

ORIGINAL ARTICLE

Novel ranking of protective and risk factors for adolescent adiposity in US females

A. Narla  and D. H. Rehkopf 

¹Division of Primary Care and Population Health, Stanford University School of Medicine, Stanford, CA, USA.

Received 4 July 2018; revised 15 October 2018; accepted 18 October 2018

Correspondence to: DH Rehkopf, ScD, MPH, Division of Primary Care and Population Health, Stanford University School of Medicine, 1070 Arastradero Road, Room 280, Stanford, CA 94304, USA.
E-mail: drehkopf@stanford.edu

Summary

Objective

Rank the importance of potentially modifiable psychosocial, dietary and environmental risk and protective factors for female adolescent obesity in order to target and inform public health prevention efforts. Utilizing the largest dataset available that captures the onset of the adolescent obesity surge in the USA, the study provides a more robust understanding of paediatric obesity risk factors.

Methods

Data were obtained from an observational, longitudinal study conducted between 1989 and 2001, the NHLBI Growth and Health Study. This study includes girls aged 9–19 years from three urban US locations, with Black and White girls generally represented equally. Data were analysed using multiple regression, random forest and propensity score matching to determine the strongest adiposity risk and protective factors during ages 9–12 predicting adiposity at age 19 with multiple methods to maximize the ability to identify possible public health interventions. Multiple linear regression and random forest analysis identified the strongest associations among 288 risk and protective factors selected from the study's literature review. For the 190 factors associated with follow-up adiposity from the data, propensity score matching was used to control for confounding of these factors.

Results

Findings suggest that highest priority interventional targets across the domains surveyed are lowering specific nutrients; eating meals with others or during activities without skipping; parents fixing evening snacks; improving perceptions of non-extremes as the healthy weight; improving self-worth, physical activity and social competence; and limiting any negative impact of dieting relatives. Similar associations were observed for Black and White girls.

Conclusion

The clinical implications of these findings allow health practitioners to target behavioural change efforts and address social and environmental factors that have demonstrated higher prioritization value for early obesity interventional efforts for adolescents.

Keywords: Adiposity, Machine learning, Obesity, Predictive analytics.

Introduction

Youth who are overweight have high odds of becoming adults with obesity, and research shows that continuous promotion of health behaviours during adolescence creates the greatest yield for adult health (1). Research addressing the transitional period of middle to late

adolescence for children in the USA can highlight intervention opportunities for targetable behaviours and influences that can be addressed before full development into adults with obesity. (2) Harnessing this transition period can alter the significant association between adolescents who develop obesity and increased disability in young adulthood and beyond (3). For female adolescents in

particular, research into clinical-based cut points for health effects of adiposity found a significant relationship with adiposity in female adolescents and cardiovascular disease risk (4). Behavioural patterns are the biggest contributors to premature death (40%) (5), indicating a research necessity to unmask significant behaviour patterns both contributing to and protecting against obesity and its associated morbidity and mortality.

The NHLBI Growth and Health Study (NGHS) is one of the most detailed and comprehensive surveys focused on risk factors for adolescent obesity ever performed in the USA (6). It longitudinally followed a cohort of female participants ($n = 2,879$) from age 9 to age 19 in three urban cities of the USA, thus providing a large sample size repeatedly measured over time with standardized surveys. Previous studies with NGHS data have identified sociodemographic obesity risk factors and have shown lower income, ineffectiveness as measured through self-efficacy surveys, and race as the top social, demographic and psychosocial risk factors, with higher obesity trends for Black participants (7). These prior studies primarily focus on prediction without comprehensive control for potential confounding factors – a critical step to best inform targeted clinical and policy interventions to decrease obesity.

The objective of the study is to use NGHS data to rank the importance of potentially modifiable psychosocial, dietary, behavioural and environmental risk and protective factors for female adolescent obesity. The study analytically takes a twofold approach. First, random forest and multiple linear regression analyses were applied to find the strongest associations with sum of skinfolds among many early adolescent social, demographic, and psychosocial risk factors. Second, using propensity score matching (PSM), the study's analysis goes beyond the associational and predictive models to rank risk factors and provide an estimate that more closely captures the potential causal effect of each factor (8). While not as unbiased as a randomized control trial, the analyses are able to better control for confounding variables and the development of models serving as an evidence base to reduce adolescent obesity, tailor public health preventive efforts and enhance efficacy of behaviour change or policy strategies. The study is also specified *a priori* to examine whether these factors differ between Black and White female adolescents.

Participants and methods

Study sample

The NGHS cohort of 2,879 female participants was intended to incorporate equal proportions of Black and

White female participants, with Black ($n = 1,213$) and White ($n = 1,166$) female participants followed from age 9 to age 19 came from three urban areas in the USA (Richmond, CA; Cincinnati, OH; and Washington, D.C.). This gave a total study sample of $n = 2,379$. The NHLBI design of sites enabled recruitment of a socioeconomically diverse study sample, obtaining written consent from participants and family members allowing for the creation of public-use deidentified data files for use in statistical analysis and reporting. Stanford Institutional Review Board provided ethics approval for this study.

Measures

There were 13,967 variables included in the publicly available NGHS dataset that could potentially be included for analysis. From these potential variables, the following criteria were used to select those to include in the study's analysis. First, established risk factors pertinent to adolescent obesity by literature review were selected for the analysis. The study included all variables examined in previous NGHS analyses of adolescent obesity, including self-worth scales, physical activity, nutrition, and family factors. The study also restricted to variables selected from years 1, 2 and 3 of the NGHS study (each year from ages 9 to 12) because this precedes the major increases in female adolescents with obesity overall as well as the divergence in adiposity trends between Black and White female adolescents (9).

The study evaluated 285 potential predictor variables based on these criteria (Table S1), cut to 241 when excluding variables likely influenced by baseline obesity such as examiner's assessment of participant's image, attempts at dieting or losing or gaining weight now or in the past 14 years, if family/friends think that the participant is too thin/fat or is a good weight, how the participant sees herself, and menarche. The predictor variables selected for analysis included psychosocial behaviours and parental/social environmental influencing physical activity, eating behaviours, and dietary intake. Psychosocial behaviours came from self-reported data from validated scales/inventories with acceptable reliability, validity, and multidimensional structure. The scales include Habitual Activity Questionnaire (10), Eating Disorders Inventory (11), Emotional Eating Index (12), Perceived Stress Scale (13), The Perceived Competence Scale for Children (14), Manifest Anxiety (15), Coping Strategies Inventory (16) and Self Assertive Efficacy (17). Physical activity was determined by validated questionnaires that a centrally trained interviewer turned into metabolic equivalent of task, and calculations for dietary intake came from a daily food diary kept for 3 days (1 day a weekend). For variables that were categorical

with more than two categories, if any single category contained less than 10% of the sample, it was recoded into an adjacent category. Chi-squared analyses for categorical variables and pairwise correlations for continuous variables were used to determine correlations across years for the variables measured annually. The variables significantly different in correlation from previous years with a correlation coefficient < 0.8 were included. For meaningful comparison of coefficients across different variables, two category variables were recoded to -1 and 1 , three category variables were recoded to -1 , 0 and 1 , and all other variables were z-scored.

The control variables for the study's initial analysis of association were selected based on literature review and standardization from previous studies with the NGHS (7). A standard set of confounders included age in months, race, income, education, social competence, athletic competence, parents in household, and self-worth scale. Additional control variables for the propensity score-matched analyses were also selected based on literature review unique to each predictive factor, because different predictors were plausibly confounded by different factors (listed in Table S2).

The outcome was % body fat (as measured by sum of skinfolds) at age 19. Sum of skinfolds was chosen as the measure of adiposity as growing paediatric patients, particularly in female populations, can relatively easily have this measured to determine % body fat when tracking and diagnosing adiposity for clinically relevant diagnostic purposes. Current clinic practices and Centers for Disease Control guidelines for paediatric populations do not recommend body mass index (BMI) as a diagnostic tool for overweight and obesity, more as a screening tool when used as a percentile for age and sex, especially for multi-ethnic populations where BMI cannot screen as well (18). Any participant who did not have an exiting sum of skinfolds measurement (year 10 of study, age 19) was excluded from the analysis ($n = 355$) and comparison of the analytic sample ($n = 2,024$), and those excluded show similarity in observed demographic characteristics and BMI (Table 1).

Statistical analysis

Analyses of the comparative association of predictor variables with adiposity were conducted through three approaches, namely, multiple regression, random forest and PSM. Multiple regression and random forest analyses were performed using RStudio Version 2.5 using the packages `randomForest`, `rpart` and `party`, and PSM was performed using RStudio Version 3.3.0 using the packages `Matchit`, `Zelig`, `optmatch` and `sandwich`.

Table 1 Demographics of excluded respondents versus analytic sample

Characteristic	Removed from analysis ($n = 355$)	Analytic sample ($n = 2,024$)
Age in years, mean	10.08	10.00
Black, n (%)	170 (48)	1,043 (52)
White, n (%)	185 (52)	981 (48)
Education ^a , n (%)		
<High school	127 (35.8)	490 (24)
College grad +	105 (29.6)	729 (36)
1–3 years post graduate +	121 (34)	804 (40)
NA	2 (0.6)	1 (0.05)
Income ^b , n (%)		
\$0–\$9,999	80 (22.5)	324 (16)
\$10,000–\$19,999	48 (13.5)	275 (13.5)
\$20,000–\$39,999	106 (30)	588 (29)
\$40,000+	98 (28)	725 (36)
NA	23 (6)	112 (5.5)
Baseline BMI, mean ^c	18.9	18.5
Year 10 BMI ^c	32.3	25.4
Male in household, n (%)	231 (65)	1,431 (71)
Number of siblings, mean	1.47	1.48

^aParticipants' education based on parent/guardian report of household income.

^bParticipants' income based on parent/guardian report of household income.

^cOnly 41 participants of the 355 excluded had end-of-study BMI values, with >21 of the 41 ending with a year 10 BMI greater than the third quartile of BMI for those included in the study. BMI, body mass index.

Linear multiple regression models, one for each predictor of interest, were used to identify specific risk factors correlated with adolescent adiposity, controlling for participant's age in months, household income, race and highest level of parental education. For the control variables, the sample mean was substituted for any missing observations. The study identified significant regression variables using adjusted p -values for the propensity score match to account for multiple comparisons.

The study *a priori* tested whether there were differences in predicting between Blacks and Whites based on prior research suggesting that predictors may differ (19,20). Out of 285 factors analysed, only 10 predictor variables were unique to Black participants (significant interaction term for race using adjusted p -value < 0.05) as shown in Table S3. Based on these findings, Black and White girls were not analysed separately for any of the subsequent analyses.

For the random forest analysis, multiple health-related factors were examined simultaneously while providing an additional way of identifying other factors that may be important for subgroups of the population. The study used `randomForest`, `party`, `base` and `Hmisc` R packages to run models on the same set of predictor variables

analysed with multiple regression to complement the multiple regression approach and test for the strength with which factor at baseline predicts later adiposity. This approach avoided biasing the study's models based on significant findings in other studies to instead ascertain the 'true extent of associations' (8). The function roughfix was applied to the predictor variables to replace missing observations with the mean, and the random forest analysis was repeated twice (once for all predictor variables and once for the subset without obesity-related predictor variables). A variable importance ranking of each random forest was created and a *p*-value determined in comparison with a manual random forest sample. One hundred and ninety of the variables were found to be predictors (pseudo *p*-value <0.05) of adiposity from the random forest analysis and were used as exposure variables in the propensity score match as a

tool for the analysis of paediatric obesity as demonstrated with other datasets (20,21).

The 190 predictor variables for the PSM came from the multiple regression and the random forest output of significant adiposity predictor variables to enable PSM to simulate randomization of treatment and control groups (exposure and no exposure) to match the groups across a variety of chosen characteristics. This method has been reported as an important practice in determining efficacy of preventive actions (22). PSM has been emphasized as a uniquely helpful analysis in observational studies to show causal effects. With the development of updated software programming and capabilities, researchers can handle larger amounts of this observational data to determine causal effects. The 190 predictors of interest could be divided broadly into three categories of exposure. (a) Diet/eating behaviours: all nutritional averages and eating practices

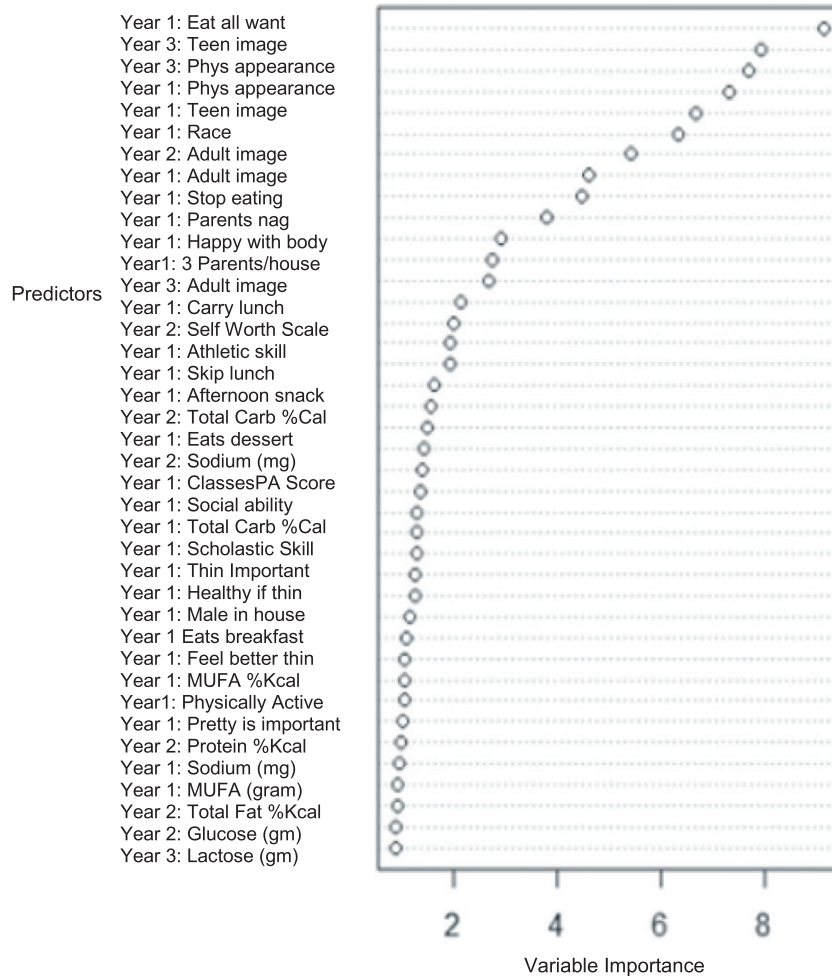


Figure 1 Random forest variable importance of adiposity predictors

were captured; (b) biologic and psychologic: scales and biologic milestones; and (c) social and environmental: family, friend, home and school influence. For the matched analyses, the risk factors were recoded to binary (Table S4) to divide into treatment and control groups for each variable. This was carried out based on literature review for standardized cut-offs associated with higher risk for adiposity or poor outcomes, which include binary divides for predictors such as total calories and vitamins (23), SFA, sodium, all types of fats and SFA limit (24).

After comparing the matched eQQ plots as a balance test to determine best matching methods for the 190 variables out of the 'nearest', 'optimal', and 'genetic' method, the 'optimal' method was chosen, with the 'full' method reserved for variables with fewer exposures after matching. The difference in means between matched treatment and control groups was used to comparatively rank the risk factors to determine the quantitative importance. Standard deviations and confidence intervals (CI) were also calculated in order to understand the range and stability of the estimates of differences. As a sensitivity analysis of the study's primary models, the models were fit to additionally control for sum of skinfolds at baseline. Population attributable risk was calculated for the three subcategories of predictor variables.

Results

The multiple linear regression models suggested that there were distinct adiposity risk factors at various points during adolescence when using the sum of skinfolds at age 19 as the measure of adolescent adiposity. The false discovery rate method was used to adjust for multiple comparisons in the analyses, as reflected in the adjusted pseudo p -value. The 157 significant predictors are presented in Table S5, with some of the strongest significant adiposity predictors ($p < 0.05$) including dieting to lose weight (coefficient of association: 27.8) and how the participant looks at age 9 (coefficient: 13.1). Strong protective factors were mom being satisfied with child's weight (coefficient: -12.2) and eating all she wants (coefficient: -9.8).

The random forest analysis broadened the scope of significant risk factors and protective factors (variable importance ranked list appears in Table S6). Figure 1 shows the top 40 predictors from the random forest analysis. Significant predictors from the multiple linear regression that were not found to be significant in the random forest analysis are listed in Table S7. Drawing from these variables resulted in the study's list of 190 significant variables ($p < 0.05$) to test using PSM. Figure S1 shows the false discovery rate-adjusted pseudo p -value histogram.

Propensity score matching ranked the order of importance of the various adiposity protective factors and adiposity risk factors controlling for baseline adiposity. Tables 2 and 3 show top results. Table S8 shows all PSM ranked results with difference scores. Results indicated high impact of certain predictors across three categories, namely, diet/eating behaviours (17 protective factors, five risk factors), biological/psychological (eight protective factors, six risk factors) and social/environmental (no protective factors, six risk factors). Protective diet/eating behaviours include at ages 9–10, eating all she wants (difference score -10.1, 95% CI -13.2 to -7.1), the next year eating at sports (difference score 9.5, 95% CI -14.6 to -3.9) and beginning teen years eating with homework (difference score -6.8, 95% CI -9.7 to -3.7). Biological/psychological protective factors included feeling physical appearance is important as early as ages 9–12. In terms of risk factors, the strongest risk factor for female adolescents comes from the social/environmental category: if the participant's mother looks overweight (difference score 18.0, 95% CI 12.9 to 22.8) followed by if the participant believes she will look overweight as a teenager (difference score 13.0, 95% CI 6.9 to 19.7) and others such as father looking overweight or relative dieting. Several diet/eating behaviours showed risk, such as excessive aspartame >18.4 mg/day (difference score 8.5, 95% CI 3.6 to 13.2), higher average of saturated fatty acids (grams) >30.3 g (difference score 5.8, 95% CI 1.4 to 10.7), skipping lunch (difference score 4.3, 95% CI 1.6 to 7.2) and consuming higher amounts of starch, cholesterol, monounsaturated fatty acids and caffeine. In each subcategory of the predictors (diet/eating behaviours, biological/psychological, and social/environmental 44–45% of tested predictors [54 total] significantly impacted adiposity.

Figure 2 plots the population attributable risk for the three subcategories of predictors after sensitivity analysis. Binary recoding necessary for the attributable risk calculation was performed per perceived risk or protection upon exposure to each predictor variable. The figure shows that around 43% (95% CI 73–90%) of female adolescent adiposity can be attributed to social and environmental factors, and 82% (95% CI 32–54%) to biological/psychological factors. The reason these add up to more than 100% is because they are not mutually exclusive causes of adolescent adiposity – social and environmental factors may impact biological and physiological factors. The study did not find evidence that diet and eating behaviour contributed substantially because of a very wide CI around this estimate caused by more substantial variance in measurement for these factors.

Table 2 PSM: significant adiposity protective factors

PSM: significant adiposity protective factors	Difference score	SD	Lower 95% confidence interval	Upper 95% confidence interval
<i>Physical Appearance Scale – Year 3</i>	–10.426	1.63	–13.57	–7.36
Eats all she wants – Year 1	–10.125	1.54	–13.20	–7.13
Eat at sports – Year 2	–9.521	2.84	–14.58	–3.88
<i>Physical Appearance Scale – Year 1</i>	–9.239	1.60	–12.49	–6.08
Eats at sports – Year 3	–7.660	2.66	–12.95	–2.23
Eats with homework – Year 3	–6.846	1.54	–9.75	–3.71
Eats big helpings – Year 2	–6.757	2.73	–12.03	–1.57
Eats with homework – Year 2	–6.201	1.43	–9.06	–3.45
Eats with friends – Year 3	–5.953	1.97	–9.99	–2.17
Eats in bedroom – Year 2	–5.764	2.33	–10.20	–1.15
Eats with TV – Year 2	–5.755	2.13	–10.14	–1.53
Eats with homework – Year 1	–5.732	2.75	–11.06	–0.09
Eats while watching TV – Year 3	–5.337	1.59	–8.49	–2.19
<i>Self-worth scale – Year 1</i>	–4.983	1.48	–7.78	–2.13
Total sugars – Year 3	–4.882	1.71	–8.17	–1.59
<i>Athletic Competence Scale – Year 3</i>	–4.658	1.68	–8.03	–1.46
<i>Self-worth scale – Year 3</i>	–4.538	1.40	–7.20	–1.81
Eat with friends – Year 2	–4.384	1.62	–7.55	–1.40
Eat as reward – Year 2	–4.280	1.38	–6.86	–1.46
Eat when very hungry – Year 3	–3.782	1.53	–6.92	–0.62
Parents fix snack – Year 3	–3.747	1.77	–7.19	–0.18
<i>Physically active – Year 1</i>	–3.571	1.36	–6.31	–1.03
<i>Think being fat is because it is natural – Year 1</i>	–3.158	1.38	–5.90	–0.53
<i>Social competence – Year 3</i>	–3.124	1.54	–5.94	–0.15
Evening snack – Year 1	–3.116	1.40	–5.85	–0.29

Subcategories of variables. Bold indicates diet/eating behaviours, and italic indicates biological/psychological. Year 2, sucrose (gm); Year 2, eats dessert; and Year 2, total carb (gm) removed from significant protective factors because they are significantly correlated with baseline sum of skinfolds. PSM, propensity score matching.

Table 3 Propensity score matching: significant adiposity risk factors

PSM: significant adiposity risk factors	Difference score	SD	Lower 95% confidence interval	Upper 95% confidence interval
Participant's image of mother – Year 2	18.028	2.54	12.89	22.78
<i>Image as teen – Year 2</i>	13.030	3.32	6.90	19.66
<i>Image as teen – Year 3</i>	13.005	3.53	6.06	20.11
<i>Image as teen – Year 1</i>	12.988	3.34	6.33	19.33
Participant's image of mother as above average – Year 3	9.925	2.19	5.97	14.31
Avg. Aspartame (grams) – Year 2	8.492	2.48	3.60	13.19
Participant's image of father – Year 3	8.253	1.79	4.58	11.76
Participant's image of father – Year 2	6.808	1.66	3.56	10.05
Participant's image of father – Year 1	6.764	1.60	3.72	9.86
<i>Healthy to be thin – Year 3</i>	4.400	1.23	1.89	6.71
Skips lunch – Year 3	4.329	1.45	1.58	7.20
Relative has dieted – Year 1	4.288	1.63	1.14	7.43
<i>Belief of No self-control if overweight – Year 1</i>	4.225	1.37	1.55	6.79
Skip lunch – Year 2	3.993	1.39	1.14	6.65
Starch (grams) – Year 1	3.916	1.93	0.23	7.88
Monounsaturated Fatty Acids (% Kcal) – Year 1	3.825	1.86	0.20	7.69
<i>Image hope to look like as adult – Year 2</i>	3.453	1.71	0.03	6.81

Subcategories of variables. Bold indicates diet/eating behaviours, and italic indicates biological/psychological. PSM, propensity score matching.

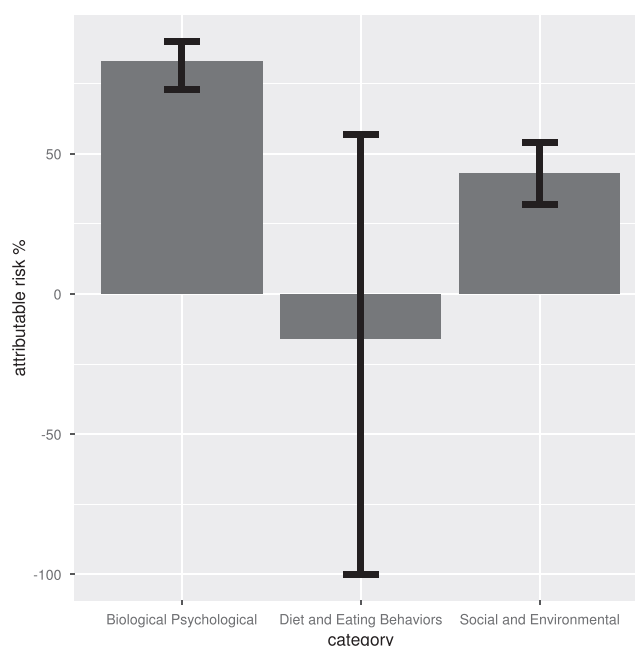


Figure 2 Population attributable risk plot – stratified by variable category

Discussion

High proportions of people with obesity impose a burden on population health and healthcare systems. With current analytic and computational tools, the ability to analyse large datasets with many potential predictor variables has increased. This allows us to examine on a more granular level how to get to the root cause of increased amounts of people dealing with the consequences of being overweight or obese at earlier ages and how to interpret data in a way that moves towards intervention studies. This study is a step in the direction of evidence generation that can inform the public and healthcare providers alike to help coordinate efforts along the life course, starting from an early age, to potentially decrease rising obesity rates. Overall, the study's results suggest that a substantial proportion of adolescent adiposity can be explained by the factors examined.

The relevant findings from this study include the determination of new harmful associations associated with adiposity in adolescents such as small aspartame doses, skipped lunches, monounsaturated fatty acids in diet and beliefs as seemingly benign as 'it is healthy to be thin' or 'overweight people lack self-control'. These risk factors are relevant to clinical and public health discussions as significant change may likely be seen with interventions for specific dietary products like aspartame, planned meal eating behaviours and reinforcing healthy mindsets and beliefs about body weight and shape. These can be discussed with adolescents and pre-adolescents through

various clinic-based, school-based or public messaging interventions taking place in one-to-one, small-group or population-based settings.

This study's findings of protection from adiposity by a female adolescent eating all she wants, eating at sports events, eating with homework, eating big helpings, eating in the bedroom and eating with TV are surprising but may deter maladaptive eating behaviour when free access to healthy foods is present (7,29–31). The behavioural distractions when food options containing balanced nutrition may allow for adolescents to focus less on common adolescent desires to eat high-fat, high-sodium and high-sugar foods and focus more on other activities instead of the eating itself as the 'treat' of sorts. Some counter-intuitive findings or protective factors like 'Total sugars' may need follow up to explore if this higher total sugar content is protective because the increased total sugars are coming from fruits and vegetables (i.e. healthier reason for higher total sugars) that can keep hunger at bay for longer with the additional fibre and complex carbohydrates. This explains a relative protection against adolescents becoming overweight or obese people because the sugars from fruits and vegetables add extra protective benefit over added sugars like high-fructose corn syrup that can cause a sugar crash and subsequent binging on non-nutritious, empty calorie foods.

This study's findings show consistency with research that has confirmed the adiposity risk factors ranked in this analysis: mothers and fathers who are overweight serve as major determinants of children who are overweight or

obese (25), how childhood and teenage obesity can set future morbidities, family dieting behaviour promoting unhealthy eating practices for the adolescents in the home, and starch (26). For protective factors, previous literature also emphasizes the beneficial impact of high self-worth, parents fixing a snack and physical activity (27,28). There are several limitations to this analysis. This study is an observational analysis limited in its inability to generalize beyond urban cities like those in the study. In addition, many risk factors that could be important could have been missed given that data collection began at age 9. Looking earlier in the life course, even epigenetically and multi-generationally, may provide additional prospects of identifying the risk factors that contribute to obesity in the USA. Additionally, using PSM does inherently come with limitations due to the way the data are analysed, such as 'model mis-specifications', categorical variables with more than two levels, difficulties in handling missing data and nonlinear relationships (8). This study's interpretation accordingly considers potential nuances of participant response lost to binary exposures and confounders not surveyed that could not be included. In addition, because of the unique nature of the study's data, there is not another independent study to confirm the study's findings, although results are consistent with prior research as mentioned earlier. Lastly, the study's inclusion of an outcome measure of sum of skinfolds to determine % body fat is limited despite its more widespread ability to be measured in clinical settings in the USA as a type of diagnostic tool. Other analyses focused on more accurate diagnostic tools like underwater weights and DEXA scan could provide further insights. For analyses related to risk factors for someone to screen positive for overweight or obesity, the outcome of BMI percentile for age and sex could be assessed to assist with understanding correlations related positive screens for overweight and obesity taking place at clinic visits.

While not from a randomized control trial, the study's findings can inform experiments that test them through obesity interventions or public health promotion practices. Data used for the analysis were obtained systematically at repeated intervals with standardized, multidimensional, reliable and valid surveys and provided a large sample size ($n = 2,024$), adding to the study's strength.

Conclusion

Clinical approaches that can be implemented during visits in paediatric practices and studied for efficacy through prospective clinical trials would be an ideal follow-up in addition to other community-based experimental interventions in partnership with obesity prevention programmes. Based on the study's analyses,

possible pan-categorical interventions include (a) enhancing perceptions of healthy weight through biopsychological interventions that deemphasize extremes in weight and promote self-worth, physical activity, social competence and acknowledgement of the relationship between physical appearance and health, (b) limiting social/environmental negative influence of family dieting behaviours or familial obesity, and (c) focusing diet and eating behaviour efforts on lowering SFA, MUFA, cholesterol, starch, caffeine and aspartame and arranging trials ensuring daily lunch, eating with others/during activities and parents fixing evening snacks. Repeating an analysis from this paper with all the included variables in the NGHS dataset could add to the robustness of insights from this analytic approach, as it could help narrow possible interventions from those this study generated by providing negative controls. Results indicating specific public health experimental intervention directions add a prioritized and multi-focused approach to prevent and help address the impact of being overweight or obese on the lives of female adolescents in the USA.

Acknowledgements

Authors thank Stanford Medical Scholars for funding, Stanford Medicine and Stanford Center on Poverty and Inequality for facilities and the National Growth and Health Survey for providing public use of the data for statistical analyses and reporting. Rehkopf is supported by a grant from the National Institute of Aging.

Funding

Stanford Medical Scholars Program (Narla); National Institute of Aging (K01 AG047280 to Rehkopf).

Financial disclosure

The authors have no financial relationships relevant to this article to disclose.

Conflict of interest statement

The authors have no conflicts of interest relevant to this article to disclose.

Contributors' statements

D. H. R. and A. N. conceptualized and designed the study, analysed the data, drafted the initial manuscript and reviewed and revised the manuscript. Both authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

References

1. Maier JH, Barry R. Associations among physical activity, diet, and obesity measures change during adolescence. *J Nutr Metab* 2015;2015:805065. <https://doi.org/10.1155/2015/805065>, 1–8
2. Huh D, Stice E, Shaw H, Boutelle K. Female overweight and obesity in adolescence: developmental trends and ethnic differences in prevalence, incidence, and remission. *J Youth Adolesc* 2012; **41**: 76–85. <https://doi.org/10.1007/s10964-011-9664-4>.
3. Lee H, Pantazis A, Cheng P, Dennisk L, Clarke PJ, Lee JM. The association between adolescent obesity and disability incidence in young adulthood. *J Adolesc Health* 2016; **59**: 472–478. <https://doi.org/10.1016/j.jadohealth.2016.05.015>.
4. Gracia-Marco L, Moreno LA, Ruiz JR, et al. Body composition indices and single and clustered cardiovascular disease risk factors in adolescents: providing clinical-based cut-points. *Progr Cardiovasc Dis* 2016; **58**: 555–564. <https://doi.org/10.1016/j.pcad.2015.11.002>.
5. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA* 2004; **291**: 1238–1245. <https://doi.org/10.1001/jama.291.10.1238>.
6. Kimm SY, Barton BA, Obarzanek E, et al. Obesity development during adolescence in a biracial cohort: the NHLBI Growth and Health Study. *Pediatr* 2002; **110**: e54–e54. <https://doi.org/10.1542/peds.110.5.e54>.
7. Rehkopf DH, Laraia BA, Segal M, Braithwaite D, Epel E. The relative importance of predictors of body mass index change, overweight and obesity in adolescent girls. The relative importance of predictors of body mass index change, overweight and obesity in adolescent girls. *Int J Pediatr Obes* 2011; **6**: e233–e242. <https://doi.org/10.3109/17477166.2010.545410>.
8. Zhao P, Su X, Ge T, Fan J. Propensity score and proximity matching using random forest. *Contemp Clin Trials* 2016; **47**: 85–92. <https://doi.org/10.1016/j.cct.2015.12.012>. 1162/003465302317331982.
9. Kimm SY, Barton BA, Obarzanek E, et al. Racial divergence in adiposity during adolescence: the NHLBI Growth and Health Study. *Pediatr* 2001; **107**: e34–e34. <https://doi.org/10.1542/peds.107.3.e34>.
10. Kimm SY, Glynn NW, Obarzanek E, et al. Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. *The Lancet* 2005; **366**: 301–307. [https://doi.org/10.1016/S0140-6736\(05\)66837-7](https://doi.org/10.1016/S0140-6736(05)66837-7).
11. Garner DM, Olmsted MP. *The Eating Disorder Inventory Manual*. Psychological Assessment Resources: Odessa, FL, 1984.
12. Arnow B, Kenardy J, Agras WS. The Emotional Eating Scale: a development of a measure to assess coping with negative affect by eating. *Int J Eat Disord* 1995; **18**: 79–90. [https://doi.org/10.1002/1098-108X\(199507\)18:1<79::AID-EAT2260180109>3.0.CO;2-V](https://doi.org/10.1002/1098-108X(199507)18:1<79::AID-EAT2260180109>3.0.CO;2-V).
13. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983; **24**: 385–396. www.jstor.org/stable/2136404.
14. Harter S. *The Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children (Manual)*. University of Denver: Denver, CO, 1985.
15. Reynolds C, Richmond B. What I think and feel: a revised measure of children's manifest anxiety. *J Abnorm Child Psychol* 1978; **6**: 271–280. <https://doi.org/10.1007/BF00919131>.
16. Tobin DL, Holroyd KA, Reynolds RV, Wigal JK. The hierarchical factor structure of the coping strategies inventory. *Cognitive Ther Res* 1989; **13**: 343–361. <https://doi.org/10.1007/BF01173478>.
17. Sabatelli RM, Anderson SA. *Assessing Outcomes in Child and Youth Programs: A Practical Handbook Revised Edition*. US Department of Justice to the State of Connecticut: Connecticut, 2005.
18. Ortega FB, Sui X, Lavie CJ, Blair SN. Body mass index, the most widely used but also widely criticized index: would a criterion standard measure of total body fat be a better predictor of cardiovascular disease mortality? *Mayo Clin Proc* 2016; **91**: 443–455. <https://doi.org/10.1016/j.mayocp.2016.01.008>.
19. Neff LJ, Sargent RG, McKeown RE, Jackson KL, Valois RF. Black–White differences in body size perceptions and weight management practices among adolescent females. *J Adolesc Health* 1997; **20**: 459–465. [https://doi.org/10.1016/S1054-139X\(96\)00273-X](https://doi.org/10.1016/S1054-139X(96)00273-X).
20. Bemile EN and Anders S. Linking diet-health behaviour and obesity using propensity score matching. International Congress. 1–13. *Economics Letters*. 2014; 120(3): 413–415. [online]. Available from: http://ageconsearch.umn.edu/bitstream/182832/2/Diet_Health_Behaviour_and_Obesity_Bemile_Anders_EAAE.pdf.
21. Brown H, Roberts J. Born to be wide? Exploring correlations in mother and adolescent body mass index. *Econ Lett* 2013; **120**: 413–415.
22. Gibson LA, Hernández Alava M, Kelly MP, Campbell MJ. The effects of breastfeeding on childhood BMI: a propensity score matching approach. *J Public Health (Oxf)* 2017; **39**: e152–e160. <https://doi.org/10.1093/pubmed/fdw093>.
23. Hebl MR, King EB, Perkins A. Ethnic differences in the stigma of obesity: identification and engagement with a thin ideal. *J Exp Soc Psychol* 2009; **45**: 1165–1172. <https://doi.org/10.1016/j.jesp.2009.04.017>.
24. Haukoos JS, Lewis RJ. The propensity score. *JAMA*. 2015; **314**: 1637–1638. <https://doi.org/10.1001/jama.2015.13480>.
25. Brown JE. *Nutrition through the Life Cycle*. Cengage Learning: Australia, 2016.
26. Koletzko B, Bhatia J, Bhatta ZA, et al. *Pediatric Nutrition in Practice*. Karger Medical and Scientific Publishers, 2015, pp. 308–315.
27. Danielzik S, Czerwinski-Mast M, Langnäse K, Dilba B, Müller MJ. Parental overweight, socioeconomic status and high birth weight are the major determinants of overweight and obesity in 5–7 y-old children: baseline data of the Kiel Obesity Prevention Study (KOPS). *Int J Obes (Lond)* 2004; **28**: 1494–1502. <https://doi.org/10.1038/sj.ijo.0802756>.
28. http://pediatrics.aappublications.org/content/101/Supplement_2/539.short
29. Sondike SB, Copperman N, Jacobson MS. Effects of a low-carbohydrate diet on weight loss and cardiovascular risk factor in overweight adolescents. *J Pediatr* 2003; **142**: 253–258. <https://doi.org/10.1067/mpd.2003.4>.
30. Strauss RS. Childhood obesity and self-esteem. *Pediatrics* 2000; **105**: e15–e15. <http://pediatrics.aappublications.org/content/105/1/e15.short>.
31. Benton D. Role of parents in the determination of the food preferences of children and the development of obesity. *Int J Obes (Lond)* 2004; **28**: 858–869. <https://doi.org/10.1038/sj.ijo.0802532>.
32. Friedman KE, Reichmann SK, Costanzo PR, Zelli A, Ashmore JA, Musante GJ. Weight stigmatization and ideological beliefs: relation to psychological functioning in obese adults. *Obes* 2005; **13**: 907–916. <https://doi.org/10.1038/oby.2005.105>.
33. Braude L, Stevenson RJ. Watching television while eating increases energy intake. Examining the mechanisms in female participants. *Appetite* 2016; **76**: 9–16. <https://doi.org/10.1016/j.appet.2014.01.005>.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: 285 Regression and Random Forest Predictor variables, NGHS data label and variable definition *Italics indicates factor is likely to be effected by prior obesity at baseline.*

Table S2: List of variables used as control variables for propensity score matching models, NGHS available upon request

Table S3: Variables with Significant Race (Black or White) Interactions (adjusted $p < 0.05$) based on regression

Table S4: Cutoffs for Binary Recode for Propensity Score Matching Exposures

Table S5: 157 Significant Predictors (adjusted p-value) from regression (year 10 Sum of Skinfolts as outcome)

Table S6. Random Forest Variable Importance Ranking for predictor variables of adiposity

Table S7: Significant Variables (adjusted p value) unique to regression compared to 289 significant Random Forest Variables

Table S8. Mean differences for all propensity score matched predictor variables, NGHS.

Figure S1: Histogram of False discovery adjusted pseudo-P values from RandomForest model predicting adiposity, NGHS.