The American Journal of Clinical Nutrition 120 (2024) 380–388

American Society for Nutrition
Excellence in Nutrition Research and Practice

The American Journal of
CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Development of a national database for dietary glycemic index and load for nutritional epidemiologic studies in the United States

Karen A Della Corte [1,*], Dennis Della Corte [2], Sean Titensor [1,2], Bo Yang [3], Simin Liu [3,**]

[1] *Department of Nutrition, Dietetics, and Food Science, Brigham Young University, Provo, UT, United States;* [2] *Department of Physics and Astronomy, Brigham Young University, Provo, UT, United States;* [3] *Departments of Epidemiology, Medicine, and Surgery, Brown University, Providence, RI, United States*

ABSTRACT

Background: The quality of carbohydrate intake, as measured by the glycemic index (GI), has not been evaluated nationally over the past 2 decades in the United States.

Objectives: We aimed to develop a comprehensive and nationally representative dietary GI and glycemic load (GL) database from 1999 to 2018 National Health and Nutrition Examination Survey (NHANES) and to examine GI and GL time trends and subpopulation differences.

Methods: We used an artificial intelligence (AI)-enabled model to match GI values from 2 GI databases to food codes from United States Department of Agriculture, which were manually validated. We examined nationally representative distributions of dietary GI and GL from 1999 to 2018 using the multistage, clustered sampling design of NHANES.

Results: Assigned GI values covered 99.9% of total carbohydrate intake. The initial AI accuracy was 75.0%, with 31.3% retained after manual curation guided by substantive domain expertise. A total of 7976 unique food codes were matched to GI values, of which soft drinks and white bread were top contributors to dietary GI and GL. Of the 49,205 NHANES adult participants, the mean dietary GI was 55.7 (95% confidence interval [CI]: 55.5, 55.8) and energy-adjusted dietary GL was 133.0 (95% CI: 132.3, 133.8). From 1999 to 2018, dietary GI and GL decreased by 4.6% and 13.8%, respectively. Dietary GL was higher among females (134.6; 95% CI: 133.8, 135.5) than among males (131.3; 95% CI: 130.3, 132.3), those with ≤high school degree (137.7; 95% CI: 136.8, 138.7) than among those with ≥college degree (126.5; 95% CI: 125.3, 127.7), and those living under the poverty level (140.9; 95% CI: 139.6, 142.1) than among those above the poverty level. Differences in race were observed (Black adults, 139.4; 95% CI: 138.2, 140.7; White adults, 131.6; 95% CI: 130.5, 132.6).

Conclusions: The national GI and GL database facilitates large-scale and high-quality surveillance or cohort studies of diet and health outcomes in the United States.

Keywords: glycemic index assignment, glycemic load, national trends, carbohydrate quality, National Health and Nutrition Examination Survey, artificial intelligence

Introduction

The glycemic index (GI) is a classification of the blood glucose–raising potential of carbohydrate-containing foods [1]. The GI of food is unitless as it is 100 multiplied by a ratio of the blood glycemic response to 50 g of available carbohydrate in a test food divided by the blood glycemic response to the same amount of available carbohydrate as glucose in water. To quantify and standardize dietary glycemic response from carbohydrate-containing foods, dietary glycemic load (GL) was introduced to characterize the quality and quantity of carbohydrates-containing foods consumed and their interactions [2]. Decades of metabolic experimental work have quantified metabolic effects for a large number of carbohydrate-containing foods [3] and their GI values are widely available in existing datasets [4–6]. These GI databases have facilitated further epidemiologic studies of disease and health outcomes concerning dietary GI and GL in selected populations [3,7–10]. Nevertheless, food GI and GL values have not been comprehensively and systematically incorporated into national food

Abbreviations: AI, artificial intelligence; CI, confidence interval; FNDDS, Food and Nutrient Database for Dietary Studies; GI, glycemic index; GL, glycemic load; LLM, large language model.

* Corresponding author.
** Corresponding author.
E-mail addresses: karen_dellacorte@byu.edu (K.A. Della Corte), simin_liu@brown.edu (S. Liu).

https://doi.org/10.1016/j.ajcnut.2024.06.001
Received 29 December 2023; Received in revised form 29 May 2024; Accepted 4 June 2024; Available online 7 June 2024

composition tables or centrally standardized to the national nutrient databases.

Adopting a rigorous sampling and design strategy, the National Health and Nutrition Examination Survey (NHANES) represents a sentinel cornerstone for the health and nutritional status of noninstitutionalized adults in the United States [11]. The wealth of information gathered through multiple NHANES surveys offers a unique opportunity to examine diet-health or disease relations among diverse and nationally representative adult populations over time. In 2001, Ford and Liu first attempted to derive dietary GI and GL variables for 1 NHANES survey based on an abbreviated food frequency questionnaire (FFQ) with limited food items and investigated their relations with plasma lipids measured [12,13]. However, the quality of carbohydrate intake, as measured by GI, has not been evaluated nationally over the past 2 decades when nearly 10,000 different food codes were used in NHANES surveys (in which 2,260,066 individual food items were reported in ten 2-y NHANES cycles between 1999 and 2018). Thus, a complex resolution of 10 NHANES cycles of food codes is urgently needed to accurately assign and update GI values from external databases (e.g., the International Tables of GI, the University of Sydney GI website, and the GI database compiled by the Diogenes study) [14–16]. To manually match food descriptions across diverse food codes and multiple databases is a time-consuming task. However, in this study, we used a novel artificial intelligence (AI)-driven methodology in developing a comprehensive and nationally representative database of dietary GI and GL for participants across 10 NHANES cycles. Specifically, we identified and matched food codes with GI values creating both dietary GI and GL variables and indentifying the top GL-contributing foods in the United States. We examined the 20-y changes of GI and GL by covariates (sex, BMI, physical activity, age group, race and ethnicity, educational level, and family income to poverty ratio) and determined changes in GI/GL by cycle of NHANES from 1999 to 2018.

Methods

Extraction of USDA food codes from NHANES

Conducted by the National Center for Health Statistics of the United States Centers for Disease Control and Prevention (CDC), NHANES uses a 4-stage, stratified, sampling design to ensure national representativeness of the civilian noninstitutionalized population. A full description of the design and data collection methods have been published elsewhere [17,18]. During the informed consent process, survey participants are assured that the data collected will be used only for stated purposes and will not be disclosed without consent. In brief, NHANES uses a validated protocol of 24-h dietary recalls as the primary method of dietary exposure assessment. This study uses 10 NHANES cycles spanning 1999–2018, where one 24-h recall was used in NHANES 1999–2002, and two 24-h dietary recalls were used from 2003 to 2018, which have been validated for estimating total energy intake, macronutrients, and micronutrients [18,19]. However, in a validation study, it was found that for normal-weight participants, underreporting of energy intake was minimal whereas for participants with obesity, underreporting was higher [17]. The recalls include information specific to each food and beverage consumed as eaten (included as directed on food packaging), the recall day, and the overall diet (i.e., type, sources, brand name, amount, combinations, meal timing and occasions) [20]. The race/ethnicity variable was derived from survey responses of participants who would self-identify as Mexican American, Other Hispanic, non-Hispanic White, non-Hispanic Black, or non-Hispanic multiracial.

The dietary intakes contained a 9628 unique USDA food codes, of which 8055 food items contained >5 g of available carbohydrates were considered for the assignment of GI values [21]. Of these, food descriptions could be retrieved for 7976 food codes (Supplemental Figure 1). Of the NHANES 1999–2018 participants, 49,205 adults (23,702 males and 25,503 females) had complete and valid dietary information (total energy intake 600–5000 kcal/d for males and 600–4500 kcal/d for females), and were included in this study.

Data source of food GI values

Food descriptions and GI values were extracted using the International Tables of Glycemic Index and Glycemic Load values (International Table) as the primary data source [15] for 4018 items with 3595 unique descriptions. New releases of GI values are updated on the website maintained by the same research group at the University of Sydney [16]. GI assignments were primarily from the International Table 2021, which is the latest edition of a series of globally recognized data sets that are used as the major source of food GI and GL values in research and other applications. To increase the number of potential matches, we used both of the International GI Tables, which differ in how rigorously studies applied the international standards of operation for GI testing. In addition, GI values assigned by the Diogenes (Diet, Obesity and Genes) study, a pan-European multicentre intervention study, were retrieved from an open-source project along with their food descriptions and English translations [22]. The systematic approach of GI assignment within the Diogenes study was largely based on published values and those measured within various participating study centers [6]. This database consisted of 18,808 entries, encompassing 7405 unique English translations. In the study, researchers tested many unique foods that complemented the International GI Tables, allowing for matches to be found for foods described in the USDA food codes. These 2 sets of GI values were aggregated, yielding a combined GI database of 10,978 unique food code descriptions (after the removal of 22 duplicates) to be assigned to 7976 USDA food codes.

For each NHANES cycle, the USDA releases Food and Nutrient Databases for Dietary Studies (FNDDS) with evolving versions [23]. We used FNDDS 1.0, FNDDS 2.0, FNDDS 3.0, FNDDS 4.1, FNDDS 5.0, FNDDS 2011–2012, FNDDS 2013–2014, FNDDS 2015–2016, and FNDDS 2017–2018 to resolve unique food codes used between 1999 and 2018 [24]. The FNDDS database has a classification scheme wherein the first 4 digits of the codes represent a category of food, and the subsequent digits specify the food further. Of 8055 relevant food codes, 7976 unique and mutually exclusive food codes could be resolved with the FNDDS databases and were assigned food descriptions. The unresolved or missing food codes were not indexed in the USDA FoodData Central website and were not included ($n = 79$).

Recent progress in natural language processing and especially vector embeddings for large language models (LLMs) prompted us to use the leading embedding model from OpenAI for the AI-supported automation of GI assignment. Specifically, the "text-embedding-ada-002" model [25] from OpenAI's API was used to calculate text embeddings for all food descriptions. Represented as 1536-dimensional vectors, these embeddings encode the description string from the first character to the second occurrence of a "," character. Cosine similarity [26] was used as a measure to gauge the similarity of vector embeddings, with larger values representing greater similarity (Figure 1). USDA food description embeddings could thus be compared with GI database food description embeddings, creating an association between the 2 entries and thereby automatically assigning GI values to USDA food codes. For manual verification, a table was generated that

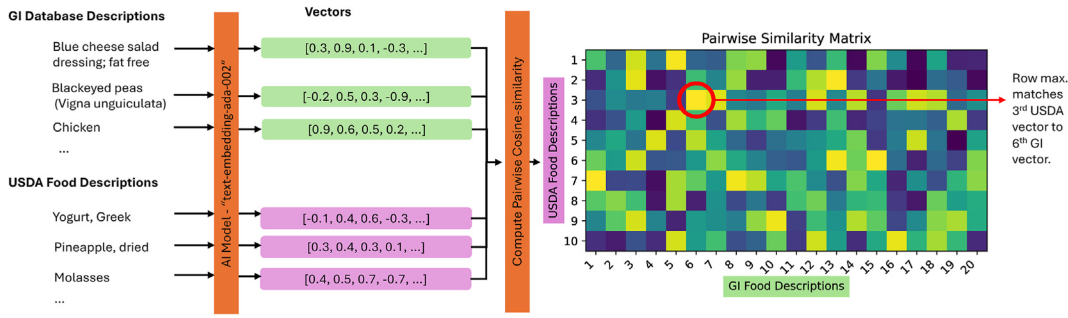

FIGURE 1. Illustration of the artificial intelligence (AI)-enabled alignment process of USDA food descriptions with GI database descriptions. In the first step, description strings are encoded through an AI model into vectors. Pairwise similarity scores are then calculated between all GI database vectors and USDA vectors. Decoding is used to select the maximum in each row of the resulting similarity matrix, which provides the mapping between GI database descriptions and USDA food descriptions. GI, glycemic index.

included the USDA food name, food frequency, food code, AI-matched GI database food name, GI value, and cosine similarity [20]. A more in-depth demonstration of this process is described in Supplemental Table 1.

Manual verification of automatically assigned GI values for USDA food codes

The automatic GI assignment helped to achieve the challenging task of making initial GI assignments to 7976 USDA food codes. Additional manual verification of 2 independent reviewers included updates of averages and meal composite GI scores. For the remaining incorrectly AI-assigned descriptions, a systematic stepwise procedure was used for assigning GI values (Supplemental Figure 2).

Some carbohydrate foods had multiple entries and values in the GI tables but were AI-assigned to only 1 possible match. In these cases, a mean GI value was calculated from all entries of the same food type (step 2 of Supplemental Figure 2B). For example, instant mashed potatoes had multiple values listed, which we averaged to 84 (for a summary of food items with averaged values assigned, see Supplemental Table 2). If food descriptions were closely related but not identical, the items with similar carbohydrate quality (i.e., types of carbohydrates, fiber, preparation method, and degrees of processing) were selected, as guided by substantial domain expertise. For mixed foods that lacked tested mixed-food matches or only had estimated mixed-food values (as estimated by Diogenes researchers), GI scores were calculated by summing the weighted GI value of the component foods using the established formula [27,28]. This calculation was done for 32% of the food codes. The only mixed foods that we did not calculate mixed GI values for were those that were formally tested and reported in the International Tables. For certain commercial food descriptions not found within the GI databases, an assignment to a similar general food group was made (e.g., various cocktails and liquors to the general group of "spirits" as found in the Diogenes database). For various unlisted brands of energy drinks, a mean of sugary drinks with similar contents was calculated. For foods with very low (<5 g available carbohydrate) or no carbohydrates, a GI value of 0 was assigned following the International GI Tables [15,29]. These included plain meat, poultry, and fish items (unbreaded and not in sauces), fats/oils, egg, unprocessed cheese, tea/coffee (unsweetened), high-fat/or high-acidic dressings, sugar-free/artificially sweetened beverages, seeds (e.g., flax and sesame), vinegar, lime juice, and certain sauces (soy, peanut, and fish). Most of these foods were also removed in the first step (i.e., available carbohydrate content ≤5 g) (Supplemental Figure 2). After the manual verification of GI average updates, more accurate matching, and meal composite GI scores, 31.3% of AI-assigned values were retained.

Calculating the dietary GI and GL

The calculation of dietary GI and GL has been described previously [30]. In brief, the dietary GL was computed by multiplying the available carbohydrate content of each food item by its respective GI (the multiplication means that a higher GI will have a greater effect at higher carbohydrate intakes). We then summed these products across all foods consumed by a participant on a given day to produce the dietary GL. This measure essentially matches available carbohydrate contents gram by gram and, thus, reflects the overall quality of carbohydrate intake in a whole diet. The majority of participants had 2 d of dietary data for which dietary GI values were averaged. We subtracted dietary fiber as unavailable carbohydrate from total carbohydrate to provide estimates of available carbohydrates. Dietary GI, a variable representing the overall quality of available carbohydrate intake for each participant, was created by dividing the dietary GL by the daily available carbohydrate intake [31]. The dietary GI and GL were calculated using formulas where GI_x is the glycemic index and g_x is the amount of available carbohydrates in grams of food x as follows (see Liu et al. [30] for details):

$$\text{Dietary GL} = \frac{\sum_{x=1}^{n} GI_x * g_x}{100}$$

$$\text{Dietary GI} = \frac{\sum_{x=1}^{n} GI_x * g_x}{\sum_{x=1}^{n} g_x}$$

Calculate 20-y weights for NHANES

We applied the National Center for Health Statistics recommended weighting scheme to account for the varying sampling designs and survey cycles [32,33]. For each participant in the merged data set, the 20-y weight was calculated for NHANES cycles 1999–2003 according to:

$$Dietary\ 20\ y\ weight = \frac{2}{10} * Dietary\ 4\ y\ sample\ weight$$

For cycles 2004–2018:

$$Dietary\ 20\ y\ weight = \frac{1}{10} * Dietary\ 2\ y\ sample\ weight$$

In this study, the dietary 2- and 4-y sample weights are variables provided by NHANES. The denominator scales each weight with respect to the the number of cycles, that is, a 4-y weight contributes 4 of the 20 and each 2-y weight 2 of the 20 to the dietary 20-y weights.

Energy adjustment: residual method

To account for the variability in energy intake among participants, all dietary values were adjusted for energy intake using the residual method [34]. We considered as a population mean the regression model value for a 2097.2 kcal energy intake, which was the weighted mean of all included participants aged 19 y and older with complete dietary information.

Calculation of subpopulation means and confidence intervals

All population means and 95% confidence intervals (CIs) were calculated using the R survey package. Taylor series linearization was used to calculate the weighted means and SEs using the dietary sample weights, masked variance units, and strata [35]. Masked variance units are constructed by aggregating secondary sampling units into pseudo-PSUs and pseudostrata to protect participant confidentiality while allowing for accurate variance estimation from the complex survey data without using the true design variables. CIs were obtained by either adding or subtracting 1.96 multiplied by the SEM from the population means. For 2-y single-cycle analyses, dietary weights were used, for all analyses covering 10 cycles the 20-y dietary weights (as calculated in "Calculate 20-y weights for NHANES" section) were used.

Calculation of GI and population intake means for food groups

Similar foods were grouped together in Tables 1 and 2 based on food code similarity (i.e., same first 5 digits in food codes) or manual assignment (e.g., "roll, white, soft" as "white bread") (Supplemental Tables 3 and 4). For each food in a food group, the available carbohydrates were weighted with NHANES 20-y weights. The percent consumption of carbohydrates is the contribution of the available carbohydrates from a given food group to the total available carbohydrate intake. For each food group, the weighted average GI value was calculated as follows:

$$\text{Weighted average GI}_{\text{food group}} = \frac{\sum_{x=1}^{n} GI_x * W20_x}{\sum_{x=1}^{n} W20_x},$$

where x is a food intake in the NHANES data set with a food code belonging to the food group in question, GI_x is the GI value assigned to that food code, and $W20_x$ the NHANES 20-y weight assigned to the reporting individual.

Results

The AI-driven procedure for alignment of GI values to foods specified to NHANES 1999 to 2018 had an initial accuracy of 75% and was followed by manual quality control (reducing the AI assignment to 31.1%), resulting in 7976 foods having assigned GI values.

A total of 150,307 daily food reports were augmented with assigned GI values and dietary GI and GL values were calculated for each participant. In total, GI values were assigned to 99.4% of all food items, covering 99.9% of total carbohydrate intake.

TABLE 1

Top 20 contributors of carbohydrate and respective percentages of total available carbohydrate and their GI values in the United States, from 1999 to 2018

Food	% of Total available CHO	Cumulative % of total available CHO[1]	Weighted average GI[2] (GI for glucose = 100)
Soft drinks	9.5	9.5	64
White bread	4.6	14.1	77
Pizza	3.1	17.2	52
Other fruit juice/drink	2.6	19.8	64
Milk	2.5	22.3	30
Rice	2.3	24.6	70
Wheat bread	2.2	26.8	67
Ice cream	1.9	28.6	60
Orange juice	1.9	30.5	52
Banana	1.7	32.2	62
Spaghetti/pasta and tomato sauce	1.6	33.8	42
Sweetened tea	1.5	35.3	43
Beer	1.5	36.8	89
Potato (French fries)	1.4	38.3	75
Tortilla (corn or wheat)	1.4	39.6	33
Corn chips and snacks	1.2	40.8	65
Sugar (table and powdered)	1.2	42.1	68
Apple (raw, sauce, processed)	1.0	43.1	39
Cheeseburger/hamburger	1.0	44.1	57
Cake	1.0	45.0	48

Abbreviation: GI, glycemic index.

[1] Running sum over percentage of the total available CHO column. Similar reported food items grouped into 20 general categories to show main carbohydrate contributors.

[2] Glucose as reference.

To characterize the main foods in the United States diet that contribute to available carbohydrate, Table 1 presents the top 20 carbohydrate sources consumed and their GI values (for the top 50 carbohydrate foods, see Supplemental Table 3). These 20 foods account for 45.0% of total available carbohydrates consumed. The top 20 foods with the highest GLs and the cumulative total GLs are listed in Table 2.

Table 3 presents mean dietary GI and crude and energy-adjusted dietary GL values by sex, BMI, physical activity, age group, race and ethnicity, educational level, family income to poverty ratio, and cycle of NHANES. Among the 1999–2018 NHANES participants, 49,205 adults, 23,702 males and 25,503 females, had available sample weights. For these, mean dietary GI was 55.7 (95% CI: 55.5, 55.8; males: 56.4; 95% CI: 56.2, 56.5; females: 55.0; 95% CI: 54.8, 55.2) and mean energy-adjusted dietary GL was 133.0 (95% CI: 132.3, 133.8; males: 131.3; 95% CI: 130.3, 132.3; females: 134.6; 95% CI: 133.8, 135.5).

Dietary GL decreased substantially between younger and older age groups. After adjustment for energy intake, changes in GL were less pronounced but still tended to decrease with increasing age (decrease of 1.5 in GL every 10 y, $P < 0.001$), except for in the oldest age group of 70 y or older where GI and energy-adjusted GL values increased. Looking at differences in race or ethnicity, the highest dietary GI and energy-adjusted dietary GL were among Black adults (57.3; 95% CI: 57.1, 57.5; 139.4; 95% CI: 138.2, 140.7, respectively) than those among White adults (55.7; 95% CI: 55.5, 55.9; 95% CI: 131.6; 130.5, 132.6, respectively).

TABLE 2
Top 20 Dietary GL-contributing foods in the United States from 1999 to 2018

Food	Total GL[1]	Cumulative GL[2]
Soft drinks	10.9	10.9
White bread	5.8	16.6
Rice	3.5	20.2
Other fruit juice/drink	3.5	23.7
Wheat bread	2.6	26.3
Pizza	2.5	28.7
Potato (French fries)	2.2	30.9
Beer	2.1	32.9
Banana	1.9	34.9
Orange juice	1.8	36.7
Sugar (table and powdered)	1.7	38.3
Ice cream	1.6	39.9
Oatmeal (instant and regular)	1.4	41.3
Corn chips and snacks	1.2	42.5
Tortilla (corn or wheat)	1.2	43.7
Spaghetti/pasta and tomato sauce	1.1	44.8
Candy	1.0	45.8
Sweet roll	1.0	46.8
Milk	1.0	47.7
Cheeseburger/hamburger	1.0	48.7

Abbreviations: GI, glycemic index; GL, glycemic load.
[1] Total available carbohydrates: GI/100.
[2] Running sum over total GL column.

Dietary GI and GL decreased 5.6 units with each increase in education level ($P < 0.001$). The highest energy-adjusted dietary GL was among those with a lower education (high school or less: 137.7; 95% CI: 136.8, 138.7) than those with higher levels of education (college graduate or above: 126.5; 95% CI: 125.3, 127.7). As the ratio between family income to poverty income level increased, dietary GI and GL decreased by 3.6 units per level ($P < 0.001$). The difference in energy-adjusted GL of those living under the poverty level (ratio < 1) than that in those with a ratio of >5 was 140.9 (95% CI: 139.6, 142.1) compared with 124.7 (95% CI: 123.3, 126.1) (Supplemental Figure 3).

Changes over time were observed between 1999 and 2018 cycles of NHANES (Figure 2). Dietary GI remained stable between 1999 and 2004 and then steadily decreased until 2018. Energy-adjusted dietary GL decreased from 1999 to 2004, held relatively stable from 2004 to 2012, then decreased further from 2012 to 2018. Overall, between the years of 1999–2018, dietary GI decreased by 0.33 units per cycle ($P < 0.001$) or 4.6% from 56.9 (95% CI: 56.3, 57.4) to 54.3 (95% CI: 53.7, 54.9) and energy-adjusted dietary GL decreased by 2.16 units per cycle ($P < 0.001$) or 13.8% from 143.1 (95% CI: 139.9, 146.2) to 123.3 (95% CI: 121.1, 125.5).

Discussion

In this study, we used an innovative AI-enabled model in the first alignment of GI values to create dietary GI and GL variables for participants across 10 cycles of NHANES. The AI initially assigned 75.0% food descriptions correctly. However, guided by domain expertise, the manual verification process updated many GI assignments for food descriptions. As a result, the final tables contain 31.3% of AI-assigned GI values. Exploratory analyses revealed substantial differences for GI and GL across sex, race/ethnicity, education, and income levels, highlighting the importance of carbohydrate quality for population health outcomes in the United States. Although this work could advance the development of targeted nutritional recommendations for large-scale national surveillance or cohort studies for dietary determinants of health outcomes in diverse populations in the United States, some limitations and the interpretation of these findings deserve further discussion.

The availability of updated international GI food composition resources that provided data of improved quantity and quality enabled the achievement of a high degree of exact or close food matching and assignment of GI values. However, we acknowledge that assigning GI values to country-specific foods using international data sources could introduce some sources of error in any database and is a potential limitation. Indeed, even the same branded foods produced in different countries may vary in composition and therefore GI values. In addition, the International Table and online databases may be subject to error in GI values, including within-subject and between-laboratory variations. However, it has been demonstrated that if the recommended physiologic methods were used, the results of GI agreed reasonably well across laboratories [36,37]. Further, although the GI databases were extensive, there were still items that were not listed, and in those cases, estimated values were assigned. Our protocols attempted to minimize error by identifying foods with high similarity in descriptions, yet efforts to generalize some unlisted brand-specific foods to similar food groups could introduce bias or error. In addition, although the GI of some mixed meals were formally tested, we made calculations to determine the GI of many mixed meals manually, albeit according to the established method. Finally, as done in other investigations on dietary GI/GL, the calculation of available carbohydrates only included the subtraction of dietary fiber from total carbohydrates and not other compounds such as unavailable oligosaccharides, modified starches, and sugar alcohols because analytic data were unavailable. Of note, in most cases, resistant starch is considered part of the total dietary fiber content of foods.

The quality of carbohydrates and their nutritional classification continue to play a significant role in cardiometabolic disease development. In the Global Burden of Disease Study of 195 countries, low intake of whole grains was the leading dietary risk factor for disability-adjusted life years among males and females and the leading dietary risk factor for mortality among females [38]. The causal role of carbohydrate quality in disease development has been controversial. Although previous meta-analyses indicate significant relations between GI and GL and cardiometabolic disease outcomes, the interpretation of these findings in the context of the Bradford-Hill criteria is necessary. A dose–response meta-analysis of prospective cohort studies reported that diets high in GI and GL were robustly associated with incident type 2 diabetes mellitus (T2D) [8]. They went further in another study to examine and interpret the causality of this association and found that all 9 of the Hill's criteria were met for GI and GL in T2D development. Neither dietary fiber nor whole grains intakes were found to be reliable surrogate measures for GI and GL [39]. Pooling systematic reviews and meta-analyses and using Bradford-Hill causality criteria, the Nutrition and Chronic Diseases Expert Group provided convincing evidence of the etiologic effects of GL on cardiometabolic outcomes including coronary artery disease, stroke, and T2D [40]. Assessing the totality and the highest-quality data globally, a recent meta-analysis of large and prospective cohorts with over 100,000 participants (6 from the United States, 1 from Europe, 2 from Asia, and 1 international) found that the consumption of high GI foods was significantly associated with increased incidence of T2D, total cardiovascular disease, diabetes-related cancer, as well as all-cause mortality [41].

Despite the well-established importance of GI and GL in understanding the blood glucose–raising potential of carbohydrate-containing foods, previous assessments have not been comprehensive

TABLE 3
Mean crude and energy-adjusted GI and GL values by sex, age group, ethnicity, educational level, family income to poverty level, and cycle for NHANES participants: 1999–2018

	n	%	Dietary GI		Dietary GL			
			Mean (crude)	95% CI	Mean (crude)	95% CI	Energy-adjusted mean	95% CI
All	49,205	100.	55.7	55.5, 55.8	133.0	132.1, 134.0	133.0	132.3, 133.8
Males	23,702	48.2	56.4	56.2, 56.5	152.8	151.4, 154.2	131.3	130.3, 132.3
Females	25,503	51.8	55.0	54.8, 55.2	114.8	113.7, 115.8	134.6	133.8, 135.5
BMI (kg/m^2)								
<25	14,674	29.8	56.1	55.9, 56.3	137.5	135.8, 139.1	136.8	135.7, 137.9
≥25 to < 30	16,213	32.9	55.6	55.4, 55.8	133.7	132.2, 135.3	132.2	131.2, 133.2
≥30	17,550	35.7	55.3	55.1, 55.5	128.8	127.4, 130.3	130.3	129.3, 131.3
Physical activity: yes	23,435	47.6	55.4	55.2, 55.5	134.1	132.9, 135.3	130.5	129.6, 131.4
Physical activity: no	12,313	25.0	56.1	55.9, 56.3	137.6	135.7, 139.4	135.6	134.3, 136.8
Age (y)								
20–29	8222	16.7	56.3	56.0, 56.6	147.5	145.2, 149.8	138.6	137.0, 140.1
30–39	8001	16.3	55.8	55.5, 56.0	144.5	142.2, 146.8	134.5	132.9, 136.0
40–49	7941	16.1	55.4	55.2, 55.7	135.6	133.6, 137.6	130.6	129.2, 132.0
50–59	7153	14.5	55.3	55.1, 55.6	126.5	124.5, 128.5	128.8	127.4, 130.1
60–69	7802	15.9	55.1	54.8, 55.4	117.5	115.7, 119.3	128.4	127.1, 129.7
≥70	8332	16.9	55.8	55.6, 56.0	111.7	110.3, 113.2	135.1	134.2, 136.0
Race and ethnicity								
Mexican	8830	17.9	53.4	53.1, 53.7	132.6	130.3, 135.0	129.9	128.5, 131.3
Other Hispanic	4003	8.1	55.4	55.1, 55.8	133.0	130.2, 135.7	138.3	136.6, 140.1
White	21,844	44.4	55.7	55.5, 55.9	133.0	131.7, 134.3	131.6	130.5, 132.6
Black	10,310	21.0	57.3	57.1, 57.5	135.8	133.7, 137.8	139.4	138.2, 140.7
Other or multiracial	4218	8.6	55.6	55.2, 56.0	129.4	127.0, 131.9	137	135.1, 139.0
Educational level								
High school or less	23,526	47.8	56.4	56.2, 56.6	134.2	132.7, 135.7	137.7	136.8, 138.7
Some college	13,511	27.5	55.6	55.4, 55.8	133.1	131.7, 134.6	132.0	131.0, 133.0
College graduate or above	10,356	21.0	54.6	54.3, 54.8	129.7	127.8, 131.7	126.5	125.3, 127.7
Ratio of family income to poverty level								
<1	9337	19.0	56.2	55.9, 56.5	136.5	134.3, 138.8	140.9	139.6, 142.1
≥1 to <2	11,974	24.3	56.2	55.9, 56.4	134.7	133.0, 136.4	138.6	137.5, 139.7
≥2 to <3	6983	14.2	56.1	55.8, 56.4	134.3	132.0, 136.5	134.6	133.2, 136.0
≥3 to <4	5154	10.5	55.9	55.6, 56.2	135.1	132.5, 137.8	132.8	131.2, 134.4
≥4 to <5	3684	7.5	55.2	54.8, 55.5	133.1	130.3, 135.9	129.3	127.2, 131.3
≥5	7903	16.1	54.9	54.7, 55.2	129.1	126.9, 131.2	124.7	123.3, 126.1
NHANES cycles: 1999–2018								
1999–2000	4325	8.8	56.9	56.3, 57.4	146	141.1, 150.8	143.1	139.9, 146.2
2001–2002	4828	9.8	57.1	56.8, 57.4	147.7	144.0, 151.5	142.6	140.2, 145.0
2003–2004	4287	8.7	56.7	56.3, 57.2	140.4	137.3, 143.5	136.9	133.9, 139.9
2005–2006	4683	9.5	56.4	55.9, 56.9	136.4	132.7, 140.1	135.2	132.7, 137.7
2007–2008	5459	11.1	56.1	55.7, 56.5	133.5	130.3, 136.7	135.6	133.1, 138.2
2009–2010	5834	11.9	55.8	55.5, 56.0	133.7	131.4, 136.1	134.2	133.0, 135.5
2011–2012	4880	9.9	55.6	55.3, 55.9	134.6	132.7, 136.5	133.5	131.7, 135.4
2013–2014	5086	10.3	54.8	54.5, 55.1	126	123.7, 128.3	128.3	126.6, 129.9
2015–2016	5062	10.3	53.9	53.5, 54.3	120.6	117.9, 123.3	123.8	121.7, 125.9
2017–2018	4761	9.7	54.3	53.7, 54.9	121.1	118.2, 123.9	123.3	121.1, 125.5

Abbreviations: BMI, body mass index; CI, confidence interval; GI, glycemic index; GL, glycemic load.
Values are weighted means and 95% CIs. Family income to poverty ratio: a ratio of <1 means that the income is less than the poverty level, and when the ratio is >1, this indicates a higher income than the poverty level. Education level is for adults aged older than 20 y. Physical activity is classified as yes/no for adhering to the federal Physical Activity Guidelines for Americans for participating in ≥150 min of moderate-intensity aerobic activity per week. GL was energy adjusted to 2097 kcal.

at the national level. In 2001, Ford and Liu [12], used an FFQ to estimate dietary GI and GL for adults who participated in the third National Health and Nutrition Examination Survey (1988–1994) and reported a higher mean GI (81.3) and GL (141.8) values using white bread as the reference standard. Using glucose as the reference, the mean dietary GI and GL values were comparable with another NHANES study of only 2 cycles (2003–2006) using two 24-h dietary recalls (overall mean dietary GI of 56.2 and GL of 138.1) [20]. National dietary GI and GL values tended to decrease over time in our study. This is comparable with a previous NHANES trends analysis that reported decreases in low-quality carbohydrates (primarily added sugar) and increases in high-quality carbohydrates (primarily whole grains) and plant protein between 1999 and 2016 [42]. However, the proportion of energy intake from low-quality carbohydrates was still high at 42%. In our study, significant disparity existed for dietary GI and GL across sex, race/ethnicity, education, and income levels, highlighting the importance of improving carbohydrate quality for population health outcomes in the United States. A study of the macronutrient ratios as well as micronutrient contents reported at different levels of GI/GL would help elucidate other aspects of the overall dietary pattern beyond carbohydrate quality. The creation of this GI database will allow such in-depth analyses to be conducted.

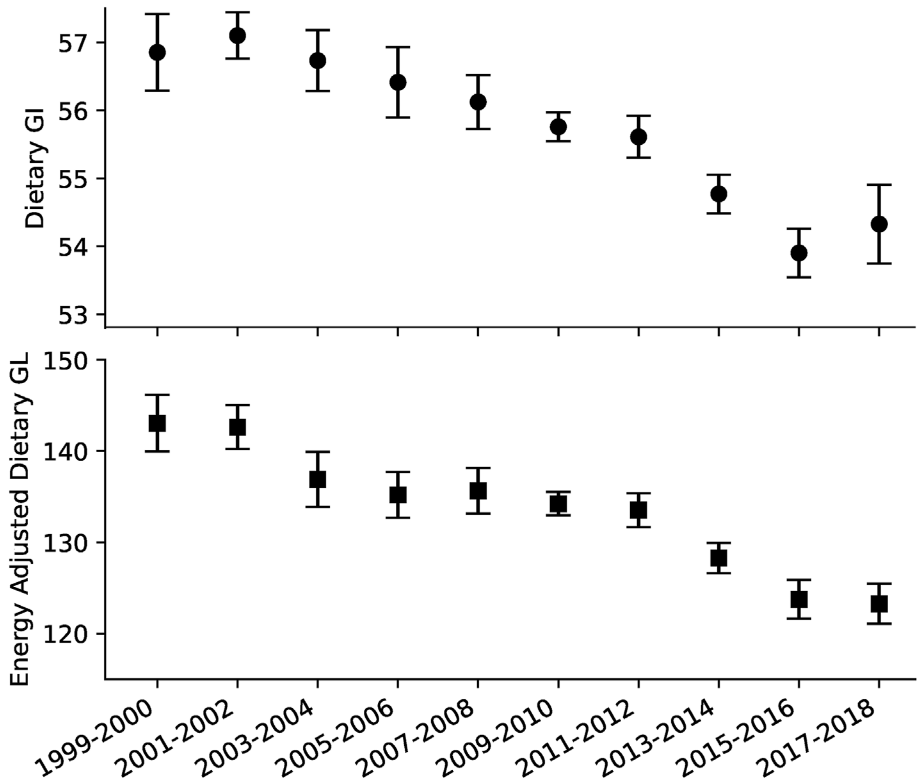

FIGURE 2. Dietary GI and energy-adjusted dietary GL changes across 10 NHANES cycles. GI, glycemic index; GL, glycemic load.

Examining changes over time in other countries, a Brazilian national study found that the proportion of individuals reporting fiber-rich, low-GI foods such as beans and fruit decreased between 2008 and 2018 [43]. In addition, it was reported that carbohydrate quality was reduced among adolescents and in rural areas over 10 years in Brazil [44]. Changes in dietary GI and GL in Australian adults were assessed through national nutrition surveys between 1995 and 2012 [45]. Overall, they reported that dietary GI and GL decreased by 5% and 12%, respectively (GI: 56.5 ± 6.2 compared with 53.9 ± 6.8; GL: 153.3 ± 62.1 compared with 135.4 ± 58.5). These changes in dietary GL are very similar to our analysis in which energy-adjusted dietary GL decreased by 13.8% from 143.1 (95% CI: 139.9, 146.2) to 123.3 (95% CI: 121.1, 125.5) between 1999 and 2018.

Recent advancements in LLMs [46] have enabled significant progress in natural language processing, allowing for large-scale analysis and the generation of human-like text. This study demonstrates that leveraging vector embeddings, components of LLMs that convert text into multidimensional numerical representations [47], can aid in the tackling of complex challenges encountered in epidemiologic research. Specifically, vector embeddings capture relationships and patterns within texts. By applying similarity measures [48] to quantify the likeness between data points, we were able to align 2 large databases, for the first time, to our knowledge, demonstrating that the recent progress in LLMs can be successfully implemented in epidemiologic studies of multiple diverse populations, especially when guided by substantive domain expertise. In this study, our observations included many group comparisons, which is of primary importance in epidemiologic assessment and interpretation.

In conclusion, this study marks the creation of the first national GI database with associated dietary GI and GL derived from 10 cycles of nationally representative surveys in the United States. Our approach extends the applicability of description mapping and advances the methodology for incorporating GI and GL values into large-scale databases of diverse surveys and cohorts. With the wealth of well-characterized individual phenotypes of health status spanning 10 NHANES cycles from 1999 to 2018 now augmented with dietary GI and GL, the groundwork is laid for continued monitoring of the nutritional classification of dietary carbohydrates as well as for in-depth and prospective assessment of the impacts of their glycemic and metabolic potentials on cardiometabolic health or disease outcomes in the United States.

Acknowledgments

We thank Kisione Taufa and Caisen Chandler for assistance in initial manual verification of AI-assigned GI values.

Author contributions

The authors' responsibilities were as follows – SL, KADC: designed the research; BY: extracted NHANES data and verified analysis; DDC: developed AI protocol, mixed meal protocol, and statistical analysis; KADC: developed manual protocol; ST, KADC: executed manual protocol; KADC, SL: interpreted and analyzed data; KADC, DDC, SL: wrote the manuscript; KADC: had primary responsibility for the content of the manuscript; and all authors: read and approved the final manuscript.

Conflict of interest

KADC, DDC, ST, and BY have no conflicts of interest. SL reports consulting payments and honoraria (or promises of the same) for scientific presentations or reviews at numerous venues, including (but not limited to) the National Institute of Health, Fred Hutchinson Cancer Center, University of Missouri, Harvard University, University of

Buffalo, Universidade Federal de São Paulo, and Guangdong Provincial Hospital and Academy of Medical Science. SL also reports honoraria from Twin Digital Health; compensation for serving on the data safety and monitoring board for several trials, including the SELECT trial sponsored by Novo Nordisk; royalties from UpToDate; and an honorarium from the American Society for Nutrition for his duties as Associate Editor. SL did not have a role in the evaluation and review of this article by *The American Journal of Clinical Nutrition*.

Funding

The authors reported no funding received for this study.

Data availability

Data described in the manuscript, code book, and analytic code will be made available on request pending application and approval. The AI-based alignment procedure and a tutorial are available under MIT license at https://github.com/dellacortelab/nhanes_gigl.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ajcnut.2024.06.001.

References

[1] T.M. Wolever, D.J. Jenkins, A.L. Jenkins, R.G. Josse, The glycemic index: methodology and clinical implications, Am. J. Clin. Nutr. 54 (5) (1991) 846–854, https://doi.org/10.1093/ajcn/54.5.846.
[2] W. Willett, J. Manson, S. Liu, Glycemic index, glycemic load, and risk of type 2 diabetes, Am. J. Clin. Nutr. 76 (1) (2002) 274S–280S, https://doi.org/10.1093/ajcn/76/1.274S.
[3] D.J. Jenkins, M. Dehghan, A. Mente, S.I. Bangdiwala, S. Rangarajan, K. Srichaikul, et al., Glycemic index, glycemic load, and cardiovascular disease and mortality, N. Engl. J. Med. 384 (14) (2021) 1312–1322, https://doi.org/10.1056/NEJMoa2007123.
[4] M.M. Van Bakel, N. Slimani, E.J. Feskens, H. Du, J.W. Beulens, Y.T. Van Der Schouw, et al., Methodological challenges in the application of the glycemic index in epidemiological studies using data from the European prospective investigation into cancer and nutrition, J. Nutr. 139 (3) (2009) 568–575, https://doi.org/10.3945/jn.108.097121.
[5] B.C. Olendzki, Y. Ma, A.L. Culver, I.S. Ockene, J.A. Griffith, A.R. Hafner, et al., Methodology for adding glycemic index and glycemic load values to 24-hour dietary recall database, Nutrition 22 (11–12) (2006) 1087–1095, https://doi.org/10.1016/j.nut.2006.07.006.
[6] L.M. Aston, D. Jackson, S. Monsheimer, S. Whybrow, T. Handjieva-Darlenska, M. Kreutzer, et al., Developing a methodology for assigning glycaemic index values to foods consumed across Europe, Obes. Rev. 11 (1) (2010) 92–100, https://doi.org/10.1111/j.1467-789X.2009.00690.x.
[7] S.N. Bhupathiraju, D.K. Tobias, V.S. Malik, A. Pan, A. Hruby, J.E. Manson, et al., Glycemic index, glycemic load, and risk of type 2 diabetes: results from 3 large US cohorts and an updated meta-analysis, Am. J. Clin. Nutr. 100 (1) (2014) 218–232, https://doi.org/10.3945/ajcn.113.079533.
[8] G. Livesey, R. Taylor, H.F. Livesey, A.E. Buyken, D.J. Jenkins, L.S. Augustin, et al., Dietary glycemic index and load and the risk of type 2 diabetes: a systematic review and updated meta-analyses of prospective cohort studies, Nutrients 11 (6) (2019) 1280, https://doi.org/10.3390/nu11061280.
[9] A.K. Dwivedi, P. Dubey, S.Y. Reddy, D.J. Clegg, Associations of glycemic index and glycemic load with cardiovascular disease: updated evidence from meta-analysis and cohort studies, Curr. Cardiol. Rep. 24 (3) (2022) 141–161, https://doi.org/10.1007/s11886-022-01635-2.
[10] D.C. Greenwood, D.E. Threapleton, C.E. Evans, C.L. Cleghorn, C. Nykjaer, C. Woodhead, et al., Glycemic index, glycemic load, carbohydrates, and type 2 diabetes: systematic review and dose–response meta-analysis of prospective studies, Diabetes Care 36 (12) (2013) 4166–4171, https://doi.org/10.2337/dc13-0325.
[11] N. Ahluwalia, J. Dwyer, A. Terry, A. Moshfegh, C. Johnson, Update on NHANES dietary data: focus on collection, release, analytical considerations, and uses to inform public policy, Adv. Nutr. 7 (1) (2016) 121–134, https://doi.org/10.3945/an.115.009258.
[12] E.S. Ford, S. Liu, Glycemic index and serum high-density lipoprotein cholesterol concentration among US adults, Arch. Intern. Med. 161 (4) (2001) 572–576, https://doi.org/10.1001/archinte.161.4.572.
[13] A. Culberson, M.R. Kafai, V. Ganji, Glycemic load is associated with HDL cholesterol but not with the other components and prevalence of metabolic syndrome in the third National Health and Nutrition Examination Survey, 1988–1994, Int. Arch. Med. 2 (1) (2009) 3, https://doi.org/10.1186/1755-7682-2-3.
[14] T.M. Larsen, S. Dalskov, M. Van Baak, S. Jebb, A. Kafatos, A. Pfeiffer, et al., The Diet, Obesity and Genes (Diogenes) Dietary Study in eight European countries—a comprehensive design for long-term intervention, Obes. Rev. 11 (1) (2010) 76–91, https://doi.org/10.1111/j.1467-789X.2009.00603.x.
[15] F.S. Atkinson, J.C. Brand-Miller, K. Foster-Powell, A.E. Buyken, J. Goletzke, International tables of glycemic index and glycemic load values 2021: a systematic review, Am. J Clin. Nutr. 114 (5) (2021) 1625–1632, https://doi.org/10.1093/ajcn/nqab233.
[16] University of Sydney, Glycemic index database [Internet] Available from, www.glycemicindex.com, 2023.
[17] A.J. Moshfegh, D.G. Rhodes, D.J. Baer, T. Murayi, J.C. Clemens, W.V. Rumpler, et al., The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes, Am. J Clin. Nutr. 88 (2) (2008) 324–332, https://doi.org/10.1093/ajcn/88.2.324.
[18] C.A. Blanton, A.J. Moshfegh, D.J. Baer, M.J. Kretsch, The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake, J. Nutr. 136 (10) (2006) 2594–2599, https://doi.org/10.1093/jn/136.10.2594.
[19] J.M. Conway, L.A. Ingwersen, A.J. Moshfegh, Accuracy of dietary recall using the USDA five-step multiple-pass method in men: an observational validation study, J. Am. Diet. Assoc. 104 (4) (2004) 595–603, https://doi.org/10.1016/j.jada.2004.01.007.
[20] C.-S. Lin, R.W. Kimokoti, L.S. Brown, E.A. Kaye, M.E. Nunn, B.E. Millen, Methodology for adding glycemic index to the National Health and Nutrition Examination Survey nutrient database, J. Acad. Nutr. Diet. 112 (11) (2012) 1843–1851, https://doi.org/10.1016/j.jand.2012.07.035.
[21] M.L. Neuhouser, L.F. Tinker, C. Thomson, B. Caan, L.V. Horn, L. Snetselaar, et al., Development of a glycemic index database for food frequency questionnaires used in epidemiologic studies, J. Nutr. 136 (6) (2006) 1604–1609, https://doi.org/10.1093/jn/136.6.1604.
[22] Assafmo. Diogenes GL calculator, 2020.
[23] J.B. Montville, J.K. Ahuja, C.L. Martin, K.Y. Heendeniya, G. Omolewa-Tomobi, L.C. Steinfeldt, et al., USDA Food and Nutrient Database for Dietary Studies (FNDDS), 5.0, Proc. Food Sci. 2 (2013) 99–112, https://doi.org/10.1016/j.profoo.2013.04.016.
[24] USDA, FNDDS databases [Internet] Available from, https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds-download-databases/, 2023.
[25] A. Neelakantan, T. Xu, R. Puri, A. Radford, J.M. Han, J. Tworek, et al., Text and code embeddings by contrastive pre-training, arXiv preprint arXiv: 220110005 (2022).
[26] F. Rahutomo, T. Kitasuka, M. Aritsugi, Semantic cosine similarity. The 7th international Student Conference on Advanced Science and Technology (ICAST), 2012, p. 1.
[27] FAO, Carbohydrates in human nutrition: report of a joint FAO/WHO expert consultation, Rome, 14–18 April 1997, FAO, Rome, 1997.
[28] T.M. Wolever, The glycaemic index: a physiological classification of dietary carbohydrate, CABI Digital Library, 2006.
[29] F.S. Atkinson, K. Foster-Powell, J.C. Brand-Miller, International tables of glycemic index and glycemic load values: 2008, Diabetes Care 31 (12) (2008) 2281–2283, https://doi.org/10.2337/dc08-1239.
[30] S. Liu, W.C. Willett, M.J. Stampfer, F.B. Hu, M. Franz, L. Sampson, et al., A prospective study of dietary glycemic load, carbohydrate intake, and risk of coronary heart disease in US women, Am. J Clin. Nutr. 71 (6) (2000) 1455–1461, https://doi.org/10.1093/ajcn/71.6.1455.
[31] S. Liu, J.E. Manson, M.J. Stampfer, M.D. Holmes, F.B. Hu, S.E. Hankinson, et al., Dietary glycemic load assessed by food-frequency questionnaire in relation to plasma high-density-lipoprotein cholesterol and fasting plasma triacylglycerols in postmenopausal women, Am. J Clin. Nutr. 73 (3) (2001) 560–566, https://doi.org/10.1093/ajcn/73.3.560.
[32] S.G. Heeringa, B.T. West, P.A. Berglund, Applied survey data analysis, 2nd Ed., Chapman and Hall/CRC, 2017.
[33] CDC, NHANES weighting tutorial. National Center for Health Statistics [Internet]. [cited 16 April 2024]. Available from: https://wwwn.cdc.gov/nchs/nhanes/tutorials/weighting.aspx.
[34] W.C. Willett, G.R. Howe, L.H. Kushi, Adjustment for total energy intake in epidemiologic studies, Am. J Clin. Nutr. 65 (Suppl 4) (1997) 1220S–1228S, https://doi.org/10.1093/ajcn/65.4.1220S.

[35] T. Lumley, Analysis of complex survey samples, J. Stat. Softw. 9 (8) (2004) 1–19, https://doi.org/10.18637/jss.v009.i08.
[36] T.M. Wolever, Effect of blood sampling schedule and method of calculating the area under the curve on validity and precision of glycaemic index values, Br. J. Nutr. 91 (2) (2004) 295–301, https://doi.org/10.1079/BJN20031054.
[37] T.M. Wolever, D.J. Jenkins, V. Vuksan, R.G. Josse, G.S. Wong, A.L. Jenkins, Glycemic index of foods in individual subjects, Diabetes Care 13 (2) (1990) 126–132, https://doi.org/10.2337/diacare.13.2.126.
[38] A. Afshin, P.J. Sur, K.A. Fay, L. Cornaby, G. Ferrara, J.S. Salama, et al., Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 393 (10184) (2019) 1958–1972.
[39] G. Livesey, R. Taylor, H.F. Livesey, A.E. Buyken, D.J.A. Jenkins, L.S.A. Augustin, et al., Dietary glycemic index and load and the risk of type 2 diabetes: assessment of causal relations, Nutrients 11 (6) (2019) 1436, https://doi.org/10.3390/nu11061436.
[40] R. Micha, M.L. Shulkin, J.L. Peñalvo, S. Khatibzadeh, G.M. Singh, M. Rao, et al., Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE), PLoS One 12 (4) (2017) e0175149, https://doi.org/10.1371/journal.pone.0175149.
[41] D.J. Jenkins, W.C. Willett, S. Yusuf, F.B. Hu, A.J. Glenn, S. Liu, et al., Association of glycaemic index and glycaemic load with type 2 diabetes, cardiovascular disease, cancer, and all-cause mortality: a meta-analysis of mega cohorts of more than 100 000 participants, Lancet Diabetes, Endocrinol 12 (2) (2024) 107–118, https://doi.org/10.1016/S2213-8587(23)00344-3.
[42] Z. Shan, C.D. Rehm, G. Rogers, M. Ruan, D.D. Wang, F.B. Hu, et al., Trends in dietary carbohydrate, protein, and fat intake and diet quality among US adults, 1999-2016, JAMA 322 (12) (2019) 1178–1187, https://doi.org/10.1001/jama.2019.13771.
[43] R.M. Rodrigues, A.d.M. Souza, I.N. Bezerra, R.A. Pereira, E.M. Yokoo, R. Sichieri, Most consumed foods in Brazil: evolution between 2008-2009 and 2017-2018, Rev. Saude. Publica. 55 (Supl 1) (2021) 4s.
[44] L.G. Cardoso, L.S. Monteiro, R. Sichieri, E.M. Yokoo, R.A. Pereira, Changes in dietary carbohydrate and lipid quality indices between 2008 and 2018: analysis of the Brazilian dietary surveys, Nutr. Bull. 48 (4) (2023) 546–558, https://doi.org/10.1111/nbu.12641.
[45] D.T.L. Kusnadi, A.W. Barclay, J.C. Brand-Miller, J.C.Y. Louie, Changes in dietary glycemic index and glycemic load in Australian adults from 1995 to 2012, Am. J. Clin. Nutr. 106 (1) (2017) 189–198, https://doi.org/10.3945/ajcn.116.150516.
[46] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, et al., Language models are few-shot learners, Adv. Neural Inform. Process. Syst. 33 (2020) 1877–1901.
[47] A. Conneau, G. Kruszewski, G. Lample, L. Barrault, M. Baroni, What you can cram into a single $&!#* vector: probing sentence embeddings for linguistic properties, arXiv preprint arXiv:180501070. (2018), https://doi.org/10.18653/v1/P18-1198.
[48] T. Kenter, M. De Rijke, Short text similarity with word embeddings. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, pp. 1411–1420.