

### RESEARCH ARTICLE

## A strategy for exploring subgroup-specific effects in nutrition science

Joshua Lambert<sup>1\*</sup> , Teenu Xavier<sup>1</sup>, Kimberly Berger<sup>1</sup> and Abigail Peairs<sup>2</sup>

<sup>1</sup>College of Nursing, Procter Hall, University of Cincinnati, Cincinnati, Ohio 45219, USA

<sup>2</sup>College of Allied Health Sciences, Department of Rehabilitation, Exercise & Nutrition Sciences, University of Cincinnati, Cincinnati, Ohio 45219, USA

(Received 2 August 2022 – Final revision received 21 October 2022 – Accepted 8 November 2022)

*Journal of Nutritional Science* (2022), vol. 11, e106, page 1 of 8

doi:10.1017/jns.2022.107

### Abstract

Nutrients are an essential part of building and maintaining optimal health. Certain nutrient exposure has been shown to be associated with many important health outcomes, although there is variability among studies. Despite the scientific efforts of many, it is unclear why some well-hypothesised nutrients lack sufficient evidence for clear association with health outcomes. One potential reason for conflicting results is that certain subgroups of patients benefit or are harmed more by adequate or inadequate exposure to certain nutrients. These subgroup-specific effects have historically not been studied, or if they are, it is often in a one-off type of approach where the investigator believes that a subgroup effect could exist based on limited previous data. In the era of big data, improvements can be made in efforts to generate new hypotheses for subgroups of patients and recommendations for precision nutrition can be made. In the present paper, we present a strategy for exploring subgroup-specific effects in nutrient-related studies. This data-driven method can be useful in secondarily exploring which subgroups are harmed/helped most by inadequate/adequate nutrient exposure and could suggest target groups for future clinical trials to test the identified hypotheses. We then present an example study utilizing the National Health and Nutrition Examination Survey (NHANES) data from the years 2001–2006. In this example, a limited selection of nutrients is protective in subgroups of participants with diabetes on their self-reported number of poor mental health days.

**Key words:** Explore: Nutrients: Subgroups: Vitamins

### Introduction

Many public health and primary prevention concerns involve complex multifaceted disease aetiologies which can go undetected or unexplained by some analytic approaches. Machine learning approaches seek to better model these complexities; however, they can lack tangible insight and explanation which is vital to pushing science forward. Nonetheless, these possible multifaceted effects linger and failing to identify them can lead to bias, misinterpretation, or incorrect public health advice and intervention<sup>(1)</sup>. By identifying subgroups of patients who exhibit complexity in the relationship between the variables that unite them in their uniqueness and the outcomes of interest, we can better understand the phenomenon and design targeted interventions.

Chronic diseases such as diabetes mellitus (DM) are significant health problems worldwide and leading causes of increased morbidity and mortality. Using diabetes as a single example, in the United States alone, approximately 34.2 million people (10.5 % of the population) have some form<sup>(2)</sup>, which imposes a significant and increasing burden on society. It has serious repercussions as a higher health care need/use, economic burden and associated societal costs have been reported among people with diabetes when compared with people without the disease<sup>(3)</sup>. While studies have consistently shown that having diabetes is associated with poorer physical and mental health outcomes<sup>(4–6)</sup>, little work has been done to understand what factors may modify their manifestation, or which subgroups are most affected by various influences on

\* Corresponding author: Joshua Lambert, email [lambcjbw@ucmail.uc.edu](mailto:lambcjbw@ucmail.uc.edu)



treatment or disease progression. One important factor known to be involved in the prevention of chronic diseases, like DM, is micronutrients<sup>(6)</sup>. Nutrition is an important element in the management of diabetes<sup>(7)</sup>, and studies have proposed that some foods and nutrients have protective or adverse effects on the management of DM<sup>(8,9)</sup>. Although medical nutrition therapy is a cornerstone of the management of diabetes, several areas of uncertainty in the dietary guidelines still exist and many areas of nutrition and diabetes need to be explored further<sup>(10)</sup>. The degree of uncertainty is especially high in the area of assessing micronutrient status and the role of micronutrients in the development of complications in diabetic patients<sup>(11–13)</sup>. Furthermore, there is limited evidence of a benefit from vitamin or mineral supplementation in people with diabetes who do not have underlying deficiencies<sup>(14)</sup>. This nuance and lack of specificity in who/what/when/where/why/how micronutrients play a role in the management and prevention of DM speak to the possible existence of complex multifaceted and subgroup-specific relationships. This is just one example of a complex chronic disease, and there are many others that may benefit from a better understanding of how micronutrient-defined subgroup-specific effects help/hurt the chance of various outcomes of interest.

In the present paper, we present an analytic strategy for exploring and identifying which micronutrient subgroups have differing effects on an outcome of interest. This strategy can be useful when the hypothesis of interest is that certain nutrients are more protective in subgroups of participants than others. Finally, an example of the proposed strategy is presented using 11,994 participants from the 2001–2006 National Health and Nutrition Examination Survey (NHANES) where we explore which nutrients are protective in subgroups of participants with diabetes on their self-reported number of poor mental health days.

## Methodology

### *A strategy for exploring subgroup-specific effects: motivation and restraint*

Variables that collectively define a subgroup-level effect can be accurately modelled, statistically, via the correct placement of an interaction term in a regression model. An interaction term models whether two (or more) explanatory variables have a significantly different effect when collectively considered (multiplicative) as compared with separately considered (additive). Including variables additively (main effects) in a regression model determines whether their effects, independently of one another, are associated with the outcome. Including variables as an interaction determines whether their effects synergistically (or antagonistically) depend on one another and cause varying levels of association on the outcome<sup>(15)</sup>.

This synergism (or antagonism) is suspected by most in theory but considered by few in practice. This lack of consideration is mostly due to the sheer volume of possible combinations to consider. For example, with ten explanatory variables and one outcome, checking all possible two-way

interactions would involve checking forty-five possible combinations (i.e.  $(10/2)$ ). Exploring all possible three-way interactions results in an even more daunting 120 possible combinations. These numbers grow exponentially larger as the number of explanatory variables increases. After all of this, the investigator should correctly consider multiple comparison correction via the false discovery rate (FDR) or the conservative Bonferroni correction. With these troubles in mind, most leave this exercise for the reader's imagination. Where does this leave the investigator who wishes to explore these possibilities, or should they even try? We argue that many consider (and are deterred by) the price of testing and making an error and yet few consider the price of not testing and failing to identify. In the right scenarios and correct understanding of the exploration under consideration, investigators can complete exploratory subgroup analysis and generate data-driven hypotheses to be scrutinised by the scientific community for validity and clinical relevance/significance and then formally tested in future well-designed powered studies.

*Overview of the proposed strategy.* The Feasible Solution Algorithm (FSA), outlined by Lambert *et al.*<sup>(16)</sup>, is a regression-based approach where investigators can explore interaction effects while adjusting for covariates and confounders. The algorithm is currently implemented in the R package rFSA<sup>(16)</sup>. In terms of exploratory interaction analyses, feasibility could be argued as identifying a set of plausible data-driven interactions that could elucidate real subgroups related to the outcome. The reality of the proposed set would not be the end goal of the analyses in hand, rather the goal would be to identify a set of interactions that have some data behind them and then to further consider the set in light of the literature and biological credibility. The FSA identifies such a set for a given outcome while adjusting for the specified covariates and confounders. FSA works for a wide variety of regression modelling strategies (e.g. linear regression, logistic regression) and can identify a set of solutions based on a criterion a user selects (e.g.  $R^2$ , AIC, Interaction  $P$ -value). The set of solutions is identified by running the algorithm many times, as a random start may land at a different optimal or semi-optimal endpoint than another random start. Guidance on how many times to run the algorithm, a detailed outline of the algorithm, and an example of how to use the R package (rFSA) can be found in the previously published works around FSA<sup>(16–18)</sup>. As a recommendation, we propose a three-step strategy to successfully implement FSA. An illustration of this strategy can be found in Fig. 1. First, the investigators would identify known covariates and confounders related to the outcome that they would like to adjust for. They would then fit a regression model with these variables and perform common statistical checks. Once the model is identified from step 1 (typically called a base model), the investigator would employ the rFSA code or utilise the web application located here: <https://shiny.as.uky.edu/mcfssa/>. From this analysis, a set of  $K$  interactions would be identified. Finally, in step 3, the set of interactions,

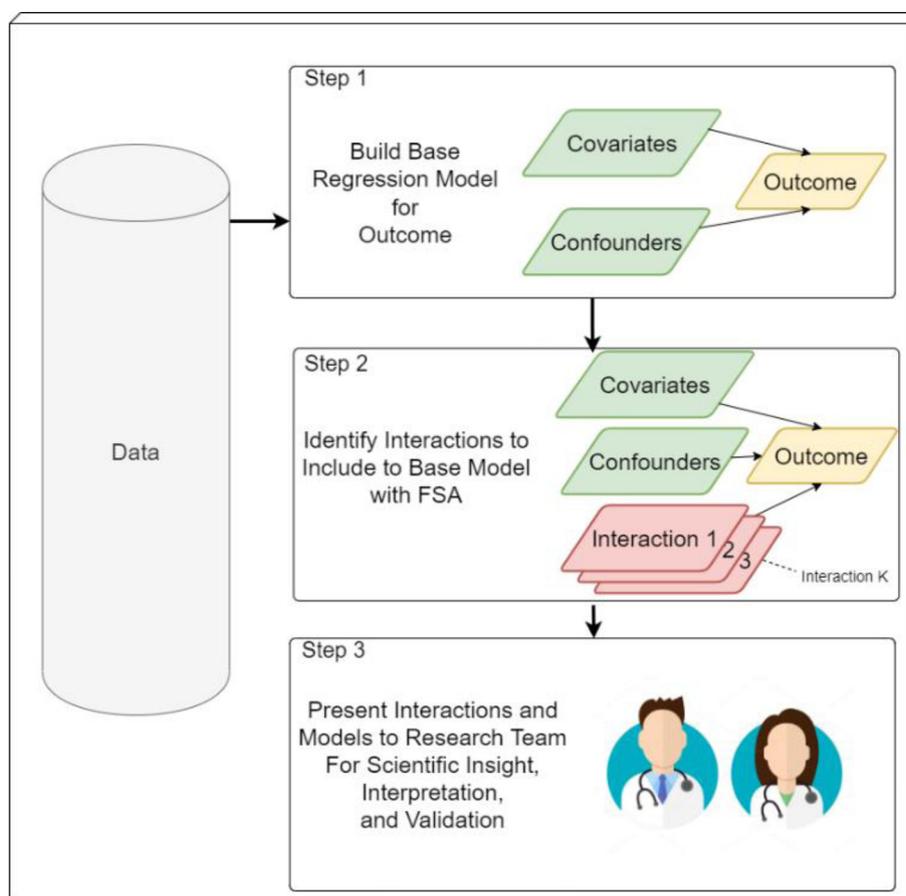


Fig. 1. Overview of the FSA approach.

parameter estimates and various visualisations of their association would be presented to the research team for scientific insight, interpretation and validation. This set of interactions is again, thought to be a suggestive set, and feedback from step 3 (research team insight) should be highly valued in how follow-up analyses and presentation of the findings are reported. In all cases, exploratory results should be restudied in separate independent datasets before conducting rigorous statistical inference.

### Example

Here, we present an example of the proposed strategy utilising the NHANES from 2001–2006. This example explores which NHANES measured nutrients are protective in subgroups of participants with diabetes on their self-reported number of poor mental health days. All R code used to download and analyse the necessary NHANES data can be found on the corresponding authors GitHub repository (<https://github.com/joshuaw Lambert/NutritionNHANES>). This may be useful for researchers looking to reproduce what has been completed in this manuscript or serve as a template for researchers who would like to explore interactions for their own data and purposes.

**Study participants and data summary.** The NHANES uses a stratified multistage probability sampling design and is meant

to reflect the civilian US population<sup>(19)</sup>. Each survey covers a 2-year time period and surveys approximately 10 000 NHANES participants. Our study used participant data for the years 2001–2006. The NHANES survey data were freely downloaded from the NHANES website<sup>(19)</sup>, using R version 3.5.1 and the *nhanesA* package<sup>(20)</sup>.

There was a total of 31 509 participants who completed NHANES from 2001–2006. Participants who were less than 21 years of age at the time of the survey, or had missing information for age, gender, race, education, family income to poverty ratio, body mass index (BMI), smoking status, the number of mental health days in last 30 d, or the number of physical health days in the last 30 d were excluded from this study. There was a total of 16 387 participants who were at least 21 years of age, and 11 994 who remained after all exclusion criteria were applied.

**Clinical, demographic and nutrient variables.** The question: ‘for how many days during the past 30 days was {your/his/her} mental health not good?’ was chosen as the primary outcome and considered a measure of mental health status.

Six variables within the NHANES database were used to define diabetes status. The variables included three verbal questions: ‘Doctor told you have diabetes’, ‘Taking insulin now’, ‘Take diabetic pills to lower blood sugar’ and three blood specimen measurements: fasting glucose (mg/dl), two-



hour glucose (mg/dl) and glycohaemoglobin (%). Any participant who answered yes to any of the verbal questions or had a fasting glucose  $\geq 126$  mg/dl, two-hour glucose  $\geq 200$  mg/dl or glycohaemoglobin  $\geq 6.5$  were classified as participants with diabetes ( $n$  1633; 13.6 %). This definition was inspired by the work of Kabadi *et al.*<sup>(7)</sup>. Participants who met this definition of having diabetes are referred to as participants with diabetes throughout the remainder of the results, discussion and conclusion.

The NHANES measured nutrient levels using high-performance liquid chromatography (HPLC) with photodiode array detection. HPLC is a method used for the analysis of serum concentrations of retinoids, carotenoids and tocopherols<sup>(21)</sup>. An overview of all nutritional variables used, time period availability, variable median cutoffs, as well as the total sample size for each variable can be found in Table 1. For the purposes of interpretability, each continuous nutritional variable was categorised by either being greater than or equal to its median or less than its median. Descriptive statistics stratified by Diabetic status and overall can be found in Table 2 with further sub-stratification by diabetes status and mental health days in Supplementary Table A1.

**Statistical Analyses.** R version 3.5.1 was used for all tables, plots and statistical analysis. Outcome variables (mental health days) were analysed as they were measured (continuous count data). Eight variables were considered for adjustment. Variables for adjustment included age, gender, race (white, black, Mexican American or other Hispanic, and other race), education (less than high school, high school, greater than high school), FIP level (less than poverty level, greater than poverty level), BMI, smoking status (former, current or non-smoker) and the number of physical health

days (0–30). Finally, diabetes status was also included as it was the main independent variable of focus for this example.

Due to the survey nature of NHANES, sample weights are included so that accurate estimation of the US adult population (twenty-one or older) can be completed. The outcome variables were both count data and are over dispersed. Both survey and over dispersed data can give too small standard errors when using traditional logistic regression, which can lead to incorrect findings<sup>(22)</sup>. To address both the survey and over dispersed nature of this data, we employed survey weighted Poisson regression (via a quasi-Poisson link)<sup>(22)</sup>. This method is fully available within the R package *survey*<sup>(23)</sup>. This approach was used to give more accurate standard errors which we hoped would lead to more accurate and interpretable findings for the US adult (twenty-one or older) population.

To build the base regression model, a two-part plan of action was employed. First, we fit the full model (all eight covariates and confounders entered the model as well as diabetes). Second, we would remove the insignificant variable(s) one at a time until a final base model was established that was interpretable and statistically valid (parameter estimate interpretability and model diagnostic checks). For the outcome, this base model would be used as the basis for which we explored subgroups of diabetes on mental health days. Using the R package, *rFSA*<sup>(16,24)</sup>, the FSA would be employed to identify nutrient and demographic two- and three-way interactions which included diabetes. All available nutrient and demographic variables were available to be explored. Once the FSA procedure was completed, the identified interactions would be examined by interpreting their parameter estimates as well as a visual aid (graph) meant to display the identified relationship. Also, interpretability considering the current literature was also completed.

## Results

### Base regression model results

From the survey-weighted logistic regression with the quasi-Poisson link for mental health, race was removed due to being insignificant ( $P$ -value  $> 0.05$ ) which left age, BMI, gender, education, FIP, smoking status and the number of physical health days as covariates for adjustment in the base model. While adjusting for these variables and including diabetes status, the model estimated that those with diabetes had a 9 % ( $P$ -value = 0.115) higher number of mental health days than those without diabetes. This model will be referred to as base model 1, and its estimates and 95 % confidence interval (CI) are contained on the left side of Supplementary Table A2.

### Interaction results

Using the mental health base model identified in 3.4, we utilised the FSA approach to explore subgroups of patients with diabetes on mental health. For this approach, all available nutrient and demographic variables were available to be explored. One result from this approach was identified as

**Table 1.** Nutrient variables medians used for cutoffs and number of missing

Variable description	Median	Missing	NHANES Years
Alpha-carotene (ug/dl)	2.80	533	01–06
<i>trans</i> -b-Carotene (ug/dl)	12.40	528	01–06
<i>cis</i> -b-Carotene (ug/dl)	0.80	865	01–06
b-Cryptoxanthin (ug/dl)	7.90	568	01–06
g-Tocopherol (ug/dl)	208.00	570	01–06
Combined lutein/zeaxanthin (ug/dl)	14.90	533	01–06
<i>trans</i> -Lycopene (ug/dl)	20.70	533	01–06
Retinyl palmitate (ug/dl)	1.60	814	01–06
Retinyl stearate (ug/dl)	0.38	942	01–06
Retinol (ug/dl)	57.60	525	01–06
Vitamin E (ug/dl)	1190.00	525	01–06
a-Cryptoxanthin (ug/dl)	2.38	8229	03–04
Total b-carotene (ug/dl)	12.52	8229	03–04
<i>cis</i> -Lycopene (ug/dl)	17.73	8229	03–04
<i>cis</i> -Lutein/zeaxanthin (ug/dl)	1.32	8229	03–04
d-Tocopherol (ug/dl)	5.00	8229	03–04
Total lycopene (ug/dl)	37.36	8229	03–04
Lutein (ug/dl)	9.54	8229	03–04
Phytofluene (ug/dl)	4.24	8229	03–04
Phytoene (ug/dl)	3.12	8229	03–04
Zeaxanthin (ug/dl)	3.68	8229	03–04
Total ( <i>cis</i> - and <i>trans</i> -)lycopene (ug/dl)	40.70	8196	05–06

**Table 2.** Descriptive statistics overall and stratified by diabetes status

	Overall	Participants without diabetes	Participants with diabetes	P-value
N	11 994	10 361	1633	
Age (mean (sd))	49.41 (18.50)	47.44 (18.35)	61.94 (14.06)	<0.001
Gender = male (%)	5812 (48.5)	4958 (47.9)	854 (52.3)	0.001
Race/Ethnicity (%)				<0.001
Black	2383 (19.9)	1982 (19.1)	401 (24.6)	
Mex Amer/Other Hispanics	2805 (23.4)	2332 (22.5)	473 (29.0)	
Other	427 (3.6)	371 (3.6)	56 (3.4)	
White	6379 (53.2)	5676 (54.8)	703 (43.0)	
Education (%)				<0.001
Greater than high school	5817 (48.5)	5225 (50.4)	592 (36.3)	
High school	2851 (23.8)	2462 (23.8)	389 (23.8)	
Less than high school	3326 (27.7)	2674 (25.8)	652 (39.9)	
FIP = Below 1 (%)	2040 (17.0)	1696 (16.4)	344 (21.1)	<0.001
BMI (mean (sd))	28.56 (6.44)	28.09 (6.20)	31.52 (7.10)	<0.001
Smoke (%)				<0.001
Current smokers	2656 (22.1)	2372 (22.9)	284 (17.4)	
Former smokers	3243 (27.0)	2653 (25.6)	590 (36.1)	
Non-smokers	6095 (50.8)	5336 (51.5)	759 (46.5)	
Nutrients (%)				
Alpha-carotene > 2.8 µg/dl	5621 (49.0)	4968 (50.2)	653 (41.6)	<0.001
trans-b-Carotene > 12.4 µg/dl	5732 (50.0)	5061 (51.1)	671 (42.8)	<0.001
cis-b-Carotene > 0.8 µg/dl	5414 (48.6)	4781 (49.8)	633 (41.6)	<0.001
b-Cryptoxanthin > 7.9 µg/dl	5700 (49.9)	5049 (51.2)	651 (41.7)	<0.001
g-Tocopherol > 208 µg/dl	5711 (50.0)	4777 (48.5)	934 (59.6)	<0.001
Lutein/zeaxanthin > 14.9 µg/dl	5712 (49.8)	4949 (50.0)	763 (48.7)	0.342
trans-Lycopene > 20.7 µg/dl	5716 (49.9)	5136 (51.9)	580 (37.0)	<0.001
Retinyl palmitate > 1.6 µg/dl	5519 (49.4)	4746 (49.2)	773 (50.6)	0.33
Retinyl stearate > 0.38 µg/dl	5497 (49.7)	4625 (48.5)	872 (57.7)	<0.001
Retinol > 57.6 µg/dl	5727 (49.9)	4842 (48.9)	885 (56.4)	<0.001
Vitamin E > 1190 µg/dl	5718 (49.9)	4840 (48.9)	878 (55.9)	<0.001
a-Cryptoxanthin > 2.38 µg/dl	1873 (49.7)	1665 (51.4)	208 (39.5)	<0.001
Total b-Carotene > 12.52 µg/dl	1882 (50.0)	1667 (51.5)	215 (40.9)	<0.001
cis-Lycopene > 17.73 µg/dl	1881 (50.0)	1673 (51.7)	208 (39.5)	<0.001
cis-Lutein/zeaxanthin > 1.32 µg/dl	1882 (50.0)	1649 (50.9)	233 (44.3)	0.006
d-Tocopherol > 5 µg/dl	1819 (48.3)	1516 (46.8)	303 (57.6)	<0.001
Total lycopene > 37.36 µg/dl	1881 (50.0)	1683 (52.0)	198 (37.6)	<0.001
Lutein > 9.54 µg/dl	1879 (49.9)	1634 (50.4)	245 (46.6)	0.11
Phytofluene > 4.24 µg/dl	1875 (49.8)	1654 (51.1)	221 (42.0)	<0.001
Phytoene > 3.12 µg/dl	1879 (49.9)	1650 (50.9)	229 (43.5)	0.002
Zeaxanthin > 3.68 µg/dl	1881 (50.0)	1639 (50.6)	242 (46.0)	0.056
Total lycopene > 40.7 µg/dl	1897 (49.9)	1689 (51.9)	208 (38.2)	<0.001
Mental health (mean (sd))	3.53 (7.55)	3.48 (7.41)	3.89 (8.41)	0.039
Physical health (mean (sd))	3.81 (7.99)	3.45 (7.53)	6.14 (10.13)	<0.001

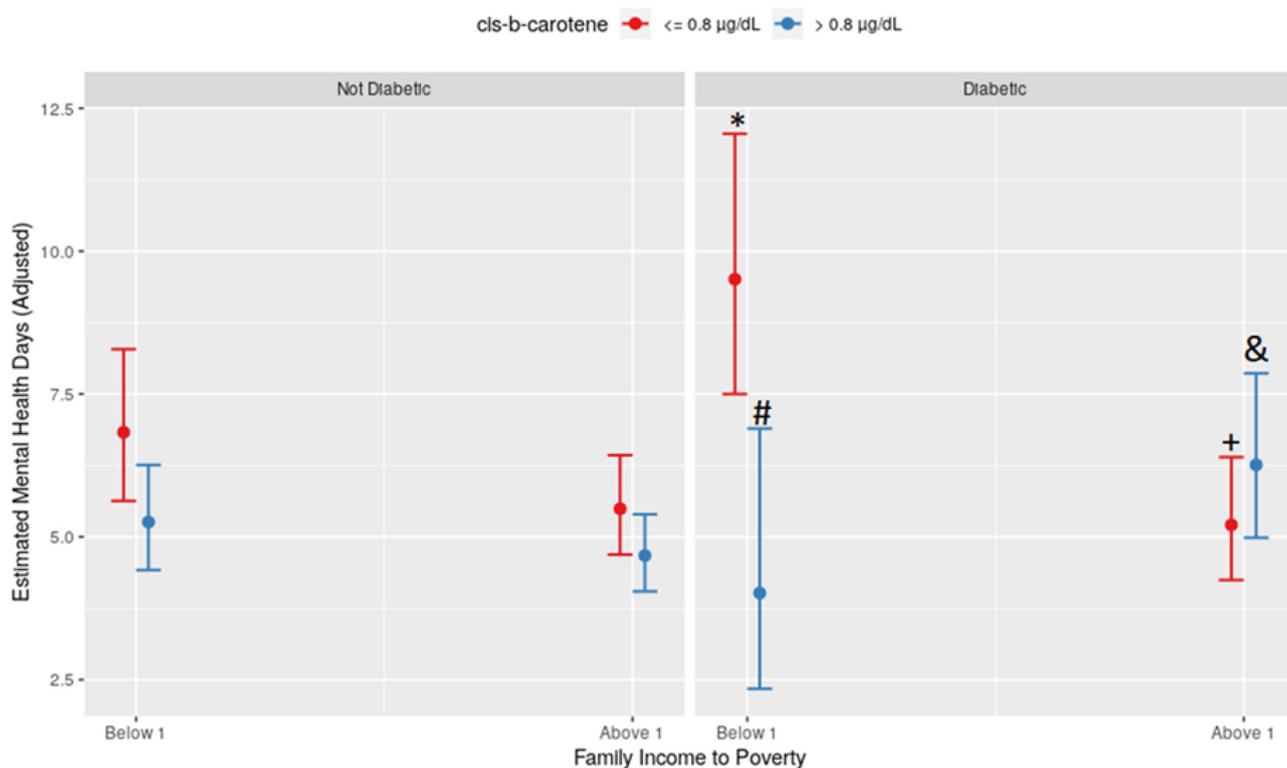
statistically valid and clinically interesting. The FSA procedure identified a three-way interaction between diabetes, *cis*-b-carotene and FIP to include in the base statistical model. The base model with interactions estimates and 95 % CI are contained on the right side of Supplementary Table A2. No significant two-way interactions with diabetes were found using the FSA approach. While adjusting for age, gender, education, FIP, BMI, smoking status and the number of physical health days in the last 30 d, the model estimated that participants with diabetes who had FIP < 1 (below poverty level) and *cis*-b-carotene  $\leq 0.8$  µg/dl are estimated to have 9.5 mental health days (95 % CI 7.5, 12) per month while those with diabetes, FIP < 1 (below poverty level), and *cis*-b-carotene > 0.8 µg were estimated to have just 4 mental health days (95 % CI 2.3, 6.9) per month. This estimated 136 % increase in mental health days can be calculated from the estimates in Supplementary Table A2 and is illustrated in Fig. 2. The figure plots the marginal effects of the interaction

term from the model with the interaction found by FSA. Within the participants with diabetes column of Fig. 2, participants with diabetes with *cis*-b-carotene  $\leq 0.8$  µg/dl and a FIP below 1 (notated by an ‘\*’ in the figure, N = 198) have a visually higher number of estimated mental health days than participants with diabetes with *cis*-b-carotene > 0.8 µg/dl and a FIP below 1 (notated by an ‘#’ in the figure, N = 133). Similar differences can be seen for other combinations of FIP and *cis*-b-carotene in the participants with diabetes group (see ‘&’ and ‘+’ notations) as well as combinations within the participants without diabetes group.

## Discussion

### Interactions results for the number of mental health days

The interaction findings in the section ‘Interaction results’ suggest that adult participants with diabetes who have a FIP < 1



**Fig. 2.** Results for modification of diabetes on mental health by *cis*-b-carotene and FIP.

may see a greater number of poor mental health days per month if they have also *cis*-b-carotene levels at or below 0.8  $\mu\text{g}/\text{dl}$  as compared with participants with diabetes who have FIP < 1 but instead have *cis*-b-carotene levels above 0.8  $\mu\text{g}/\text{dl}$ . Participants with diabetes who can increase their *cis*-b-carotene levels may see a decrease in their number of poor mental health days per month. The result with FIP is not surprising and was found prior to identifying the interaction in the base model. What is interesting is that the effect of FIP < 1 in participants with diabetes on the number of mental health days may be made better by increasing amounts of *cis*-b-carotene.  $\beta$ -Carotene is a type of carotenoid that acts as a precursor of vitamin A and functions as a ligand to regulate the expression of genes involved in metabolic processes and is mostly found in fruits and vegetables<sup>(25)</sup>. Recent studies have shown controversial findings regarding the protective effect of  $\beta$  carotene Type 2 diabetes. The findings in the present study are supported by studies from Polidori *et al.*<sup>(26)</sup> and Abahusain *et al.*<sup>(27)</sup> which both reported that  $\beta$  carotene was significantly lower in individuals with type 2 diabetes as compared with non-diabetic control participants. Evidence from prospective cohort studies also supports an inverse association between the incidence of type 2 diabetes and dietary or plasma concentrations of b-carotene<sup>(28)</sup>. However, studies by Song *et al.*<sup>(29)</sup> and Kataja-Tuomola *et al.*<sup>(30,31)</sup> complicate the understanding of the relationship as they reported no statistical evidence that supplementation with  $\beta$ -carotene had an effect on the incidence of type 2 diabetes and/or fasting plasma glucose.

The plausibility of this relationship stems from a few possibilities, first the antioxidant effect that beta-carotene is known to have. They can effectively eliminate reactive oxygen

species and other free radicals and protect from oxidative stress and subsequent damage<sup>(32)</sup>. Studies have found that the development of depression is related to imbalance between antioxidant and prooxidants levels and is often accompanied by oxidative stress which, in turn, causes DNA damage<sup>(33)</sup>. There is also fundamental evidence that the antioxidant property of beta-carotene has a protective role from the oxidative stress and subsequent damage often caused in depression. In addition, research has suggested that the antidepressant properties of beta-carotene may be related to its ability to increase brain-derived neurotrophic factor (BDNF) and decrease levels of tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) and interleukin-6 (IL-6)<sup>(34)</sup>. Several studies have investigated the effects of beta-carotene on mental health. Ge *et al.* examined the relationship between dietary intakes of carotenoids and the risk of depressive symptoms using data from NHANES during 2009–16 and they found a similar association. They reported that beta-carotene intake may be inversely associated with the risk of depressive symptoms in the US adults<sup>(35)</sup>. The findings of the present study are also supported by the findings of the meta-analysis by Zhang *et al.* who reported that beta-carotene intake is inversely associated with depression<sup>(36)</sup>.

It is also possible that higher serum beta-carotene is serving as a surrogate for higher fruit/veggie intake, indicating a higher diet quality. However, there are many other carotenoids and antioxidant nutrients in fruits and vegetables included in the current analysis (i.e. lutein, lycopene, b-cryptoxanthin and tocopherols) that were not identified as having the same relationship with mental health days as beta-carotene, leading us to believe that there is something special about this nutrient.



### Limitations

There are various limitations to our example study. No longitudinal data was collected on these participants related to our phenomenon of interest and micronutrient measurements were not collected at multiple time points (which may be necessary to assure they are accurately measured). Finally, unobserved heterogeneity between those classified as participants with diabetes and those without may not have been adequately controlled for with our covariates and confounders.

### Implications

This proposed methodology can help us to realise the goal of using precision nutrition to reduce morbidity and mortality associated with many chronic diseases. Major health organisations (USDA, NIH) have recognised the important role that individualised nutrition can play in accomplishing this goal<sup>(37)</sup>. While it has been recognised that individuals with certain phenotypes respond better to certain medications, it is also becoming apparent that some individuals are more sensitive to certain nutrients, and thus, they may experience a greater benefit from a specialised nutrition intervention.

However, the challenge has been in the identification of subgroups who are most likely to benefit from specific nutrients, beyond those individuals who are deficient in those nutrients. Nutrition studies using traditional data analysis methods have been largely unsuccessful in identifying subgroups, as the focus has been on the average response rather than individual responses. There is often wide variability among individuals in these nutrition studies and all facets of biological differences (genetics, gut microbiome, gender, disease status, etc.) affect an individual's response to a nutritional intervention<sup>(37)</sup>. In fact, it has been suggested that participants recruited for nutrition studies may not be the ones most likely to experience a benefit. For example, many studies have tested the ergogenic potential of antioxidant supplements, however, the wide range of oxidative status of participants likely plays a role in the overall lack of response to antioxidants that has been reported<sup>(38)</sup>. There also appear to be individual responses to consumption of macronutrients, due to characteristics not commonly assessed in nutrition studies (such as genetics and gut microbiota), making it even more apparent that nutrition is not a one size fits all strategy<sup>(39)</sup>.

There is clearly a need for statistical methods which can identify subgroups of people with enhanced sensitivity to nutritional status, in order to be able to use nutrition as medicine to prevent and manage chronic diseases.

### Conclusion

In the present paper, we present a strategy for exploring subgroup-specific effects in studies focused on nutrition science. We present an example of utilising the method and exploring the effects of micronutrients and demographic factors in participants with diabetes on the number of mental health days they were experiencing.

While there are limitations to exploratory analysis, there are many potential impacts from completing an exploratory

analysis meant to uncover subgroup-specific effects. One strength of our proposed method is that the investigator can control known potential covariates and confounders and use a regression technique appropriate for outcome characteristics and distribution (e.g. survey-weighted Poisson). The FSA procedure can harness this regression adjustment while illuminating interesting and workable subgroups to follow up with and investigate further. Finally, if the effects of certain vitamins or micronutrients are only present in subgroups of participants with the outcome of interest, then taking the time to explore interactions in statistical models or completing a specified subgroup analysis is required to identify their effects. The proposed methodology and FSA procedure enables the investigator to complete this initial exploration.

### Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/jns.2022.107>.

### Acknowledgements

We would like to thank the reviewer for the valuable feedback and comments they provided.

J. L., K. B., T. X. and A. P. drafted the manuscript. J. L., K. B. and T. X. researched the literature. J. L. contributed to the statistical analysis. J. L. collected the data from NHANES. J. L., K. B., T. X. and A. P. contributed to editing the manuscript. J. L. conceived and supervised the study.

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

The authors declare that they have no conflict of interest.

### References

1. Vatcheva KP, Lee M, McCormick JB, *et al.* (2015) The effect of ignoring statistical interactions in regression analyses conducted in epidemiologic studies: an example with survival analysis using Cox proportional hazards regression model. *Epidemiology (Sunnyvale)* **6**(1), 216.
2. Centers for Disease Control and Prevention CDC (2020) *National Diabetes Statistics Report*. Centers for Disease Control and Prevention, US Department of Health and Human Services. [cited 2 July 2020]. Available from: <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf>
3. Yang W, Zhao W, Xiao J, *et al.* (2012) Medical care and payment for diabetes in China: enormous threat and great opportunity. *PLoS ONE* **7**, e39513.
4. Campbell HM, Khan N, Cone C, *et al.* (2011) Relationship between diet, exercise habits, and health status among patients with diabetes. *Res Soc Adm Pharm* **7**, 151–161.
5. Li C, Ford ES, Mokdad AH, *et al.* (2008) Clustering of cardiovascular disease risk factors and health-related quality of life among US adults. *Value Health* **11**, 689–699.
6. Bonnefont-Rousselot D (2004) The role of antioxidant micronutrients in the prevention of diabetic complications. *Treat Endocrinol* **3**, 41–52.
7. Kabadi SM, Lee BK & Liu L (2012) Joint effects of obesity and vitamin D insufficiency on insulin resistance and type 2 diabetes: results from the NHANES 2001–2006. *Diabetes Care* **35**, 2048–2054.



8. Bazzano LA, Li TY, Joshipura KJ, *et al.* (2008) Intake of fruit, vegetables, and fruit juices and risk of diabetes in women. *Diabetes Care* **31**, 1311–1317.
9. Pan A, Sun Q, Bernstein AM, *et al.* (2011) Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am J Clin Nutr* **94**, 1088–1096.
10. Franz MJ, Bantle JP, Beebe CA, *et al.* (2003) Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes Care* **26**, S51–S61.
11. Mooradian AD, Failla M, Hoogwerf B, *et al.* (1994) Selected vitamins and minerals in diabetes. *Diabetes Care* **17**, 464–479.
12. Mooradian A, McLaughlin S, Boyer C, *et al.* (1999) Diabetes care for older adults. *Diab Spec* **12**, 70–77.
13. Dubey P, Thakur V & Chattopadhyay M (2020) Role of minerals and trace elements in diabetes and insulin resistance. *Nutrients* **12** (6), 1864.
14. Chehade JM, Sheikh-Ali M & Mooradian AD (2009) The role of micronutrients in managing diabetes. *Diabetes Spectrum* **22**, 214–218.
15. Jaccard J & Turrisi R (2003) *Interaction Effects in Multiple Regression*. USA: SAGE Publications, Inc.
16. Lambert J, Gong L, Elliott C F, *et al.* (2019) rFSA: an R package for finding best subsets and interactions. *R J* **10**, 295.
17. Elliott CF, Lambert JW, Stromberg AJ, *et al.* (2021) Feasibility as a mechanism for model identification and validation. *J Appl Stat* **48**, 2022–2041.
18. Janse SA & Thompson KL (2019) Properties of the number of iterations of a feasible solutions algorithm. In *Modern Statistical Methods for Spatial and Multivariate Data*, pp. 95–105 [Diawara N, editor]. Cham: Springer International Publishing.
19. Centers for Disease Control and Prevention CDC (2020) *National Health and Nutrition Examination Survey Data*. [cited 1 July 2020]. Available from <https://www.cdc.gov/visionhealth/vehss/data/national-surveys/national-health-and-nutrition-examination-survey.html>
20. Endres C (2018) *nhanesA: NHANES Data Retrieval*, 0.7.1 CRAN.
21. Laboratory Procedure Manual [Internet]. [cited 19 Mar 2021]. Available from [https://www.cdc.gov/nchs/data/nhanes/nhanes\\_03\\_04/l45vit\\_c\\_met\\_vitAE\\_carotenoids.pdf](https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/l45vit_c_met_vitAE_carotenoids.pdf)
22. Ver Hoef JM & Boveng PL (2007) Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* **88**, 2766–2772.
23. Lumley T (2020) *survey: Analysis of Complex Survey Samples*, 4.0 CRAN.
24. Lambert J, Gong L & Elliott C (2020) *rFSA: Feasible Solution Algorithm for Finding Best Subsets and Interactions*, 0.9.6 CRAN.
25. Ross AC, Zolfaghari R & Weisz J (2001) Vitamin A: recent advances in the biotransformation, transport, and metabolism of retinoids. *Curr Opin Gastroenterol* **17**, 184–192.
26. Polidori MC, Mecocci P, Stahl W, *et al.* (2000) Plasma levels of lipophilic antioxidants in very old patients with type 2 diabetes. *Diabetes Metab Res Rev* **16**, 15–19.
27. Abahusain MA, Wright J, Dickerson JW, *et al.* (1999) Retinol, alpha-tocopherol and carotenoids in diabetes. *Eur J Clin Nutr* **53**, 630–635.
28. Montonen J, Knekt P, Jarvinen R, *et al.* (2004) Dietary antioxidant intake and risk of type 2 diabetes: response to granado-lorenco and olmedilla-alonso. *Diabetes Care* **27**, 1845–1846.
29. Song Y, Cook NR, Albert CM, *et al.* (2009) Effects of vitamins C and E and beta-carotene on the risk of type 2 diabetes in women at high risk of cardiovascular disease: a randomized controlled trial. *Am J Clin Nutr* **90**, 429–437.
30. Kataja-Tuomola MK, Kontto JP, Männistö S, *et al.* (2010) Effect of alpha-tocopherol and beta-carotene supplementation on macrovascular complications and total mortality from diabetes: results of the ATBC study. *Ann Med* **42**, 178–186.
31. Kataja-Tuomola MK, Kontto JP, Männistö S, *et al.* (2011) Intake of antioxidants and risk of type 2 diabetes in a cohort of male smokers. *Eur J Clin Nutr* **65**, 590–597.
32. Fiedor J & Burda K (2014) Potential role of carotenoids as antioxidants in human health and disease. *Nutrients* **6**, 466–488.
33. Sarandol A, Sarandol E, Eker SS, *et al.* (2007) Major depressive disorder is accompanied with oxidative stress: short-term antidepressant treatment does not alter oxidative-antioxidative systems. *Hum Psychopharmacol* **22**, 67–73.
34. Kim N-R, Kim H-Y, Kim M-H, *et al.* (2016) Improvement of depressive behavior by Sweetme Sweet Pumpkin™ and its active compound, β-carotene. *Life Sci* **147**, 39–45.
35. Ge H, Yang T, Sun J, *et al.* (2020) Associations between dietary carotenoid intakes and the risk of depressive symptoms. *Food Nutr Res* **64**, 3920.
36. Zhang Y, Ding J & Liang J (2022) Associations of dietary vitamin A and beta-carotene intake with depression: a meta-analysis of observational studies. *Front Nutr* **9**, 881139.
37. Stover PJ & King JC (2020) More nutrition precision, better decisions for the health of our nation. *J Nutr* **150**, 3058–3060.
38. Margaritelis NV, Paschalis V, Theodorou AA, *et al.* (2018) Antioxidants in personalized nutrition and exercise. *Adv Nutr* **9**, 813–823.
39. San-Cristobal R, Navas-Carretero S, Martínez-González MÁ, *et al.* (2020) Contribution of macronutrients to obesity: implications for precision nutrition. *Nat Rev Endocrinol* **16**, 305–320.