This is certainly true for consuming sugar-added drinks and ultra-processed foods, frequent eating (“snacking”), and sedentary activities, as shown in Table 1. Also, all of these are modifiable risk factors as opposed to immutable genetic factors, so the weight of the former for health interventions in schoolchildren is of utmost value.

Lunch is the main meal in Brazil, with rice and beans traditionally accompanying meat or chicken. Brazilian schoolchildren who consumed such a lunch which concentrated most of the daily energy intake had lower obesity risk compared to other dietary patterns based on the time of day of eating events [33]. This is in line with the present study finding that frequent eating (“snacking”) increases the risk of overweight/obesity

and some reviews shared this concern, especially in the context of social isolation during the coronavirus disease (COVID) epidemic [32–34].

The consumption of ultra-processed foods increased worldwide, including in the two largest countries in the Americas, the USA [35] and Brazil [36], and was associated with lower-quality diets in children and adults [37]. The present study did not find consistent evidence of a causal link between these foods and overweight/obesity, so further research is needed to verify if a longer follow-up to late adolescence and adulthood may confirm this link.

In line with the present study, schoolchildren's sedentary time, especially their screen time, was associated with overweight/obesity in other studies [38, 39]. Other studies also found an impact of consuming SSB on excess weight in children and adults [40–43]. In Brazil, the consumption of SSB increased over the weekend by more than a third compared to the weekday [36]. The problem is compounded by the increasing tendency of both consuming SSB and sedentary activities worldwide [40–50].

The strengths of the present study include a large sample size and consequently a high statistical power, the application of robust statistical analysis of causal effects, with a probe into their sensitivity to broader *versus* the narrower definition of excess weight and to misreporting of dietary intake and PA. Also, the covariates controlling for other variables associated with the outcomes covered principal risk factors established in the literature. The best available statistical methods were applied for causal analysis. The survey questionnaire has been thoroughly tested for over fifteen years, and its reliability and validity are at least as good as other similar instruments, thus providing confidence in its capacity to evaluate schoolchildren's diet and PA. Intraclass correlations between survey reports from the children in the same classroom were also taken into account, thus making statistical inference more robust.

The present study's limitations include memory error, inevitable in self-reports, only one day of the week being asked about for each child, and not knowing the quantities of the foods/beverages consumed [51]. The survey application across different days of the week attenuated the issue of the representativeness of the results, as did a virtually total coverage of the target population. These limitations stem from the trade-off between the feasibility of the questionnaire designed for monitoring and surveillance of schoolchildren's diet and PA on the one hand, and its accuracy and precision on the other hand. Moreover, statistical adjustment for multiple comparisons was not made, so the interpretation of isolated *P*-values should be cautious. For example, the effect of consuming dinner (Table 3) was significant in only 1 ($P = 0.016$) of 4 probes that combined 2 outcomes and 2 data quality adjustment methods. Also, despite a wide coverage of the principal factors known to be associated with excess weight, residual confounding cannot be ruled out. For example, genetic and early-life influences [52, 53] were not accounted for in the present study.

It is beyond the scope of this work to discuss the strengths and limitations of a variety of causal analysis designs and methods. However, it is worth noticing that the application of Mendelian randomization has been growing exponentially in medical research, including the field of nutrition [54, 55]. When the genetic analysis is unavailable, other designs that improve similarities between the baseline characteristics of the groups under comparison are still available [5, 6, 9, 31, 32]. The best design to evaluate the effectiveness of the interventions to prevent overweight/obesity—a matter of ongoing debate [56–58]—may benefit from a causal analysis, despite some skeptical views held by those who think confounding is ubiquitous in observational studies [2]. Cost-benefit analysis of these interventions is also affected by causal analysis. The aforementioned issues are bound to dominate future research in nutrition epidemiology and related disciplines.

The lack of a standardized definition of implausible dietary records has hampered more rigorous between-study comparisons, so the present study followed the recommendation to use both covariate adjustment and the exclusion of such records [59]. Some nationally representative surveys found that under-reporters were more likely to be overweight and conceal consumption of dietary sugar, including SSB [59]. Web-CAAFE validation studies showed that more than 30% of schoolchildren misreport dietary intake to some degree [15], especially for rarely and frequently consumed items [33]. Despite examining

two extreme scenarios for Web-CAAFE misreporting, residual confounding cannot be ruled out, although its impact on the present study conclusions is likely small.

It is worth emphasizing that “causal” here means “unbiased” or “not confounded” plausible effect whose mechanism is often only partially known. In the last two decades, statistical advances and software development have brought causal analysis within the reach of researchers who are not specialists in this area. Contemporary information technology allows fast retrieval of big observational data where confounding is a primary concern because often the researchers did not have a say in the design of the data collection. Contrary to the widespread opinion, the statistical theory has firmly established the unbiasedness of causal analysis, even for cross-sectional studies, given the conditions similar to those required for traditional regression methods [6]. Accurate estimates of effect size provide a huge advantage for better preventive actions and evaluating their costs and benefits.

In conclusion, this work has pointed to the causal impact of consuming the SSB, eating ten or more foods per day, skipping breakfast, and sedentary behavior on increasing the odds of overweight/obesity. Under the right conditions, this type of analysis can be used in observational studies in nutrition sciences.

Abbreviations

BMI: body mass index

CI: confidence interval

METs: metabolic equivalents

ORs: odds ratios

PAs: physical activities

RCTs: randomized controlled trials

SSB: sugar-sweetened beverages

Declarations

Acknowledgments

The author thanks Maria Alice Altenburg de Assis for her leading role as the principal investigator during data collection in the nutritional surveys analyzed here, and for useful advice in joint publications using these data.

Author contributions

EK: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing.

Conflicts of interest

The author declares that he has no conflicts of interest.

Ethical approval

Not applicable for the present study as it is a secondary data analysis. The original studies followed the guidelines of the Code of Ethics of the World Medical Association (Declaration of Helsinki) and were approved by the Human Research Ethics Committee of the Federal University of Santa Catarina under protocol numbers 108.386 and 1.410.381.

Consent to participate

Informed consent to participate in the study was obtained from parents or legal guardians of the children who took part in the original studies that provided the data for the present study.

Consent to publication

Not applicable.

Availability of data and materials

The data will not be shared as they belong to a corporate owner whose permission to do so is not available at the moment.

Funding

Not applicable.

Copyright

© The Author(s) 2023.

References

1. Hardy LL, Mhrshahi S. Elements of effective population surveillance systems for monitoring obesity in school aged children. *Int J Environ Res Public Health*. 2020;17:6812.
2. Ejima K, Li P, Smith DL Jr, Nagy TR, Kadish I, van Groen T, et al. Observational research rigour alone does not justify causal inference. *Eur J Clin Invest*. 2016;46:985–93.
3. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–92.
4. Aragno M, Mastrocola R. Dietary sugars and endogenous formation of advanced glycation endproducts: emerging mechanisms of disease. *Nutrients*. 2017;9:385.
5. Mazzocchi M, Capacci SB, Biondi B. Causal inference on the impact of nutrition policies using observational data. *Bio-based Appl Econ*. 2022;11:3–20.
6. Cunningham S. *Causal inference: the mixtape*. New Haven, CT: Yale University Press; 2021.
7. Pearl J. *Causality: models, reasoning, and inference*. 2nd ed. Cambridge: Cambridge University Press; 2009.
8. Pearl J. An introduction to causal inference. *Int J Biostat*. 2010;6.
9. Bennett DA, Du H. An overview of methods and exemplars of the use of mendelian randomisation in nutritional research. *Nutrients*. 2022;14:3408.
10. Lohman TG, Roche AF, Martorelli R. *Anthropometric standardization reference manual*. Champaign, IL: Human Kinetics; 1991.
11. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ*. 2007;85:660–7.
12. da Costa FF, Schmoelz CP, Davies VF, Di Pietro PF, Kupek E, de Assis MA. Assessment of diet and physical activity of brazilian schoolchildren: usability testing of a web-based questionnaire. *JMIR Res Protoc*. 2013;2:e31.
13. Perazi FM, Kupek E, Assis MAA, Pereira LJ, Cezimbra VG, Oliveira MT, et al. Effect of the day and the number of days of application on reproducibility of a questionnaire to assess the food intake in schoolchildren. *Rev Bras Epidemiol*. 2020;23:e200084. Portuguese.
14. Davies VF, Kupek E, de Assis MA, Engel R, da Costa FF, Di Pietro PF, et al. Qualitative analysis of the contributions of nutritionists to the development of an online instrument for monitoring the food intake of schoolchildren. *J Hum Nutr Diet*. 2015;28:65–72.
15. Davies VF, Kupek E, de Assis MA, Natal S, Di Pietro PF, Baranowski T. Validation of a web-based questionnaire to assess the dietary intake of Brazilian children aged 7–10 years. *J Hum Nutr Diet*. 2015;28:93–102.
16. de Assis MAA, Kupek E, Guimarães D, Calvo MCM, de Andrade DF, Bellisle F. Test-retest reliability and external validity of the previous day food questionnaire for 7–10-year-old school children. *Appetite*. 2008;51:187–93.
17. Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T, et al. The food intake recording software system is valid among fourth-grade children. *J Am Diet Assoc*. 2002;102:380–5.

18. Diep CS, Hingle M, Chen TA, Dadabhoj HR, Beltran A, Baranowski J, et al. The automated self-administered 24-hour dietary recall for children, 2012 version, for youth aged 9 to 11 years: a validation study. *J Acad Nutr Diet*. 2015;115:1591–8.
19. Kupek E, de Assis MA, Bellisle F, Lobo AS. Validity of WebCAAFE questionnaire for assessment of schoolchildren's dietary compliance with Brazilian Food Guidelines. *Public Health Nutr*. 2016;19:2347–56.
20. Jesus GM, Assis MAA, Kupek E. Validity and reproducibility of an Internet-based questionnaire (Web-CAAFE) to evaluate the food consumption of students aged 7 to 15 years. *Cad Saude Publica*. 2017;33:e00163016. Portuguese.
21. Rede Ipea plataforma de pesquisa em rede [Internet]. Ipea - Instituto de Pesquisa Econômica Aplicada; [cited 2023 Jan 25]. Available from: https://www.ipea.gov.br/redeipea/index.php?option=com_content&view=article&layout=edit&id=118
22. Giacomelli SC, de Assis MAA, de Andrade DF, Schmitt J, Hinnig PF, Borgatto AF, et al. Development of a food-based diet quality scale for Brazilian schoolchildren using item response theory. *Nutrients*. 2021;13:3175.
23. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr*. 1991;45:569–81.
24. Ridley K, Ainsworth BE, Olds TS. Development of a compendium of energy expenditures for youth. *Int J Behav Nutr Phys Act*. 2008;5:45.
25. Leal DB, Assis MAA, Conde WL, Lobo AS, Bellisle F, Andrade DF. Individual characteristics and public or private schools predict the body mass index of Brazilian children: a multilevel analysis. *Cad Saude Publica*. 2018;34:e00053117.
26. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005;61:962–73.
27. Kanyenji GM, Oluoch-Kosura W, Onyango CM, Ng'ang'a SK. Does the adoption of soil carbon enhancing practices translate to increased farm yields? A case of maize yield from Western Kenya. *Heliyon*. 2022;8:e09500.
28. StataCorp. Stata 16 [software]. [cited 2023 Jan 25]. Available from: <https://www.stata.com/stata16/>
29. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc*. 2010;105:493–505.
30. Bruhn CA, Schuck-Paim C, Kürüm E, Taylor RJ, Simonsen L, Weinberger DM. Improving assessments of population-level vaccine impact. *Epidemiology*. 2017;28:233–6.
31. Harris ML, Oldmeadow C, Hure A, Luu J, Loxton D, Attia J. Stress increases the risk of type 2 diabetes onset in women: a 12-year longitudinal study using causal modelling. *PLoS One*. 2017;12:e0172126.
32. Dolton PJ, Tafesse W. Childhood obesity, is fast food exposure a factor? *Econ Hum Biol*. 2022;46:101153.
33. Kupek E, Lobo AS, Leal DB, Bellisle F, de Assis MA. Dietary patterns associated with overweight and obesity among Brazilian schoolchildren: an approach based on the time-of-day of eating events. *Br J Nutr*. 2016;116:1954–65.
34. Mattes RD. Snacking: a cause for concern. *Physiol Behav*. 2018;193:279–83.
35. Jha S, Mehendale AM. Increased incidence of obesity in children and adolescents post-COVID-19 pandemic: a review article. *Cureus*. 2022;14:e29348.
36. Cena H, Fiechtner L, Vincenti A, Magenes VC, De Giuseppe R, Manuelli M, et al. COVID-19 pandemic as risk factors for excessive weight gain in pediatrics: the role of changes in nutrition behavior. A narrative review. *Nutrients*. 2021;13:4255.

37. Wang L, Martínez Steele E, Du M, Pomeranz JL, O'Connor LE, Herrick KA, et al. Trends in consumption of ultraprocessed foods among US youths aged 2-19 years, 1999–2018. *JAMA*. 2021;326:519–30.
38. Monteiro LS, Hassan BK, Estima CCP, Souza AM, Verly E Junior, Sichieri R, et al. Food consumption according to the days of the week – national food survey, 2008-2009. *Rev Saude Publica*. 2017;51:93.
39. Liu J, Steele EM, Li Y, Karageorgou D, Micha R, Monteiro CA, et al. Consumption of ultraprocessed foods and diet quality among U.S. children and adults. *Am J Prev Med*. 2022;62:252–64.
40. Martins AP, Levy RB, Claro RM, Moubarac JC, Monteiro CA. Increased contribution of ultra-processed food products in the Brazilian diet (1987-2009). *Rev Saude Publica*. 2013;47:656–65. Portuguese.
41. Shao T, Wang L, Chen H. Association between sedentary behavior and obesity in school-age children in China: a systematic review of evidence. *Curr Pharm Des*. 2020;26:5012–20.
42. Sun X, Zhao B, Liu J, Wang Y, Xu F, Wang Y, et al. A 3-year longitudinal study of the association of physical activity and sedentary behaviours with childhood obesity in China: the childhood obesity study in China mega-cities. *Pediatr Obes*. 2021;16:e12753.
43. Mihrshahi S, Gow ML, Baur LA. Contemporary approaches to the prevention and management of paediatric obesity: an Australian focus. *Med J Aust*. 2018;209:267–74.
44. Ponce-Blandón JA, Deitos-Vasquez ME, Romero-Castillo R, da Rosa-Viana D, Robles-Romero JM, Mendes-Lipinski J. Sedentary behaviors of a school population in Brazil and related factors. *Int J Environ Res Public Health*. 2020;17:6966.
45. Keane E, Li X, Harrington JM, Fitzgerald AP, Perry IJ, Kearney PM. Physical activity, sedentary behavior and the risk of overweight and obesity in school-aged children. *Pediatr Exerc Sci*. 2017;29:408–18.
46. Cárdenas Sánchez DL, Calvo Betancur VD, Flórez Gil S, Sepúlveda Herrera DM, Manjarrés Correa LM. Consumption of sugary drinks and sugar added to beverages and their relationship with nutritional status in young people of Medellín (Colombia). *Nutr Hosp*. 2019;36:1346–53. Spanish.
47. Livingstone KM, McNaughton SA. A health behavior score is associated with hypertension and obesity among Australian adults. *Obesity (Silver Spring)*. 2017;25:1610–17.
48. Chen CH, Tsai MK, Lee JH, Wen C, Wen CP. Association of sugar-sweetened beverages and cardiovascular diseases mortality in a large young cohort of nearly 300,000 adults (age 20–39). *Nutrients*. 2022;14:2720.
49. Bleich SN, Vercammen KA, Koma JW, Li Z. Trends in beverage consumption among children and adults, 2003-2014. *Obesity (Silver Spring)*. 2018;26:432–41. Erratum in: *Obesity (Silver Spring)*. 2019;27:1720.
50. Beck AL, Martinez S, Patel AI, Fernandez A. Trends in sugar-sweetened beverage consumption among California children. *Public Health Nutr*. 2020;23:2864–9. Erratum in: *Public Health Nutr*. 2021;24:376.
51. Gibson RS, Charrondiere UR, Bell W. Measurement errors in dietary assessment using self-reported 24-hour recalls in low-income countries and strategies for their prevention. *Adv Nutr*. 2017;8:980–91.
52. Kerr JA, Long C, Clifford SA, Muller J, Gillespie AN, Donath S, et al. Early-life exposures predicting onset and resolution of childhood overweight or obesity. *Arch Dis Child*. 2017;102:915–22.
53. Mihrshahi S, Baur LA. What exposures in early life are risk factors for childhood obesity? *J Paediatr Child Health*. 2018;54:1294–8.
54. Zulyniak MA, Fuller H, Iles MM. Investigation of the causal association between long-chain n-6 polyunsaturated fatty acid synthesis and the risk of type 2 diabetes: a mendelian randomization analysis. *Lifestyle Genom*. 2020;13:146–53.
55. Barning F, Abarin T. Assessing the causality factors in the association between (abdominal) obesity and physical activity among the Newfoundland population—a mendelian randomization analysis. *Genet Epigenet*. 2016;8:15–24.

56. Beets MW, Brazendale K, Weaver RG, Armstrong B. Rethinking behavioral approaches to compliment biological advances to understand the etiology, prevention, and treatment of childhood obesity. *Child Obes.* 2019;15:353–8.
57. Baranowski T, Motil KJ, Moreno JP. Public health procedures, alone, will not prevent child obesity. *Child Obes.* 2019;15:359–62.
58. Bahia L, Schaan CW, Sparrenberger K, Abreu GA, Barufaldi LA, Coutinho W, et al. Overview of meta-analysis on prevention and treatment of childhood obesity. *J Pediatr (Rio J).* 2019;95:385–400.
59. Lioret S, Touvier M, Balin M, Huybrechts I, Dubuisson C, Dufour A, et al. Characteristics of energy under-reporting in children and adolescents. *Br J Nutr.* 2011;105:1671–80.