Does ultra-processed food consumption significantly increase obesity risk in low-socioeconomic (SES) groups in Australia?

NUTM/STAT3888 Project 1

Semester 2, 2021

490432473[1], 480131317[2], 490294826[3]

490284045[4], 490050978[5], 490199440[6]

[1] NUTM3888, The University of Sydney NSW 2006
[2] NUTM3888, The University of Sydney NSW 2006
[3] NUTM3888, The University of Sydney NSW 2006
[4] STAT3888, The University of Sydney NSW 2006
[5] STAT3888, The University of Sydney NSW 2006
[6] STAT3888, The University of Sydney NSW 2006

Abstract

Consumption of ultra-processed foods (UPF) and obesity prevalence have been increasing simultaneously [1]. Research has shown that UPF consumption increases obesity risks among low socioeconomic (SES) populations in the United States [2]. When compared to nutritious food, UPF is more enticing to those with low SES background due to the affordability [3]. The current study provides a novel investigation into whether UPF consumption significantly contributes to greater obesity prevalence in low-SES groups in Australia. By using the methods of classification tree, Random Forest, Logistic Regression, K-Nearest Neighbour and NOVA food classification system. Although results revealed that low SES individuals consume more UPF in Australia, increasing consumption of UPF was not found to significantly contribute to obesity. This presents implications for government initiatives targeting increased UPF consumption by suggesting shifts in resource allocation from low-SES focused policies (e.g., subsidies on healthy foods) to more broadly encourage healthier food environments.

Introduction

Obesity is a chronic disease caused by energy imbalance when energy intake exceeds expenditure over a prolonged period, driven by individual behaviour [4] In Australia, obesity has become an epidemic as two in three Australian adults are considered overweight (67%) or obese (38%) [5] Obesity is elaborately defined by body mass index (BMI) which measures a person's height in correspondence to their weight, with one considered obese when BMI is 30 and above [4]. However, waist circumference-to-height ratio (WHtR) is a more effective indicator in identifying obesity, compared to BMI which does differentiate between muscle and fat in an individual's body weight[6]. Instead, WHtR takes into account a person's central and visceral fat by dividing an individuals' waist circumference with height [7]. A person is considered healthy if their WHtR is smaller than 0.52, overweight (0.52-0.63) and ratio larger than 0.63 were defined as obese [6].

In Australia, consumption of ultra-processed foods (UPF) and obesity prevalence have been increasing simultaneously[1]. UPF accounted for 40% of total energy intake among Australian adults [8]. They are industrially formulated products that have been subjected to specific forms of food processing, which are products that are ready-to-eat and often include additives and flavours, including sugar-sweetened beverages, biscuits and instant-noodles [3]. UPF has a poor nutritional profile, generally high in salt, sugar, calories and low in protein, micronutrients and fibre [9]. These characteristics make them hyperpalatable and nutritionally imbalanced and higher consumption of UPF was positively associated with the prevalence of obesity[3,10]. The 'protein leverage hypotheses' posits that humans will regulate the intake of macronutrients (i.e., carbohydrates and fats) and prioritise the absolute intake of protein over other dietary nutrients [11]. Accordingly, UPF tends to have low protein content, encouraging food overconsumption until protein demands are fulfilled, resulting in excess energy intake and potentially contributing to the obesity pandemic [7,12].

UPF are prevalent among adolescents and young adults with lower education and income levels in Australia [13]. Socio-economic status is measured in terms of income, education, or occupation [14]. Within developed industrial countries, individuals with low socioeconomic status (SES) are more likely to become overweight or obese than those with high SES [2]. As such, people with a low SES background have been found to have a higher risk of developing obesity than those with high SES in Australia [15]. Low SES is associated with poor nutrition and consumption of high calorie foods, which contributes to the accumulation of excess body fats [2]. On the other hand, people with high SES tend to afford healthy foods, which are nutritional balanced [13]. People with low socio-economic background have higher consumption of UPF and higher energy intake [14]. Since low SES individuals are found to have higher consumption of UPF and higher energy intake [14], UPF are suggested to be more appealing to low SES individuals due to their affordability and convenience [3].

Additionally, UPF consumption increases the risk of obesity among low SES population [1]. However, researchers have demonstrated that in middle and low-income countries, individuals with high SES have higher consumption of UPF, whereas in high-income countries, those with low SES have higher consumption of UPF [13,16]. More research needs to be conducted to certify these claims. Consequently, it is important to carry out the current study to inform appropriate measures to address the increased prevalence of obesity among the low SES population in Australia. Although previous cross-sectional studies on the Australian adult population have investigated the links between dietary share of UPFs and obesity indicators (i.e., BMI and waist circumference), as well as dietary share of UFPs and socioeconomic status, there has not been a study investigating the significance of UPF consumption to obesity risk in low-SES populations [8,14]. Thus, this project aimed to investigate whether ultra-processed food consumption increases obesity risk differently among socio-economic groups in Australia.

Following aims will be explored: to determine whether low-socioeconomic groups consume more ultra-processed food compared to other socioeconomic groups, whether increased ultra-processed food ratio leads to lower protein and increased energy intake ratio among the low-socioeconomic group and whether ultra-processed food consumption is a significant contributor to obesity. It is hypothesised that ultra-processed food consumption increases obesity risk in low-socioeconomic groups in Australia.

Results

Aim 1: Whether low SES groups consume more UPF compared to other SES groups

In investigation of whether low SES groups consume more UPF than high SES groups, an overview of UPF consumption across SES groups divided into low and high levels, without any adjustments to important confounding variables is considered **(Figure 1)**. The density plots are separated into obese and non-obese groups for both high and low SES groups. Findings consistently demonstrate higher UPF consumption in low compared to high SES groups, irrespective of obesity status.

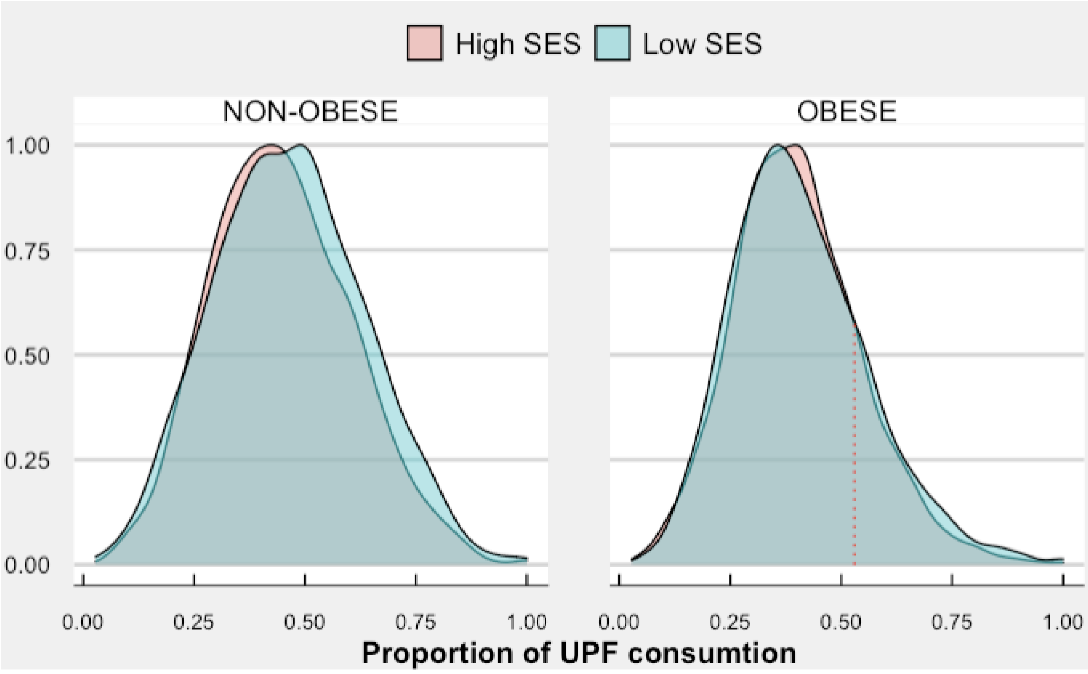

Figure 1 Scaled Density Plot of Between-Group UPF Consumption

Aim 2：Whether increased UPF ratio leads to lower protein and increased energy intake ratio among the low-SES group

Increasing dietary proportion of UPF consumption was found to result in decreasing protein intake, demonstrated by the smoothed line **(Figure 2).** This indicated that increasing UPF consumption in the diet tends to dilute dietary protein intake.

Moreover, increasing dietary proportion of UPF consumption was also associated with increasing energy intake, indicated by the smoothed line **(Figure 3)**. This indicated that increasing dietary UPF consumption tends to increase energy intake. However, these findings were based on EDA and the relations were pairwise without any adjustments from other important confounding variables. Future studies should consider validating these findings by applying formal statistical tests.

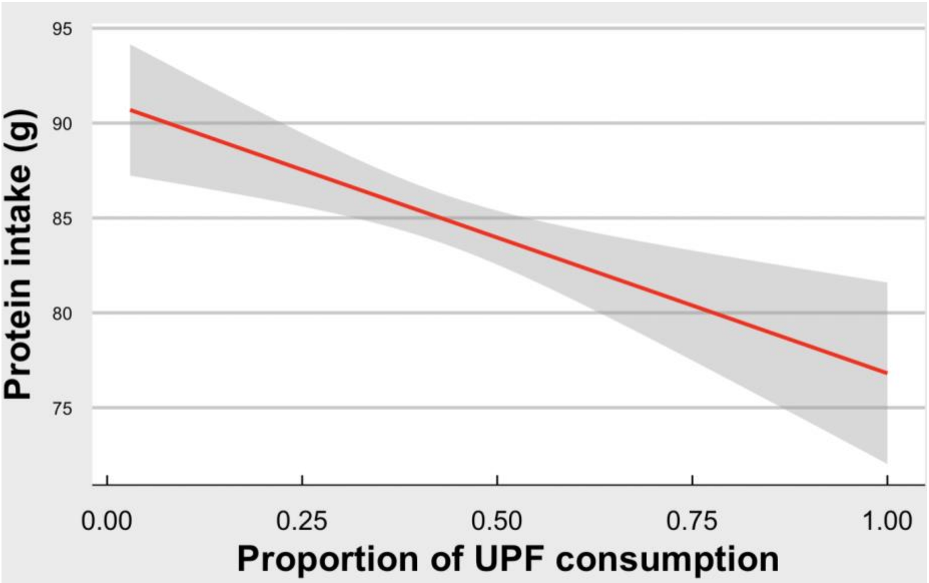

Figure 2 Relationship (smoothed) between UPF consumption and protein intake

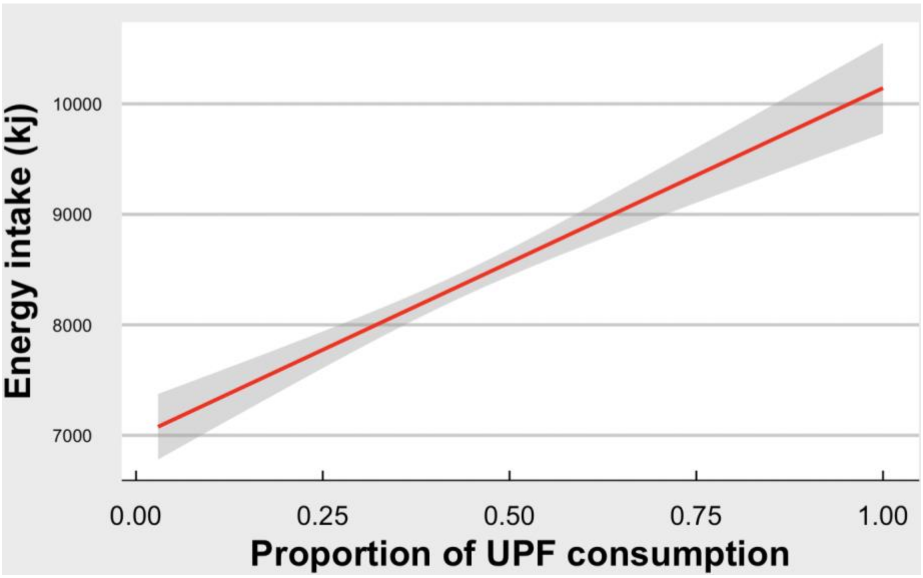

Figure 3 Relationship (smoothed) between UPF consumption and energy intake

According to the protein leverage hypothesis, macronutrient regulation of relatively constant protein increased energy intake, potentially resulting in increased obesity risk (Martinez et al., 2018). Based on figures above, the protein leverage hypothesis was supported.

Aim 3: Whether UPF consumption is a significant contributor to obesity in low-SES groups

Classification Trees

The classification tree was firstly constructed to find if any biomedical variables significantly contributed to obesity within the low SES group.

Table 1 Classification trees based on different complexity parameter (cp)

CP	nsplit	rel.error	xerror	xstd
0.256	0	1.000	1.000	0.031
0.204	1	0.744	0.744	0.028
0.053	2	0.540	0.590	0.026
0.024	3	0.487	0.530	0.024
0.013	6	0.408	0.477	0.023
0.008	7	0.395	0.452	0.023
0.005	**8**	**0.387**	**0.445**	**0.023**
0.005	9	0.382	0.453	0.023
0.005	14	0.358	0.453	0.023
0.004	18	0.340	0.453	0.023
0.004	19	0.336	0.453	0.023
0.003	22	0.325	0.461	0.023
0.003	25	0.316	0.478	0.023
0.002	32	0.298	0.480	0.023
0.002	36	0.290	0.497	0.024
0.001	45	0.269	0.505	0.024
0.001	52	0.260	0.518	0.024
0.000	57	0.256	0.542	0.025

[a] Sample size (n) = 2767

Table 1 suggests the suitability of choosing a cp value of 0.005 to plot the classification tree. This was used for classification tree model 1, with variable importance presented in Table 2 (left). The model was then generalised via 10-fold cross validation, which showed a preferred cp around 0.004 in Figure 4. This was used in model 2, with the variable importance presented in Table 2 (right).

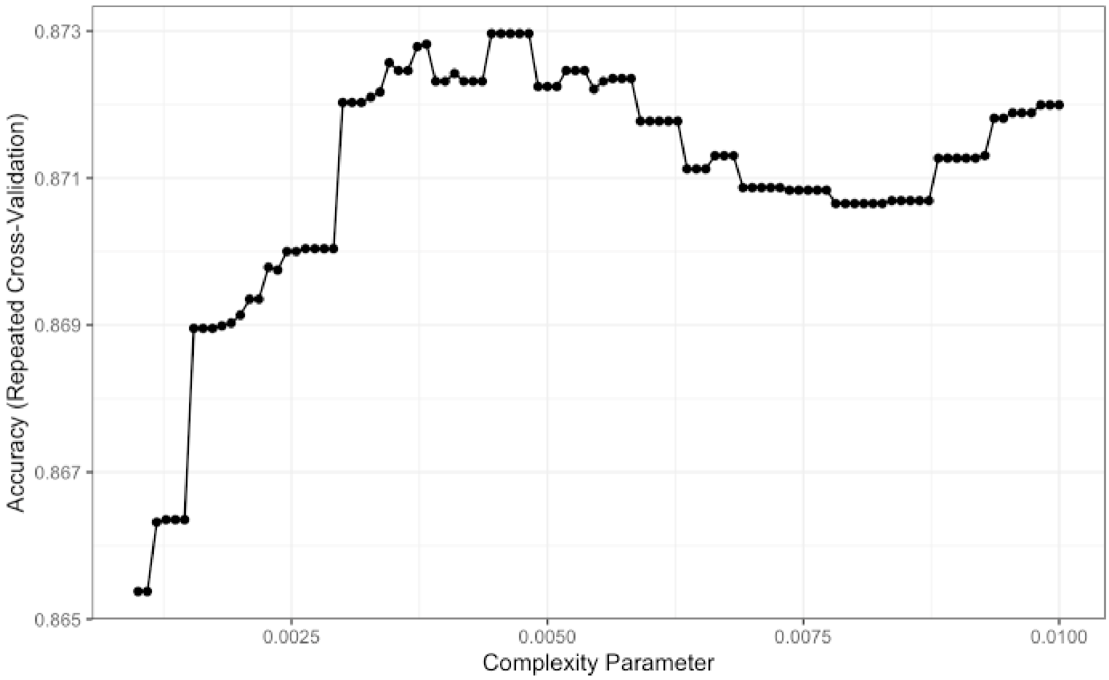

Figure 4 Model accuracy given different cp

Models built on these two slightly different cps are summarised in Table 2. BMI, measured weight, self-perceived body mass and systolic blood-pressure were identified as the five most important variables, whereas UPF consumption ranked eleventh with a relatively low importance of 10.45.

Table 2 The first 15 important variables from classification trees

Variable Description (Model 1)	Importance (Model 1)	Variable Description (Model 2)	Importance (Model 2)
Body mass index (BMI)	644.86	Body mass index (BMI)	662.49
Measured weight (kg)	411.43	Measured weight (kg)	436.49
Self-perceived body mass	304.04	Self-perceived body mass	306.13
Age of person (y)	147.05	Age of person (y)	160.41
Systolic blood pressure (mmHg)	121.23	Systolic blood pressure (mmHg)	123.86
Diastolic blood pressure (mmHg)	77.92	Diastolic blood pressure (mmHg)	79.70
Sex of person	46.37	Sex of person	52.70
Equivalised income of household: deciles	29.12	Equivalised income of household: deciles	37.19
Whether has hypertensive disease	26.02	Whether has hypertensive disease	26.02
Whether has high cholesterol	16.21	Whether has high cholesterol	16.21
Ultra Processed Foods consumption	**10.45**	**Ultra Processed Foods consumption**	**14.05**
Whether has diabetes mellitus	6.61	Whether has diabetes mellitus	7.50
Total mins undertaken physical activity in last week	3.11	Total mins undertaken physical activity in last week	7.38
Sleep duration in minutes on day prior to interview	1.89	Usual daily serves of vegetables	6.95
Whether currently on a diet	1.68	Usual daily serves of fruit	3.50

[a] Sample size (n) = 2767
[b] Three variables that descripe ID, weight and height of the sample were removed
[c] Model 1: The model given by the complexity parameter that returns the smallest error.
[d] Model 2: The model given by 10-fold cross validation

Random Forest

In Table 3, we extracted the importance of each variable which was either derived from impurity-based method (left) or permutation-based method (right). Mean error of generated models was 0.125, implying the acceptability of importance ranking from the random forest model.

Table 3. Variable importance from Random Forest

Variable Description (Model 1)	Importance (Model 1)	Variable Description (Model 2)	Importance (Model 2)
Body mass index (BMI)	335.63	Body mass index (BMI)	0.178
Age of person (y)	100.67	Age of person (y)	0.038
Measured weight (kg)	66.14	Measured weight (kg)	0.026
Self-perceived body mass	65.25	Self-perceived body mass	0.013
Systolic blood pressure (mmHg)	25.86	Systolic blood pressure (mmHg)	0.012
Ultra Processed Foods consumption	**23.75**	**Ultra Processed Foods consumption**	**0.004**
Diastolic blood pressure (mmHg)	22.28	Diastolic blood pressure (mmHg)	0.004
Sex of person	22.16	Sex of person	0.003
Total mins undertaken physical activity in last week	21.44	Total mins undertaken physical activity in last week	0.002
Total mins spent sitting or lying down	20.55	Total mins spent sitting or lying down	0.002
Equivalised income of household: deciles	20.42	Equivalised income of household: deciles	0.001
Sleep duration in minutes on day prior to interview	15.33	Sleep duration in minutes on day prior to interview	0.001
Total mins undertaken vigorous physical activity in last week	12.40	Total mins undertaken vigorous physical activity in last week	0.001
Usual daily serves of vegetables	9.65	Usual daily serves of vegetables	0.001
Usual daily serves of fruit	9.43	Usual daily serves of fruit	0.001

[a] Sample size (n) = 2767
[b] Three variables that descripe ID, weight and height of the sample were removed
[c] Model 1: Impurity-based variable importance
[d] Model 2: Permutation-based variable importance

As shown above, these two methods in random forest reached a consensus on the five most crucial variables, aligning with the variables identified in the classification trees showing partial ranking difference. However, the ranking of UPF consumption does rise from rank 11 in classification trees to 5 in random forest, implying UPF consumption was more significant after the removal of potential bias of a single classification tree in the random forest model.

Logistic Regression

We compared the mean error of each generated model. As shown in Figure 5, the Stepwise Bayesian Information Criterion (BIC) and the Forward BIC returned similar smallest mean error and extreme values. Ridge regression presented the worst model.

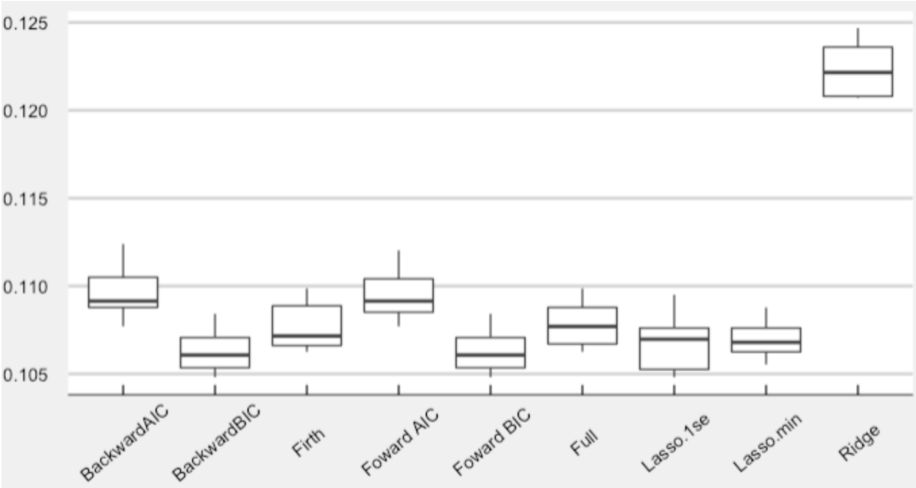

Figure 5 Mean error of each model

Summary of Backward BIC and Forward BIC models are presented below (Table 4). ‘BMI’, ‘Age’, ‘Sex’, ‘Measured weight’ and ‘Physical activity duration’ were included. Statistically significant p-values were identified for these variables in their association with the target variable (Obesity). The odds ratio of Measured weight was less than 1, meaning one unit increase in it predicts a decrease in the odds of obesity occurring. Regarding sex, females are demonstrated to have increased risks of obesity. Additionally, a larger BMI was also shown to increase obesity risk. However, this model did not demonstrate significant associations between UPF consumption and obesity.

Table 4 Summary of Backward BIC and Forward BIC

Backward BIC

	obesity		
Predictors	*Odds Ratios*	*CI*	*p*
Intercept	0.00	0.00 – 0.00	**<0.001**
BMI	2.43	2.20 – 2.68	**<0.001**
Age	1.06	1.05 – 1.07	**<0.001**
Measured weight	0.94	0.92 – 0.96	**<0.001**
Physical activity	1.00	0.99 – 1.00	**<0.001**
Sex	2.93	2.01 – 4.29	**<0.001**
Observations	2767		
R^2 Tjur	0.616		

Forward BIC

	obesity		
Predictors	*Odds Ratios*	*CI*	*p*
Intercept	0.00	0.00 – 0.00	**<0.001**
BMI	2.43	2.20 – 2.68	**<0.001**
Age	1.06	1.05 – 1.07	**<0.001**
Sex	2.93	2.01 – 4.29	**<0.001**
Measured weight	0.94	0.92 – 0.96	**<0.001**
Physical activity	1.00	0.99 – 1.00	**<0.001**
Observations	2767		
R^2 Tjur	0.616		

Note: Sex in models above represents whether the input is a female

Neural Networks

The prediction test of established neural networks is presented in Table 4. This demonstrated that FALSE predictions were always greaterthanTRUE predictions, indicating the unsuitability of the neural network model.

Table 5. Results of prediction test

	FALSE	TRUE
NON-OBESE	225	197
OBESE	594	430
[a] Sample size (n) = 2892		

Figure 6 suggests the second (S2) and fourth (S4) hidden neurons are not as significant as other neurons in impacting obesity. Therefore, we focused on S1, S3, S5 and S6, where the important variables recommended by neural networks were self-perceived weight (underweight, acceptable weight and overweight), sex, UPF consumption, diastolic blood pressure and Systolic blood pressure according to Figure 7.

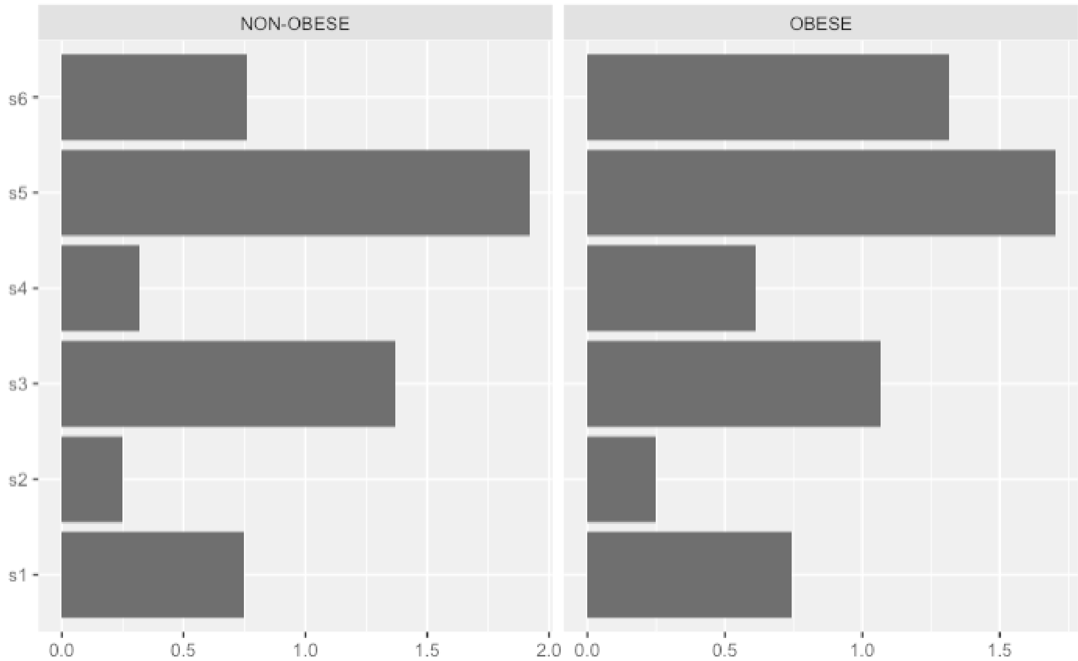

Figure 6 Importance of hidden neurons

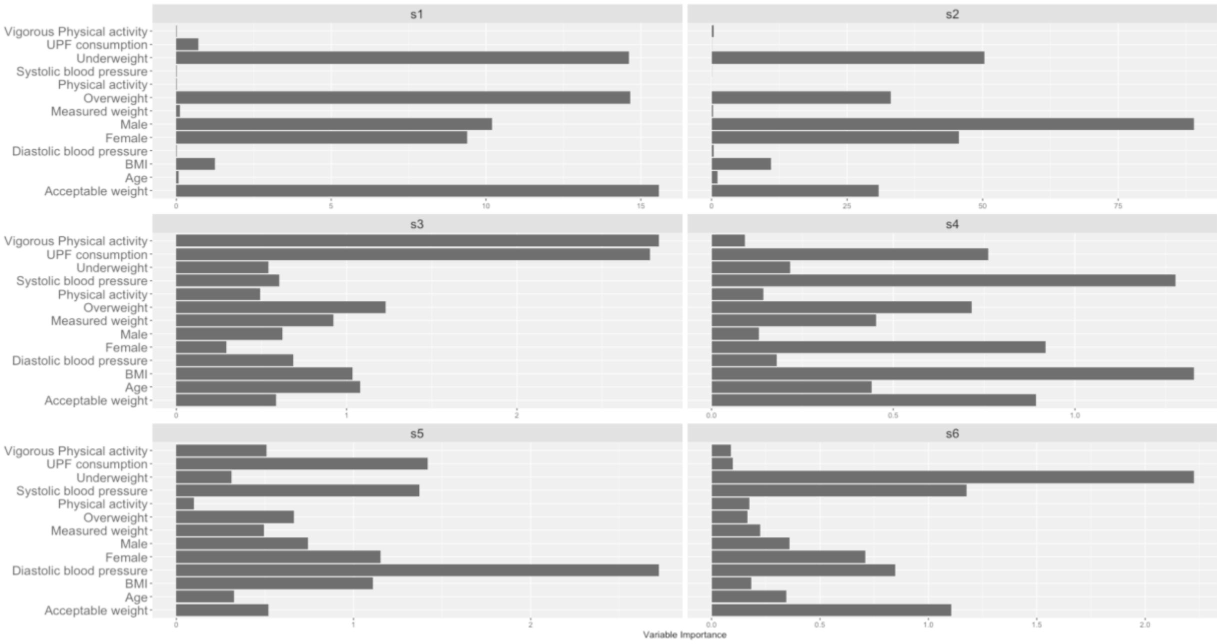

Figure 7 Importance of each variable in each neuron

K Means Clustering

K-means clustering was utilised to separate the dataset in order to control major confounding variables. Figure 8 indicated that 5 clusters were sufficient to capture similarities existing in the confounding variables listed in figure 7.

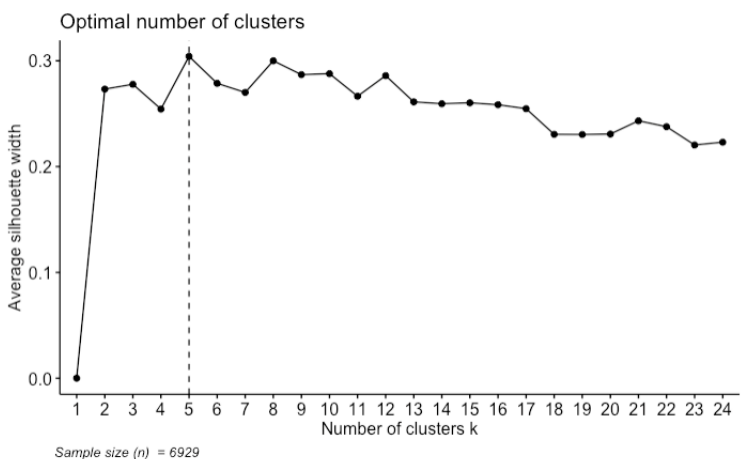

Figure 8. Number of clusters suggested

Figure 9 visualised these 5 clusters, suggesting great overlap among clusters 3, 4 and 5. Thus, confounding variables within these clusters are not controlled.

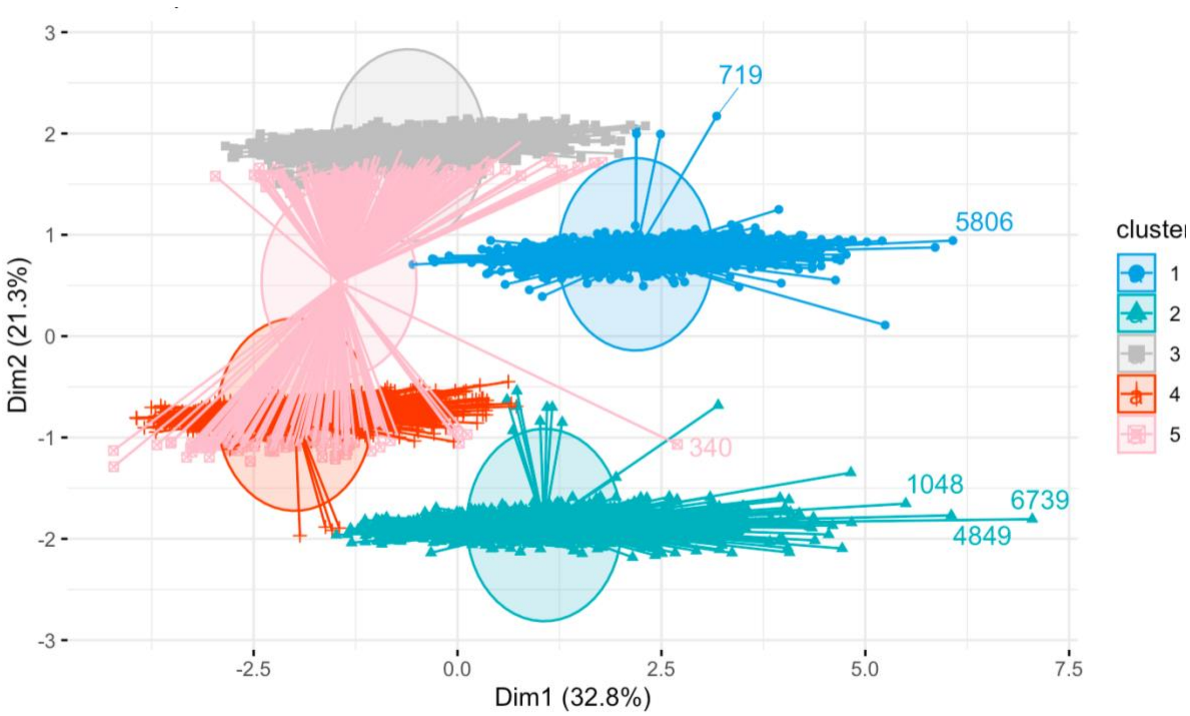

Figure 9 Visualisation of clusters

Direct link between UPF consumption and obesity in low SES group

The final plot between UPF consumption and the number of obese people, with confounding variables controlled within each cluster is presented in Figure 6. For each cluster, as UPF consumption increases, the number of obese people (blue) is not necessarily greaterthan that of non-obese people.

Results conclusion: Is UPF consumption significant in determining obesity?

Variable importance deriving from classification trees, logistic regression models and neural networks do not seem to imply UPF consumption is a significant contributor while the reverse is shown for models from random forest. The density plot was used to study the relationship between the number of obese or non-obese people and UPF consumption (Figure 10). After controlling for major confounding variables within each cluster, increasing UPF consumption was not shown to significantly contribute to increasing the number of obese people. Instead, there is a greater number of non-obese compared to obese people at high level of UPF consumption. Combined with the previous models, this leads to the rejection of our hypothesis.

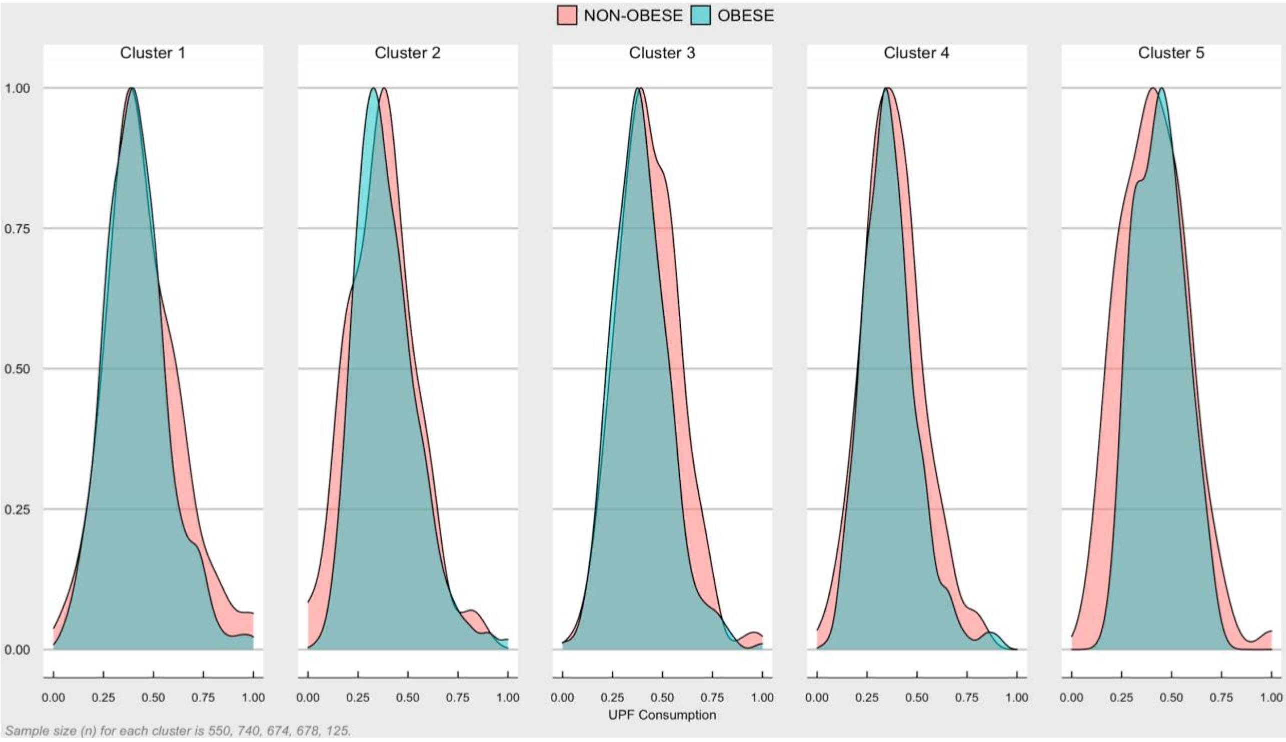

Figure 10 Density plot between UPF consumption and the number of obese people in low SES group

Discussion

The present analysis of national health survey data investigated the contribution of UPF consumption to increased obesity prevalence previously found for low-SES individuals compared to other SES groups in Australia [17,19]. Findings indicated non-significant associations between UPF consumption and obesity risk in low-SES groups in Australia, resulting in the rejection of our hypothesis. This has significant implications for resource allocation in government initiatives addressing socioeconomic inequalities in health, suggesting focus should be shifted from UPF-related policies aimed at low-SES communities (e.g., UPF taxes and subsidies on healthy foods) to target food environment factors more broadly across the population [20].

In our first aim, greater UPF consumption as a proportion of dietary intake was found for low-SES groups compared to high-SES groups in Australia. This aligns with trends found in the United States of America (USA), suggesting considerable influence of socioeconomic drivers in increased UPF consumption [21,22]. To understand UPF associations with obesity, our second aim demonstrated support for the protein leverage hypothesis, as increased UPF was associated with lower protein and higher energy intake [23]. Although mechanisms behind UPF and calorie intake associations are complex and multidimensional (e.g., other possible factors include ultra-palatability) [20], the present study validates the potential of dietary protein dilution from high UPF consumption in activating compensatory feeding for protein, driving excess calorie intake [11,20]. Considering positive correlations between cost of supermarket foods and their protein content combined with the prevalence of low-cost, low-protein and high-energy density UPF, greater consumption of protein-diluted diets is suggested to be encouraged in low-SES groups suggested [23]. Consequent expectations of excess energy intake caused by greater UPF consumption may explain increased prevalence of obesity in low-SES groups [17].

However, investigation into our final aim found that within low-SES populations, increased UPF consumption was a small, non-significant contributor to greater obesity risk, after controlling for major confounding variables of body mass index (BMI), age and measured weight (kg), resulting in the rejection of our hypothesis. Although dose-dependent links between dietary proportion of UPF and obesity indicators among Australian adults were previously found, present results indicate that increased UPF consumption in low-SES groups does not significantly inform their greater obesity prevalence compared to high-SES groups

[8,17]. Subsequently, future research should investigate the influence of other obesity risk factors including low physical activity and smoking, which are found to disproportionately influence low-SES individuals and communities [14,24]. Additionally, non-obese individuals were found to consume a greater dietary share of UPF than obese individuals across the population. Due to associations between increased UPF intake with obesity and chronic diseases (e.g., diabetes) risk, this presents population-wide concerns of poorer health outcome trajectories [14].

Consequently, future directions for public health policies targeting UPF consumption are presented, including shifts from low-SES targeted interventions (e.g., price-reductions of healthy foods) to encouragement of healthier food environments across all socioeconomic groups [20]. Indeed, a comparison of low-income households, indicated by those participating in the U.S. food assistance program (Supplemental Nutrition Assistance Program; SNAP) and non-participants demonstrated the considerable influence of UPF in rational considerations of meal planning and budgeting compared to the cost of healthy foods [22]. UPF selections were strongly mediated by their familiarity and long shelf-life due to their role in alleviating fears of money wastage on foods likely to be rejected by children and/or spoil rapidly [22]. Consequently, only increasing the availability and even lowering prices of healthier options (e.g., vegetables) for low-SES individuals may not significantly impact UPF purchases [22]. Instead, population-wide encouragement of healthier food environments (e.g., in schools) and regulating UPF marketing in media, stores and food packaging are proposed as more effective strategies for improving diet quality and health outcomes [22].

There are several strengths to this study. The study provides a novel investigation into whether UPF consumption significantly contributes to greater obesity prevalence in low-SES groups in Australia. The use of nationally representative data allows increasing generalisability of findings to the Australian population. Additionally, the use of the NOVA food classification system [25] as a basis of our analyses ensured relevant approaches to linking UPF consumption with obesity. Recognised by UN agencies, the NOVA system provided objective, clear and standardised criteria for classification.

Limitations

1. Manual categorisation of food variables into NOVA criteria may be prone to human error and incorrect classification.
2. Despite the strengths of the NOVA classification system, it also presents limitations due to its inability to separate out healthier reformulations of ultra-processed foods. Indeed, some nutrient-dense, low-cost ultra-processed foods can improve the overall eating quality when following national dietary guidelines, particularly for low and medium SES groups [26].
3. Not all confounding variables were controlled since some models are computationally expensive to plot. Inaccurate models (e.g., neural networks), may unavoidably be established and their importance ranking is unreliable. More time was needed to allow the computer to conduct relevant building.
4. When building clusters, even though categorical variables were converted to dummy variables before applying k-means clustering method, we potentially lost some signals from this conversion. A variation of k-means known as k-modes introduced by Zhexue Huang should be further considered [27].

Conclusions

Contrary to the hypothesis, this study found small, non-significant effects of UPF consumption on obesity prevalence in low-SES groups in Australia. Despite low-SES groups being previously found to have higher obesity prevalence [17] and consume a greater dietary share of UPFs in Australia [14], the present study negates the existence of significant associations between disparate UPF consumption in low-SES groups and obesity. Considering the general increase of UPF sales in Australia and associations with poor diet quality [8], this study presents policy implications by proposing shifts in resource allocation from UPF-related health initiatives aimed at low-SES groups (e.g. subsidising healthy foods) to ones which target the population more broadly.

Methods

Overview

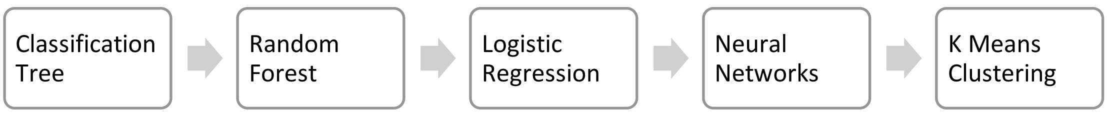

Figure 11. Main models and workflow

Since initial exploratory analysis in Aim 1 and 2 was only descriptive and effects of important confounding variables are not controlled, further analysis is conducted. Through controlling the effects of other important confounding variables, adjusted effects of low SES groups on the consumption of UPF were investigated. Consequently, multiple models were constructed, including classification tree, logistic model, random forest, neutral networks and K-means clustering.

Myung et al. (2019) suggested $\frac{Waist\ circumference}{Height}$ as the better measurement of obesity than the prevalently used BMI indicator [6] . Thus, we applied the same ratio in our study. The codes of consumed foods during the experiment period were recorded and classified into five different levels, i.e., unprocessed, minimally processed, processed culinary, processed and ultra-processed based on the NOVA food classification system. The proportion of UPF consumption was then obtained. Low SES was defined by the first and decile of an index which measures relative socio-economic disadvantage. The rest of the population was defined as high SES.

The research question was approached through the first four main models listed in Figure 11. The ranking of variable importance was utilised as the criteria for confounding variables. Then K means clustering allows data grouping based on similarities of those confounding variables. Subsequently, effect of confounding variables on obesity could be removed within each cluster when plotting the variance innumber of obese and non-obese people within the low SES group with increasing UPF consumption (Figure 6).

Methods used in each model

Classification Trees

Classification trees were built upon biomedical dataset with a new variable UPF consumption calculated, as introduced above. The sample was then refined to only include people in the low SES group. Two different complexity parameters presented similar models, one of which was obtained from 10-fold cross validation and the other guaranteed a model which returned the smallest error.

Random Forest

The input list of random forest was the same as those in classification trees. To perform a more comprehensive random forest, a grid of hyperparameters was needed to tune the model. The specific settings were inspired by an article issued by University of Cincinnati [24].

A cartesian grid search was then executed and returned a total of 90 models. Those models were run again, enabling to obtain importance ranking derived from two measures (i.e., impurity-based and permutation-based). Even though both approaches possessed its own drawbacks, considering that the purpose of constructing models in this study focused more on the importance of other confounding variables rather than specific quantitative inferences of how much UPF consumption affects obesity, it was sensible to refer to both rankings.

Logistic regression

Significant variables (i.e, those rank before UPF consumption) in classification trees and random forest were selected to fit a logistic regression model. We started with a full or null model and introduced Akaike's Information Criteria (AIC) and BIC. To further generalise our model, penalties were also brought in to fit ridge regression and two Lasso models built on two different λ, one of which returned the smallest prediction error and the other one returned a simpler model but lies within one standard error of the optimal value of λ. Also, 10-fold cross validation was used to generate mean error figure which was the criteria for choosing the best model. Finally, the **optimal lasso logistic model** shows that only part of variables of interest have been selected and UPF consumption was not included.

Neural Networks

We extracted variables from classification trees and random forest that presented more importance than UPF consumption. However, as many variables as possible were excluded to reduce the size of the input list. These variables included income, hypertensive disease and high cholesterol which presented similar importance to UPF consumption. The neural network was built on only 1 hidden layer with 6 neurons embedded. Followed by standardisation of these input variables, we divided the data into 2 samples, one of which would be used to train the data and the other one was used to test the model. Since all variables had been standardised, the weights in the neural network could be seen as the importance.

K Means Clustering

Important variables found from previous models were defined as confounding variables in terms of determining obesity. Their importance was the only criteria and we finally ended up with the following confounding variables, BMI, age, measured weight, self-perceived weight, sex, diastolic as well as systolic blood pressure. Additionally, according to the silhouette method, the dataset was classified into 5 clusters based on variables above to ensure they were controlled, meaning they would not significantly impact obesity as UPF consumption increased within each cluster. Therefore, we could understand how UPF consumption affected obesity in the low SES group.

Direct link between UPF consumption and obesity

The final density plot compares relative amount of obese people and non-obese people under each UPF consumption ratio, from which the relationship between UPF consumption and obesity was available.

Author contribution

Author (Name/SID)	Contributions
Ethan Li 490294826	Conceptualization, Validation, Investigation, Writing- original draft, Writing- review & editing, Project administration
Jasmine Leung 480131317	Conceptualization, Validation, Investigation, Writing- original draft, Writing- review & editing, Project administration
Rui Jing Zoe Zhang 490432473	Conceptualization, Methodology, Investigation, Writing- original draft, Writing- review & editing, Project administration

Mingze Gong 490284045	Methodology, Software, Validation, Formal analysis, Data curation, Writing-original draft, Writing- review & editing, Visualization
Haoru Yang 490199440	Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - Original Draft
Aoni Fu 490050978	Methodology, Software, Data curation, Visualization, Validation, Writing-original draft, Formal analysis

References

[1] Beslay M, Srour B, Méjean C, Allès B, Fiolet T, Debras C, et al. Ultra-processed food intake in association with BMI change and risk of overweight and obesity: A prospective analysis of the French NutriNet-Santé cohort. PLoS Med 2020;17. https://doi.org/10.1371/JOURNAL.PMED.1003256.

[2] Houle B. How Obesity Relates to Socioeconomic Status | PRB. Colorado, USA: 2013.

[3] Monteiro CA, Cannon G, Levy RB, Moubarac JC, Louzada MLC, Rauber F, et al. Ultra-processed foods: what they are and how to identify them. Public Health Nutr 2019;22:936–41. https://doi.org/10.1017/S1368980018003762.

[4] Brewer CJ, Balen AH. The adverse effects of obesity on conception and implantation. Reproduction 2010;140:347–64. https://doi.org/10.1530/REP-09-0568.

[5] Thorburn AW. Prevalence of obesity in Australia. Obes Rev 2005;6:187–9. https://doi.org/10.1111/J.1467-789X.2005.00187.X.

[6] Myung J, Jung KY, Kim TH, Han E. Assessment of the validity of multiple obesity indices compared with obesity-related co-morbidities. Public Health Nutr 2019;22:1241–9. https://doi.org/10.1017/S1368980019000090.

[7] Austin GL, Ogden LG, Hill JO. Trends in carbohydrate, fat, and protein intakes and association with energy intake in normal-weight, overweight, and obese individuals: 1971-2006. Am J Clin Nutr 2011;93:836–43. https://doi.org/10.3945/AJCN.110.000141.

[8] Machado PP, Steele EM, Levy RB, da Costa Louzada ML, Rangan A, Woods J, et al. Ultra-processed food consumption and obesity in the Australian adult population. Nutr Diabetes 2020 101 2020;10:1–11. https://doi.org/10.1038/s41387-020-00141-0.

[9] Gibney MJ. Ultra-Processed Foods: Definitions and Policy Issues. Curr Dev Nutr 2019;3. https://doi.org/10.1093/CDN/NZY077.

[10] Vandevijvere S, De Ridder K, Fiolet T, Bel S, Tafforeau J. Consumption of ultra-processed food products and diet quality among children, adolescents and adults in Belgium. Eur J Nutr 2019;58:3267–78. https://doi.org/10.1007/S00394-018-1870-3.

[11] Martínez Steele E, Raubenheimer D, Simpson SJ, Baraldi LG, Monteiro CA. Ultra-processed foods, protein leverage and energy intake in the USA. Public Health Nutr

2018;21:114–24. https://doi.org/10.1017/S1368980017001574.

[12] Derbyshire E. Are all ‘ultra-processed’ foods nutritional demons? A commentary and nutritional profiling analysis. Trends Food Sci Technol 2019;94:98–104. https://doi.org/10.1016/J.TIFS.2019.08.023.

[13] Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism 2019;92:6–10. https://doi.org/10.1016/J.METABOL.2018.09.005.

[14] Marchese L, Livingstone KM, Woods JL, Wingrove K, Machado P. Ultra-processed food consumption, socio-demographics and diet quality in Australian adults. Public Health Nutr 2021:1–11. https://doi.org/10.1017/S1368980021003967.

[15] Australian Institute of Health and Welfare. Australia’s Health 2016 4.4 Overweight and obesity (Australia’s health 2016) (AIHW). 2016.

[16] Shim JS, Shim SY, Cha HJ, Kim J, Kim HC. Socioeconomic characteristics and trends in the consumption of ultra-processed foods in Korea from 2010 to 2018. Nutrients 2021;13. https://doi.org/10.3390/NU13041120.

[17] Ghosh A, Charlton KE, Batterham MJ. Socioeconomic disadvantage and its implications for population health planning of obesity and overweight, using cross-sectional data from general practices from a regional catchment in Australia. BMJ Open 2016;6:e010405. https://doi.org/10.1136/BMJOPEN-2015-010405.

[18] Martínez Steele E, Raubenheimer D, Simpson SJ, Baraldi LG, Monteiro CA. Ultra-processed foods, protein leverage and energy intake in the USA. Public Health Nutr 2018;21:114–24. https://doi.org/10.1017/S1368980017001574.

[19] James A, Mendolia S, Paloyo AR. Income-based inequality of adolescent obesity in Australia. Econ Lett 2021;198:109665. https://doi.org/10.1016/J.ECONLET.2020.109665.

[20] Darmon N, Drewnowski A. Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. Nutr Rev 2015;73:643–60. https://doi.org/10.1093/NUTRIT/NUV027.

[21] Baraldi LG, Martinez Steele E, Canella DS, Monteiro CA. Consumption of ultra-processed foods and associated sociodemographic factors in the USA between 2007 and 2012: evidence from a nationally representative cross-sectional study. BMJ Open 2018;8:e020574. https://doi.org/10.1136/BMJOPEN-2017-020574.

[22] Moran AJ, Khandpur N, Polacsek M, Rimm EB. What factors influence ultra-processed food purchases and consumption in households with children? A comparison between participants and non-participants in the Supplemental Nutrition Assistance Program (SNAP). Appetite 2019;134:1–8. https://doi.org/10.1016/J.APPET.2018.12.009.

[23] Raubenheimer D, Simpson SJ. Nutritional Ecology and Human Health. Http://DxDoiOrg/101146/Annurev-Nutr-071715-051118 2016;36:603–26. https://doi.org/10.1146/ANNUREV-NUTR-071715-051118.

[24] Kim Y, Landgraf A, Colabianchi N. Living in High-SES Neighborhoods Is Protective against Obesity among Higher-Income Children but Not Low-Income Children: Results from the Healthy Communities Study. J Urban Health 2020;97:175–90. https://doi.org/10.1007/S11524-020-00427-9.

[25] Monteiro CA, Cannon G, Lawrence M, Laura Da Costa Louzada M, Machado PP. Ultra-processed foods, diet quality, and health using the NOVA classification system Prepared by. Rome: 2019.

[26] Blumfield M, Starck C, Keighley T, Petocz P, Roesler A, Abbott K, et al. Diet and Economic Modelling to Improve the Quality and Affordability of the Australian Diet for Low and Medium Socioeconomic Households. Int J Environ Res Public Heal 2021, Vol 18, Page 5771 2021;18:5771. https://doi.org/10.3390/IJERPH18115771.

[27] Huang Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Min Knowl Discov 1998;12:283–304.

[28] Random Forests · UC Business Analytics R Programming Guide. UC R Program Univ Cincinnati n.d. https://uc-r.github.io/random_forests (accessed November 10, 2021).

Supplementary Material