SPECIFIC AIMS

Substantial research has shown a significantly negative association between screen time and diet quality[1-5]. Heavy television viewers have higher intakes of nutrient-poor and energy-dense foods[1], and have less preference for fresh fruits, vegetables, whole grain, and calcium-rich food[2]. Poor diet behavior can increase cardio-metabolic risk[3] and cause adverse chronic conditions[6]. High-sugar and high-fat foods are the risk factors for high blood pressure, abnormal BMI, high sympathetic arousal, and cortisol dysregulation, which impair stress regulation[7]. However, most of the studies only focus on screen time's impact on children, adolescents, and young adults, while the analysis of older adults has been markedly ignored. Moreover, little research investigates the interconnection between screen time, diet quality, and overall health. Therefore, our team intends to explore the relationship between screen time and diet quality, and how diet quality further affects overall health among older adults aged 65 and older.

In our preliminary exploratory analysis, we analyzed data from UK Biobank, which recruited 490,966 UK participants, aged 37 to 73 years, from 2006 to 2010[8]. Assuming 2008 as the investigation year, we have discovered that (1) There are 10476 participants aged 65 and older, and 27 nutrient intakes with less than 50% missing values. (2) There are two variables: time spent on television and time spent on the computer can measure participants' daily screen time. (3) The median time that older adults spent on television are 3 hours/day, while the median time spent on the computer is 1 hour/day. Utilizing this information, we can divide the 10476 observations into two groups based on the median of total screen time - 4 hours/day. (4) Both time spent on television and time spent on the computer are negatively correlated to fresh fruit intake and positively correlated to process meat intake, which indicates a negative association between screen time and diet quality. (4) Fresh fruit intake is positively correlated and salt added to food is negatively correlated to overall health rating, providing essential evidence of the positive relationship between diet quality and health outcomes.

Given these findings, we hypothesized that among older adults aged 65 and older (65+), increased daily screen time will impact their diet behaviors and impair their diet qualities, which further causes adverse health outcomes. Specifically, the more time they spend on television or the computer, the fewer fresh fruits and vegetables they are likely to intake, and the more salted, high-sugar, and high-fat food they would intake. Adverse diet patterns will directly result in poor health outcomes.

These hypotheses will be addressed in different statistical and machine learning methods based on secondary analyses of the UK Biobank data with the following Specific Aims:

Specific Aim 1. To investigate the association between time spent on television or computer among older adults and their dietary quality.
Multiple diet quality indices will be used to measure the mean diet score of each observation. ANOVA test will be conducted to compare the mean diet score difference between older adults who spent more than 4 hours of screen time per day and older adults with less than 4 hours per day. Multivariable and mediation analysis will be constructed to examine the relationship by considering confounders: age, sex, and social economic status.

Specific Aim 2. To investigate the association between dietary quality and overall health among older adults aged 65+.
Classification methods will be developed to train the data set based on diet quality indices and other baseline covariates with the labeled overall health rating. Multivariable analysis will be used to investigate the relationship between diet quality score and overall health rating.

We anticipate that this study will reveal the adverse impact that increased screen time has on diet quality, and poor diet quality affects older adults' health outcomes. If the hypothesis is proven, we can control older adults' daily screen time, and purposely adjust the diet pattern for heavy screen time users.

RESEARCH STRATEGY

1 SIGNIFICANCE

The proportion of internet users, among American older adults 65+, has been dramatically increasing from 14% to 73% for the past two decades, and on average, they spent more than half of their daily leisure time, about 4 hours and 16 minutes, in front of the screen[9]. Increased sedentary time and decreased physical activity are directly linked to adverse physical and psychological health outcomes[10,11]. Prolonged exposure to screen time may disrupt sleep patterns and lead to a decrease in the time and energy spent on preparing and consuming healthy meals. Based on previous literature, sedentary time can be categorized as television viewing, computer use, and sitting time[12]. Since sitting time is not able to be obtained from the UK Biobank, we will use television and computer use as screen time measurements. We have previously mentioned that television and computer use is associated with an increased intake of energy-dense, nutrient-poor foods, and beverages, which can negatively impact dietary quality. Therefore, if our hypothesis is proven effective, we can improve older adults' diet quality and overall health conditions by reducing screen time.

2 INNOVATION

During the literature review, we were astonished by the limited study about the screen time impact on diet quality among older adults 65+. Most of the studies concern more about children's or teenagers' screen time and diet issues[1,2]. We assume that researchers are more concerned about adolescents' addiction issues to television and computer use, and neglect the fact that older adults have more opportunities to access electronic devices when they have more leisure time after retirement. Our study will fill the gap in older adults' diet and health problems caused by extended screen viewing. From our study's anticipated results, clinicians can pay more attention to the diet intake of older adults with long screen time, adjust their deficient diet behaviors and improve their health outcomes.

3 RESEARCH PLAN

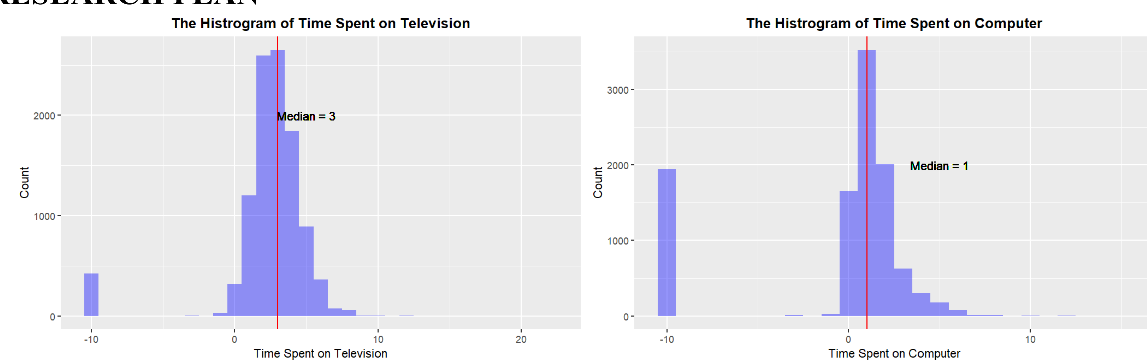

Figure 1: The histograms of screen time distribution among older adults 65+

Analysis Preparation. The screen time data, collected from the UK Biobank, is self-reported by participants. The distribution of daily screen time is shown in Figure 1. Since the median time of television watching is 3hr/day and computer use is 1hr/day, it is reasonable to split the targeted observations into two experimental groups at 4hr/day. During our preliminary analysis, we also found that screen time was significantly associated with age. That is to say people become more sedentary, as they aged.

Diet Quality Score Measurement. The most challenging task of this project is to quantitatively measure diet quality. Since no single method can quantify the 27 nutrition intakes, we propose to use multiple verified diet quality indices to create a new diet quality score, considering the variation in diet, each nutrition measurement score, and relative weight to combine all factors. Based on previous research on US Biobank's diet data, we will utilize Recommended Food Score (RFS), Mediterranean Diet Score (MDS), Healthy Diet Indicator (HDI), and Dietary Approaches to Stop Hypertension (DASH) as the nutrition measurement indices[13-16], while exploring WELL diet score as the variation in participants' diet[17].

Statistical Methods. One-way ANOVA (analysis of variance) has been employed in numerous research to compare the means of two groups for one dependent variable[18], which in our situation, can test if the mean diet scores among heavy and light screen time users are significantly different. Multivariable analysis like linear regression and multi-level logistics regression, which are widely used regression models to estimate the effect between exposure and outcome[13,19], will be used to further explore the association between screen time and diet score, and between diet score and overall health, adjusted for age, sex, and social economic status.

Data Mining Classification Models. Among various machine learning algorithms, tree-based models like Decision-Tree and Random-Forest are well suited for multi-categories clinical classification problems. Besides coping with missing values, tree-based models combine heterogeneous data types into a single model[20]. Since the overall health rating in UK Biobank has five different rating classes, tree-based models will have satisfactory performance on the diet data. Compared to tree-based models, SVMs, which can work as linear and non-linear classifiers, are better at dealing with high-dimensional data[21]. However, SVM is not initially designed for multi-classes problems. We will split the classification issue into five SVM classifiers.

Specific Aims 1. To investigate the association between time spent on television or computer among older adults and their dietary quality.

Hypothesis. Older adults aged 65+ who spend more than 4 hours per day watching television or using a computer will have lower dietary quality scores, as measured by the diet quality index, compared to those who spend less than 4 hours per day.

Rationale. To quantify the diet quality, we can construct a diet quality index or improve established indices used in other literature[13-16] to provide a comprehensive measure of overall diet quality, which considers various aspects of diet quality, including the consumption of fruits, vegetables, whole grains, dairy, protein foods, added sugars, saturated fats, and sodium, among others. Using a diet quality index allows for a standardized and objective measure of overall diet quality, which can be compared across different populations and time periods. To examine the relationship between time spent on television or computer and dietary quality, we can conduct hypothesis testing to check the difference in dietary quality between two groups based on statistical inference. Then, we can conduct regression analysis to quantify the influence of time on television or computer on dietary quality, adjusting for confounders. Furthermore, we can perform mediation analysis to examine the underlying mechanisms by which time on TV/PC affects dietary quality. We can check whether the relationship between those two is partially or fully mediated by one or more intermediate variables, e.g., activity time.

Experimental Approach. In this portion of the study, we first calculate the diet quality score for each participant based on the improved dietary quality index. Then we conduct descriptive statistics to summarize the characteristics of two target groups. We examine their age, sex, and other socioeconomic statuses, if possible, physical activity level coupled with average dietary score and time spent on TV/PC. Then, we can do ANOVA test to assess the association between time spent and diet quality. We divide the participants into two groups based on their time spent: one group with more than 4 hours per day and the other with less than 4 hours per day. Then, we calculate the mean diet quality score for each group and compare the means using ANOVA. Next, we conduct multivariable analysis using linear or multilevel logistic regression based on the data type of dietary quality. We use the calculated dietary quality score as the dependent variable and indicator variable of time as the independent variable including any potential confounders, such as age, sex, education level, socioeconomic status, and physical activity. To address the confounding problem, we may use stratification or regression methods. Furthermore, we can conduct the mediation analysis with physical activity as the mediator between time spent on TV/PC and diet quality. We first estimate the direct effect of time on diet quality, controlling for physical activity using a regression model, where physical activity is a covariate. Next, we estimate the indirect effect of time on diet quality, through a physical activity using a mediation model, where physical activity is the mediator. Then, to determine whether physical activity mediates this relationship, we test the significance of the indirect effect using a bootstrapping procedure. We can generate multiple bootstrap samples from the original data, estimate the indirect effect for each sample, and calculate a 95% confidence interval for the indirect effect. If the CI does not include zero, this indicates the indirect effect is statistically significant, and physical activity mediates the relationship between time and diet quality. Finally, we can calculate the proportion of the total effect of time on diet quality that is mediated by physical activity.

Interpretation of Results. For descriptive statistics, we can examine the descriptive statistics to identify any patterns or trends that may help explain the results, such as differences in specific dietary components or physical activity levels between the two groups. For ANOVA, a low p-value indicates a significant difference between the two groups. For Multivariable regression, a negative coefficient estimate for time spent on TV/PC would suggest that spending more time on TV/PC is associated with lower dietary quality scores. The p-value associated with the coefficient estimate would indicate whether the association is statistically significant. For Mediation analysis, a significant indirect effect and a high proportion of the total effect mediated by physical activity indicate physical activity plays a significant role in the relationship between time spent on TV/PC and diet quality.

Potential Problems and Alternative Approaches. The relationship between diet quality and health outcomes may not be linear. Instead of using a continuous numeric variable, we can categorize diet quality into ordinal or nominal variables since it can help capture potential non-linear associations and patterns that may be missed when using a continuous variable. To fully understand the difference in diet patterns among participants spending time on TV/PC over 4 hours per day and participants spending time on TV/PC less than 4 hours per day, we can apply the PCA method on 20+ nutrition intake variables to extract several components and identify the pattern differences using PCA graph.

Specific Aims 2. To investigate the association between dietary quality and overall health among older adults aged 65+.

Hypothesis. There is a significant association between diet quality and overall health among older adults aged 65+.

Rationale. To estimate the association between diet quality and overall health status, we rely on a framework of causal inference. Our approach involves fitting a random forest model that uses several variables related to overall health and diet quality as covariates, with overall health as the response. We chose the random forest method for its flexibility in capturing non-linear effects and complex interactions between covariates. After fitting the model, we will evaluate the counterfactual overall health status for each individual and calculate the average treatment effect to assess the association between diet quality and overall health status.

Experimental Approach. We have adopted the causal random forest model to investigate the relationship between diet quality and overall health status[22]. Since overall health status can be associated with several factors in the table, we intend to select baseline covariates in addition to the diet quality index to improve the fitting. Next, we will use a random forest model to fit the overall health status and other selected covariates. The fitting parameters will be chosen through cross-validation. Once the model is fitted, we will predict the counterfactual health status of each individual. This will involve manually inputting a counterfactual diet quality index for each individual and predicting the counterfactual health status using the fitted model while keeping the other baseline covariates constant. We will evaluate the conditional average treatment effect (CATE) for each individual by determining the difference between their true health status and the predicted counterfactual health status. We will calculate the average CATE of all individuals to determine the average treatment effect (ATE), which will be used to assess the causal relationship. Both CATE and ATE will be evaluated for misclassification error rate, which will be calculated using bagging[23].

Interpretation of Results. The evaluation of the results will be based on multiple metrics, primarily focusing on the misclassification rate as health status is a categorical variable. A large average treatment effect (ATE) with an acceptable misclassification error rate will be considered evidence of a strong association between the health quality index and overall health status. On the other hand, an individual-level analysis will be carried out using the conditional average treatment effect (CATE), which provides a more detailed perspective. The estimated distribution of CATE will be examined to determine if it aligns with our expectations. For instance, if one's diet quality index is significantly altered, we anticipate a significant change in overall health status. We expect to observe a substantial association between the health quality index and overall health status in the majority of individuals.

Potential Problems and Alternative Approaches. One of the potential problems is that there are unknown confounders[16]. It is hard to ensure unconfoundedness, which is an important assumption that our analysis method is based on. We will try multiple methods to select the covariates involved and we may also conduct sensitivity analysis to ensure unconfoundedness assumptions. Another potential problem with the study is that the assessment of self-reported overall health relies on a single-item question, which may be ambiguous and result in participants using inconsistent standards[24]. Discrepancies between self-reported overall health rating and actual health status are expected. To address this, it may be necessary to collect additional data on participants' medical history, which could be used to extract more reliable information on health status and to adjust the self-reported overall health[25]. Besides, overfitting is a common problem for tree-based methods. Hyperparameter tuning for the regression tree model can also be very time-consuming. Alternatively, multilevel logistic regression can be utilized as a reference.

REFERENCES

1. Shea BE, Harvey-Berino JR, Johnson RK. Watching television: how does it influence the dietary quality of children? *Nutrition Bulletin.* 2010;35(2):165-171.
2. Barr-Anderson DJ, Larson NI, Nelson MC, Neumark-Sztainer D, Story M. Does television viewing predict dietary intake five years later in high school students and young adults? *International Journal of Behavioral Nutrition and Physical Activity.* 2009;6(1):7.
3. Wagnild JA-O, Pollard TA-O. How is television time linked to cardiometabolic health in adults? A critical systematic review of the evidence for an effect of watching television on eating, movement, affect and sleep. (2044-6055 (Electronic)).
4. Myszkowska-Ryciak J, Harton A, Lange E, et al. Reduced Screen Time is Associated with Healthy Dietary Behaviors but Not Body Weight Status among Polish Adolescents. Report from the Wise Nutrition—Healthy Generation Project. *Nutrients.* 2020;12(5):1323.
5. Shqair AQ. Screen time, dietary patterns and intake of potentially cariogenic food in children: A systematic review. *Journal of dentistry.*86:17-26.
6. Fanelli SM, Jonnalagadda SS, Pisegna JL, Kelly OJ, Krok-Schoen JL, Taylor CA. Poorer Diet Quality Observed Among US Adults With a Greater Number of Clinical Chronic Disease Risk Factors. *Journal of Primary Care & Community Health.* 2020;11:215013272094589.
7. Lissak G. Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental research.*164:149-157.
8. Foster HME, Ho FK, Sattar N, et al. Understanding How Much TV is Too Much. *Mayo Clinic Proceedings.* 2020;95(11):2429-2441.

9. LIVINGSTON G. Americans 60 and older are spending more time in front of their screens than a decade ago. 2019; https://pewrsr.ch/2ZlCbpn.
10. Hsueh M-C, Lin C-Y, Huang P-H, Park J-H, Liao Y. Cross-Sectional Associations of Environmental Perception with Leisure-Time Physical Activity and Screen Time among Older Adults. *Journal of Clinical Medicine.* 2018;7(3):56.
11. Zusman EZ, Dawes M, Fleig L, et al. Older Adults' Sedentary Behavior and Physical Activity After Hip Fracture: Results From an Outpatient Rehabilitation Randomized Controlled Trial. *Journal of Geriatric Physical Therapy.* 2019;42(2):E32-E38.
12. Rezende LFMD, Rey-López JP, Matsudo VKR, Luiz ODC. Sedentary behavior and health outcomes among older adults: a systematic review. *BMC Public Health.* 2014;14(1):333.
13. Livingstone KM. Associations between three diet quality indices, genetic risk and body composition: A prospective cohort study. *Clinical nutrition (Edinburgh, Scotland).*41(9):1942-1949.
14. Zhang J, Cao X, Li X, et al. Associations of Midlife Diet Quality with Incident Dementia and Brain Structure: Findings from the UK Biobank Study. In: Cold Spring Harbor Laboratory; 2022.
15. Schulz C-A, Weinhold L, Schmid M, Nöthen MM, Nöthlings U. Analysis of associations between dietary patterns, genetic disposition, and cognitive function in data from UK Biobank. *European Journal of Nutrition.* 2023;62(1):511-521.
16. Ding D. Physical activity, diet quality and all-cause cardiovascular disease and cancer mortality: a prospective study of 346 627 UK Biobank participants. *British journal of sports medicine.*56(20):1148-1156.
17. Springfield S, Cunanan K, Heaney C, Peng K, Gardner C. The WELL diet score correlates with the alternative healthy eating index-2010. *Food Science & Nutrition.* 2020;8(6):2710-2718.
18. Ross A, Willson VL. One-Way Anova. In: *Basic and Advanced Statistical Tests.* SensePublishers; 2017:21-24.
19. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in Medicine.* 2017;36(20):3257-3277.
20. Shaikhina T. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomedical signal processing and control.*52:456-462.
21. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology.* 2010;63(8):826-833.
22. Stefan Wager SA. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association.* 2018:1228-1242.
23. Imbens GW, Rubin DB. *Causal inference for statistics, social, and biomedical sciences: An introduction.* New York, NY, US: Cambridge University Press; 2015.
24. Luo J, Lee RYW. Opposing patterns in self-reported and measured physical activity levels in middle-aged adults. *European Journal of Ageing.* 2022;19(3):567-573.
25. Rutten-Jacobs LC, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ.* 2018:k4168.