AMERICAN UNIVERSITY OF BEIRUT

DEVELOPING AND VALIDATING A MODEL FOR PREDICTING CHILD OBESITY IN LEBANON

by
HAZAR SAMIR SHAMAS

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Epidemiology and Population Health
of the Faculty of Health Sciences
at the American University of Beirut

Beirut, Lebanon
April 2023

AMERICAN UNIVERSITY OF BEIRUT

DEVELOPING AND VALIDATING A MODEL FOR PREDICTING CHILD OBESITY IN LEBANON

by
HAZAR SAMIR SHAMAS

Approved by:

Signature
Dr. Stephen McCall, Assistant Professor — Advisor
Department of Epidemiology and Population Health

Signature
Dr. Hala Ghattas, Associate Professor — Co-Advisor
Department of Epidemiology and Population Health
Department of Health Promotion, Education and Behavior
Arnold School of Public Health

Signature
Dr. Khalil El Asmar, Assistant Professor — Member of Committee
Department of Epidemiology and Population Health

Signature
Dr. Lara Nasreddine, Professor — Member of Committee
Department of Nutrition and Food Sciences

Date of thesis defense: April 27th, 2023

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name: _ Shamas_______________Hazar_____________Samir __________
Last First Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

☐ As of the date of submission

☒ One year from the date of submission of my thesis.

☐ Two years from the date of submission of my thesis.

☐ Three years from the date of submission of my thesis.

Hazar

______________________ Wednesday May 3rd, 2023_____________

Signature Date

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisors Dr. Stephen McCall and Dr. Hala Ghattas for their constant support and guidance throughout this study. It was a privilege to work with you on this study and work with the SCALE study dataset. Your expertise has helped me develop my understanding of nutritional epidemiology. I also extend my thanks to my committee members Dr. Khalil El Asmar and Dr. Lara Nasreddine for their insightful comments and expertise which greatly enhanced the quality of this research.

I thank the Center for Research on Population and Health team for constant support and informative suggestions.

May this study be the path to aid clinicians and children in need.

ABSTRACT
OF THE THESIS OF

Hazar Samir Shamas for Master of Science
Major: Epidemiology

Title: Developing and Validating a Model for Predicting Child Obesity in Lebanon

Introduction Child obesity, defined as the BMI at or above the 95th percentile for children of same age and sex is steadily climbing the ladder of public health concern. With child obesity manifesting in our society, it is crucial to predict risk of child obesity so that health interventions and context-specific policies can be implemented. Thus, the aim of this study is to develop and internally validate a prediction model for child obesity in Lebanon. ***Methods*** This is a cross-sectional study of 2,125 school students from the SCALE study. The SCALE study employed a two-stage sampling method aiming to include a representative sample of 8-12 years old students in Greater Beirut. The first stage used a random sample of 50 schools stratified into public, private, and private free schools, which were identified by a list provided by the Ministry of Education. The second stage included randomly assigning 50 students from grades 4,5,6 from each of the 50 schools that accepted to join the study in the first stage after parents signed the consent forms. This study produced prediction model discrimination and calibration slope for models developed using backward logistic regression as a statistical approach and LASSO, Ridge, Elastic net as machine learning approach. Two binary outcomes were assessed in this study: obese versus non-obese and obese or overweight versus normal or thin. Seventeen predictors were included in the prediction models: age, gender, food insecurity, nutrition knowledge, school type, crowding index, parent marital status, mother education, screen time, TV time, eating while on screen, physical activity, fast food consumption, fruit availability, vegetable availability, sugar sweetened beverages availability at home, having an obese mother. ***Results*** The sample size included 1,409 participants of median age 11 years (10-12). The best performing model is that of Lasso adaptive with discrimination of 0.632 (0.60-0.66) and C-slope 0.968 (0.72-1.21) with outcome obese or overweight. Ten predictors were selected by that model where older age, being female, having married parents, adequate availability of fruits at home, crowding index less than 3 are protective factors and eating while on screen, child nutrition knowledge, tv viewing time more than 2 hours, vegetable availability plus having an obese mother as risk factors for child obesity or overweight***. Interpretation*** This study showed a well calibrated predictive model with moderate discrimination. Such prediction model in clinical settings could be used to prevent the risk of child obesity and thus reduce the risk for other potential non-communicable diseases while utilizing fewest resources possible. A range of policies could be implemented by parents, schools, and the government as a product of this study. Parents are requested to limit their child's TV time and encourage their child to consume fruits. The strength of this study includes measuring height and weight in duplicate to prevent misclassification, using close ended questionnaires, usage of sampling weights in analysis, defining variable cutoffs based of

systematic reviews and expert opinion. Limitations of this study includes absence of some variables present in the literature, presence of variables with missing data above 10%, potential recall bias and differential non-response bias.

TABLE OF CONTENTS

ILLUSTRATIONS

Figure

TABLES

Table

ABBREVIATIONS

SCALE: School and Community Drivers of Child Diets in Arab Cities; Identifying Levers for Intervention

BMI: Body Mass Index

JBI: Joanna Briggs Institute

ACE: Adverse Childhood Experience

OR: Odds Ratio

PPV: Positive Predicted Value

NPV: Negative Predicted Value

AUC: Area under curve/ discrimination

MI: Mutual information

SS: Sample Space

RF: Random Forest

DT: Decision Trees

BN: Bayesian Networks

SVM: Support Vector Machines

SVM-RF: Support Vector Machine Recursive Feature Elimination

GBN: Gradient Boosting Machines

KNN: k-nearest neighbors

RBF: Radial Basis Function

ANN: Artificial Neural Networks

O/N_O: Obese/Non-Obese

OW/N_OW: Overweight/ Non-Overweight

TN: True Negative

TP: True Positive

MCC: Matthews Correlation Coefficient

EHR: Electronic Health Records

NA: Not applicable

FF: Fast food

SSB: Sugar Sweetened Beverages

CHAPTER 1

INTRODUCTION

1.1. Global Perspective on Child Obesity

Child obesity, defined as a BMI at or above the 95^{th} percentile for children of same age and sex is steadily climbing the ladder of public health concern [1].Worldwide, there has been an increase from 45 million to more than 124 million obese children aged 5-16 from the year 2000 to 2016 especially in the presence of globalization accompanied with rapid technological advancement in the food industry [2]. Thus, declared as a leading public health problem, obesity is likely to stay with those children till adulthood leading to the development of non-communicable diseases such as type 2 diabetes or cardiovascular disease earlier in life [3].

1.2. Child obesity in Lebanon

Apart from the global perspective of child obesity, Lebanon has also its fair share of troubles and health epidemics. Since 2019, and after the Lebanese uprising movements that lead to temporary closures of schools, companies and work, Lebanon was also hit with the COVID 19 pandemic that changed children's lifestyle from going regularly to school to being taught online at home. This deterioration in the economy along with the child's sedentary behavior due to staying at home for long hours facing the screen for weeks in the time where children at that age are encouraged to be physically active and eat nutritious food likely aided the increase in food insecurity and child obesity. Trends in child obesity in Lebanon have been shown to be on the increase; two national cross-sectional surveys administered on Lebanese school children aged 6-9 years old between

1997 and 2009 found the odds of child obesity in 2009 to be 2 times that of child obesity in 1997 [4]. This number is expected to further increase due to the increase in factors, such as screen time and high caloric diet.

1.3 Rationale for this study

As child obesity is manifesting in our society, it will be important to develop tools that allow the prediction of child obesity risk so that relevant health interventions and context-specific policies can be implemented. However, with the economic deterioration that Lebanon is facing along with the COVID-19 epidemic that transformed learning from schools to learning from home, most of the studies of school child obesity in Lebanon are now outdated. In addition, no prior studies have focused on developing a prognostic model to predict child obesity including factors at the household and child level.

It is to be noted that this study uses data from the SCALE (School and Community Drivers of Child Diets) study conducted at the Center for Research on Population and Health (CRPH) in the Faculty of Health Sciences (FHS) at the American University of Beirut (AUB) [5].

CHAPTER 2

LITERATURE REVIEW

To develop a tool that permits the prediction of child obesity risk at the household and child level, two literature reviews were conducted, the first focusing on finding variables associated with child obesity while the second examines commonly used prognostic models of child obesity. Results of both reviews aided the fitting and analysis of the model.

2.1. First Literature Review

2.1.1. Objective of First Literature Review

The purpose of the first literature review was to identify risk factors associated with child obesity.

2.1.2. Risk Factors of Child Obesity

2.1.2.1. Literature Review Methods

Due to the large number of articles that examine risk factors of obesity, this literature review includes only systematic reviews that examine the association between child obesity and its predictors or risk factors. A search was conducted in three databases: PubMed, MEDLINE, and Embase without date restriction. Example of the search terms included: "risk factors" OR "predictors" OR "associations" AND "childhood obesity" OR "child obesity". Systematic reviews were included if they met the following criteria: English language, full text available online, outcome is child obesity or BMI for children aged 6-18. This review focuses on children aged 6-18 years as this is the age range that the data source (SCALE study) for the present analysis falls within.

Systematic reviews found were first assessed based on their titles and abstract where 659 were excluded using EndNote either due to: not actually being a systematic review, no available full text, no proper definition of variable, BMI was measured only for less than 6-year-old children, or due to not mentioning association between variable and obesity/BMI. Risk factors available in included systematic reviews are presented in Table 1. Table 1 also indicates which variables from the literature were also included in the SCALE study questionnaire.

The quality of each included systematic review used to provide evidence on the association of the variables present in the SCALE study and child obesity was assessed using Joanna Briggs Institute (JBI) critical appraisal checklist for systematic reviews and research syntheses (Appendix 1).

2.1.2.2. Quality Assessment

JBI critical appraisal checklist for systematic reviews is a quality appraisal checklist that assesses using 11 items the possibility of bias in the study's design, conduct and analysis. For each item answered with a "Yes" the JBI score of the systematic review increases by 1 point on a total score of 11. The scoring of each systematic review can be found in Appendix 2. Though no studies were removed on the basis of JBI checklist, the importance of the JBI critical appraisal lies in assessing the quality of the systematic reviews and variable definition which aided interpretation of Table 3 [6].

2.1.2.3. Literature Review Results

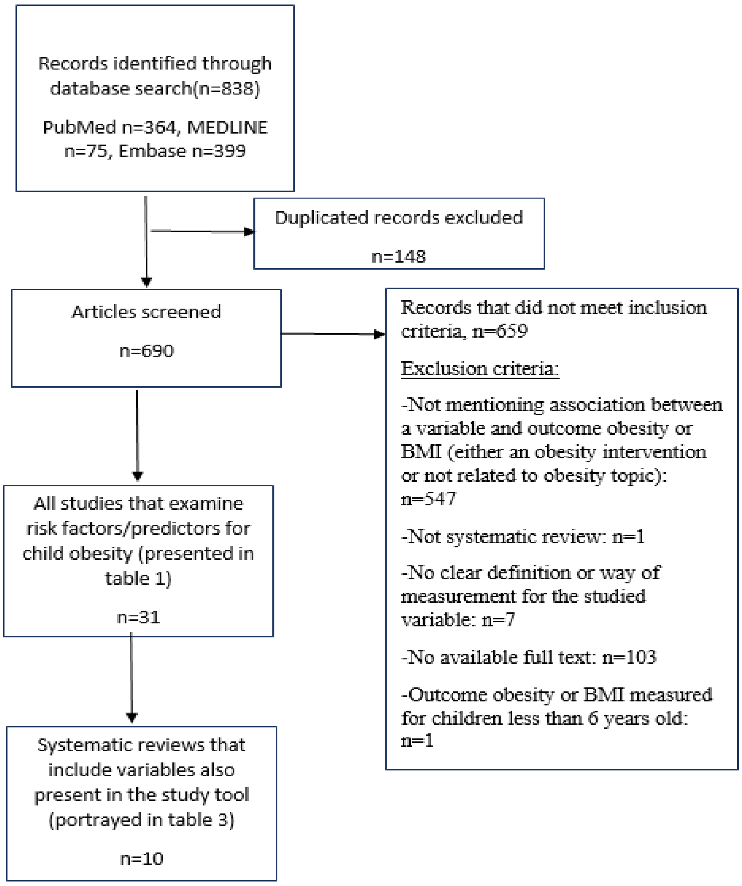

Figure 1. Flowchart representing selection of systematic reviews related to child obesity

Variables were named according to their availability in systematic reviews and the SCALE study questionnaires.

All systematic reviews had good score (greater or equal to 6). However, the majority of the results could not present a pooled odds ratio (OR) or relative risk (RR) due to heterogeneity between the studies and variable definition.

Table1. Table of all predictors associated with child obesity established from systematic reviews

Predictors	Reference number in List	Available in SCALE study
1. Socio-Demographic: *a. <u>Child Related:</u>*		
Age	[7], [8]	✓
Gender	[7], [8]	✓
School Type	[7], [8]	✓
Food insecurity	[9]	✓
b. <u>Parent Related:</u>		
Education level	[7]	✓
Marital Status	[10]	✓
Crowding Index	[7], [8]	✓
2. Lifestyle factors: a. <u>Child Related:</u> ***i.*** <u>Sedentary Behaviour, Physical Activity and Nutrition Knowledge:</u>		
Screen time	[7], [8], [11]	✓
Watching Television	[7], [8]	✓
Eating while on screen	[8]	✓
Physical Activity	[7]	✓
Video Gaming	[12]	
Short sleep duration	[7], [13]	
Child Nutrition Knowledge	[14]	✓
Adverse Childhood Experience (sexual abuse or violence)	[15], [16],[17]	
ii. <u>Eating Habits and Food Availability at Home:</u>		
Consumption of fast food and calorie high dense food	[7],[18],[19]	✓
Sugar sweetened beverages Availability in household	[7],[19], [20]	✓
Vegetable Availability in household	[7], [8],[18]	✓
Fruit Availability in household	[7], [8], [18]	✓
Meal Skipping	[18],[19]	
iii. <u>Health:</u>		
Chronic disease (Asthma)	[21]	
High Birthweight	[22]	
b. <u>Parent Related:</u>		
Low Maternal Education	[23]	
Maternal Obesity (during pregnancy)	[10]	
Preconception Obesity	[24]	
Maternal symptoms of depression	[25]	
Maternal Smoking	[23] ,[26]	

3. Family History of Obesity:		
Obese Parent	[27]	✓
Presence of DNA methylation	[28]	
4. School and Neighbourhood Environment:		
Food outlets near schools	[29]	✓
Convenience Stores/Supermarket in neighbourhoods	[30],[31]	
Physical activity facility in neighbourhood	[32]	
5. Other Factors:		
Air Pollution	[33],[34] [35],[36]	
Bisphenol exposure	[37]	
Birthed through C-section	[38]	

2.1.2.4. Predictors in the Literature

Using the literature, 35 identified potential predictors were associated with child obesity (Table 1).These predictors include: age (continuous), sex (male/female), type of school (private/public), child nutrition knowledge (pass/fail), food insecurity (secure/insecure), parent education level (intermediate and below/secondary and above), parent marriage status (single/married), crowding index (≤ 3/ >3)[39], screen time (<2 hours per day /≥2 hours per day), watching television (<2 hours per day /≥2 hours per day), times of eating while on screen (continuous), physical activity (<3 days per week/≥3 days per week), video game time (<2 hours per day/ ≥ 2 hours per day), short sleep duration (<6 hours per day/ ≥ 6 hours per day), adverse childhood experience (any form of ACE, categorical), consumption of fast food (< 3 times per week/ ≥3 times a week), consumption of sugar sweet beverages (<4 times per week/≥ 4 times per week), availability of vegetables (limited/adequate), availability of fruits (limited/ adequate), meal skipping- number of meals skipped (continuous), chronic disease (yes/no), birthweight (continuous), maternal education (intermediate and below/secondary and above), preconception obesity (yes/no), maternal symptoms of depression (yes/no), maternal smoking- number of cigarettes per day (continuous), obese parent (yes/no), presence of DNA methylation (yes/no), food outlets near schools (yes/no), convenience stores in child neighborhood (yes/no), physical activity facility in neighborhood (continuous) ,air pollution (ex: NO_2 equal or more/less than 10 μg/m^3), bisphenol exposure (< 2 μg/L in urine/ ≥2 μg/L in urine),C- section birth (yes/no).

2.1.2.5. Definition of predictors in this study according to the SCALE study

Common predictors present in the literature and the SCALE study will be used for child obesity model development. To ensure the reliability of this study, definitions of predictors presented in the SCALE study that are common with the literature are presented as follows:

-Screen time: Time spent facing the screen for homework, chat, and surfing (does not include TV-time)

-Physical activity: activity that makes a person breathe hard and increases a person's heart rate.

-Sweetened sugar beverages: Liquids sweetened with various forms of sugar such as honey or brown sugar. Example: boxed juice, sweetened tea, and soft drinks.

-Child Nutrition Knowledge: Assessed using a set of 7 general nutrition questions in the SCALE study derived from the General Nutrition Knowledge questionnaire and adapted into the local context [40]

-Crowding: Number of people living in the household divided by the number of rooms in the household excluding bathrooms, kitchen, garage, and unclosed balcony.

2.2. Second Literature Review

2.2.1. Objective of Second Literature Review

The second literature review aim is to identify commonly used prognostic models and measures of model fit for child obesity.

2.2.2. Prognostic Models of Child Obesity

2.2.2.1. Literature Review Methods

A second literature review was conducted to examine commonly used prognostic models and prognostic model measures of model fit to predict child obesity. A search was conducted in three databases: PubMed, MEDLINE, and Embase without date restriction. Example of search terms included: “prediction models” AND “child BMI” OR “child obesity”. Studies were included if they met the following criteria: English language, full text available, outcome is child obesity or BMI for 6–18-year-old children, and at least one statistical or machine learning model was used.

2.2.2.2. Literature Review Results

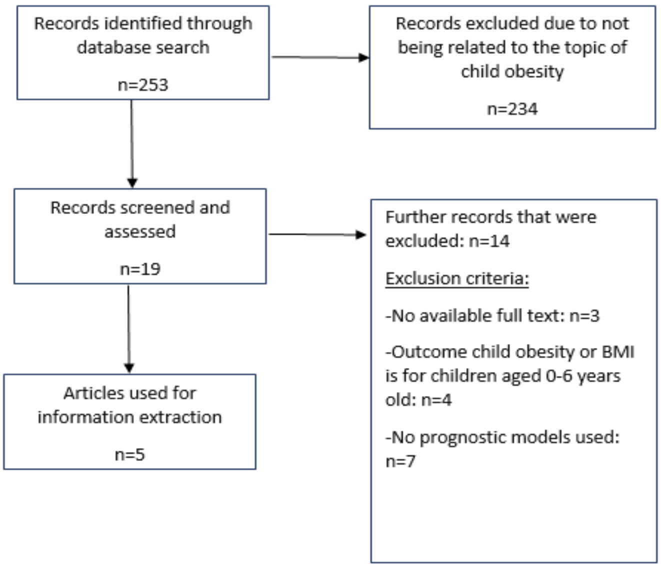

Figure 2. Flowchart representing prognostic model article selection

The search resulted in 253 studies. These 253 studies were first assessed based on their titles and abstract using EndNote where 234 were excluded due to not being related to the child obesity topic in the first place. Another 14 articles were excluded for: no available full text, outcome child obesity or BMI is only for children less than 6 years old, and no prognostic model used. Out of the 5 remaining studies, 1 was a systematic review that presented statistical and machine learning models to predict child obesity and a comparison between these two methods. Variables presented in the final models and model performance measurements extracted from the 5 studies are presented in Table 2.

Table2. Findings on child obesity prognostic models and prognostic model measurement

Reference	**Sample Age and Size**	**Data Source**	**Statistical Method**	**Machine Learning Method (ML)**	**Variables in Final Model**	**Outcome and Age at Outcome**	**Measurements**	**Results including (AUC/C-slope)**
Pang et al. [41](2021)	Age: age >2 and <8 years Sample Size: 27,203	Pediatric Big Data derived from HER at the Children's Hospital of Philadelphia (CHOP)	NA	Compared 7 ML models: DT, GNB, NN, BNB, SVM, RBF, XGB	Present in Appendix 3	O/ N_O (age/sex adjusted BMI>95th percentile) for 2-7 years old	-Models compared using Cochran's Q test and post-hoc McNemar pairwise testing. -Sensitivity analysis, stratified by race and geographic location -AUROC -Sensitivity -Precision -FI-score, -Accuracy -Specificity	-KNN imputation and using a proxy category for missing values yielded comprehensive results for not MCAR variables -Sensitivity analysis yielded new factors such as respiratory rate and head circumference -XGBoost outperformed other models in predicting child BMI with 0.81 AUC
Potter et al. [42] (2018)	Care givers: 69 Children (age 5-11 years) 148 Total: 217	From: 1-National Institute of health (NIHR) 2-Opportunity sample of overweight and lean children from a local interactive center	Multiple Linear Regression	NA	-Child IP (Kcal) -Parent IP (Kcal) -Child MP (Kcal) -Parent MP(Kcal) -Parent BMI	BMI (continuous) for 5-11 years old	Adjusted R^2 (Model fit)	Bigger portion Sizes is a predictor for child obesity

Shi et al. [43](2021)	Age: 2-18 years Sample Size: 426,813	Osakidetza databases of the Public Health Provider in the Basque Country (Spain)	NA	BFSMR representatives: -Filter (MI)/SVM-RFE -LASSO/Ridge -RF	Present in Appendix 4	OW/N_OW (age/sex adjusted BMI>90th percentile) for 2-18 years old	-Accuracy -F-score	-Variables with more than 10% missing for numeric variables were replaced by 0 ,and dropped for categorical variables -LASSO, Ridge and Filter had highest F score of 0.912 and 0.915 -SVM-RFE had highest accuracy 0.845
Hammond et al. [44] (2019)	Children (pre-pregnancy through age two) 52,945 Mothers: 36,244	EHR data from patients in a safety net health system called Family Health Centers at NYU Langone from years 2008-2016 retrospective cohort	Logistic regression	-LASSO -RF -Gradient Boost	-Ethnicity -Race -Marriage status -Maternal Birthplace -Maternal Diagnosis of illness -Infant Diagnosis of illness	**Statistical:** -O/ N_O(age/sex adjusted BMI>95th percentile) **ML:** -BMI(continuous) for 5-6 years old	-Sensitivity -Specificity -PPV -AUC -FI score -MCC -N(Obese) -N not obese (TN+TP)	-LASSO achieved highest AUC (0.81 for girls and 0.76 for boys)
Colmenarejo [45] (2020) (Systematic review)	NA	NA	-Logistic regression -Stepwise logistic regression	RF,DT,NB,SVM,LASSO,GBN,KNN,RBF	-Parent BMI -Sex -Birthweight -Maternal Smoking -Parent Education -Recent weight gain -Breastfeeding period	**Statistical**: BMI (continuous) **ML:** -BMI (continuous) -O/ N_O (age/sex adjusted BMI>95th percentile) Ranging from 2-17 years old	-Bootstrap -External validation -Cross validation -AUROC -PPV -NPV -Sensitivity	-ML in general had better prediction accuracy that statistical methods -ANN had largest accuracy -SVM largest sensitivity

PPV: Positive Predicted Value, NPV: Negative Predicted Value, AUC: Area under curve; SS: Sample Space; MI: Mutual information: Random Forest, DT: Decision Trees, BN: Bayesian Networks, SVM: Support Vector Machines, SVM-RFE Support Vector Machine Recursive Feature Elimination, GBN: Gradient Boosting Machines, KNN: k-nearest neighbors, RBF: Radial Basis Function, ANN: Artificial Neural Networks; O/N_O: Obese/Non-Obese, OW/N_OW: Overweight/ Non-Overweight, TN: True Negative, TP: True Positive, MCC: Matthews Correlation Coefficient EHR: Electronic Health Records

2.2.2.3. Common Model Performance Methods and Variables Used

None of the prediction models listed provided predictors that were always included in the models. Variation in prognostic models used for child obesity existed in these 5 studies where 1 study used only statistical model, 2 used only machine learning,1 compared between both machine learning and statistical models and 1 systematic review pooled the study results and compared performance of statistical models with machine learning models. Articles and systematic reviews that used machine learning models reported model discrimination using AUC (ranging from 0.63-0.89) and performed sensitivity analysis for the final model. However, none reported a calibration slope.

Common variables present in this literature review and SCALE study include age (continuous), sex (male/female), obese parent (yes/no), physical activity (<3 days per week/≥3 days per week), screen time (<2 hours per day /≥2 hours per day), parent marital status (single/married), parent education (intermediate and below/secondary and above).

2.3. Aim and Objectives

The aim of the present study is to develop and internally validate a prediction model for child obesity in Lebanon at the household and child level using candidate predictors collected through the SCALE study.

Objectives:

1- To develop a prediction model for child obesity at household and child level using standard regression tools such as multiple logistic regression.

2- To develop a prediction model for child obesity using LASSO, Ridge and Elastic net regression and compare their output to that of backwards logistic regression

CHAPTER 3

METHODS

3.1. Study design and study location

This study uses secondary data from a cross-sectional study entitled SCALE which was conducted in Lebanese schools.

3.2. Sampling design and study population

The SCALE study employed a two-stage sampling method aiming to include a representative sample of 8-12 years old students living in Greater Beirut.

The first stage used a random sample of 50 schools stratified into public, private and private free schools, which were identified by a list provided by the Ministry of Education.

The second stage included randomly assigning 50 students from each school in grade 4,5,6 from the schools that agreed to join the study in the first stage after parent signed the consent forms.

3.3. Data Collection

To collect data, the SCALE study used structured questionnaires that include standardized tools and validated scales such as the GNKQ and the Child Food Security Questionnaire administered to the students and students' parents [46].

3.4. Studied Outcome

In the SCALE study, height and weight of the child were measured in duplicate according to standard protocols using a standard stadiometer and calibrated weighing scale[5]. In case of a 1cm or 1kg difference between 2 measures, height and weight were measured a third time. The average between the closest value of the first two height and weight measures with the third measures was then calculated. Overweight and obesity are defined based on sex and age specific +1 and +2 BMI *z*-scores according to the WHO new growth standards respectively. The WHO AnthroPlus 2007 software was used to calculate BMI *z*-score for each specific age and sex [47].

Two outcomes were studied:

Primary outcome: Obese/ Non-Obese (BMI-for-age Z-score>2 vs Z-score ≤2)

Secondary Outcome: Obese or Overweight vs Normal or Thin (BMI-for-age Z-score>1 vs Z-score ≤1)

3.5. Study Predictors

SCALE study includes 18 of the variables identified in the literature at the household and child level (Table 3). However, though the SCALE study has food outlets near schools variable, this study focuses on predictors at the household and child level, thus food outlets near schools variable won't be included meaning that 17 variables will be present in the analysis.

These 17 variables measured in the SCALE study have been collected using structured questionnaires which include: age (continuous), sex (male/female), type of school (private/public), child nutrition knowledge (fail/pass), food insecurity (secure/insecure), parent education level (intermediate and below/secondary and above), parent marital status (single/married), crowding index (≤ 3/ >3 persons/room), screen time (<2 hours per day /≥2 hours per day), watching television (<2 hours per day /≥2 hours per day), eating while on screen (continuous), physical activity (<3 days per week/≥3 days per week), consumption of fast food (< 3 times per week/ ≥3 times a week), consumption of sugar sweet beverages (<4 times per week/≥ 4 times per week), availability of vegetables (limited/adequate), availability of fruits (limited/ adequate), obese parent (yes/no).

Mother education and obese mother were used as proxies for parent education and obese parent variables.

It is to be noted that the study predictors' cutoffs were decided based on their mostly used cutoff point in the systematic reviews.

Table 3. Predictors associated with child obesity extracted from systematic reviews and are available in SCALE study questionnaires

Predictors:	Reference number in List
1. Socio-Demographic: *a. Child Related:*	
Age	[7], [8]
Gender	[7], [8]
School Type	[7], [8]
Food insecurity	[9]
b. Parent Related:	
Education level	[7]
Marital Status	[10]
Crowding Index	[7], [8]
2. Lifestyle factors: *a. Child Related:* *i.* Sedentary Behaviour, Physical Activity and Nutrition Knowledge*:*	
Screen time	[7], [8], [11]
Watching Television	[7], [8]
Eating while on screen	[8]
Physical Activity	[7]
Child Nutrition Knowledge	[14]
ii. Eating Habits and Food Availability at Home:	
Consumption of fast food and calorie high dense food	[18],[19]
Sugar sweetened beverages Availability	[7],[19] ,[20]
Vegetable Availability	[7],[8],[18]
Fruit Availability	[7],[8], [18]
iii. Health**:**	
b. Parent Related:	
3. Family History of Obesity:	
Obese Parent	[27]

3.6. Missing Data

The largest amount of missing data was 41% in the obese mother variable. Missing data was tested to be missing at random, so a subset analysis was performed to keep the events per variable ratio (EPV) above 10. However, with the obese mother variable still having 16% missing, 2 approaches were undertaken to tackle this issue. The first "Case 1" was to remove participants with missing data on the obese mother variable and the other "Case 2" was to replace the missing by a proxy category "99" aiming to retain data and check what could have happened had this variable had complete data.

3.7. Statistical Analysis

3.7.1. Cluster and Intraclass Correlation Testing

The intraclass correlation for mixed model outcome obese with school cluster and mixed model with school and class cluster were $1.83 * 10^{-0.37}$ and $2.24 * 10^{-0.34}$ which are less than 0.4 respectively [48]. Additionally, a non-significant likelihood ratio test was present between a: 1) mixed model with school cluster and a 2) mixed model with school plus class cluster indicating that there is no need to account for clustering.

3.7.2 Machine Learning Models and Definitions

Since child obesity is a main modifiable risk factor for the development of multiple non-communicable diseases in the future, identifying the most important predictors provides a mean to inform targeted interventions. Machine learning models have the potential to become extremely useful in predicting child obesity especially in such situations where we have plenty of predictors with various levels of influence on the outcome. Thus, machine learning models such as Least Absolute Shrinkage and Selection

Operator- Lasso (cross validation), Lasso (Adaptive), Lasso (BIC), Ridge, and Elastic net were implemented along with backwards logistic regression.

The common 17 variables present in the literature along with the SCALE study have been fit into backward logistic model, Elastic net, Ridge, LASSO (cv, adaptive and BIC) models for both outcomes obese and obese plus overweight. Within the same outcome, the model with a calibration slope closest to one and has a higher AUC was chosen as the optimal model.

Lasso Definition:

Least Absolute Shrinkage and Selection Operator (LASSO) in all types (Cross validation, adaptive, BIC) is a type of regression that relies on shrinkage by selecting from great number of variables that leads to a more parsimonious model. This selection is performed by shrinking the coefficients of the variables to zero thus eliminating the variable from the model using L1-norm penalty term (sum of absolute coefficients) and keeping the best subset of variables to predict the outcome [45]. It is to be noted that the shrinkage requires the selection of a tuning parameter (lambda) that determines the amount of shrinkage.

Lasso Adaptive Definition:

Lasso adaptive uses two-stage process to find the proper penalty for each variable. In the first stage, a preliminary estimate of the regression coefficients is obtained by using a relatively small penalty parameter but applied to all predictors. However, in the second stage, the penalty parameter is adjusted using a weight vector based on the magnitude of the preliminary estimate. Specifically, larger penalties are applied to the less important variables, and smaller penalties are applied to the more important variables.

3.7.3. Analysis with Backward logistic regression

All predictors identified were entered into a multivariable logistic regression and removed using stepwise backwards method using a $P < 0.157$, a proxy for the Akaike Information Criterion (AIC) where predictors are removed to obtain the lowest AIC [49]. Multicollinearity of variables was assessed beforehand using the variance inflation factor (VIF) were VIF greater than 5 indicated collinearity. The final model's selection of predictors, discrimination, and calibration estimates were internally validated using bootstrap methods, in which 500 bootstrap samples with replacement were used to validate the model selection process and generate an estimate of optimism, an optimism-adjusted estimates of C statistic, and an optimism-adjusted calibration plot.

Bootstrap shrinkage was applied to the final apparent model. Internal validation of the model aids to prevent overfitting and ensure that the model produced measures what it is intended to measure.

3.7.4. Model Measure

All 17 candidate predictors were fit into the machine learning models. It is to be noted that variables were binary except for age and eating in front of the screen which were found to have a linear association with the outcome obese and the outcome obese or overweight versus normal or thin.

The final model's performance resulting from backward logistic regression, lasso, ridge, and elastic net were assessed through their discrimination capabilities using the

Area Under the Receiver Operating Curve (AUC), that ranges from 0.5 to 1.0 where a value of 1.0 represents perfect discriminative ability between those with and without the outcome and 0.5 denotes a discriminative ability equal to chance. The calibration of the final model, which describes the agreement between observed and model's prediction was also evaluated through the C-slope and calibration plot that categorizes children into 10 groups according to predictive probabilities, where the mean predicted risk within each of these groups is plotted against the mean observed proportion of events. In the case of perfect calibration, the graph shows a diagonal line with 0 as an intercept and 1 as a slope. However, C-slope less than 1 suggests overfitting in the model, meaning that respondents with high risk of the outcome have overestimated risk predictions while those with low risk of the outcome have underestimated risk predictions and vice versa. Model selection was primary based on having a calibration slope close to 1, and in case multiple models of the same outcome had a calibration slope close to 1 within the same outcome, the model with the higher discrimination/area undercurve (AUC)/ c-statistic was chosen.

A sensitivity analysis included the outcome being obese and overweight versus normal and thin as the outcome was performed. All analyses were conducted using Stata/SE statistical software version 17 (STATA Corp).

CHAPTER 4

RESULTS

4.1. Subset Chosen

A subset of the 2,125 students who participated in the SCALE study was chosen. Of the 2,125 students, 659 students were excluded due to the absence of parent interviews and further 57 students were excluded due to the absence of data on their BMI which decreases the sample size to 1,409 students in total. However, the obese mother variable still had 16% missing.

A table including the characteristics of the original sample size 2,125 students with the weighted percentages whose BMI z-score has been calculated is presented in Supplementary table 4. The characteristics of the sample of the secondary outcome are also included in Supplementary table 5.

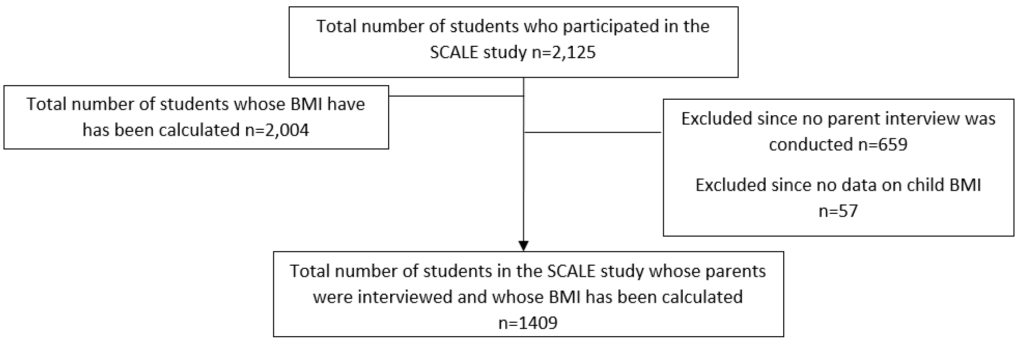

Figure 3. Flow chart representing the students included in the study

4.2. Unadjusted Analysis

Out of the 1,409 participants, 241 students were obese. The study sample included 677 boys and 732 girls where girls had lower odds of being obese (OR: 0.38; CI: 0.25-0.56) and to be obese or overweight (OR: 0.49; CI: 0.39-0.63). Moreover, having married parents acted as a protective factor for child obesity (OR: 0.31; CI: 0.17-0.54). Additionally, consuming fast food and having an obese mother increased the odds of being an obese or overweight child (OR: 2.18; CI: 1.2-3.96) and (OR: 1.79; CI: 1.08-3.23) respectively.

Table 4. Characteristics of students in the study sample size

	Total n=1,409	Non-Obese (z-score ≤ 2) n=1,168(83.5)	Obese (z-score>2) n=241(16.5)	Odds Ratio¹	95% CI	Odds Ratio²	95% CI
	n(%)	n(%)	n(%)				
Socio-demographics							
Age median (IQR)	11(10-12)	11(10-12)	11(10-12)	0.81	0.70-0.94	0.91	0.81-1.02
Child Sex							
Boy	677(51.6)	522(77.4)	155(22.6)	1		1	
Girl	732(48.4)	646(90.0)	86(10.0)	0.38	0.25-0.56	0.49	0.39-0.63
School Type							
Public	896(32.2)	739(82.5)	157(17.5)	1		1	
Private	513(67.8)	429(84.0)	84(16.0)	0.89	0.61-1.31	1.04	0.79-1.37
Child Food insecurity							
Secure	949(81.7)	786(84.2)	163(15.8)	1		1	
Insecure	341(18.3)	285(82.3)	56(17.7)	1.14	0.72-1.83	0.98	0.59-1.63
Missing		97	22				
Nutrition Knowledge							
Fail	885(59.2)	733(83.0)	152(17.0)	1		1	
Pass	514(40.8)	428(84.6)	86(15.4)	0.88	0.67-1.16	1.18	0.89-1.55
Missing		7	3				
Mother Education							
Intermediate & Below	544(25.7)	449(82.5)	95(17.5)	1		1	
Secondary & Above	805(74.3)	668(83.9)	137(16.1)	0.9	0.56-1.43	1.25	0.83-1.88
Missing		51	9				

Parent Marital Status							
Single (Widowed or Divorced)	87(5.1)	68(62.8)	19(37.2)	1		1	
Married	1,296(94.9)	1,078(84.4)	218(15.6)	0.31	0.17-0.54	0.53	0.27-1.06
Missing		22	4				
Crowding Index							
≤3people/room	1,293(97.4)	1,069(83.5)	224(16.5)	1		1	
>3 people/room	109(2.6)	94(86.1)	15(13.9)	0.81	0.45-1.48	0.76	0.42-1.36
Missing		5	2				
Lifestyle Factors							
Time spent on screen							
< 2hrs/day	1,052(73.3)	887(84.7)	165(15.3)	1		1	
≥ 2hrs/day	337(26.7)	264(80.3)	73(19.7)	1.36	0.92-2.01	1.23	0.91-1.69
Missing		17	3				
Watching TV							
< 2hrs/day	1,046(74.7)	863(83.7)	183(16.3)	1		1	
≥ 2hrs/day	349(25.3)	294(83.5)	55(16.5)	1.04	0.67-1.52	1.32	0.92-1.89
Missing		11	3				
Eating on Screen(days/week) median(IQR)	2(0-5)	2(0-4)	2(0-7)	1.04	0.95-1.15	1.04	0.97-1.12
Missing		22	7				
Physical Activity							
<3days/week	801(51.5)	669(84.7)	132(15.3)	1		1	
≥3days/week	587(48.5)	483(82.4)	104(17.6)	1.18	0.83-1.67	1.08	0.85-1.37
Missing		16	5				
Fruit Availability							
Limited	547(24.1)	456(81.7)	91(18.3)	1		1	
Adequate	861(75.9)	711(84.1)	150(15.9)	0.84	0.51-1.39	0.95	0.72-1.25
Missing		1					
Vegetable Availability							
Limited	348(14.7)	297(84.5)	51(15.5)	1		1	
Adequate	1,061(85.3)	871(83.3)	190(16.7)	1.09	0.75-1.59	1.35	0.88-2.05

SSB Availability							
Limited	999(63.3)	833(84.6)	166(15.4)	1		1	
Adequate	409(36.7)	334(81.6)	75(18.4)	1.24	0.75-2.04	1.09	0.71-1.66
Missing		1					
FF Consumption							
<3days/week	1,333(95.7)	1,105(83.9)	228(16.1)	1		1	
≥3days/week	55(4.3)	48(78.5)	7(21.5)	1.41	0.69-2.9	2.18	1.2-3.96
Missing		15	6				
History of Obesity							
Obese Mother							
No	1,015(90.1)	854(84.0)	161(16.0)	1		1	
Yes	162(9.9)	116(74.5)	46(25.5)	1.79	0.88-3.63	1.79	1.01-3.23
Missing		198	34				

Odds ratio ¹: Odds of Obese versus Non-Obese
Odds ratio ²: Odds of Obese or Overweight versus Normal or Thin
SSB: Sugar Sweetened Beverages
FF: Fast food

4.3. Backward logistic versus Lasso, Ridge and Elastic Net

The area under curve and the calibration slopes of the regularized logistic models that include Lasso (cross validation), Lasso (Adaptive), Lasso (BIC), Ridge, and Elastic net along with backwards logistic regression were implemented in both primary and secondary outcome with both cases. Results of the area under curve and the C-slopes are presented in table 5.

Table 5. Area under Curve and Calibration Slope of the statistical and machine learning models in the four cases

Model Used	**Backward Logistic**	**Lasso (CV)**	**Lasso (Adaptive)**	**Lasso (BIC)**	**Ridge**	**Elastic Net**
Outcome 1: Obese/Non-Obese						
Case 1						
Discrimination/ AUC	0.615(0.57-0.66)	0.645(0.60-0.69)	0.639(0.60-0.68)	0.640(0.60-0.68)	0.662(0.62-0.70)	0.641(0.60-0.68)
C-slope	0.749(0.56-0.99)	1.368(0.96-1.78)	0.963(0.68-1.25)	1.605(1.12-2.10)	1.601(1.15-2.04)	1.542(1.07-2.01)
Case 2						
Discrimination/ AUC	0.622(0.58-0.66)	0.645(0.61-0.68)	0.645(0.60-0.68)	0.640(0.60-0.68)	0.665(0.62-0.70)	0.645(0.61-0.68)
C-slope	0.754(0.58-0.98)	1.278(0.92-1.63)	0.925(0.67-1.18)	1.369(0.98-1.75)	1.535(1.13-1.93)	1.428(1.03-1.83)
Outcome 2: Obese or Overweight/ Normal or Thin						
Case 1						
Discrimination/ AUC	0.601(0.57-0.63)	0.638(0.60-0.67)	0.632(0.60-0.66)	0.601(0.57-0.63)	0.645(0.61-0.68)	0.637(0.60-0.68)
C-slope	0.764(0.60-0.97)	1.117(0.84-1.40)	0.968(0.72-1.21)	1.873(1.30-2.43)	1.245(0.94-1.56)	1.136(0.85-1.42)
Case 2						
Discrimination/ AUC	0.594(0.56-0.62)	0.628(0.60-0.66)	0.618(0.60-0.65)	0.598(0.57-0.63)	0.635(0.60-0.67)	0.628(0.60-0.66)
C-slope	0.766(0.60-0.98)	1.245(0.93-1.56)	0.979(0.72-1.23)	1.764(1.26-2.27)	1.415(1.07-1.76)	1.311(0.98-1.64)

Case 1: Participants with missing data on obese mother removed

Case 2: Participants with missing data on obese mother retained as a proxy category

4.4. Optimal Model Chosen

Model selection was primary based on having the highest calibration slope for model replicability.

In case calibration slopes of the models within the same outcome were close, the model chosen would be based on having the higher area under curve (AUC).

In this case, the highest two calibration slopes that were close to each other were for the lasso (adaptive) selection method with the outcome obese or overweight vs normal or thin. Their respective C-slopes are: 0.968 (0.72-1.21) and 0.979(0.72-1.23).

Therefore, since the calibration slopes were similar, the model selection was based on having the higher discrimination/AUC. Thus, the model with the calibration slope 0.968 and AUC 0.632 was chosen to predict child obese or overweight versus normal or thin. Calibration plot of that model is presented in figure 4, predictors chosen by that model are presented in Table 6.

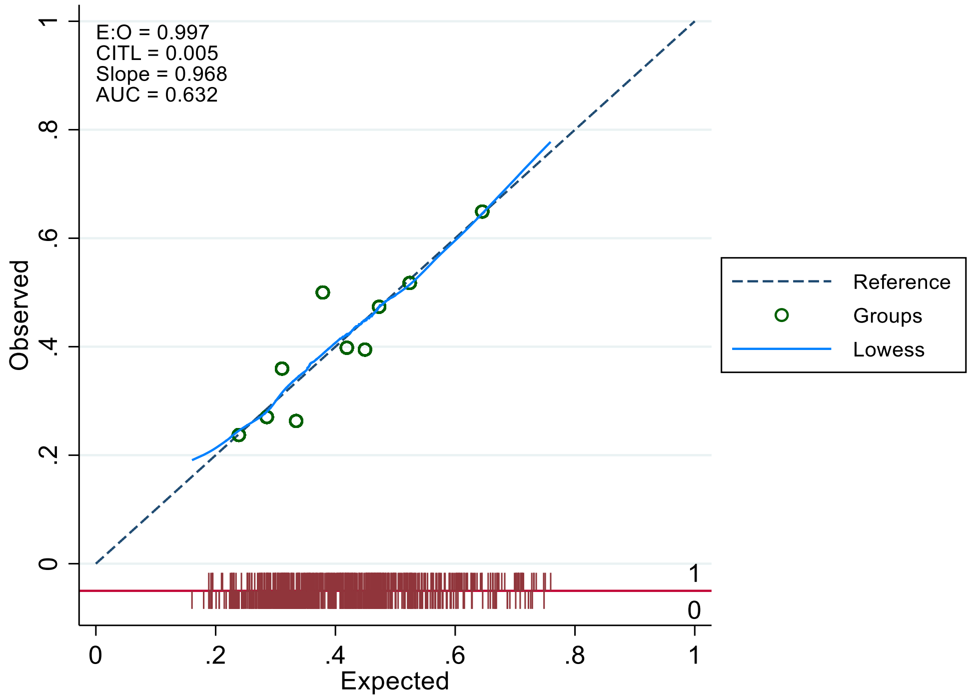

Figure 4. Calibration Plot of the model chosen to predict obese or overweight versus normal or thin

Table 6. Predictors chosen by the best machine learning model to predict obese or overweight vs normal or thin

Variable	Age	Gender	Nutrition Knowledge	Parent Marriage Status	Crowding Index	Tv time	Eating on Screen	Fruit Availability	Vegetable Availability	Obese Mother	Intercept
Coefficient	-0.02	-0.60	0.16	-0.53	-0.21	0.15	0.02	-0.23	0.50	1.03	0.12

4.5. Predicted Risk of Obese or Overweight vs Normal and Thin

Predicted risk for obesity or overweight for children with various predictors met are presented in Table 6. To illustrate, the predicted risk for an obese or overweight school student, is a girl student who is 12 years old, doesn’t have the proper nutrition knowledge , has her parents single(divorced or widowed), lives in a house with a crowding index of 3.5, watches TV for 1 hour a day, eats 1 time per week on screen, has adequate fruit available and limited vegetable available in her home plus doesn’t have an obese mother is 24.2%. On the other hand, the predicted risk for a boy who is 10 years old, has proper nutrition knowledge, single parents, lives in a home with a crowding index of 2, watches tv for 1 hour a day and eats 4 times per week in front of the screen, has limited fruit availability but adequate vegetable availability at home and an obese mother is 84.4%.

Table 7. The predicted risk of obese or overweight for children with various characteristics

Age	Gender	Child Nutrition Knowledge	Parent Marriage Status	Crowding Index	Tv time	Eating on Screen	Fruit Availability	Vegetable Availability	Obese Mother	Predicted Risk of Outcome
10	Girl	Fail	Married	1.5	4	3	Adequate	Adequate	No	36.5%
11	Boy	Fail	Single	4	3	2	Limited	Limited	Yes	70.9%
10	Boy	Pass	Single	2	1	4	Limited	Adequate	Yes	84.4%
12	Girl	Fail	Single	3.5	1	1	Adequate	Limited	No	24.2%

CHAPTER 5

DISCUSSION

5.1. Study Summary

The study aimed to develop and internally validate a prediction model for child obesity in Lebanon and identify the most important predictors for the outcome. Out of 1,409 participants selected from the original 2,125 participants, 241 students were obese, and 569 students were obese or overweight. All machine learning models performed better than the backwards logistic regression even after bootstrap shrinkage in terms of calibration and discrimination. The best model out of the 5 machine learning models in the 4 cases was that of Lasso adaptive with the outcome obese or overweight vs normal or thin with calibration slope of 0.968 and discrimination of 0.632. This model showed good calibration and moderate discrimination. Ten predictors were retained where older age, being female, having married parents, adequate availability of fruits at home, crowding index less than 3 are protective factors and eating while on screen, child nutrition knowledge, tv time more than 2 hours, vegetable availability plus having an obese mother as risk factors for child obesity or overweight.

5.2. Strengths

This study has several strengths. To begin with, the SCALE study implemented a two-stage sampling approach where schools stratified by Public, Private and Private free were randomly selected and then students from those 50 schools were randomly selected which ensures a more representative sample that isn't biased towards a particular characteristic and gives a more precise estimates of population parameters than single-stage sampling. Second, trained data collectors were used to record students' anthropometric measurements such as height and weight 2 times on average which decreases misclassification bias and ensures students were classified into the correct BMI z-score group. Third, sampling weights were used to analyze survey data while calculating univariate statistics and prevalence ratios. Furthermore, variable cutoffs chosen were based on systematic reviews and experts' opinions. Finally, this study used machine learning models which are powerful tools to improve performance and interpretability of prediction models especially in such situations where there are multiple predictors associated with the outcomes of interest.

5.3. Limitations

There are some limitations to this study. First, regarding the study participants, there might be some selection bias due to differential non-response between those who agreed to join versus those who did not agree to join the study. Moreover, results aren't representative of children in Greater Beirut as a whole, as there are children who don't attend schools which could either overestimate or underestimate the results depending on those children weight status. Second, this is a cross-sectional study so temporality is not met and there could be reverse causality. Third, there might be some information bias regarding the obese mother variable since the weight and height of the patents were self-reported. Furthermore, not all predictors present in the literature review were assessed in the SCALE study, this could explain why the model chosen had a moderate discriminative ability of 0.632.

5.4. Interpretation

5.4.1. Model Construction and Performance with Reference to Previous Studies

This is the first study that presents a LASSO adaptive model along its calibration slope to predict the risk for child obesity or overweight. According to the second literature review, a range of machine learning models were used to predict child obesity. Results are consistent with previous studies where machine learning models performed better than regular statistical models [45], [44].

One of the main reasons why LASSO performed better than other models is because LASSO is designed to handle low events per variable (EPV <10) where there are plenty of predictors [50].

Handling missing data was arbitrary in models used to predict child obesity in the literature. One of the articles that used machine learning models tested whether the data is missing completely at random (MCAR), for variables that were not MCAR, both imputation using K- nearest neighbor (KNN) and retaining data using a proxy category was implemented and results were compared. It was mentioned that imputation using KNN produced uncomprehensive results, thus the study proceed by replacing missing data with proxy category [41]. Moreover, another study, reported replacing missing values in variables with more than 10% missing by 0 in case of a continuous variable and dropping participants with missing data on that variable in case the variable was binary/categorical [43]. Therefore, this study implemented both methods (dropping participants with more than 10% missing data and retaining the participants through proxy category) used to account for missing data.

In terms of the outcome chosen for the machine learning models, this is the first study to perform sensitivity analysis with two binary outcomes Obese/ Non-Obese or

Obese Overweight/ Normal Thin. Studies in literature preferred to stick to one of these outcomes. Studies that chose to perform sensitivity analysis while having one of the binary outcomes as their primary outcome used BMI continuous as a secondary outcome where none of the models with BMI continuous outcome outperformed models with binary outcomes[45] ,[41], [43], [44].

Regarding the model performance measures, the performance of models with binary outcome was measured using the area under curve (AUC) primarily followed by F1 score and Mathews correlation coefficient. None of the models reported a calibration slope though measuring a calibration slope is important to make sure that the model's predicted probabilities are compatible with the probability of having the outcome in study[45],[41], [43], [44].

In the literature, the discrimination (AUC) of the prediction models ranged between 0.630 to 0.81 with Lasso (cv) having the highest area under curve of 0.81 for girls' model when data was stratified by gender indicating that the model chosen in this study was on the lower borderline in terms of discrimination [44]. However, models that used machine learning to predict obesity had at least a sample size of 27,203 and were able to include more than 17 predictors into their model while maintaining an event per variable ratio greater than 10 which possibly explains their high discrimination. Moreover, common predictors for child obesity or overweight chosen by this study's model and the literature include age, gender, parent marriage status, nutrition education and having an obese mother. Other prediction models either additionally included maternal predictors, physical activity, parent education or screentime.

5.4.2. Model Predictors

Age

This study highlighted older age to be a protective factor against obesity or overweight. However, the odds ratio for age in this study is 0.98 which is very close to 1. This could be due to the narrow age range of children that the SCALE study studied. Thus, had this model been used in a clinical setting it would be better to eliminate the age variable from the model. Age is one of the variables that machine learning models usually selected. For example, a study that used the Osakidetza data of 426,813 participants less than 18 years old had lasso and ridge select age as one of the most important 10 risk factors for childhood obesity [43]. However, a systematic review that included 44 studies examining the association between screen time and obesity or overweight explained that age could be a protective factor where younger children have less structured time and are more prone to watch TV and stay on the screen [8].

Gender

Results regarding the association between gender and obesity or overweight is conflicting. In this study, being female is a protective factor against obesity which is also present in studies conducted in Lebanon. This may be a result of a social cultural on adolescent girls in this study's age group to maintain an acceptable body image[51]. However, in the same systematic review, one cross-sectional study in Saudi Arabia with 2822 adolescent participants found that males have significant higher odds to become overweight than females with an odds ratio of 1.67. On the contrary, another cross-sectional study in Saudi Arabia with 1869 participants reported females having higher odds to be overweight or obese with a significant odds ratio of 1.37. Additionally, the study that used the Osakidetza data of 426,813 participants less than 18 years old had lasso and ridge select female gender to be a protective factor against obesity or overweight with no coefficients being reported which is consistent with the results of this study[7].

Child Nutrition Knowledge

It is important to note that this study found child nutrition knowledge to be a risk factor of obesity or overweight. This finding is in contrast with most of results in the literature that examined association between diet/ nutrition knowledge and child obesity. Thus, if this model had to be used in clinical settings, such variable should be eliminated from the model. Overall, studies that tackled the association between child nutrition knowledge and child obesity were in the form of school intervention programs provided by teachers and school nurses during or after classes that lasted between 6 to 12 months. Moreover, in a systematic review using the 14 studies that reported outcome based on BMI z-score a pooled estimated of change in the BMI z-score between intervention and the control group was -0.06(-0.1- - 0.03) [14]. A possible explanation for the results of this study could be reverse causality due to this study design being cross sectional. Obese or overweight children could have developed the interest in learning more about nutrition and dietary habits aiming to reduce their weight.

Parent Marital Status

This study found that children with divorced or widowed parents have 1.7 times the odds of being obese or overweight than that of married parents. This finding is consistent with other studies. For example, a cross sectional study among 3,166 third grade participants revealed that 1.54 times more prevalent among children of divorced parents compared with children of married parents. One of the consequences due to having divorced parents is having less time for domestic work and reliance on fast food high in calories. Another reason could be related to emotional stress and adverse childhood experience which is highly associated with obesity [15], [16],[17],[52].

Crowding

Socio-economic status had various proxies in the literature where family income and crowding index are mostly used depending on whether the countries were high or middle-low-income countries. Given that this study uses data on children in Lebanon (a low-middle income country), the literature that tackled low-middle income countries was mostly consistent with the findings of this study. Lower crowding index in Lebanon is associated with child obesity and overweight which explains why crowding is a protective factor with an odds ratio of 0.81 for a crowding index more than 3[53]. A meta-analysis of 3 studies revealed a statistically significant positive association between higher family income/ lower crowding index and the odds of child obesity with a pooled odds ratio of 1.57 (1.30-1.91) [7]. This association could be explained by the fact that families with higher socio-economic status could order and purchase fast food rich in calories unlike families with low socio-economic status. Moreover, with crowding index of 3 being an indicative of high poverty at a time of economic and food insecurity crisis, this finding that low socio-economic status is protective of obesity in Greater Beirut could portray early stages of nutrition transition.

Tv- time and Eating While on Screen

Similar to previous studies, eating while on screen and spending time on tv especially for more than 2 hours were found to be associated with obesity. An example is a systematic review and meta-analysis with 44 included cross sectional studies and a total of 112,489 participants reveled that eating while on screen is a risk factor for obesity. Moreover, the most important obesogenic screen turned out to be TV watching with a significant odds ratio of 1.81 (1.42-2.35) [8]. Such result is expected, as the child burns few calories during prolonged sedentary time which triggers weight gain especially if the child is eating on the screen. Moreover, TV is packed filled with fast food commercials that could grasp the child's attention to try new food rich in fat. This illustrates the importance of alerting parents to set limits for the time the child watches TV and encourage alternative activities such as practicing hobbies.

Fruit Availability

It has been constantly shown throughout systematic reviews and this study that fruit consumption and fruit availability at home are protective against overweight and obesity. A systematic review of 4 studies from the Middle East indicated that low micronutrient intake due to low fruit consumption was associated with obesity[18]. It was reported that fruit availability and consumption by the child increases satiety and decreases fat storage which triggers weight loss [8]. This exemplifies the importance of having an environment that promotes healthy eating.

Vegetable Availability

Vegetable availability and consumption were highly inconsistent in the literature. In the same systematic review, one of the studies that measured vegetable consumption and availability in child's home reported that vegetable availability or consumption especially is significantly associated with overweight and obesity where it was important to look at the vegetable preparation/cooking method. In contrast, another study noted the absence of a significant association between vegetable availability and being obese or overweight [7].

Obese Mother

According to the Lasso adaptive model in this study, having an obese mother increases the odds of the child being overweight or obese by 2.8 times. However, this result was expected as strong associations are presented in systematic reviews with pooled odds ratio of 1.97[27]. This indicates that the risk of a child being overweight or obese is greatly influenced by the mother's weight status where unmeasured genetic factors could also play a role. Parents have an important role in reducing that risk as much as possible. Further investigation portrayed the presence of an unhealthy promoting lifestyle and eating habits induced by the mother which influences the child's eating and sedentary behavior making the child at risk of being overweight or obese.

5.4.3. Study Implications and Potential Clinical Use

While it is important to note that few child obesity prediction models were developed around the world, this is the first prediction model that incorporates child socio-demographics, child nutrition knowledge, family socio-economic and parent marital status in Lebanon. Predicting the risk of a child being obese or overweight with a parsimonious calibrated model could be particularly useful in clinical dietetics and pediatric settings. Using predictive models in clinical settings could prevent the outcome of interest and reduce risk for other potential non-communicable diseases while maintaining the time of the clinician and utilizing less resources.

A range of short term and long-term interventions could be explored targeting parents, schools, and the government to reduce the prevalence of overweight and obese children. These include interventions that aim to limit TV watching behavior in general, and during mealtimes specifically, encourage healthy family meals rich in micronutrients and ensure availability and affordability of fruits.

5.5. Conclusion

In conclusion, child obesity is an issue of public health concern. With the manifestation of an obesogenic environment and rapid technological and food industry advancements, child obesity is expected to rise and present a greater problem to society and its economy. This prognostic study highlights the main predictors of child obesity are associated with child's sociodemographic, sedentary behavior and home environment (parent marriage status and availability of fruits plus vegetables). This study provides a first steppingstone for future larger studies to develop child obesity prediction models. The predictors identified in this study could allow health care professionals to identify children at risk of child obesity and provide the necessary interventions and assistance to reduce that risk.

APPENDIX 1

JBI Critical Appraisal Checklist for Systematic Reviews and Research Syntheses

Reviewer_________________________________ Date_________________________________

Author _________________________________ Year_____________ Record Number______

		Yes	No	Unclear	Not applicable
1.	Is the review question clearly and explicitly stated?	☐	☐	☐	☐
2.	Were the inclusion criteria appropriate for the review question?	☐	☐	☐	☐
3.	Was the search strategy appropriate?	☐	☐	☐	☐
4.	Were the sources and resources used to search for studies adequate?	☐	☐	☐	☐
5.	Were the criteria for appraising studies appropriate?	☐	☐	☐	☐
6.	Was critical appraisal conducted by two or more reviewers independently?	☐	☐	☐	☐
7.	Were there methods to minimize errors in data extraction?	☐	☐	☐	☐
8.	Were the methods used to combine studies appropriate?	☐	☐	☐	☐
9.	Was the likelihood of publication bias assessed?	☐	☐	☐	☐
10.	Were recommendations for policy and/or practice supported by the reported data?	☐	☐	☐	☐
11.	Were the specific directives for new research appropriate?	☐	☐	☐	☐

APPENDIX 2

Supplementary Table 1. Scoring of articles present in the literature and the SCALE study using Joanna Briggs Institute (JBI) critical appraisal checklist for systematic reviews

Item / Reference	1	2	3	4	5	6	7	8	9	10	11	Total JBI score
Farrag et al.[7]	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	No	Yes	7
St Pierre et al.[9]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	8
Haghjoo et al. [8]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	10
Albataineh et al.[18]	Yes	Yes	Yes	Yes	Yes	No	No	Unclear	No	Yes	Yes	7
Lindsay et al.[19]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	10
Frantsve-Hawley et al.[20]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	9
Ziauddeen et al. [10]	Yes	Yes	Yes	Yes	Unclear	No	Yes	Unclear	No	Yes	Yes	7
Jacob et al.[14]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11
Lee et al.[27]	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes	7
Liberali et al. [11]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Unclear	Yes	9

APPENDIX 3

Supplementary Table 2: Variables present in final prognostic model of Pang et al. [41]

Variable Name	
Body height Measured	
Body weight Measured	Chloride serum/plasma
Head Occipital-frontal circumference by Tape measure	Calcium serum/plasma serum/plasma
Body temperature	Creatinine serum/plasma
Hemoglobin	Monocytes [#/volume] in Blood
Lead [Mass/volume] in Capillary blood	Basophils [#/volume] in Blood
Oxygen saturation in Arterial blood by Pulse oximetry	Cells Counted Total [#] in Blood
White Blood cell (WBC) count (leukocyte)	Urea nitrogen serum/plasma
Erythrocytes [#/volume] in Blood by Automated count	Total Bilirubin serum/plasma
Hematocrit	Aspartate aminotransferase serum/plasma
MCV	Alanine aminotransferase serum/plasma
Erythrocyte distribution width [Ratio] in Cord blood	Indirect bilirubin serum/plasma
Erythrocyte mean corpuscular hemoglobin [Entitic mass] by Automated count	Specific gravity of Urine
Erythrocyte mean corpuscular hemoglobin concentration [Mass/volume] by Automated count	Variant lymphocytes/100 leukocytes in Blood by Manual count
Platelet count	Alkaline phosphatase serum/plasma
Heart rate	Color of Urine
Specimen type	Lipoprotein lipase [Enzymatic activity/volume] in Serum or Plasma
Respiratory rate	Cholesterol [Mass/volume] in Serum or Plasma
Streptococcus pyogenes Ag [Presence] in Throat	Hemoglobin A1c (Glycated)
Platelet mean volume [Entitic volume] in Blood by Automated count	Cholesterol in HDL [Mass/volume] in Serum or Plasma
Neutrophils [#/volume] in Blood	
Lymphocytes %	
Monocytes %	
Eosinophils %	
Basophils %	
Lymphocytes [#/volume] in Blood	
Segmented neutrophils/100 leukocytes in Blood	
BP diastolic	
BP systolic	
Glucose lab	
Sodium serum/plasma	
Potassium serum/plasma	
Bicarbonate (Carbon dioxide serum/plasma)	

APPENDIX 4

Supplementary Table 3. Variables present in final prognostic model of Shi et al. [43](2021)

	Filter (MI)	SVM-RFE	Ridge	Lasso	RandomForest
1	Age	MoDietEducation	Age	Age	SystolicPressure
2	Sleep_Normal (–)	MoRDType_LowSalt	Sex (–)	Sex (–)	MoDiastolicPressure (–)
3	BFType_Maternal (–)	RDType_2000 cal	Tobacco_No (–)	Tobacco_No (–)	MoSystolicPressure (–)
4	DiastolicPressure (–)	AdeDKnowledge	DietEducation	DietEducation	Sex
5	MoSystolicPressure	MoPE_Inadequate (–)	MoTobacco_Yes	MoDietEducation	Birthyear (–)
6	MoNumberCigarettes	DietCompliesAdvice	BFType_Maternal (–)	BFType_Maternal (–)	Tobacco_No (–)
7	Birthheight (–)	MoRDType_Free (–)	PE_Inadequate	Birthyear (–)	MoExerciseAdvice (–)
8	MoBMI	MoPEHour	MoDiabetes_No (–)	MoNumberCigarettes	MoAlcohol_No (–)
9	Birthweight (–)	DiastolicPressure (–)	PE_Adequate(–)	PE_Inadequate	PE_Inadequate
10	MoDiastolicPressure (–)	SystolicPressure	MoDietEducation	DCExecution_No	MoTobacco_Ex

APPENDIX 5

Supplementary Table 4. Characteristics of all students with BMI z-score data from the SCALE study

	Total n=2,004	Thin n=26(1.27%)	Normal n=1,172(57.1%)	Overweight n=476(25.43%)	Obese n=330(16.2%)	Odds Ratio [1]	95%CI	Odds Ratio [2]	95%CI
	n(%)	n(%)	n(%)	n(%)	n(%)				
Sociodemographic									
Child Age median(IQR)	11(10-12)	11(10-13)	11(10-12)	11(10-12)	11(10-12)	0.84	0.71-1.01	0.9	0.82-0.99
Child Sex									
Boy	917(48.5)	14(1.2)	469(49.0)	231(28.1)	203(21.7)	1		1	
Girl	1,087(51.5)	12(1.3)	703(64.7)	245(23.0)	127(11.0)	0.44	0.32-0.6	0.52	0.39-0.67
School Type									
Public	1,275(30.1)	18(1.9)	767(59.6)	275(21.2)	215(17.3)	1		1	
Private	729(69.9)	8(1.1)	405(56.0)	201(27.2)	115(15.7)	0.89	0.62-1.28	1.20	0.96-1.49
Child Food insecurity									
Secure	1,328(81.9)	15(1.3)	773(57.8)	322(25.2)	218(15.7)	1			
Insecure	492(18.1)	10(1.5)	297(57.8)	108(24.2)	77(16.5)	1.06	0.64-1.75	0.99	0.68-1.43
Missing	184(9.1%)	1	102	46	35				
Child Nutrition Knowledge									
Fail	1,241(59.1)	18(1.2)	740(59.2)	277(23.4)	206(16.2)	1		1	
Pass	740(40.9)	8(1.4)	420(54.8)	194(28.3)	118(15.5)	0.95	0.76-1.18	1.18	0.94-1.49
Missing	23(1.1%)		12	5	6				
Mother Education Level									
Intermediate & Below	544(25.7)	10(1.5)	328(62.0)	111(19.0)	95(17.5)	1		1	
Secondary & Above	805(74.3)	10(1.4)	453(56.7)	205(25.9)	137(16)	0.9	0.56-1.43	1.25	0.83-1.88
Missing	655(32.7%)	6	391	160	98				
Parent Marital Status									
Single (Widowed or	87(5.1)	0	46(45)	22(17.8)	19(37.2)	1		1	
Divorced)	1,296(94.9)	20(1.5)	761(58.8)	297(24.1)	218(15.6)	0.31	0.17-0.54	0.53	0.27-1.06
Married	621(30.9%)	6	365	157	93				
Missing									
Crowding Index									
≤3people/room	1,293(97.4)	16(1.4)	749(57.9)	304(24.2)	224(16.5)	1		1	
>3 people/room	109(2.6)	4(2.5)	67(63.1)	23(20.5)	15(13.9)	0.81	0.45-1.48	0.76	0.42-1.36
Missing	602(30%)	6	356	149	91				
Lifestyle factors									
Time spent on screen									
< 2hrs/day	1,470(73)	19(1.4)	873(57.6)	356(26)	222(15)	1		1	
≥ 2hrs/day	492(27)	7(1.2)	272(55.9)	111(24)	102(18.9)	1.31	0.94-1.82	1.08	0.87-1.33
Missing	42(2%)		27	9	6				
Watching TV									
< 2hrs/day	1,482(76.1)	22(1.4)	870(57.4)	349(24.9)	241(16.3)	1		1	
≥ 2hrs/day	494(23.9)	4(0.9)	287(57.1)	120(26.9)	83(15.1)	0.91	0.66-1.25	1.03	0.81-1.31
Missing	28(1.3%)		15	7	6				

Eating on Screen(day/w) [3]	2(0-5)	1(0-3)	2(0-4)	1(0-5)	2(0-5)	1		1	
median (IQR)	51(2.5%)		28	13	10	1.05	0.97-1.13	1.02	0.97-1.08
Missing									
Physical Activity									
<3days/week	1,141(52.7)	18(1.4)	674(57.8)	260(24.9)	189(15.9)	1		1	
≥3days/week	828(47.3)	7(1.1)	482(57.0)	206(25.9)	133(16)	0.99	0.64-1.54	1.04	0.83-1.31
Missing	35(1.7%)	1	16	10	8				
Availability of Fruit									
Limited	547(24.1)	11(2)	326(56.6)	119(23.1)	91(18.3)	1		1	
Adequate	861(75.9)	9(1.2)	494(58.5)	208(24.4)	150(15.9)	0.84	0.51-1.39	0.95	0.72-1.25
Missing	596(29.7%)	6	352	149	89				
Availability of Vegetables									
Limited	348(14.7)	4(1.3)	221(64.1)	72(19.1)	51(15.5)	1		1	
Adequate	1,061(85.3)	16(1.4)	599(57.0)	256(24.9)	190(16.7)	1.09	0.75-1.59	1.35	0.88-2.05
Missing	595(29.6%)	6	352	148	89				
Availability of Sugar-Sweetened Beverages									
Limited	999(63.3)	13(1.4)	591(58.9)	229(24.3)	166(15.4)	1		1	
Adequate	409(36.7)	7(1.5)	229(56.6)	98(23.5)	75(18.4)	1.24	0.75-2.04	1.09	0.71-1.66
Missing	596(29.7%)	6	352	149	89				
Fast Food Consumption									
<3days/week	1,888(96.2)	25(1.3)	1,108(57.8)	445(25.2)	310(15.7)	1		1	
≥3days/week	81(3.8)	1(0.6)	46(44.5)	24(32.9)	10(22.0)	1.52	0.6-3.8	1.7	0.98-3.16
Missing	35(1.7%)		18	7	10				
History of Obesity									
Obese Mother									
No	1,015(90.1)	18(1.8)	611(58.7)	225(23.4)	161(16.1)	1		1	
Yes	162(9.9)	0	68(46.1)	48(28.4)	46(25.5)	1.79	0.88-3.63	1.79	1.01-3.23
Missing	827(41.2%)	8	493	203	123				

APPENDIX 6

Supplementary Table 5. Characteristics and Environmental factors of thin or normal weight compared to overweight or obese school children from the SCALE study

		Total n=1,409	Thin & Normal Weight n=840(59.5)	Overweight & Obese n=569(40.5)
		n(%)	n(%)	n(%)
Sociodemographic				
Child Age	**median (IQR)**	11(10-12)	11(10-12)	11(10-12)
Child Sex				
Boy		677(51.6)	357(51.4)	320(48.6)
Girl		732(48.4)	483(68.0)	249(32.0)
School Type				
Public		896(32.2)	543(60.2)	353(39.8)
Private		513(67.8)	297(59.1)	216(40.9)
Child Food insecurity				
Secure		949(81.7)	561(60.2)	388(39.8)
Insecure		341(18.3)	213(60.5)	128(39.5)
Missing			66	53
Child Nutrition Knowledge				
Fail		885(59.2)	546(61.3)	339(38.7)
Pass		514(40.8)	290(57.3)	224(42.7)
Missing			4	6
Mother Education Level				
Intermediate & Below		544(25.7)	338(63.5)	206(36.5)
Secondary & Above		805(74.3)	463(58.1)	342(41.9)
Missing			39	21
Parent Marital Status				
Single (Widowed or Divorced)		87(5.1)	46(45.0)	41(55.0)
Married		1,296(94.9)	781(60.3)	515(39.7)
Missing			13	13

Crowding Index				
≤3people/room		1,293(97.4)	765(59.3)	528(40.7)
>3 people/room		109(2.6)	71(65.6)	38(34.4)
Missing			4	3
Child Lifestyle Factors				
Time spent on screen				
< 2hrs/day		1,052(73.3)	638(61.0)	414(39)
≥ 2hrs/day		337(26.7)	191(55.8)	146(44.2)
Missing			11	9
Watching TV				
< 2hrs/day		1,046(74.7)	631(61.4)	415(38.6)
≥ 2hrs/day		349(25.3)	203(54.6)	146(45.4)
Missing			6	8
Eating on Screen(days/week)	**median (IQR)**	2(0-5)	2(0-4)	2(0-5)
Missing			15	14
Physical Activity				
<3days/week		801(51.5)	486(60.6)	315(39.4)
≥3days/week		587(48.5)	345(58.6)	242(41.4)
Missing			9	12
Availability of Fruit				
Limited		547(24.1)	337(58.7)	210(41.3)
Adequate		861(75.9)	503(59.7)	358(40.3)
Missing				1
Availability of Vegetables				
Limited		348(14.7)	225(65.5)	123(34.5)
Adequate		1,061(85.3)	615(58.4)	446(41.6)
Availability of Sugar-Sweetened				
Beverages		999(63.3)	604(60.3)	395(39.7)
Limited		409(36.7)	236(58.1)	173(41.9)
Adequate				1
Missing				
Fast Food Consumption				
<3days/week		1,333(95.7)	800(60.4)	533(39.6)
≥3days/week		55(4.3)	30(41.1)	25(58.9)
Missing			10	11
History of Obesity				

Obese Mother			
No	1,015(90.1)	629(60.5)	386(39.5)
Yes	162(9.9)	68(46.0)	94(54)
Missing		143	89

REFERENCES

1. Lakshman, R., C.E. Elks, and K.K. Ong, *Childhood obesity.* Circulation, 2012. **126**(14): p. 1770-1779.
2. Ahmad, Q.I., C.B. Ahmad, and S.M. Ahmad, *Childhood obesity.* Indian journal of endocrinology and metabolism, 2010. **14**(1): p. 19.
3. Hobbs, M. and D. Radley, *Obesogenic environments and obesity: a comment on 'Are environmental area characteristics at birth associated with overweight and obesity in school-aged children? Findings from the SLOPE (Studying Lifecourse Obesity PrEdictors) population-based cohort in the south of England'.* BMC medicine, 2020. **18**(1): p. 1-3.
4. Nasreddine, L., et al., *Trends in overweight and obesity in Lebanon: evidence from two national cross-sectional surveys (1997 and 2009).* BMC public health, 2012. **12**(1): p. 1-11.
5. Ghattas, H., et al., *School and community drivers of child diets in two Arab cities: The SCALE protocol and innovative tools to assess children's food environments.* Plos one, 2022. **17**(7): p. e0264963.
6. Aromataris, E., et al., *Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach.* JBI Evidence Implementation, 2015. **13**(3): p. 132-140.
7. Farrag, N.S., L.J. Cheskin, and M.K. Farag, *A systematic review of childhood obesity in the Middle East and North Africa (MENA) region: Prevalence and risk factors meta-analysis.* Adv Pediatr Res, 2017. **4**.

8. Haghjoo, P., et al., *Screen time increases overweight and obesity risk among adolescents: a systematic review and dose-response meta-analysis.* BMC primary care, 2022. **23**(1): p. 1-24.

9. St Pierre, C., et al., *Food Insecurity and Childhood Obesity: A Systematic Review.* Pediatrics, 2022. **150**(1).

10. Ziauddeen, N., et al., *Use of maternal and early life risk factors to predict childhood overweight and obesity: A systematic review.* The Lancet, 2017. **390**(SPEC.ISS 1): p. S100.

11. Liberali, R., et al., *Latent Class Analysis of Lifestyle Risk Factors and Association with Overweight and/or Obesity in Children and Adolescents: Systematic Review.* Child Obes, 2021. **17**(1): p. 2-15.

12. Kracht, C.L., E.D. Joseph, and A.E. Staiano, *Video Games, Obesity, and Children.* Current Obesity Reports, 2020. **9**(1).

13. Ruan, H., et al., *Habitual Sleep Duration and Risk of Childhood Obesity: Systematic Review and Dose-response Meta-analysis of Prospective Cohort Studies.* Sci Rep, 2015. **5**: p. 16160.

14. Jacob, C.M., et al., *A systematic review and meta-analysis of school-based interventions with health education to reduce body mass index in adolescents aged 10 to 19 years.* International Journal of Behavioral Nutrition and Physical Activity, 2021. **18**(1): p. 1-22.

15. Schroeder, K., et al., *The association between adverse childhood experiences and childhood obesity: A systematic review.* Obesity reviews : an official journal of the International Association for the Study of Obesity, 2021. **22**(7): p. e13204.

16. Hughes, K., et al., *The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis.* The Lancet Public Health, 2017. **2**(8): p. e356-e366.

17. Midei, A.J. and K.A. Matthews, *Interpersonal violence in childhood as a risk factor for obesity: a systematic review of the literature and proposed pathways.* Obes Rev, 2011. **12**(5): p. e159-72.

18. Albataineh, S.R., E.F. Badran, and R.F. Tayyem, *Dietary factors and their association with childhood obesity in the Middle East: A systematic review.* Nutr Health, 2019. **25**(1): p. 53-60.

19. Lindsay, A.C., et al., *Non-responsive feeding practices, unhealthy eating behaviors, and risk of child overweight and obesity in southeast asia: A systematic review.* International Journal of Environmental Research and Public Health, 2017. **14**(4).

20. Frantsve-Hawley, J., et al., *A systematic review of the association between consumption of sugar-containing beverages and excess weight gain among children under age 12.* J Public Health Dent, 2017. **77 Suppl 1**: p. S43-s66.

21. Shan, L.S., Q.L. Zhou, and Y.X. Shang, *Bidirectional Association Between Asthma and Obesity During Childhood and Adolescence: A Systematic Review and Meta-Analysis.* Frontiers in Pediatrics, 2020. **8**.

22. Matthews, E.K., J. Wei, and S.A. Cunningham, *Relationship between prenatal growth, postnatal growth and childhood obesity: A review.* European Journal of Clinical Nutrition, 2017. **71**(8): p. 919-930.

23. Albers, L., et al., *Maternal smoking during pregnancy and offspring overweight: is there a dose–response relationship? An individual patient data meta-analysis.* International Journal of Obesity, 2018. **42**(7): p. 1249-1264.
24. Heslehurst, N., et al., *Maternal BMI is an early life predictor of childhood obesity: A systematic review and meta-analysis.* Obesity Facts, 2019. **12**: p. 54-55.
25. Marco, P.L., et al., *Systematic review: Symptoms of parental depression and anxiety and offspring overweight.* Revista de saude publica, 2020. **54**: p. 49.
26. Oken, E., E.B. Levitan, and M.W. Gillman, *Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis.* Int J Obes (Lond), 2008. **32**(2): p. 201-10.
27. Lee, H.J., et al., *The global relationship between parent and child obesity: A systematic review and metaanalysis.* Journal of Pediatric Gastroenterology and Nutrition, 2021. **72**(SUPPL 1): p. 1264-1265.
28. Alfano, R., et al., *Perspectives and challenges of epigenetic determinants of childhood obesity: A systematic review.* Obesity reviews : an official journal of the International Association for the Study of Obesity, 2022. **23 Suppl 1**: p. e13389.
29. Williams, J., et al., *A systematic review of the influence of the retail food environment around schools on obesity-related outcomes.* Obesity Reviews, 2014. **15**(5): p. 359-374.
30. Matsuzaki, M., et al., *Food environment near schools and body weight-A systematic review of associations by race/ethnicity, gender, grade, and socio-*

economic factors. Obesity reviews : an official journal of the International Association for the Study of Obesity, 2020. **21**(4): p. e12997.

31. Zhou, Q., et al., *Neighborhood supermarket access and childhood obesity: A systematic review.* Obesity reviews : an official journal of the International Association for the Study of Obesity, 2021. **22 Suppl 1**: p. e12937.
32. Mei, K., et al., *State-of-the-art of measures of the obesogenic environment for children.* Obesity Reviews, 2021. **22**(S1).
33. Huang, C., et al., *The Association between Childhood Exposure to Ambient Air Pollution and Obesity: A Systematic Review and Meta-Analysis.* International journal of environmental research and public health, 2022. **19**(8).
34. Malacarne, D., et al., *The built environment as determinant of childhood obesity: A systematic literature review.* Obesity Reviews, 2022. **23**(S1).
35. Parasin, N., T. Amnuaylojaroen, and S. Saokaew, *Effect of Air Pollution on Obesity in Children: A Systematic Review and Meta-Analysis.* Children (Basel), 2021. **8**(5).
36. Wang, Z., et al., *Traffic-related environmental factors and childhood obesity: A systematic review and meta-analysis.* Obesity Reviews, 2021. **22**(S1).
37. Kim, K.Y., E. Lee, and Y. Kim, *The association between bisphenol a exposure and obesity in children—a systematic review with meta-analysis.* International Journal of Environmental Research and Public Health, 2019. **16**(14).
38. Li, H.T., Y.B. Zhou, and J.M. Liu, *The impact of cesarean section on offspring overweight and obesity: a systematic review and meta-analysis.* Int J Obes (Lond), 2013. **37**(7): p. 893-9.
39. Organization, W.H., *WHO housing and health guidelines.* 2018.

40. Gao, Z., et al., *Development and validity of a general nutrition knowledge questionnaire (Gnkq) for Chinese adults.* Nutrients, 2021. **13**(12): p. 4353.

41. Pang, X., et al., *Prediction of early childhood obesity with machine learning and electronic health record data.* International Journal of Medical Informatics, 2021. **150**: p. 104454.

42. Potter, C., et al., *Parental beliefs about portion size, not children's own beliefs, predict child BMI.* Pediatric obesity, 2018. **13**(4): p. 232-238.

43. Shi, X., et al., *An ensemble-based feature selection framework to select risk factors of childhood obesity for policy decision making.* BMC medical informatics and decision making, 2021. **21**(1): p. 1-13.

44. Hammond, R., et al., *Predicting childhood obesity using electronic health records and publicly available data.* PLoS One, 2019. **14**(4): p. e0215571.

45. Colmenarejo, G., *Machine learning models to predict childhood and adolescent obesity: a review.* Nutrients, 2020. **12**(8): p. 2466.

46. Connell, C.L., et al., *Food security of older children can be assessed using a standardized survey instrument.* The Journal of Nutrition, 2004. **134**(10): p. 2566-2572.

47. Onis, M.d., et al., *Development of a WHO growth reference for school-aged children and adolescents.* Bulletin of the World health Organization, 2007. **85**(9): p. 660-667.

48. Koo, T.K. and M.Y. Li, *A guideline of selecting and reporting intraclass correlation coefficients for reliability research.* Journal of chiropractic medicine, 2016. **15**(2): p. 155-163.

49. Heinze, G., C. Wallisch, and D. Dunkler, *Variable selection–a review and recommendations for the practicing statistician.* Biometrical journal, 2018. **60**(3): p. 431-449.

50. Pavlou, M., et al., *How to develop a more accurate risk prediction model when there are few events.* Bmj, 2015. **351**.

51. Sibai, A.M., et al., *Prevalence and covariates of obesity in Lebanon: findings from the first epidemiological study.* Obesity research, 2003. **11**(11): p. 1353-1361.

52. Biehl, A., et al., *Parental marital status and childhood overweight and obesity in Norway: a nationally representative cross-sectional study.* BMJ open, 2014. **4**(6): p. e004502.

53. Nasreddine, L., et al., *Prevalence and correlates of preschool overweight and obesity amidst the nutrition transition: findings from a national cross-sectional study in Lebanon.* Nutrients, 2017. **9**(3): p. 266.