frontiers | Research Topics

# Databases and nutrition, volume III

**Edited by**
Alessandra Durazzo, Igor Pravst and
Massimo Lucarini

**Published in**
Frontiers in Nutrition

ISSN 1664-8714
ISBN 978-2-8325-7214-6
DOI 10.3389/978-2-8325-7214-6

**Generative AI statement**
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Databases and nutrition, volume III

**Topic editors**
Alessandra Durazzo — Research Centre for Food and Nutrition, Council for Agricultural Research and Economics, Italy
Igor Pravst — Institute of Nutrition, Slovenia
Massimo Lucarini — Research Centre for Food and Nutrition, Council for Agricultural Research and Economics, Italy

**Citation**
Durazzo, A., Pravst, I., Lucarini, M., eds. (2025). *Databases and nutrition, volume III*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7214-6

# Table of contents

frontiers | Frontiers in Nutrition

TYPE Editorial
PUBLISHED 23 October 2025
DOI 10.3389/fnut.2025.1675443



OPEN ACCESS

EDITED AND REVIEWED BY
Angela M. Zivkovic,
University of California, Davis, United States

*CORRESPONDENCE
Alessandra Durazzo
alessandra.durazzo@crea.gov.it
Igor Pravst
igor.pravst@nutris.org
Massimo Lucarini
massimo.lucarini@crea.gov.it

RECEIVED 29 July 2025
ACCEPTED 22 September 2025
PUBLISHED 23 October 2025

CITATION
Durazzo A, Pravst I and Lucarini M (2025) Editorial: Databases and nutrition, volume III. *Front. Nutr.* 12:1675443.
doi: 10.3389/fnut.2025.1675443



# Editorial: Databases and nutrition, volume III

Alessandra Durazzo[1]*, Igor Pravst[2]* and Massimo Lucarini[1]*

[1]CREA-Research Centre for Food and Nutrition, Rome, Italy, [2]Institute of Nutrition, Ljubljana, Slovenia

KEYWORDS

**food data, food groups, nutrients, natural substances, dietary supplements, classification, categorization, food composition database**

Editorial on the Research Topic
Databases and nutrition, volume III

## Introduction

Public health nutrition is the promotion of the nutrition-related health of populations. Food composition databases have an essential role in the assessment, analysis, and action phases of public health nutrition. The food composition database provides comprehensive information on the energy content, various nutrients, and other bioactive constituents in food products obtained from agriculture, fisheries, and livestock. The country-specific food composition databases are developed to include the composition data of foods consumed by the population and represent essential tools for assessing national nutritional status, thus being critical to advance nutritional research and policy. The management of food composition programmes includes the maintenance and continuous updating of food composition information, as this is a useful tool for estimating nutrient intake at the national, regional, and/or certain population levels.

Accurate, country-specific food composition databases that reflect the national food supply are essential for estimating nutrient intake and conducting reliable dietary assessments, thereby serving as a key tool for evaluating and monitoring diets. Indeed, food composition databases are utilized to meet the supply and demand of agricultural products, assess the quality of exported products in international trade, public health campaigns, nutrition programmes and strategies, and boost innovations in the food industry.

Food composition databases provide reliable data on nutrient composition and bioactive profiles, supporting diverse applications such as clinical nutrition, epidemiological research, health surveys, diet therapy and planning, dietary guidelines, nutrition policies, food development, nutrition recommendations, nutrition education, and food labeling regulations. Therefore, food composition databases are fundamental for a broad user base, i.e., researchers, dietitians, clinical dietitians, and other health professionals, government policymakers, consumers, marketing professionals, and other policymakers. These databases are therefore also used in a wide variety of organizations—from academia to various industries, including food businesses, IT providers, and governments.

The integration and harmonization of food composition data and modern omics technologies is an ongoing challenge. Beyond the macro- and micronutrient information provided by national databases, resources for food composition data are increasingly focused on high-resolution analyses aimed at capturing the full spectrum of small, potentially bioactive molecules present in foods. The availability of standardized, harmonized, and integrated large-scale food composition data and mass spectra resources will be fundamental for future directions in the perspective of data integration and interoperability. Quality control of analytical procedures is a key element for the accuracy, precision, and reliability of data for inclusion in food composition databases.

Safe food represents a key aspect of food security, and, consequently, food traceability along the supply chain represents a fundamental component. Data traceability starts from data collection and continues through to analysis results; its role is to ensure data reproducibility along the food chain, from raw material production to transportation to logistics. The analytical data and the development of a food safety assessment system produce useful information and represent key elements to obtain an effective traceability system and guarantee efficiency in the management of the entire supply chain. Emerging technologies such as cloud computing, digital platforms, mobile tools, and artificial intelligence offer new opportunities to build smart food traceability systems that integrate across the agri-food supply chain. These systems can monitor food supply and population-level dietary data, thereby improving data quality and safety while supporting the development of integrated food data infrastructures. Particularly, the use of artificial intelligence is currently emerging as a key part of the management of food composition databases.

There is a need to improve the international harmonization of food composition databases to meet expectations for international research and comparisons. The classification and harmonization of foods are essential to the development of connectable database systems. The growing availability of standardized data facilitates integration across sources, as future analyses increasingly rely on data harmonization and interoperability. A key current challenge is linking environmental and food composition databases, connecting nutritional and environmental entries in order to identify more sustainable food options. Furthermore, there is a need for additional data regarding food waste and by-products and, consequently, for databases to include information on chemical composition, origin, and quantities of by-products from the agri-food sector.

The availability of branded food databases also brings new opportunities and challenges. By providing detailed and up-to-date nutritional information specific to branded products, these databases improve the reliability of data for applications such as nutrient intake assessment and food reformulation monitoring.

In this context, the present Special Section, *Databases and nutrition, volume III*, brings together nine contributions that address these themes from different perspectives. Concerning the development of automatic procedures in database management, the study of Westenbrink et al. addressed the development of an automated approach to identify fortified foods in the Dutch branded food database LEDA (short for LEvensmiddelenDAtabank). An automated procedure, based on a stepwise approach conforming to European food labeling legislation, using a list of rules and search terms, was developed and resulted in the identification of 1,817 foods, fortified with one or more of the selected nutrients in the LEDA dataset (0.94%; Westenbrink et al.).

The study of Bardon et al. described the development and evaluation of the FNS-Cloud data quality assessment tool for dietary intake datasets.

The study of Valenčič et al. presented NutriBase, a novel database and knowledge management system designed to advance the science of food composition through improvements in harmonization, data quality, reduction of missing data, and interoperability.

Regarding uses and applications of databases, the study of Fazzino et al. quantified the prevalence of hyper-palatable food (HPF) in the Italian food system and compared the hyper-palatability of similar foods across Italy and the United States (US), which has wide HPF saturation: HPF comprise less than one third of the Italian food system, indicating the Italian food system may confer protection from HPF exposure. Findings also revealed key differences in HPF products between Italy and the US, with HPF from Italy tending to have lower palatability-inducing nutrients and higher satiety-promoting nutrients relative to comparable US products (Fazzino et al.). Moreover, authors highlighted that food companies in Italy and the US should consider reducing the sodium, refined carbohydrates, and fat in salty snacks, frozen pizzas, industrial breads, and protein/cereal bars to reduce the hyperpalatability of these commonly consumed foods in Italy and the US.

Wang X. et al. investigated the association between plain water intake (PWI) and the risk of osteoporosis among middle-aged and elderly people in the United States by a cross-sectional study: results suggested that among middle-aged and elderly people, a greater PWI was connected with a moderately lower osteoporosis risk. The study of Wang Y. et al. is focused on the application of food composition data; their work was focused on the exploration of the links between consumption of Sugar-sweetened beverages (SSB) and specific health-related outcomes and lifestyle parameters.

Kraemer et al. have discussed methodological evolution and challenges of in-store census methods for assessing the composition of branded foods, and they characterized a Brazilian food label database.

Terro et al. present the IsoFoodTrack database—a comprehensive, scalable, and flexible platform designed to manage isotopic and elemental composition data for a wide range of food commodities. Brinkley et al. conducted an integrative review of 35 data attributes across 101 FCDBs from 110 countries, highlighting emerging opportunities and recommendations.

Contributions in Volume III of Databases and Nutrition showcase cutting-edge efforts to develop and update comprehensive and dedicated food databases, emphasizing rigorous standardization, harmonization, and interoperability across data sources—from analytical measurements to literature-derived values, labeling, and calculated data. The adoption of robust quality evaluation indices, consistent food description systems, and semi-automated matching and alignment procedures reflects the growing implementation of nutritional data infrastructures. These resources serve not only to support food composition research but also to underpin interdisciplinary applications spanning health, environmental science, policy, and beyond.

## Author contributions

AD: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. IP: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. ML:

Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

We would like to acknowledge the authors and reviewers of the publications in this Research Topic for their invaluable contributions and efforts. We are also grateful to the editorial board members and support staff of the journal for their kind support during the preparation of this Research Topic.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 10 April 2024
DOI 10.3389/fnut.2024.1366083

Check for updates

OPEN ACCESS

EDITED BY
Massimo Lucarini,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Mojca Korošec,
University of Ljubljana, Slovenia
Urška Blaznik,
Institute of Public Health of the Republic of Slovenia, Slovenia
Simone Eussen,
Danone Nutricia Research,
Netherlands

*CORRESPONDENCE
Marga Ocké
✉ marga.ocke@rivm.nl

†Deceased

RECEIVED 05 January 2024
ACCEPTED 18 March 2024
PUBLISHED 10 April 2024

CITATION
Westenbrink S, Tenhagen CJ, Toxopeus I, Verkaik-Kloosterman J, Feskens EJM and Ocké M (2024) Procedure to identify fortified foods in the Dutch branded food database. *Front. Nutr.* 11:1366083.
doi: 10.3389/fnut.2024.1366083



# Procedure to identify fortified foods in the Dutch branded food database

Susanne Westenbrink[1†], Cyrelle J. Tenhagen[1], Ido Toxopeus[1], Janneke Verkaik-Kloosterman[1], Edith J. M. Feskens[2] and Marga Ocké[1,2]*

[1]National Institute for Public Health and the Environment, Bilthoven, Netherlands, [2]Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, Netherlands

**Introduction:** Information on fortified foods is needed for multiple purposes, including food consumption research and dietary advice. Branded food databases are a valuable source of food label data. European labeling legislation prescribes that food fortification should be indicated in the ingredient list, and nutrient values should be declared under certain conditions. This creates the potential to identify fortified foods in branded food databases, though it is not straightforward and labor-intensive. The aim of our study was to develop an automated approach to identify fortified foods in the Dutch branded food database called LEDA.

**Methods:** An automated procedure, based on a stepwise approach conforming with European labeling legislation, using a list of rules and search terms, was developed to identify fortified foods. Fortification with calcium, folic acid, vitamin B12, and zinc was studied as an example. The results of a random stratified sample with fortified and not-fortified foods were validated by two experts.

**Results:** The automated approach resulted in identifying 1,817 foods fortified with one or more of the selected nutrients in the LEDA dataset (0.94%). The proportions of fortified foods per nutrient were below 0.7%. The classification of fortified/non-fortified foods matched manual validation by experts for the majority of the foods in the sample, i.e., sensitivity and specificity indicating the probability of correctly identifying fortified and non-fortified foods was high (>94.0%).

**Conclusion:** The automated approach is capable of easily and quickly identifying fortified foods in the Dutch branded food database with high accuracy, although some improvements to the automated procedure could be made. In addition, the completeness, correctness, and consistency of the LEDA database can be improved. To fully benefit from this automated approach, it needs to be expanded to cover all micronutrients that may be added to foods.

KEYWORDS

**automated approach, branded foods, branded food database, decision tree, food fortification, LEDA, label data, Netherlands**

# 1 Introduction

Healthy and safe diets providing adequate nutrient intake are essential to maintain good health. Several authors reported low intakes of micronutrients for various population groups and identified the possible contribution of fortified foods to improve intakes and related health outcomes (1–6). In a review of European evidence from 2013, it was concluded that voluntary fortification by food manufacturers can reduce the risk of sub-optimal intakes of a range of micronutrients at a population level, whereas small proportions of the population, especially children, may exceed the upper intake level for some micronutrients (6). Information on fortified foods is needed for multiple purposes, including food consumption research, personalized dietary advice, public health information, development and monitoring of food fortification strategies, and enforcement of legislation related to fortification (7–13).

In Europe, food fortification is regulated by national and European legislation to ensure safe and necessary fortification practices. Adding vitamins or minerals should at least result in significant amounts as defined by the European labeling legislation (14). On the food label, added nutrients need to be declared in the ingredient list. The total nutrient content (naturally present plus added as fortificants or other food additives) is mandatory in the nutritional panel if present in significant amounts as defined in the EU labeling legislation. Amounts are considered significant when reaching 7.5% of the dietary reference intake (DRI) for drinks per 100 mL or 15% for solid foods per 100 g and per single portion packs. Mandatory declaration of nutritional values on the label also applies in case of nutritional or health claims (14, 15). European legislation has not yet defined maximum levels for fortification. In the Dutch legislation, the maximum level is set at 100% of the reference intake (4, 16, 17), except for vitamins A and D, folic acid, iodine, selenium, copper, and zinc intakes for which fortification is prohibited to prevent excessive intake. There are, however, generic and specific exemptions for food fortification with these nutrients. For vitamin D and folic acid, a maximum of 4.5 µg /100 kcal and 100 µg /100 kcal are set as generic exemptions, respectively. In addition, for some micronutrients (e.g., zinc and copper), addition to food is allowed for restoration or substitution purposes. In the Netherlands, fortification is always on a voluntary basis, although the addition of vitamins A and D in plant-based fat products (such as margarine) and iodized salt in bread are encouraged by covenants between the food industry and the government. Legislation on food fortification in the Netherlands is summarized by de Jong et al. (4).

Most generic national food composition databases contain no or limited information on branded foods, and information on the fortification of foods is often lacking (18, 19), among others, because fortification is generally brand-specific. The growing number of branded food databases worldwide can fill this gap [e.g., (20–27)], provided that information on relevant foods, nutrient values, and fortification is present, correct, and up to date. Some authors report “manual” identification of fortified foods by experts for (subsets of) their databases, for example (11, 28–30).

The Dutch national branded food database LEDA contains food label data provided by food producers, including ingredient lists and nutritional values for energy, macronutrients, salt, and some data on micronutrients (31). The LEDA database is hosted at the Dutch Nutrition Centre. In 2022, the total number of branded foods was nearly 200,000, and it was estimated that 75% of the retail market was represented (31). The size of this branded food database and the rapid changes make identifying fortification by researchers on a food-by-food basis very labor intensive. We are not aware of any automated approach to identify fortified foods in branded food databases. Therefore, the aim of our study was to develop and validate an automated, standardized procedure to identify fortified foods in the Dutch national branded food database LEDA. Foods fortified with added calcium, zinc, vitamin B12, or folic acid were taken as an example.

# 2 Methods

## 2.1 Definition of fortified foods

To identify fortified foods, our definition is based on the European labeling legislation (EU 1169/2011). This means that we consider food to be fortified when the micronutrient, or its chemical form, is mentioned in the ingredient list and the total amount present is declared in the nutritional panel on the food label, if significant according to the European labeling legislation (14, 15). Nutrients may also be added for restoration (to make up for losses during processing) or substitution (to replicate the content of another food). European legislation does not differentiate between reasons for adding nutrients to foods, and this information cannot be retrieved from food labels. For this study, all information on added micronutrients on the food label (if complying with the labeling rules) is considered fortification. Foods and formulae for infants and young children, foods for specific medical purposes, and foods for total diet replacement often contain added micronutrients but are considered ineligible in this study because other legislation applies and EU 1169/2011 cannot be followed (4).

## 2.2 Selection of nutrients

To develop and test the automated procedure for identifying fortified foods, four nutrients were selected that can be added in multiple chemical forms. In the selection, we choose a mineral and a vitamin that can, according to the Dutch legislation, be added up to 100% of the RDA per reasonable daily consumption (calcium and vitamin B12) and a mineral and a vitamin for which addition is only allowed in lower amounts (folic acid) or specific cases [zinc is only allowed for the substitution or restauration purposes and in a specific type of menthol pastille and specific dairy products (32, 33)].

## 2.3 Procedure to identify fortified foods from the LEDA database

The European legislation on food labeling and the addition of micronutrients, as well as overarching Dutch legislation (15–17, 34, 35), was the basis for the procedure to identify if a branded food is fortified with the micronutrient(s) under study. The automated procedure is built as a decision tree (Figure 1) consisting of seven

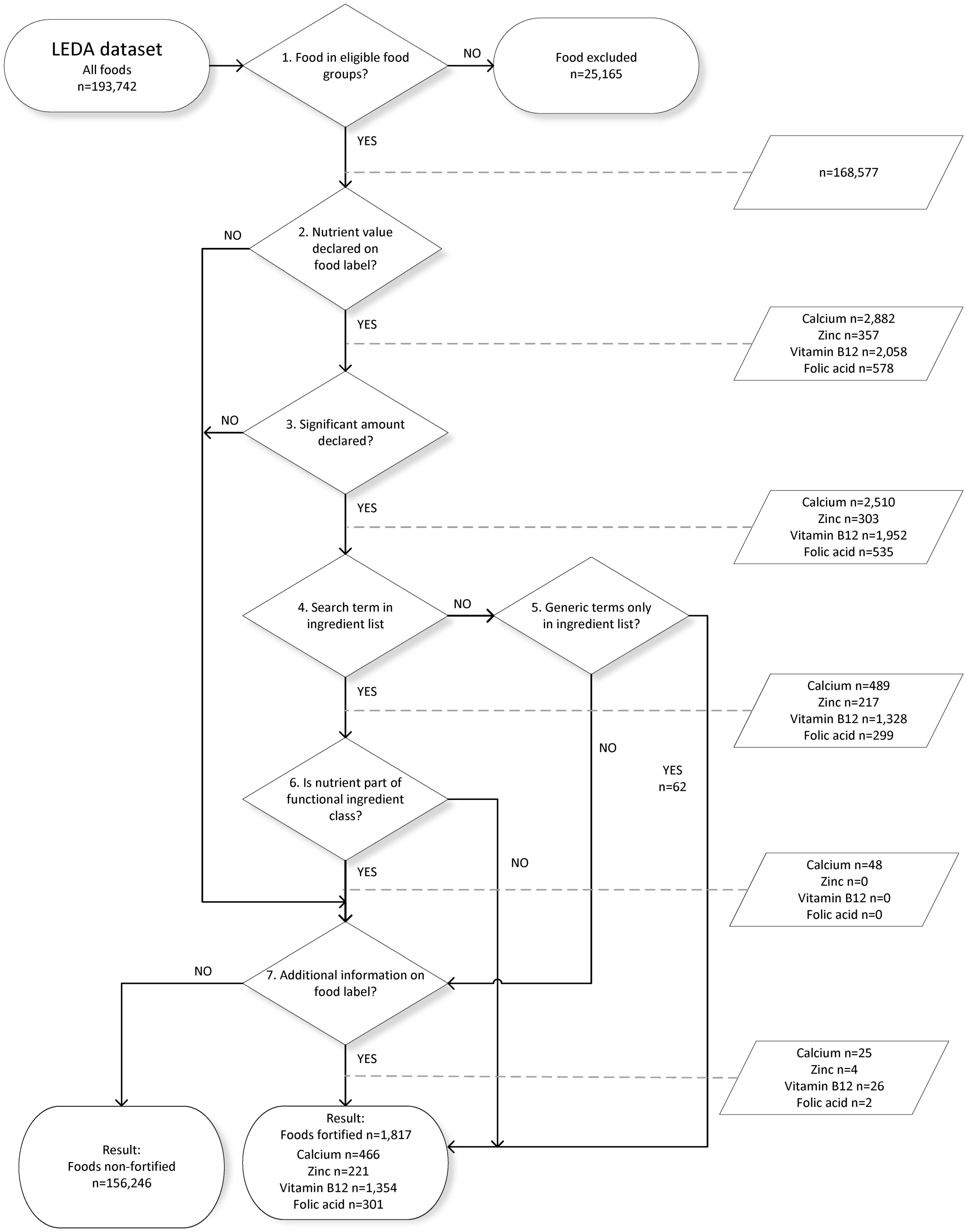



FIGURE 1
Decision tree with steps to identify fortified foods.

steps. In successive steps, foods are classified as ineligible, fortified, or non-fortified for each micronutrient (see Section 2.3.1). Currently, four nutrients are included, and the script can be extended to other micronutrients of interest if present in the database. For each micronutrient studied, relevant search terms need to be added. The decision tree was converted to a script in R (36).

The development of the automated procedure followed an iterative approach, particularly for the inclusion of the appropriate search terms. Vitamin formulation and mineral substances that may be added to foods were taken as the starting point (14), and ingredient lists were scrutinized for nutrient names, food additive names, and known synonyms. All search terms were in Dutch since this language is used in the LEDA database and were translated to English for this publication. Mixtures of substances are likely to be written as one word in Dutch, while in English, two separate words are used. Dutch search terms were treated as separate words by the script, resulting in all possible combinations with other compounds. The results of previous versions were verified and resulted in adaptations to improve the results. For example, searching for the wording "*fortified with*" and "*added*" not combined with specific search terms in the legal name and mandatory particulars (see EU legislation 1169/2011) in step 7 was removed as it yielded too many false positive results based on nutrients that were out of the scope.

### 2.3.1 Decision tree

#### 2.3.1.1 Step 1: Is the food eligible?

Food groups for which legislation other than EU 1925/2006 on the addition of vitamins and minerals and other substances applies, and thus the constraints of the generic EU labeling legislation 1169/2011 cannot be followed, are considered ineligible. These included foods and formulae for infants and young children, foods for specific medical purposes, and foods for total diet replacement for weight control. In addition, food supplements and foods not classified in a food group were considered ineligible. All other foods go to step 2. Selections are based on food group classifications used in the Dutch national branded food database LEDA.

#### 2.3.1.2 Step 2: Is a nutrient value declared on the food label?

Eligible foods with nutrient values reported for the selected nutrient(s) are classified as potentially fortified and go to step 3 for additional evaluation. When nutrient values are missing, foods go to step 7.

#### 2.3.1.3 Step 3: Is the declared nutrient value significant?

The nutritional values from the potentially fortified foods, identified in step 2, were checked for significance according to the EU labeling legislation (15). It was assumed that the indicated nutrient values refer to the food in the state as sold. The legislation allows for nutritional information on the label after preparation (e.g., cooking) if clearly indicated, but the LEDA database does not give this differentiation. The cutoff values for significance differ between beverages and non-beverages and single portions (see Table 1). The definition of a beverage is not given in the legislation. We considered as beverages all foods in the group of beverages (soft drinks, juices, lemonade, water, coffee, tea, and alcohol) as well as the following subcategories of other food groups: milk, chocolate milk, condensed milk, coffee milk/cream, buttermilk, dairy drinks, and liquid breakfast based on fruit juice or dairy. The script did not differentiate between non-beverage and single-portion foods. Values equal to the cutoff value were considered significant as the legislation does not indicate that the values need to be larger than this percentage. In the LEDA database, folic acid or dietary folate equivalents (DFE) values can be reported. We decided to assume that if DFE values were available, the best option would be to consider these as an indicator of folic acid fortification. No information was available on how food producers derived DFE values. If the declared nutrient value is significant, the food is considered potentially fortified and will be further evaluated in step 4. If not significant, the foods go to step 7.

#### 2.3.1.4 Step 4: Is the nutrient mentioned in the ingredient list?

The ingredient lists of all potentially fortified foods resulting from step 3 are searched for selected nutrients using the search terms as defined (Table 2). For foods with significant nutrient values but without any of the search terms in the ingredient list, go to step 5 for further evaluation. Foods with significant nutrient values and one of the search terms in the ingredient list were considered potentially fortified and go to step 6 for further evaluation.

#### 2.3.1.5 Step 5: Is the generic term vitamin(s) and/or mineral(s) included in the ingredient list?

When specific search terms were not found in the ingredient list, the next step was to search for generic terms such as *vitamin(s)/ mineral(s)/vitamin(s) and mineral(s),* not combined with any other micronutrient name in the ingredient lists (Table 2). EU labeling legislation (1169/2011) allows this generic wording when three or more micronutrients are added to the food. Foods with significant nutritional values for calcium, folic acid, vitamin B12, or zinc and one of these generic search terms in the ingredient list are classified as fortified. For the remaining foods, go to step 7 to check for additional information on the label.

#### 2.3.1.6 Step 6: Was the nutrient used as a food additive, or was the nutrient naturally present at high levels?

Micronutrients, in particular minerals, may also be used as part of food additives (e.g., stabilizers, emulsifiers, and acidity regulators)

TABLE 1 Significant values for selected nutrients according to EUR-Lex (15).

| Nutrient | Unit | DRI[a] | Significant amount for beverages[b] | Significant amount for non-beverages and single portion packages[c] |
|---|---|---|---|---|
| Folic acid | μg | 200 | 15 | 30 |
| Vitamin B12 | μg | 2.5 | 0.1875 | 0.3750 |
| Calcium | mg | 800 | 60 | 120 |
| Zink | mg | 10 | 0.75 | 1.50 |

[a]Daily reference intake for adults according to EU labeling legislation.
[b]7.5% of the nutrient reference values are supplied by 100 g or 100 mL in the case of beverages.
[c]15% of the nutrient reference values are supplied by 100 g or 100 mL in the case of products other than beverages or per portion if the package contains only a single portion.

instead of fortification. The European labeling legislation requires that food additive categories be mentioned in the ingredient list, and they need to be followed by the ingredient name, which may be a chemical structure that includes one of the nutrients of interest. Of note, the legislation does not consider fortificants as a food additive category. When nutrient names (calcium, zinc, vitamin B12, folic acid, or synonyms) in the ingredient list are combined with a food additive category, e.g., antioxidant or stabilizer (Table 3), this is considered a food additive, rather than a fortificant, and these foods go to step 7 for further evaluation. Similarly, when the nutrient name is mentioned in an additional remark within the ingredient list informing the consumer that the food is a *source of [search term]* or *rich in [search term],* the food is added to the list of non-fortified foods because *source of* and *rich in* are considered to represent the natural content of the nutrient. When the specific search term is not found in combination with a food additive category or a remark about a natural high content, the food is classified as fortified food.

#### 2.3.1.7 Step 7: Is additional information available on the food label?

Foods not identified as potentially fortified in steps 2, 3, 5, and 6 are checked for the selected search terms (Table 2) in the legal name or mandatory particulars of EU legislation 1169/2011. Foods for which selected search terms are found in combination with the wording *added* or *fortified* are identified as fortified foods. When the search terms are found in combination with the wording *source of [search term]* or *rich in [search term]*, the food is classified as non-fortified, as explained in step 6. Foods with significant nutrient values (step 3) but without any specific or generic search terms in the legal names or mandatory particulars are considered to contain natural amounts of the nutrients under study and are classified as non-fortified foods.

### 2.3.2 Applying the automated procedure to the LEDA dataset

A data file was extracted from the LEDA database (version LEDA_20220404) in CSV format with UTF-8 encoding and contained 193,742 food items. The following variables from the database were used: food group classification (as coded by the host organization), food name, ingredient list, nutrient name, nutrient values (calcium, folic acid or dietary folate equivalents, vitamin B12, and zinc), legal food name, and mandatory particulars as provided by the food producers. Data from the food producers were considered correct. For each nutrient, the automated procedure classified each food as fortified, non-fortified, or ineligible and stored detailed information about the outcome of each step in the decision tree.

Food groups most frequently fortified with the selected nutrients are illustrated in pie charts using the food group classification shown in Table 4. To highlight the most relevant food groups for fortification, food groups with less than five fortified foods were added to the group of miscellaneous foods.

## 2.4 Validation

Two validation steps were undertaken: a random validation for the entire procedure and a targeted validation for foods classified as fortified in step 5.

TABLE 2 Search terms[a] used (generic and for the 4 selected nutrients).

| Category | Dutch search term | English search terms[b,c,d,e,f,g] |
|---|---|---|
| Generic | Vitamine | Vitamin |
| | Vitaminen | Vitamins |
| | Vitamines | Vitamins |
| | Mineralen | Minerals |
| | Mineraal | Mineral |
| | Vitamines en mineralen | Vitamins and minerals |
| | Vitaminen en mineralen | Vitamins and minerals |
| | Vitamine en mineralen | Vitamin and minerals |
| Folic acid | Foliumzuur | Folic acid |
| | B9 | B9 |
| | B11 | B11 |
| | Folaat | Folate |
| | Tetrahydrofolaat | Tetrahydrofolate |
| | Polyglytamaat | Polyglytamate |
| | Pfteroylmonoglutaminezuur | Pfteroyl monoglutamic acid |
| | Folic acid | Folic acid |
| | Folinezuur | |
| | Foliumzout | |
| B12 | B12 | B12 |
| | Cobalamine | Cobalamin |
| | Cyanocobalamine | Cyanocobalamin |
| Zinc | Zink | Zinc |
| | Zinklactaat | Zinc lactate |
| | Zink lactaat | Zinc lactate |
| | Zinksulfaat | Zinc sulfate |
| | Zink sulfaat | Zinc sulfate |
| | Zinkoxide | Zinc oxide |
| | Zink oxide | Zinc oxide |
| | Zinkgluconaat | Zinc gluconate |
| | Zink gluconaat | Zinc gluconate |
| | Zinkcitraat | Zinc citrate |
| | Zink citraat | Zinc citrate |
| Calcium | Calcium | Calcium |
| | Calciumcarbonaat | Calcium carbonate |
| | Calciumfosfaat | Calcium phosphate |
| | Dicalciumfosfaat | Dicalcium phosphate |
| | Calciumlactaat | Calcium lactate |
| | Tricalciumcitraat | Tricalcium citrate |
| | Calciumcitraat | Calcium citrate |
| | Calciumzouten van orthofosforzuur | Calcium salt of orthophosphoric acid |

*(Continued)*

TABLE 2 (Continued)

| Category | Dutch search term | English search terms[b,c,d,e,f,g] |
|---|---|---|
| | Dicalciumdicitraat | Dicalcium dicitrate |
| | Calciumhydroxide | Calcium hydroxide |

[a]All search terms were in Dutch and were translated into English for the purpose of this publication.
[b]Capitals in the text are neglected.
[c]Mixtures of substances are likely to be written as one word in Dutch. The script treated the Dutch search terms as separate words, resulting in all possible combinations, e.g., calcium and zinc with other compounds.
[d]Combinations of calcium with another compound may be in the ingredient lists as added nutrients or as functional ingredients.
[e]Foods with minerals listed in combination with one of the functional ingredient classes are searched for and classified as non-fortified. For functional ingredient classes, see Table 3.
[f]Specific vitamins and minerals are also searched in combination with the wording mineral, vitamin and vitamin and mineral using the generic search terms for these and including delimiters as.
[g]Calcium is excluded when used as calcium-D-pantothenate, indicating pantothenic acid.

Random validation. A sample of 500 foods was taken for validation. The sample consisted of four random samples of 100 foods classified with each of the four nutrients and a random sample of the other foods (non-fortified or non-eligible foods). The results were validated by two experienced dietitians. They determined whether each food in the sample was eligible and, if so, whether it was fortified or not. The experts received specific instructions and written documentation with background information on the constraints of the EU labeling legislation. They were not aware of the outcome of the automated procedure and were not informed of the details of the decision tree. The experts worked independently from each other. The experts' results were compared, and they were asked to reach a consensus on those foods with discrepancies in classification. The discrepancies were caused by uncertainty about whether foods were classified in the correct food groups (e.g., infant formula classified as milk product) and incorrect, unclear, inconsistent, or incomplete information in the dataset. For each of the four nutrients, a two-way contingency table was created with classes fortified, non-fortified, and ineligible foods assessed by the experts and the automated procedure. For all proportions, simultaneous 95% confidence intervals for multinomial proportions according to the methods of Sison and Glaz were calculated (37). Sensitivity and specificity were included, indicating the probability that the automated procedure correctly returned, respectively, fortified foods (true positive rate) and non-fortified foods (true negative rate). Statistical analyses were conducted in R (36).

Targeted validation. For the complete LEDA dataset, all foods were classified as fortified because a generic search term was found in the ingredient list in step 5 and was manually checked by an expert.

## 3 Results

### 3.1 Fortified foods in the LEDA dataset

For each step, the number of foods that are potentially fortified is reported in Figure 1. The final number of foods fortified with a specific nutrient can be derived by subtracting the number of "yes" for step 6 from the number of yes from step 4 and adding the number of yes from step 7. Table 5 shows the coverage of variables in the LEDA dataset and the results of the automated procedure to identify fortification with calcium, zinc, vitamin B12, or folic acid. Coverage gives the proportion of foods for which the information is available in LEDA. With 92.5%, ingredient information is considered complete for several food groups (e.g., fresh meat, fruit, and vegetables), and this information is non-mandatory. Food group classification was not yet fully added by the hosting organization (13% missing).

The automated procedure identified 1,817 foods as fortified with one or more of the selected nutrients (0.94% of all 193,742 foods). This total percentage does not reflect the full spectrum of fortification in the LEDA database since only four nutrients were included in this study. The results per selected nutrient varied between 0.11 and 0.70% of all foods. The numbers of fortified foods per food group are shown in Figures 2A–D. For calcium fortification, dairy products and snacks (sweet snacks such as biscuits, ice cream, and sweets) were the main food groups. For folic acid fortification, cereal products were the most important food group, followed by drinks, snacks, and fats (margarine-type products). Vitamin B12 was most frequently added to meat replacers, followed by cereal products, drinks, and dairy products. Zinc was most frequently added to meat replacers (allowed in case of substitution) and to drinks (not allowed unless an exemption is given).

TABLE 3 Functional ingredient class names, according to EUR-Lex (15).

| Functional ingredient class | Functional ingredient class continued |
|---|---|
| Acid | Foaming agent |
| Acidity regulator | Gelling agent |
| Anti-caking agent | Glazing agent |
| Anti-foaming agent | Humectant |
| Antioxidant | Modified starch |
| Bulking agent | Preservative |
| Color | Propellent gas |
| Emulsifier | Raising agent |
| Emulsifying salts | Sequestrant |
| Firming agent | Stabilizer |
| Flavor enhancer | Sweetener |
| Flour treatment agent | Thickener |

TABLE 4 Food groups in the LEDA database used to identify fortified foods.

| Food group | Food group continued |
|---|---|
| Bread | Milk, milk products, and milk replacers |
| Bread filling | Miscellaneous |
| Cereals and cereal products | Nuts and seeds |
| Cheese and cheese substitutes | Oils and fats |
| Composite meals | Potatoes and other tubers |
| Drinks | Pulses |
| Eggs | Sauces |
| Fish, shellfish, crustacean | Snacks (sweet and savory) |
| Fruit | Soup |
| Meat replacers | Vegetables |
| Meat, cold cuts, and poultry | |

TABLE 5 Coverage in LEDA dataset and results of the automated procedure to identify fortification with selected nutrients.

| Variable description | Data type | Coverage | | Fortified | |
|---|---|---|---|---|---|
| | | *n* | % | *n* | %[a] |
| Total number of foods | | 193,742 | 100% | | |
| Food group classification | Text | 168,577 | 87.0% | | |
| Ingredient list | Text | 179,243 | 92.5% | | |
| Legal food name | Text | 86,254 | 44.5% | | |
| Mandatory particulars | Text | 18,482 | 9.5% | | |
| Calcium values (mg) | Number | 3,741 | 1.9% | 466 | 0.24% |
| Folic acid + DFE values (μg) | Number | 963 | 0.5% | 301 | 0.16% |
| Vitamin B12 values (μg) | Number | 2,975 | 1.5% | 1,354 | 0.70% |
| Zinc values (mg) | Number | 824 | 0.4% | 221 | 0.11% |
| Total number of foods fortified with one or more of the selected nutrients (calcium, folic acid, vitamin B12, or zinc) | | | | 1,817 | 0.94% |

[a]The percentages were calculated with the total number of foods (*n* = 193,742) as the denominator.

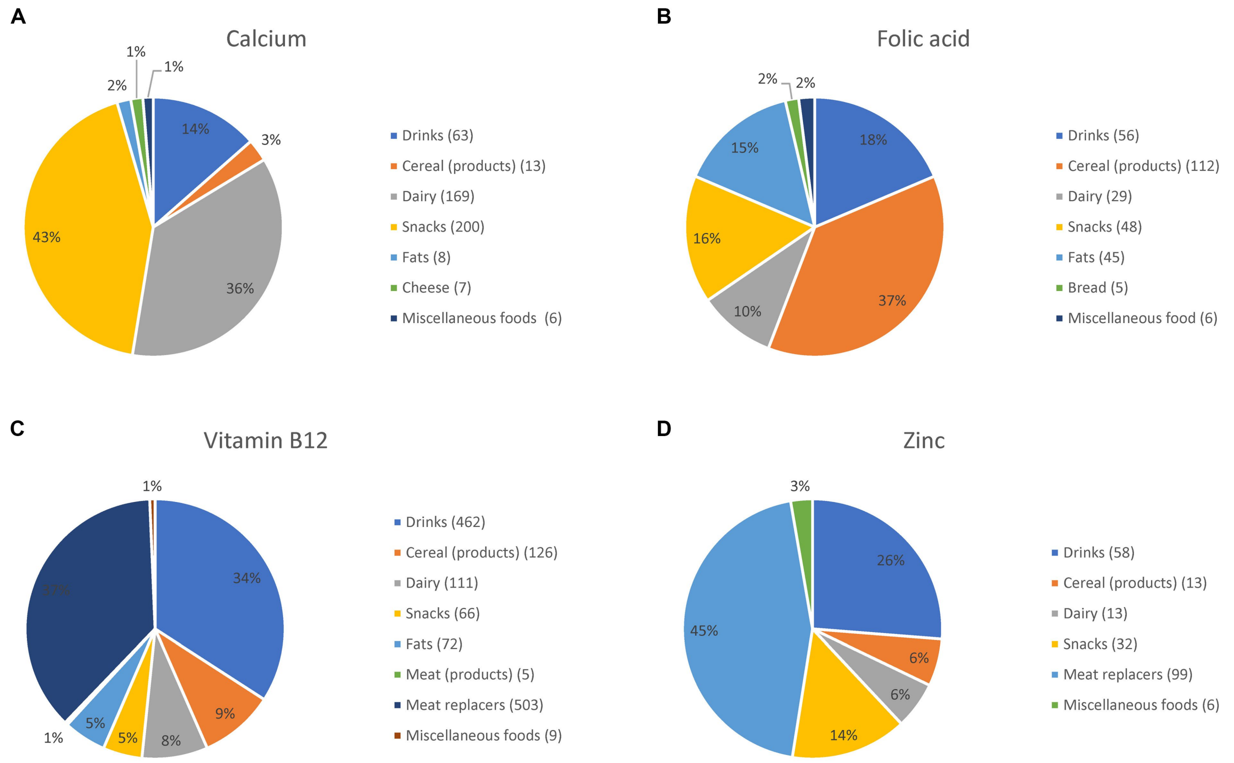



FIGURE 2
Foods fortified with calcium **(A)**, folic acid **(B)**, vitamin B12 **(C)**, and zinc **(D)** per food group; the number of fortified foods per food group is given in brackets.

## 3.2 Validation of results

The comparison of the results from experts and the automated procedure for 500 randomly selected foods is shown in Table 6.

Classification of fortified and non-fortified foods by experts and the automated procedure agreed for more than 94% of the foods included. The large number of foods fortified with vitamin B12 can be explained by the large proportion of sampled foods fortified with folic acid or zinc that were also fortified with vitamin B12 (89 of 117 folic acid-fortified foods and 77 of 103 zinc-fortified foods). In the total validation sample for each nutrient, the percentages of false positive and false negative results by the automated procedure were small (0–4%), with calcium producing the most false-positive results (2.6% with 95% CI 0.8–5.0%). The false-negative results ranged from 3.6% (with 95% CI 0.9–6.7%) for folic acid, 3.8% (with 95% CI 0.9–7.7%) for calcium, to 4% (with

TABLE 6 Differences between experts and automated procedures for random samples of 400 fortified and 100 non-fortified foods in the LEDA dataset based on the final script.

| Nutrients | | Total | Procedure: fortified | Procedure: non-fortified | Procedure: non-eligible |
|---|---|---|---|---|---|
| | | *n* | *n* (%) confidence interval (CI) | *n* (%) confidence interval (CI) | *n* (%) confidence interval (CI) |
| Calcium | Experts: fortified | 106 | 101 (95.3%) CI: 92.5–99.2% | 4 (3.8%) CI: 0.9–7.7% | 1 (0.9%) CI: 0.0–4.9% |
| | Experts: non-fortified | 379 | 10 (2.6%) CI: 0.8–5.0% | 357 (94.2%) CI: 92.4–96.5% | 12 (3.2%) CI: 1.3–5.5% |
| | Experts: non-eligible | 15 | 12 (80.0%) CI: 66.7–100.0% | 0 (0.0%) CI: 0.0–21.4% | 3 (20.0%) CI: 6.7–41.4% |
| Folic acid | Experts: fortified | 111 | 107 (96.4%) CI: 93.7–99.4% | 4 (3.6%) CI: 0.9–6.7% | 0 (0.0%) CI: 0.0–3.1% |
| | Experts: non-fortified | 374 | 1 (0.3%) CI: 0.0–2.1% | 360 (96.3%) CI: 94.7–98.1% | 13 (3.5%) CI: 1.9–5.3% |
| | Experts: non-eligible | 15 | 9 (60.0%) CI: 40.0–85.8% | 3 (20.0%) CI: 0.0–45.8% | 3 (20.0%) CI: 0.0–45.8% |
| Vitamin B12 | Experts: fortified | 253 | 242 (95.7%) CI: 93.7–98.1% | 10 (4.0%) CI: 2.0–6.4% | 1 (0.4%) CI: 0.0–2.9% |
| | Experts: non-fortified | 232 | 0 (0.0%) CI: 0.0–2.9% | 220 (94.8%) CI: 92.7–97.8% | 12 (5.2%) CI: 3.0–8.1% |
| | Experts: non-eligible | 15 | 9 (60.0%) CI: 40.0–85.8% | 3 (20.0%) CI: 0.0–45.8% | 3 (20.0%) CI: 0.0–45.8% |
| Zinc | Experts: fortified | 91 | 91 (100.0%) CI: 100.0–100.0% | 0 (0.0%) CI: 0.0–1.8% | 0 (0.0%) CI: 0.0–1.8% |
| | Experts: non-fortified | 394 | 0 (0.0%) CI: 0.0–1.6% | 381 (96.7%) CI: 95.2–98.3% | 13 (3.3%) CI: 1.8–4.9% |
| | Experts: non-eligible | 15 | 12 (80.0%) CI: 66.7–100% | 0 (0.0%) CI: 0.0–21.4% | 3 (20.0%) CI: 6.7–41.4% |

95% CI 2.0–6.4%) for vitamin B12. In the case of calcium, 9 out of 10 false positive results could be explained by the natural calcium content of mineral water that was mentioned (including the word calcium) in the ingredient list. It was more difficult to explain the relatively high percentage of false-negative results for vitamin B12 and folic acid. The most likely explanation is that experts made the wrong decision in cases where vitamin B12 was declared in the ingredient list and the nutritional panel, but the amount was below the level of significance. This was the case for meat substitutes, which are likely to be fortified or substituted with vitamin B12. For folic acid, in 3 of 4 cases, the decision of experts was correct, and the source of error in the automated procedure could not be identified. In the fourth case, the automated procedure and experts disagreed on whether the food was a beverage. Overall, sensitivity and specificity, indicating the probability that the automated procedure correctly classified foods as fortified and non-fortified, were high for all four nutrients. Sensitivity ranged from 96.0 to 100.0%, with all lower limits of 95% confidence intervals more than 92.5%. Specificity ranged from 94.2 to 96.6%, with all lower limits of 95% confidence intervals more than 92.4% (Table 6).

Most discrepancies between the automated procedure and experts are related to the classification of non-eligible food groups. Experts considered 15 foods as ineligible (3%) (foods and formulae for infants and young children, foods for specific medical purposes, and foods for total diet replacement), of which 12 were incorrectly classified as another eligible food group in the LEDA database. The automated procedure considered 16 foods as ineligible (3.2%), of which 13 foods were not classified in any food group in the LEDA database, whereas the experts concluded that based on the available information, the foods were eligible.

A targeted validation was done for all foods in the complete database classified as fortified based on step 5. Step 5 yielded 62 foods that were classified as fortified based on declared significant nutrient values without specific search terms but with generic search terms such as *vitamin(s)* and/or *mineral(s)* in the ingredients list. Manual checking showed that 35 of these foods were correctly classified as fortified and 27 were not, although, in some situations, information was not fully clear (e.g., incomplete ingredient lists and unexpected wording such as B(1)(2) instead of B12, that were not included as search terms).

## 4 Discussion

### 4.1 Main findings

A seven-step decision tree, aligned with European food labeling legislation, was developed to identify fortified foods in the Dutch branded food database LEDA. Steps were integrated into a script for automated application using relevant search terms. When label data correctly follows the constraints of the labeling legislation, the automated procedure successfully identifies if foods are fortified or not. Nearly 1% of the foods in the LEDA database were fortified with one to four of the micronutrients studied (calcium, folic acid, vitamin B12, and zinc). Validation showed over 94% agreement between the automated procedure and experts to identify fortified and non-fortified foods. The percentage of false-positive or false-negative results compared to the expert opinion was low (0–4%). Calcium produced the most false-positive results and vitamin B12 the most false-negative results. The remainder of the disagreements between the script and experts were for foods considered ineligible by either the script or experts (about 3%).

### 4.2 Challenges

#### 4.2.1 Data

Working with the LEDA dataset showed that identifying fortified foods is not straightforward due to the lack of specific variables to indicate fortification, complex labeling legislation, and the addition of micronutrients as food additives rather than fortification. Most challenges to developing an automated procedure were found in the LEDA data and related to the large variation in wording (including typing errors), the structure of ingredient lists, incomplete ingredient lists, wrong or missing food group classifications, data not fully in line with the European labeling legislation, and the difficulty to capture all optional search terms used in ingredient lists, legal name, and mandatory particulars. Some ingredient data were not complete or seemed to be truncated during data transfer from the food producer to the LEDA database, as complete information could be found on the food producers' websites. As a result, some foods could not be correctly identified as fortified or non-fortified. It needs to be noted that food producers are responsible for providing complete and correct label data for the LEDA database but not for assigning the correct food classification.

Nutritional values in the LEDA dataset were supposed to be correct, but checking values was not the purpose of this study. However, errors may occur and have an impact as the values are checked for significance according to the European labeling legislation. This can be exemplified by values that were 1,000-fold too high due to decimal point or unit errors. Errors could also be related to nutritional information before and after cooking/preparing. The LEDA dataset only contained one set of nutritional values, which was assumed to represent the food as sold. However, information on nutritional composition before and after preparation is expected to become available in the LEDA database.

### 4.2.2 Validation

For some foods, the experts' classifications were different from the automated procedure due to their ability to combine data differently. In addition, experts had access to further information, e.g., on product websites, and this explains some of the false-positive or false-negative results.

When data aligned with labeling rules, steps 1 to 4 and 6 of the decision tree worked well. When data were less clear, steps 5 and 7 were needed. Considering all options used on labels was impossible due to the large variation in structure and wording. Currently, when generic terms such as *vitamin(s)* and/or *mineral(s)* are used in the ingredient list, legal name, or mandatory particulars, this sometimes refers to nutrients that are not in the scope of this study. This explains some of the errors found when manually checking 62 results (if yes) from step 5 and implies that the approach cannot currently be fully automated unless misclassifications are accepted. Including all nutrients that may be added is expected to limit this problem, as the generic search terms found in steps 5 and 7 will then refer to at least one of the nutrients added. As legislation may change and allow for adding other compounds containing micronutrients, updating the search terms will be needed. The option to better distinguish between the nutritional composition of raw and prepared versions of food is also expected to improve the results of the automated procedure.

Foods with significant natural levels of calcium, folic acid, vitamin B12, or zinc, and one of the generic search terms [vitamin(s)/mineral(s)] in the ingredient list due to the addition of other micronutrients would be classified as fortified if no other details were present. However, this combination did not occur in the LEDA dataset.

### 4.2.3 Nutrients

The European labeling legislation states that nutritional values for vitamins and minerals may only be declared if significant, regardless of whether they are added or naturally present. The LEDA database also contains insignificant values provided for the database but are not shown on the food label. In combination with other information regarding added vitamins or minerals, this can complicate the decision as to whether the food should be classified as fortified, especially when judged by humans, who may tend to deviate from strict rules.

#### 4.2.3.1 Calcium

Calcium, added to fortify food, can either be mentioned in the ingredient list as *calcium* or as one of many chemical forms, e.g., *calcium carbonate, dicalcium phosphate, calcium lactate, calcium acetate,* and *calcium propionate*. Calcium is also often added for technological reasons as part of a food additive, e.g., thickener or preservative, representing various chemical forms, or it may be part of the chemical structure of vitamins, such as calcium-D-pantothenate. Food additive category names need to be declared, followed by the name of the food additive. When the food additive category is erroneously missing from the ingredient list, the food will be identified as fortified with this component if the amount is significant.

#### 4.2.3.2 Folic acid

*Folic acid* was mostly listed in the ingredient lists as such or as *vitamin B9, B11,* or *pteroylmonoglutamic acid*, and it was not found as part of food additives. According to the EU labeling legislation, folic acid may be added to foods as the synthetic form of folate or calcium-L-methylfolate. Although the total amount present in the food needs to be declared as significant, the EU legislation does not give information on how to deal with bioactivity levels of natural and synthetic forms (1.7 * natural folate). As a result, it is unclear if food producers sum folic acid and natural folate with or without conversion factors to calculate total folate of folate activity or if they only declare the amount of the added folic acid. Due to this unclearness, our procedure may interpret the significance of the values incorrectly and draw incorrect conclusions about the fortification. The US Nutrition and Supplement Facts are more clear, stating that folate and folic acid need to be declared as dietary folate equivalents (DFE) on food labels (38). Clear instructions in the European labeling legislation, as exemplified in the US, would solve this problem.

#### 4.2.3.3 Vitamin B12

Vitamin B12 can be found in the ingredient lists as *B12* or *vitamin B12* or in wordings including *cobalamin or cyanocobalamin*. Vitamin B12 was not detected as part of any food additive. Foods expected to be fortified with vitamin B12, such as meat substitutes, but with an insignificant nutritional value were confusing for experts.

#### 4.2.3.4 Zinc

Zinc is listed in the ingredient lists as *zinc* or in combination with other chemical compounds, e.g., *zinc lactate, zinc gluconate,* and *zinc sulfate*. Zinc was not found as part of food additives. In the Netherlands, adding zinc is allowed for restoration or substitution purposes only (4), and in that situation, it is not obliged to declare zinc in the ingredient list. However, it is not clear from food labels if nutrients are added for restoration, substitution, or fortification, and the automated procedure cannot make this distinction either. Although a limited number of exemptions to fortify foods with zinc are valid in the Netherlands, we found zinc added in significant amounts for many more foods and brands. Zinc added to meat substitutes may be added to substitute zinc as present in meat. Most drinks with added zinc were lemonade and vitamin water, and restoration or substation did not seem to be the reason for adding it.

## 4.3 Strengths and limitations

The main strength of our automated procedure to identify fortified foods is that it can be run quickly and as frequently as needed, with over 90% correct classification for the selected micronutrients in the validated sample. This is important because identifying fortification manually, food by food, is very time-consuming due to the size of the LEDA database and its rapid changes. In the Slovenian branded food dataset, 80% of the foods had disappeared from the market between 2011 and 2020 (27) and there is no reason to expect differently for the Netherlands. The script can easily be adapted to include additional micronutrients by adding or replacing search terms. The decision tree and the search terms can be re-used, and the script can be adapted for datasets other than the LEDA dataset. Other possible extensions of the script include estimating the amount of added micronutrients and distinguishing between the different chemical compounds of a fortificant, which can be useful for estimating bioavailability.

Another strong point was that validation was done in duplicate. Two experts independently evaluated all sampled foods and used expert opinion as the reference. Experts were flexible in combining information from multiple variables and could use additional information. On the contrary, experts were also in doubt in some cases due to the confusing, incomplete, or inconsistent information not fully complying with the rules of legislation and may have taken incorrect decisions, as exemplified by the results for vitamin B12.

A limitation of the automated procedure is that including all optional search terms is almost impossible. Although we took vitamin formulations and mineral substances that can be added to foods as a starting point and supplemented this with additional terms discovered while carefully examining ingredient lists for nutrient names, we cannot rule out the possibility that we missed some search terms. It seems more likely that we missed discovering a fortificant due to typing errors in the ingredient list. Based on the small percentages of differences between the automated procedure and expert judgment, this impact proved limited. Furthermore, fortified foods were oversampled for validation because finding fortification was expected to be more difficult than finding non-fortification. The limited sample for validation and the small number of nutrients studied only allow conclusions for the nutrients under study. Nonetheless, several options for improvement of the data and the automated procedure were identified.

## 4.4 Usability and quality of the automated procedure

Researchers may need information on fortification for food policy development, food fortification strategies, and enforcement of fortification legislation. Other use cases are personalized dietary advice and public health information. Although the incidence of fortifications in the LEDA database is low, frequent consumption of fortified food will greatly impact individual nutrient intake. The need for details may depend on the intended use; however, complete, correct, and up-to-date information on fortification, including the amounts added, is important for all users.

The validation showed that agreement between experts and the automated procedure was high (>94%), and the percentage of false-positive or false-negative results was low (0–4%). In our opinion, the automated procedure can be used to identify if branded foods are fortified. Even though there are some uncertainties and possible errors, this procedure can significantly reduce the amount of manual work needed to a manageable level. In specific situations, users may want to apply additional data-checking steps. Suggestions for improvement are mentioned under data challenges and recommendations.

Per single portion package, the European labeling legislation uses the same levels of significance as for non-beverages (15% of RDI). Calculated per 100 g or ml of food, as in LEDA, for portion sizes smaller than 100 g or ml, this results in nutritional values higher than the level of significance; for portion sizes larger than 100 g or ml, this is the other way around. Furthermore, the European labeling legislation is not fully clear if, for single-portion beverages, the RDI of 15% also applies instead of 7.5%. These limitations were not considered, as all nutritional information in the LEDA database is given per 100 g or 100 mL of food. In case of any errors in the database, the level of significance of the values may have been misinterpreted, with single-portion packages of drinks assigned higher thresholds for significance than other drinks. The LEDA dataset contained data on 113 beverages listed as single-portion packages, of which 7 were classified as fortified. The impact of possible errors was small.

Ideally, the branded food database does not contain errors in nutritional values. Automated validation, e.g., on outliers, would help to correct values during data entry. Moreover, the nutritional value in the database may deviate from the real value because food producers often add a higher amount of micronutrients to allow for losses during processing and shelf life (28, 39). Not knowing the actual amounts added or remaining and, in some cases, using incorrect values is a challenge for our approach.

In addition to the levels of significance in the EU labeling legislation, maximum fortification levels are needed to secure a safe level of intake. In Europe, maximum levels have not yet been defined, and national legislation prevails. For example, a general exemption is given in the Netherlands to fortify with folic acid to a maximum level of 100 µg/100 kcal. Checking if nutrient levels remain within the maximum level allowed is not included in the automated procedure; however, this can be monitored using the results of our approach.

If an ingredient list declares a fortified ingredient, it depends on the nutrient value (significant or not) and whether the food is considered fortified by the automated procedure. An example is the ingredient wheat flour enriched with iron, folic acid, and niacin. In the LEDA database, none of the foods containing this ingredient were classified as fortified with folic acid due to insignificant levels caused by "dilution" by other ingredients.

## 4.5 Recommendations

To improve the usability of the LEDA data to identify all fortified foods, the coverage of 75% of the food supply needs to be extended by creating liaisons with new data providers. Food group classification in the LEDA database must be completed for each food to allow the identification of fortification for all foods. Control procedures are needed to ensure that ingredient lists are uploaded without any missing information. Additional and, if possible, mandatory variables to mark fortifications at food and nutrient levels are needed both at the data provider side and in the LEDA software to allow for easier identification. Such variables could be fortified yes/no at the food level and fortified yes/no for each individual nutrient, with multiple choice options for the allowed chemical forms of food fortificants. Ideally, this would make the current approach redundant. Harmonized formats and control steps for data entry will help the food industry to improve data quality. The same applies to instructions on how to deal with substitution versus fortification on the food label. Lessons can be learned from the USDA Global Branded Food Products Database, where several so-called hard validations are in place during data entry, and if not met, further data entry is impossible (25).

To allow complete identification of fortified and non-fortified foods, all micronutrients that may be added to foods need to be added to the script. Moreover, an updated version of the script could consider significant values for single-portion foods if legal considerations are clearer.

The European labeling legislation can be further improved by (a) giving clearer information on the definition of beverages and the use of DRI for single portion packages, (b) requesting detailed information on fortification per component using dedicated variables on the food label, (c) making added micronutrients for fortification a food additive category, for which the class name needs to precede the list of individual micronutrients added, and (d) providing detailed instructions on how to define and declare the added micronutrients, in particular when conversion factors related to bioavailability are available as for folic acid and dietary folate equivalents.

## 4.6 Conclusion

A step-wise approach, including a decision tree to define if foods in the Dutch national branded food database LEDA are fortified, was developed and applied. Validation by experts showed that agreement between experts and the automated procedure was above 94% for each nutrient, and the percentages of false-positive and false-negative results were limited.

When the food label correctly followed the EU labeling legislation, the automated procedure was able to identify fortification correctly. For some foods, missing information on the label (in ingredient lists or nutritional values) led to false negative results compared to classification by experts. Inconsistent ways of presenting information on the label (wording, brackets, etc.) make it difficult to include all optional search terms, and more standardization on labels is expected to lead to better results.

To include all micronutrients that may be added to foods, the script needs to be extended. This is expected to improve results as any notification of vitamins or minerals added will then refer to one of the nutrients included in the search terms.

Considering the limitations posed by the unclear legislation and label information, this automated procedure allows quick identification of fortifications present in branded foods in the Netherlands.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: data cannot be made publicly available because of contractual obligations with the data providers. Requests to access these datasets should be directed to IT, ido.toxopeus@rivm.nl.

## Author contributions

SW: Conceptualization, Methodology, Writing – original draft. CT: Conceptualization, Data curation, Writing – review & editing. IT: Data curation, Writing – review & editing. JV-K: Writing – review & editing. EF: Supervision, Writing – review & editing. MO: Conceptualization, Data curation, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is funded by the Dutch Ministry of Health, Welfare and Sport.

## Acknowledgments

The authors acknowledge Agnes Roos and Annick Lindeboom for performing the validation on the random sample of branded foods, Hugo Albers for the artwork in Figure 1, and Mark Roe for English proofreading and correcting the English spelling. The National Nutrition Centre (Voedingscentrum) is acknowledged for collecting and maintaining the LEDA data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References


1. Tam E, Keats EC, Rind F, Das JK, Bhutta ZA. Micronutrient supplementation and fortification interventions on health and development outcomes among children under-five in low-and middleincome countries: a systematic review and meta-analysis. *Nutrients*. (2020) 12:289. doi: 10.3390/nu12020289

2. Mkambula P, Mbuya MNN, Rowe LA, Sablah M, Friesen VM, Chadha M, et al. The unfinished agenda for food fortification in low-and middle-income countries: quantifying progress, gaps and potential opportunities. *Nutrients*. (2020) 12:354. doi: 10.3390/nu12020354

3. Pachón H, Reynolds B, Duong M, Tsang BL, Childs L, Luthringer CL, et al. The potential contribution of fortified maize flour, oil, rice, salt, and wheat flour to estimated average requirements and tolerable upper intake levels for 15 nutrients in 153 countries. *Nutrients*. (2021) 13:1–14. doi: 10.3390/nu13020579

4. de Jong MH, Nawijn EL, Verkaik-Kloosterman J. Contribution of voluntary fortified foods to micronutrient intake in the Netherlands. *Eur J Nutr*. (2022) 61:1649–63. doi: 10.1007/s00394-021-02728-4

5. Newman JC, Malek AM, Hunt KJ, Marriott BP. Nutrients in the US diet: naturally occurring or enriched/fortified food and beverage sources, plus dietary supplements: NHANES 2009-2012. *J Nutr*. (2019) 149:1404–12. doi: 10.1093/jn/nxz066

6. Hennessy A, Walton J, Flynn A. The impact of voluntary food fortification on micronutrient intakes and status in European countries: a review. *Proc Nutr Soc*. (2013) 72:433–40. doi: 10.1017/S002966511300339X

7. Bruins MJ, Mugambi G, Verkaik-Kloosterman J, Hoekstra J, Kraemer K, Osendarp S, et al. Addressing the risk of inadequate and excessive micronutrient intakes: traditional versus new approaches to setting adequate and safe micronutrient levels in foods. *Food Nutr Res*. (2015) 59:26020–10. doi: 10.3402/fnr.v58.26020

8. Hennessy A, Hannon EM, Walton J, Flynn A. Impact of voluntary food fortification practices in Ireland: trends in nutrient intakes in Irish adults between 1997-9 and 2008-10. *Br J Nutr*. (2015) 113:310–20. doi: 10.1017/S0007114514003651

9. Kloosterman J, Bakker MI, De Jong N, Ocké MC. Framework for intake simulation of functional ingredients. *Public Health Nutr*. (2008) 11:279–87. doi: 10.1017/S1368980007000316

10. Kloosterman J, Fransen HP, De Stoppelaar J, Verhagen H, Rompelberg C. Safe addition of vitamins and minerals to foods: setting maximum levels for fortification in the Netherlands. *Eur J Nutr*. (2007) 46:220–9. doi: 10.1007/s00394-007-0654-y

11. Krušič S, Hribar M, Hafner E, Žmitek K, Pravst I. Use of branded food composition databases for the exploitation of food fortification practices: a case study on vitamin D in the Slovenian food supply. *Front Nutr*. (2022):8. doi: 10.3389/fnut.2021.775163

12. Verkaik-Kloosterman J, Beukers M, Buurma-Rethans E, Verhagen H, Ocké MC. Evaluation of the Dutch general exemption level for voluntary fortification with folic acid. *Food Nutr Res*. (2012):56. doi: 10.3402/fnr.v56i0.5443

13. Verkaik-Kloosterman J, McCann MT, Hoekstra J, Verhagen H. Vitamins and minerals: issues associated with too low and too high population intakes. *Food Nutr Res*. (2012):56. doi: 10.3402/fnr.v56i0.5728

14. EUR-Lex. Regulation (EC) no 1925/2006 of the European Parliament and of the Council of 20 December 2006 on the addition of vitamins and minerals and of certain other substances to foods. Retrieved April 2020. Available at: https://eur-lex.europa.eu/homepage.html, L404 (2006).

15. EUR-Lex. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers. Available at: https://eur-lex.europa.eu/homepage.html, retrieved April 2020, (2011).

16. VWS. Convenant Vitaminering van Smeerbare vetproducten (covenant vitamisation of spreadable fat products), Ministry of Public Health and sports and spreadable fat producers. *Staatscourant*. (2000) 17:12.

17. VWS. Convenant gebruik bakkerszout (covenant use bakery salt), Ministry of Public Health and sports and Dutch bakery organisations. *Staatscourant*. (2008) 209:1–3.

18. Euro FIR. Euro FIR, the European food information resource. Available at: https://www.eurofir.org/food-information/, retreived December 2022 (2022)

19. DFI. Compilers' toolbox. Danish Food Information. Avilable at: http://toolbox.foodcomp.info/ToolBox.asp, retrieved December 2022. (2022)

20. Menard C, Dumas C, Goglia R, Spiteri M, Gillot N, Combris P, et al. OQALI: a French database on processed foods. *J Food Compos Anal*. (2011) 24:744–9. doi: 10.1016/j.jfca.2010.09.001

21. NUBEL. InterNubel, Database of trade names, Belgium. Available at: http://www.internubel.be/, retrieved May 2022 (2022)

22. USDA. USDA Global Branded Food Products Database Enhancing Information Transparency and Safeguarding Data Quality. Retrieved May 2020. Available at: https://fdc.nal.usda.gov/docs/USDA_BFPD_DataQuality_Aug2019.pdf. (2019)

23. GS1. GS1 Data Source for the exchange of product data. Available at: https://www.gs1.nl/en/industries/reliable-product-data-food-health-and-beauty/gs1-data-source-for-the-exchange-of-product, retrieved March 2020. (2020).

24. Mintel. Mintel global new products database. Available at: https://www.mintel.com/global-new-products-database, retrieved March 2020. (2020)

25. Larrick B, Kretser A, McKillop K. Update on "a Partnership for Public Health: USDA global branded food products database". *J Food Compos Anal*. (2022) 105:104250. doi: 10.1016/j.jfca.2021.104250

26. Perrin C, Battisti C, Chambefort A, Digaud O, Duplessis B, Volatier JL, et al. A comparison of the nutritional content of processed foods available on the French market, according to the type of brand, and potential impact on nutrient intakes—an Oqali study. *Food Sci Nutr*. (2018) 6:1410–21. doi: 10.1002/fsn3.655

27. Pravst I, Hribar M, Žmitek K, Blažica B, Koroušić Seljak B, Kušar A. Branded foods databases as a tool to support nutrition research and monitoring of the food supply: insights from the Slovenian composition and labeling information system. *Front Nutr*. (2022):8. doi: 10.3389/fnut.2021.798576

28. Flynn A, Kehoe L, Hennessy Á, Walton J. Estimating safe maximum levels of vitamins and minerals in fortified foods and food supplements. *Eur J Nutr*. (2017) 56:2529–39. doi: 10.1007/s00394-016-1288-8

29. Allen RE, Dangour AD, Tedstone AE. Update of the vitamin D content of fortified foods and supplements in the UK National Diet and nutrition survey nutrient databank. *Nutr Bull*. (2014) 39:247–52. doi: 10.1111/nbu.12099

30. Martin CL, Clemens J, Moshfegh AJ. Items designated as fortified: food and nutrient database for dietary studies (FNDDS), 2015–2016. *J Food Compos Anal*. (2019) 84:103255. doi: 10.1016/j.jfca.2019.103255

31. Westenbrink S, van der Vossen-Wijmenga W, Toxopeus I, Milder I, Ocké M. LEDA, the branded food database in the Netherlands: data challenges and opportunities. *J Food Compos Anal*. (2021) 102:104044. doi: 10.1016/j.jfca.2021.104044

32. Anonymous. Besluit van de Minister voor Medische Zorg van 27 maart 2019, kenmerk 1506796–188807-VGP, houdende onthefng aan Katjes Fassin GmbH + Co. KG te Emmerich van artikel 5 van het Warenwetbesluit Toevoeging micro-voedingsstoffen aan levensmiddelen (Decision of the Minister for Medical care of March 27th 2019, reference 1506796–188807-VGP, handling exemption to Katjes Fassin GmbH + Co. KG in Emmerich of article 5 of the Commodity Act decision Addition of Micronutrients to foods). Staatscourant. 2019; 1838: 1-2.

33. Anonymous. Besluit van de Minister voor Medische Zorg van 7 januari 2020, kenmerk 1632453–195489-VGP, houdende onthefng aan Arla Foods B.V. te Nijkerk van artikel 5 van het Warenwetbesluit Toevoeging micro-voedingsstofen aan levensmiddelen (Decision of the Minister for Medical care of January 7th, 2020, reference 1632453–195489-VGP, handling exemption to Arla Foods B.V. in Nijkerk of article 5 of the Commodity Act decision Addition of Micronutrients to foods). Staatscourant. 2020; 2337: 1-2.

34. European Commission. Guidance document for competent authorities for the controle of compliance with EU legislation with regard to the setting of tolerances for nutrient values declared on a label. Retrieved May 2020. Avaulable at: https://ec.europa.eu/food/sites/food/files/safety/docs/labelling_nutrition-vitamins_minerals-guidance_tolerances_1212_en.pdf. (2012).

35. EUR-Lex. Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods. Retrieved April 2020. Available at: https://eur-lex.europa.eu/homepage.html, L404 (2006).

36. Posit. RStudio: Integrated development environment for R. Posit Software, PBC, Boston, MA. Available at: http://www.posit.co/. (2023)

37. Sison CP, Glaz J. Simultaneous confidence intervals and sample size determination for multinomial proportions. *J Am Stat Assoc*. (1995) 90:366–9. doi: 10.1080/01621459.1995.10476521

38. McBurney MI, Hartunian-Sowa S, Matusheski NV. Implications of US nutrition facts label changes on micronutrient density of fortified foods and supplements. *J Nutr*. (2017) 147:1025–30. doi: 10.3945/jn.117.247585

39. Casala E, Matthys C, Péter S, Baka A, Kettler S, McNulty B, et al. Monitoring and addressing trends in dietary exposure to micronutrients through voluntarily fortified foods in the European Union. *Trends Food Sci Technol*. (2014) 37:152–61. doi: 10.1016/j.tifs.2014.03.001

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 17 April 2024
DOI 10.3389/fnut.2024.1364695



OPEN ACCESS

EDITED BY
Alessandra Durazzo,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Saniye Bilici,
Gazi University Faculty of Health Sciences, Türkiye
Taygun Dayı,
Near East University, Cyprus

*CORRESPONDENCE
Tera L. Fazzino
✉ tfazzino@ku.edu

RECEIVED 02 January 2024
ACCEPTED 08 April 2024
PUBLISHED 17 April 2024

CITATION
Fazzino TL, Summo C and Pasqualone A (2024) The Italian food environment may confer protection from hyper-palatable foods: evidence and comparison with the United States.
*Front. Nutr.* 11:1364695.
doi: 10.3389/fnut.2024.1364695



# The Italian food environment may confer protection from hyper-palatable foods: evidence and comparison with the United States

Tera L. Fazzino[1,2]*, Carmine Summo[3] and Antonella Pasqualone[3]

[1]Department of Psychology, University of Kansas, Lawrence, KS, United States, [2]Cofrin Logan Center for Addiction Research and Treatment, University of Kansas, Lawrence, KS, United States, [3]Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Bari, Italy

**Background:** Multi-national food corporations may saturate country-level food systems with hyper-palatable foods. However, the degree to which global food corporations have been integrated into country-level food systems may vary. Italy has largely retained local food production and may have low hyper-palatable food (HPF) availability in the food supply. The study quantified the prevalence of HPF in the Italian food system and compared the hyper-palatability of similar foods across Italy and the United States, which has wide HPF saturation.

**Methods:** A national food system dataset was used to characterize HPF availability in Italy. A representative sample of foods commonly consumed in both Italy and the US were collected and compared. Foods represented six categories: cookies/biscotti, cakes/merendine, salty snacks, industrial bread, frozen pizza and protein/cereal bars. A standardized definition from Fazzino et al. identified HPF.

**Results:** Less than one third (28.8%) of foods in the Italian food system were hyper-palatable. US HPF items had significantly higher fat, sugar, and/or sodium across most food categories ($p$ values = 0.001 to 0.0001). Italian HPF items had higher fiber and/or protein relative to US HPF from the same category ($p$ values = 0.01 to 0.0001).

**Conclusion:** The Italian food system may confer protection from HPF exposure. HPF products in Italy had lower palatability-related nutrients and higher satiety-promoting nutrients.

KEYWORDS

food environment, carbohydrate, fat, sodium, sugar, fiber

## 1 Introduction

The industrialization of food systems globally has yielded substantial changes in country-level food environments and population health indices related to obesity and metabolic disease (1). However, there is variability in the degree to which globally produced foods have been integrated in different country-level food systems. Some countries, such as the United States (US), have developed a highly industrialized food environment run by several multi-national food companies (2) that have saturated the market with hyper-palatable foods (3). The food environment, combined with a structural environment that promotes limited physical activity and a culture that embraces convenience and eating on-the-go (4, 5) has yielded a US adult obesity rate 42.2%, the highest globally (6). However, other countries, such as Italy, have retained more local/national food companies and agricultural production and have experienced slower integration of industrialized foods into the food supply. Italy also has a structural environment that promotes physical activity and values food quality over convenience (7, 8). As such, Italy has one of the lowest adult obesity rates in the European Union (10.4%) (9, 10) and may represent an environment that confers greater protection from obesogenic foods.

To inform global obesity prevention efforts, it is important to identify food environments that are highly saturated with hyper-palatable foods, as well as those that confer protection from such foods. Hyper-palatable foods (HPF) contain combinations of palatability-inducing nutrients (fat, sugar, sodium, and/or carbohydrates) at thresholds that do not occur in nature, which yield a highly rewarding eating experience (11). HPF can excessively activate brain reward neurocircuitry, the same neurocircuitry activated by psychoactive drugs, and slow engagement of physiological satiety mechanisms (11–13). As a result, HPF may be difficult to stop eating and when consumed repeatedly over time, may increase the risk for weight gain and obesity. Prior research has identified the US food environment as being highly saturated with HPF. As of 2018, HPF comprised 68.9% of available foods in the US food system (3). Given the extensive HPF availability and the high rate of obesity among the US adult population (6), the US food environment may be considered obesogenic and therefore may substantially increase obesity risk.

In contrast, the Italian food system may represent a food environment that yields protection from HPF; however the availability of HPF in the Italian food system has not been quantified. Although not immune to the influence of the global food industry, the Italian food system has largely retained local and national-level food production, including food producers that specialize in key foods including breads, cheeses, and fruits and vegetables (7, 8). The structure of the food environment may facilitate the provision of high quality, fresh foods that are relatively inexpensive, consistent with an Italian cultural tradition that values fresh high quality foods and ingredients and that relies in a limited way on ready-to-eat, industrial foods (7, 8). Furthermore, southern regions of Italy have been recognized globally for their cultural dietary roots in the Mediterranean diet (14, 15), which is comprised of whole grains, legumes, and fruits and vegetables, combined with locally available fish, and limited intake of richer cheeses and cured meats (16, 17). Thus, the Italian food system may have a higher prevalence of whole, fresh foods that are not HPF (and correspondingly a lower prevalence of HPF). Additionally, evidence has indicated that some foods (e.g., fast foods) sold in different countries by the same parent company may have substantially different nutrient contents, and may be tailored to the country's expectations surrounding taste preferences and health (18). Thus, HPF available in Italy may differ in their nutrient contents from the HPF available in the US, a premise that should be tested.

The purpose of the current study was to (1) quantify the prevalence of HPF in the Italian food system using nationally representative data obtained from the Banca Dati di Composizione Degli Alimenti per Studi Epidemiologici in Italia (BDA) (19); and (2) to compare the hyper-palatability of food products from categories that are commonly consumed in Italy and the US, using representative data collected from grocery stores in Italy and the US.

## 2 Materials and methods

The study was conducted in two parts and consisted of (1) analysis of an Italian food system dataset to quantify HPF availability nationally; and (2) data collection and analysis to compare the hyper-palatability of a representative sample of foods from Italian and US grocery stores. Procedures are detailed below. All data were processed and analyzed using R statistical software (20).

### 2.1 Processing and analysis of the national dataset

The study analyzed a dataset considered representative of the Italian food system, the Banca Dati di Composizione Degli Alimenti per Studi Epidemiologici in Italia (BDA) (19), to quantify HPF availability. The BDA was developed by a collaborative working group of researchers with the purpose of creating a database for use in epidemiological research (21). As such, the BDA is comprised of selected foods deemed to be representative of the Italian diet and the Italian food system on the national level (21). At the time of the study, the most recent update of the BDA was conducted in 2015 (19). The BDA provided detailed nutrient and ingredient data for a total of $N = 978$ food and beverage items. The BDA was processed in accordance with procedures from Fazzino et al (11) to apply the HPF definition to all foods. Beverages were removed before analysis, as the HPF definition does not apply to beverages (11). Thus, a total of $N = 857$ food items were analyzed. Total sugar was calculated by summing the values of glucose, fructose, galactose, sucrose, maltose, and lactose. Percent calories from fat, sugar, and carbohydrates (following the subtraction of fiber and sugar) were calculated. Salt was converted to sodium and then calculated as percent sodium per food weight in grams. Items that met criteria for at least one of the following were classified as HPF: (1) fat and sodium, FSOD (> 25% kcal from fat, ≥ 0.30% sodium), (2) fat and simple sugars, FS (> 20% kcal fat, > 20% kcal sugar), and (3) carbohydrate and sodium, CSOD (> 40% kcal carbohydrates, ≥ 0.20% sodium) (11). The percentage of HPF available in the BDA was calculated as n total HPF items divided by N total items. The percentage of each HPF group was also calculated using the same procedures.

### 2.2 Data collection and comparative analysis of Italian and US food samples

To directly compare similar food items available in Italy and the US, and to address a limitation of the BDA that it may not contain some prepared foods, we collected a representative sample of foods available in grocery stores from six categories of foods that were commonly consumed across both countries. Specifically, cookies/biscotti were considered dry sweet snacks that are commonly consumed in Italy (22–24) and the US (5, 25). Cakes/merendine were identified as moist sweet snacks and are commonly consumed at breakfast and as afternoon snacks across both countries (5, 26). Salty snacks included crackers, pretzels, breadsticks, and other crunchy savory items that are consumed as snacks or pre-meal appetizers across both countries (5, 24). Industrially produced breads were selected as a category because industrial breads are the standard bread product consumed in the US (4, 27), and while the consumption of artisanal breads is most common in Italy, the use of industrial breads is emerging (28). Similarly, frozen pizza was chosen as a category because it is widely consumed in the US (29) and its availability and consumption in Italy has grown in recent years (30). Finally, protein and cereal bars were selected because they have experienced wide expansion and consumption in both the US (31, 32) and Italian food markets (32, 33) and are marketed as a 'healthier' snack option than other available sweet or salty snacks (34). The foods across the six categories had similar methods of preparation, and were therefore directly comparable across the countries. Data were collected from grocery stores selected in the US and Italy, and data were collected on all products in the stores that aligned with the aforementioned six food categories. To collect food and nutrient data, researchers used mobile phones to photograph the front and back of all food items. Photographs were downloaded and food item and nutrient data were entered into excel spreadsheets using a standardized double-entry process.

Following data collection and entry, data were processed in preparation for applying the HPF definition. Percent calories from fat, sugar, and carbohydrates (following the subtraction of fiber and sugar) were calculated. Items that met criteria for at least one of the following were classified as HPF: (1) fat and sodium, FSOD (> 25% kcal from fat, ≥ 0.30% sodium), (2) fat and simple sugars, FS (> 20% kcal fat, > 20% kcal sugar), and (3) carbohydrate and sodium, CSOD (> 40% kcal carbohydrates, ≥ 0.20% sodium) (11).

A series of Fisher's exact tests were used to compare the proportion of HPF items across Italian and US samples by food category. Fisher's exact tests can be used to compare samples with different proportions and cell sizes (35), as was the case between Italian and US samples. In alignment with study aim 2 to examine differences in the hyper-palatability of items across Italian and US samples, we examined whether the Italian and US items had significantly different median values that contribute to HPF designation, specifically % kcal from fat, sugar, and carbohydrates, and % sodium per food weight in grams. Additionally, protein and fiber in grams were examined to understand potential differences in satiety-promoting nutrients across food categories and countries. The variables had different distributions and therefore Mood's test of medians was used to compare Italian and US sample values for each nutrient of interest.

# 3 Results

## 3.1 National analysis

Findings indicated that 28.8% (247/857) of food items in the BDA met criteria as HPF, suggesting that less than one third of foods available in the Italian food system are hyper-palatable. The most common type of HPF was fat and sodium HPF (61.5%; 152/247). About a quarter of HPF items were fat and sugar HPF (24.5%; 61/247) and less than a quarter were carbohydrate and sodium HPF (20.6%; 51/247). Foods that were fat and sodium HPF were primarily preserved meats (e.g., cured pork) and cheeses (68.4%; 104/152). Items that were fat and sugar HPF were most commonly cookies and cakes (54.1%; 33/61) and items that were most commonly carbohydrate and sodium HPF were industrially produced breads and crackers (78.4%; 40/51).

## 3.2 Comparative analysis of Italian and US food products

Cookies/biscotti and cakes/merendine from Italy had a significantly lower percentage of items that were HPF relative to cookies/biscotti and cakes/merendine from the US ($p$ values <0.001; Table 1). There were no other significant differences in the proportion of HPF for salty snacks, frozen pizza, industrial breads, and protein and cereal bars ($p$ values = 0.081 to 0.999; Table 1) across countries.

Table 2 presents the food categories across HPF groups by country. Patterns across food categories were distinct; some food categories aligned primarily with one HPF group, others aligned with multiple HPF groups, and some patterns differed by country (Table 2). Across both countries, industrial breads were most commonly classified as carbohydrate and sodium HPF (Table 2). Cookies/biscotti and salty snacks from both countries were commonly classified as two HPF groups (Table 2). Furthermore, cakes/merendine from Italy were most commonly classified as fat and sugar HPF, whereas cakes/merendine from the US were commonly classified as both fat and sodium HPF and fat and sugar HPF (Table 2).

When examining HPF items specifically, US HPF items had significantly higher median values for at least one palatability-related nutrient, with the exception of industrial breads (Table 3). US HPF items had significantly higher % kcal (calories) from fat (salty snacks and frozen pizza) and/or % sodium (cakes/merendine, frozen pizza, cereal/protein bars; Table 3). US cookies/biscotti that were HPF also had significantly higher % kcal from sugar than Italian cookies that were HPF (35.8% vs. 25.3%; Table 3). Italian HPF items among cookies, salty snacks, and frozen pizza had significantly higher % kcal from carbohydrates compared to US HPF (Table 3). Italian industrial breads that were HPF had significantly higher % kcal from fat than did US industrial breads that were HPF (16.4% vs. 12.0%; Table 3). Regarding satiety-promoting nutrients, Italian cookies/biscotti and

TABLE 1 Prevalence of hyper-palatable foods among Italian and US samples.

| | Italian | US | |
|---|---|---|---|
| | % HPF (n/N) | % HPF (n/N) | $p$ value[a] |
| Food categories | | | |
| Cookies/biscotti | 52% (14/27) | 96% (196/205) | <0.00001 |
| Cakes/merendine | 77% (23/30) | 100% (57/57) | 0.0003 |
| Salty Snacks | 98% (54/55) | 93% (654/700) | 0.244 |
| Industrial breads | 94% (15/16) | 95% (186/195) | 0.554 |
| Frozen pizza | 100% (37/37) | 98% (121/124) | 0.999 |
| Cereal and protein bars | 68% (15/22) | 84% (121/144) | 0.081 |

[a]$p$ value from fisher's exact test.
HPF, hyper-palatable food.

TABLE 2 Prevalence of hyper-palatable food groups among Italian and US samples.

| | Italian | | | US | | |
|---|---|---|---|---|---|---|
| | %(n/N) | | | %(n/N) | | |
| Food type | FSOD | FS | CSOD | FSOD | FS | CSOD |
| Cookies/biscotti | 54% (7/13) | 62% (8/13) | 23% (3/13) | 63% (123/196) | 84% (164/196) | 9% (18/196) |
| Cakes/merendine | 4% (1/23) | 83% (19/23) | 17% (4/23) | 72% (41/57) | 96% (55/57) | 0% (0/57) |
| Salty snacks | 43% (23/54) | 0% (0/54) | 98% (53/54) | 83% (546/654) | 5% (32/654) | 61% (400/654) |
| Industrial breads | 27% (4/15) | 0% (0/15) | 100% (15/15) | 9% (16/186) | 0% (0/186) | 94% (175/186) |
| Frozen pizza | 95% (35/37) | 0% (0/37) | 43% (16/37) | 99% (120/121) | 0% (0/121) | 17% (21/121) |
| Cereal and protein bars | 7% (1/15) | 87% (13/15) | 33% (5/15) | 42% (51/121) | 79% (95/121) | 12% (15/121) |

FSOD, fat and sodium hyper-palatable food group; FS, fat and sugar hyper-palatable food group; CSOD, carbohydrate and sodium hyper-palatable food group.

cakes/merendine that were HPF had significantly higher fiber than US items (Table 3), and Italian cookies/biscotti, cakes/merendine, and salty snacks that were HPF had significantly higher protein than US HPF items (Table 3).

## 4 Discussion

The study examined the availability of hyper-palatable foods in the Italian food system and conducted the first comparative analysis of HPF across two countries, Italy and the United States. Findings revealed that less than one third of foods in the Italian food system were HPF, indicating that the Italian food system may confer some degree of protection from HPF exposure. Of the foods that were HPF, the majority were classified as HPF with elevated fat and sodium, and were typically cured meats and cheeses. A comparison of HPF among six categories of commonly consumed foods indicated that Italian cookies/biscotti and cakes/merendine had significantly lower proportions of items that were HPF, relative to US cookies/biscotti and cakes/merendine. Our findings also identified differences in the nutrient contents of HPF products across countries, with US products typically containing higher fat, sodium, and/or sugar, and Italian products typically containing higher carbohydrates and more fiber and protein. Taken together, findings indicated that HPF comprise less than one third of the Italian food system, and that HPF items from Italy tended to have lower palatability-inducing nutrients and higher satiety-promoting nutrients relative to US products that were classified as HPF.

Among the 28.8% of foods that were classified as HPF using the Italian national data, HPF items most commonly contained elevated fat and sodium, and were typically in cured meat products and cheeses. Most of the meat items that were classified as HPF had elevated sodium, which may have been necessitated by food safety considerations in the preparation process. Most HPF meats were prepared in a manner that involved slow aging of meat (e.g., curing) without direct cooking. To prevent the growth of bacteria or pathogens, sodium levels between 3.0 and 5.0% are typically required in cured meat products (36). Notably, the sodium level is in excess of the fat and sodium HPF criterion (≥0.30% sodium) and therefore it may not be surprising that many cured meats were classified as HPF. Overall, the finding that fat and sodium HPF was the most common type of HPF is consistent with prior studies conducted in the US food system (3, 11). Studies of the US food system also reported that meats and cheeses were commonly fat and sodium HPF; however most of the US produced meats were cooked and did not require high sodium content for food safety purposes, which may represent a difference across countries. Overall, evidence from two countries indicates that fat and sodium HPF may be the most commonly available type of HPF, and highlights meats and cheeses as commonly fat and sodium HPF. However, more work is needed to support this premise across countries globally.

Our findings overall revealed that most foods in the Italian food system do not have nutrient combinations that exaggerate their palatability, indicating that the Italian food system may confer some protection from HPF exposure. The finding is in stark contrast to the prevalence of HPF in the US food system, which demonstrated that as of 2016 (the year most closely matched to BDA 2015), 62% of foods in the US food supply were HPF (11). Thus, Italy had less than half of the HPF availability for the same time frame relative to the US. This study therefore presents the first evidence of different HPF availability across country-level food systems. Overall, the relatively low prevalence of HPF in the Italian food system and the high availability of whole fresh foods may protect the population from regular exposure to and consumption of HPF. The availability of non-HPF whole foods is consistent with Southern Italy's cultural dietary roots in the Mediterranean diet (14, 15), which largely comprises whole grains, legumes, and fruits and vegetables, combined with locally available fish (16, 17). The low availability of HPF and adoption of the Mediterranean diet may promote higher diet quality and lower obesity, metabolic disease, and related chronic disease risk among the Italian population, which has been observed in the literature (15, 37). Furthermore, other characteristics of Italian societal structure and culture, including a built environment that facilitates physical activity (e.g., centralized towns built for walking, strong public transportation system), cultural preferences for high quality (non-HPF) food (7, 8), and limited reliance on eating outside of the home (7, 8) may contribute to the lower chronic disease rates as well. Overall, findings of the current study revealed the limited prevalence of HFP in the Italian food environment, a factor that is consistent with Italian dietary values and practices (7, 8), and may confer protection from obesity and chronic disease risk (15, 37).

In addition to analyzing nationally representative data, we also collected representative data from grocery stores in Italy and the US to compare products from six food categories identified a prior that are typically consumed in both countries that have similar preparation. Overall, there were substantially lower percentages of HPF among

TABLE 3 Comparison of hyper-palatable food items across Italian and US food samples.

| | Italian | US | |
|---|---|---|---|
| | Median (IQR) | Median (IQR) | *p* value[a] |
| **Cookies/biscotti** | | | |
| % kcal fat | 34.7 (6.5) | 38.6 (11.3) | 0.107 |
| % kcal sugar | 20.3 (5.7) | 28.6 (11.4) | 0.014 |
| % kcal carbohydrates | 31.0 (12.6) | 26.3 (9.3) | 0.006 |
| % sodium | 0.28 (0.15) | 0.32 (0.12) | 0.450 |
| Total protein (g/100 g) | 7.3 (1.8) | 3.9 (3.3) | 0.0003 |
| Total fiber (g/100 g) | 3.5 (1.8) | 1.8 (3.3) | <0.0001 |
| **Cakes/merendine** | | | |
| % kcal fat | 38.7 (11.3) | 41.3 (13.1) | 0.323 |
| % kcal sugar | 25.3 (14.3) | 35.8 (17.1) | 0.138 |
| % kcal carbohydrates | 23.9 (12.5) | 20.0 (7.8) | 0.048 |
| % sodium | 0.20 (0.09) | 0.35 (0.12) | <0.0001 |
| Total protein (g/100 g) | 7.2 (3.2) | 3.6 (1.8) | <0.0001 |
| Total fiber (g/100 g) | 2.6 (1.8) | 1.0 (2.1) | 0.002 |
| **Salty snacks** | | | |
| % kcal fat | 24.5 (20.9) | 45.0 (20.3) | <0.0001 |
| % kcal sugar | 2.9 (3.0) | 1.4 (5.3) | <0.0001 |
| % kcal carbohydrates | 55.8 (13.5) | 44.3 (15.9) | <0.0001 |
| % sodium | 0.70 (0.31) | 0.70 (0.39) | 0.440 |
| Total protein (g/100 g) | 10.1 (3.0) | 7.1 (3.3) | <0.0001 |
| Total fiber (g/100 g) | 3.5 (4.7) | 3.6 (4.1) | 0.021 |
| **Industrial breads** | | | |
| % kcal fat | 16.4 (12.5) | 12.0 (7.2) | 0.001 |
| % kcal sugar | 4.9 (6.8) | 10.0 (5.8) | 0.059 |
| % kcal carbohydrates | 54.1 (12.5) | 60.0 (10.3) | 0.537 |
| % sodium | 0.55 (0.11) | 0.47 (0.12) | 0.073 |
| Total protein (g/100 g) | 7.7 (1.3) | 9.5 (2.3) | 0.009 |
| Total fiber (g/100 g) | 3.4 (2.8) | 2.7 (4.8) | 0.578 |
| **Frozen pizza** | | | |
| % kcal fat | 33.4 (8.6) | 41.7 (8.0) | <0.0001 |
| % kcal sugar | 5.3 (2.8) | 5.2 (3.4) | 0.860 |
| % kcal carbohydrates | 39.3 (9.0) | 33.1 (8.2) | 0.0004 |
| % sodium | 0.47 (0.08) | 0.52 (0.10) | <0.0001 |
| Total protein (g/100 g) | 10.0 (0.8) | 10.1 (2.9) | 0.014 |
| Total fiber (g/100 g) | 2.0 (0.8) | 1.5 (0.3) | 0.001 |
| **Cereal and protein bars** | | | |
| % kcal fat | 26.6 (10.7) | 30.0 (14.3) | 0.714 |
| % kcal sugar | 26.4 (15.2) | 28.0 (9.0) | 0.538 |
| % kcal carbohydrates | 20.6 (20.1) | 28.0 (21.7) | 0.584 |
| % sodium | 0.17 (0.18) | 0.33 (0.15) | 0.001 |
| Total protein (g/100 g) | 6.7 (6.9) | 9.1 (14.7) | 0.627 |
| Total fiber (g/100 g) | 4.3 (2.2) | 4.2 (4.6) | 0.809 |

[a]*p* value from Mood's median test.

Italian cookies/biscotti and cakes/merendine relative to US cookies/biscotti and cakes/merendine. However, there were no significant differences in the proportion of HPF across Italian and US salty snacks, industrial breads, frozen pizzas, and protein and cereal bars. Findings regarding the substantially lower percentage of HPF among Italian sweet snacks (cookies/biscotti and cakes/merendine) may reflect a recent focus in Italy on ways to formulate products consumed by children to reduce child obesity risk. In a recent report by the World Health Organization, the overweight and obesity prevalence of Italian children was identified as among the highest of countries in the European Union (10), which has been attributed to a decreased adherence to the Mediterranean diet and reductions in physical activity (38). Thus, Italian food companies have focused on formulating products with greater care to help prevent child obesity, and their efforts may be reflected in these findings. However, we did not find any significant differences in the percentage of HPF among Italian and US salty snacks, industrial breads, frozen pizzas, and protein and cereal bars, many of which may also be consumed by children and contribute to obesity among children in Italy and the US. Thus, our findings highlight areas for potential improvement in both the Italian and US food industries regarding product formulation to promote health and reduce availability of HPF.

Our findings also indicated that HPF are not created equally, as evidenced by substantial differences across Italian and US foods that were classified as HPF. HPF items from the US had significantly higher contents of at least one palatability-related nutrient (fat, sugar, and/or sodium) across five of the six food categories, relative to Italian HPF items. HPF from Italy had significantly higher carbohydrates among three categories (cookies/biscotti, salty snacks, and frozen pizza), relative to US HPF items. Our characterization of carbohydrates in this study was focused on starchy carbohydrates, and did not include sugar or fiber. Therefore, our findings indicate that Italian HPF items had higher starchy carbohydrates in three of the six product categories relative to US HPF items. This finding is overall consistent with the Italian diet, which typically includes high quantities of starchy carbohydrates, such as pasta and bread (26). Therefore, starchy carbohydrates may be more accepted in packaged products such as cookies/biscotti and salty snacks as well. Furthermore, in the US and other European countries, low carbohydrate diets have become popular and starchy carbohydrate reduction may be a focus for consumers (39, 40). However, in contrast, Italians perceive a low carbohydrate diet as very far from their traditional food habits (41), a point that may also contextualize the differences in starchy carbohydrates across Italian and US products. Finally, Italian HPF had significantly higher fiber and/or protein across most food categories relative to US HPF items. Thus, Italian HPF items tended to have more satiety-promoting nutrients relative to US HPF items. Overall, findings indicated that Italian HPF had lower palatability-inducing nutrients and higher satiety promoting nutrients, relative to US HPF items.

The study had several limitations. First, the most recent nationally available data representing the Italian food system was from 2015. Therefore, it is unclear whether estimates of HPF availability may be different for today's food environment. In addition, the BDA may have limited representation of prepared foods, which may lead to an underestimation of HPF availability. However, to address this limitation in the national data, we collected representative data from foods available in grocery stores in Italy and the US to directly compare foods from categories that may be underrepresented in the BDA, and that are commonly consumed across both cultures. Furthermore, to maintain methodological rigor in comparing foods across countries, we limited our comparisons to food categories for which the preparation was the same and for which the nutrient values would not change when cooked (e.g., frozen pizza).

In conclusion, our results indicate that HPF comprise less than one third of the Italian food system, indicating the Italian food system may confer protection from HPF exposure. Findings also revealed key differences in HPF products from Italy vs. the US, with HPF from Italy tending to have lower palatability-inducing nutrients and higher satiety promoting nutrients relative to US products of the same type. However, our findings suggest that food companies in Italy and the US should consider reducing the sodium, refined carbohydrates, and fat in salty snacks, frozen pizzas, industrial breads, and protein/cereal bars to reduce the hyper-palatability of these commonly consumed foods in Italy and the US.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

TLF: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. CS: Methodology, Investigation, Writing – review & editing. AP: Conceptualization, Methodology, Investigation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding from the University of Bari “Aldo Moro” Visiting Scholar Program funded TF’s time during the study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. *Lancet*. (2011) 378:804–14. doi: 10.1016/S0140-6736(11)60813-1

2. Hendrickson M., Howard P., Miller E. *The Food System: Concentration and Its Impacts. A Special Report to the Family Farm Action Alliance*. Family Farm Action Alliance (2020).

3. Demeke S, Rohde K, Chollet-Hinton L, Sutton C, Kong KL, Fazzino TL. Change in hyper-palatable food availability in the US food system over 30 years: 1988 to 2018. *Public Health Nutr*. (2022) 26:182–9. doi: 10.1017/S1368980022001227

4. Hoy M., Clemens J., Murayi T., Moshfegh A. Consumption of Restaurant Foods by US Adults: What We Eat in America, NHANES 2017–2018. Available at: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/48_Restaurant_food_consumption_by_Adults_1718.pdf (2022).

5. Morton S., Rhodes D., Moshfegh A. Convenience stores: Source of food/beverages among children: What we eat in America, NHANES 2015-2018. (2021). Available at: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/42_Convenience_Stores_Food_and_Beverages_among_Children_1518.pdf.

6. Fryar C., Carroll M., Afflu J. Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2013–2014. (2021). Available at: https://www.cdc.gov/nchs/data/hestat/obesity-adult-17-18/obesity-adult.htm.

7. Ferretti MP, Magaudda P. The slow pace of institutional change in the Italian food system. *Appetite*. (2006) 47:161–9. doi: 10.1016/j.appet.2006.05.005

8. Parasecoli F: A history of food in Italy. Reaktion Books. EAN: 9781780232768 (2014).

9. Azzolini E., Ricciardi W. (2019). TACKLING THE OBESITY CHALLENGE IN ITALY. https://apps.who.int/iris/bitstream/handle/10665/332528/Eurohealth-25-1-17-20-eng.pdf?sequence=1&isAllowed=y

10. World Health Organization. WHO European Regional Obesity Report 2022. (2022). Available at: https://apps.who.int/iris/bitstream/handle/10665/353747/9789289057738-eng.pdf

11. Fazzino TL, Rohde K, Sullivan DK. Hyper-palatable foods: development of a quantitative definition and application to the US food system database. *Obesity*. (2019) 27:1761–8. doi: 10.1002/oby.22639

12. DiFeliceantonio AG, Coppin G, Rigoux L, Edwin Thanarajah S, Dagher A, Tittgemeyer M, et al. Supra-additive effects of combining fat and carbohydrate on food reward. *Cell Metab*. (2018) 28:33–44.e3. doi: 10.1016/j.cmet.2018.05.018

13. Small DM, DiFeliceantonio AG. Processed foods and food reward. *Science*. (2019) 363:346–7. doi: 10.1126/science.aav0556

14. Lăcătușu C-M, Grigorescu E-D, Floria M, Onofriescu A, Mihai B-M. The Mediterranean diet: from an environment-driven food culture to an emerging medical prescription. *Int J Environ Res Public Health*. (2019) 16:942. doi: 10.3390/ijerph16060942

15. Renna M, Rinaldi VA, Gonnella M. The Mediterranean diet between traditional foods and human health: the culinary example of Puglia (southern Italy). *Int J Gastron Food Sci*. (2015) 2:63–71. doi: 10.1016/j.ijgfs.2014.12.001

16. Vitiello V, Germani A, Dolcetta EC, Donini LM. The new modern Mediterranean diet Italian pyramid. *Ann Ig Med Prev E Comunita*. (2016) 28:179–86. doi: 10.7416/ai.2016.2096

17. del Balzo V, Diolordi L, Pinto A, Giusti AM, Vitiello V, Cannella C, et al. Mediterranean diet pyramids: towards the Italian model. *Ann Ig*. (2012) 24:443–7.

18. Ziauddeen N, Fitt E, Edney L, Dunford E, Neal B, Jebb SA. Variability in the reported energy, total fat and saturated fat contents in fast-food products across ten countries. *Public Health Nutr*. (2015) 18:2962–9. doi: 10.1017/S1368980015000336

19. Instituto Europeo di Oncologia-Milano & Universita degli Studi Udine. *Banca Dati di Composizione degli Alimenti per Studi Epidemiologici in Italia*. Italy: Instituto Europeo di Oncologia (2015).

20. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2021).

21. Salvini S., Parpinel M., Gnagnarella P., Maisonneuve P., Turrini A. *Banca Dati Di Composizione Degli Alimenti per Studi Epidemiologici in Italia*. Instituto Europeo di Oncologia- Milano (1998).

22. Dall'Asta M, Rosi A, Angelino D, Pellegrini N, Martini D. Evaluation of nutritional quality of biscuits and sweet snacks sold on the Italian market: the food labelling of Italian products (FLIP) study. *Public Health Nutr*. (2020) 23:2811–8. doi: 10.1017/S1368980020000853

23. Caponio F, Summo C, Delcuratolo D, Pasqualone A. Quality of the lipid fraction of Italian biscuits. *J Sci Food Agric*. (2006) 86:356–61. doi: 10.1002/jsfa.2357

24. Leclercq C, Arcella D, Piccinelli R, Sette S, le Donne C, Turrini A, et al. The Italian National Food Consumption Survey INRAN-SCAI 2005-06: main results in terms of food consumption. *Public Health Nutr*. (2009) 12:2504–32. doi: 10.1017/S1368980009005035

25. Sebastian R., Wilkinson Enns C., Martin C., Goldman J., Moshfegh A. (2020). Sweet Foods Consumption by Adults in the U.S.: What We Eat in America, NHANES 2015–2018. Available at: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/33_Sweet_foods_adults_1518.pdf

26. Pounis G, Bonanni A, Ruggiero E, di Castelnuovo A, Costanzo S, Persichillo M, et al. Food group consumption in an Italian population using the updated food classification system FoodEx2: results from the Italian nutrition & HEalth survey (INHES) study. *Nutr Metab Cardiovasc Dis*. (2017) 27:307–28. doi: 10.1016/j.numecd.2017.01.004

27. Bowman S., Clemens J., Friday J. (2021). Food Pattern Group and Macronutrient Intakes of Adults: WWEIA, NHANES 2003–2004 to 2017–2018. Available at: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/35_Trend_Analysis_of_Adults_Macronutrient_and_Food_Pattern_Food_Intakes_0318.pdf

28. Carcea M, Narducci V, Turfani V, Aguzzi A. A survey of sodium chloride content in Italian artisanal and industrial bread. *Foods Basel Switz*. (2018) 7:181. doi: 10.3390/foods7110181

29. Rhodes D., Adler M., Clemens J., LaComb R., Moshfegh A. (2014). Consumption of Pizza: What We Eat in America, NHANES 2007–2010. Available at: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/11_consumption_of_pizza_0710.pdf

30. Future Market of Pizza. Which attributes do they matter? *AGRIS-Line Pap Econ Inform*. (2016). doi: 10.22004/ag.econ.254026

31. Agriculture and Agri-Food Canada. (2013). Snack, Cereal and Nutrition Bars in the United States. Available at: https://publications.gc.ca/collections/collection_2015/aac-aafc/A74-2-2013-21-eng.pdf

32. Fortune Business Insights. (2019). Protein Bar Market Size, Share & Covid-19 Impact Analysis, by Source (Plant-Based and Animal-Based), by Type (Sports Nutritional Bars, Meal-Replacement Bars, and Others), by Distribution Channel (Mass Merchandisers, Specialty Stores, Convenience Stores, Online Channels, Others), and Regional Forecast, 2022–2029. Available at: https://www.fortunebusinessinsights.com/industry-reports/protein-bar-market-100805

33. Lombardi A, Verneau F, Lombardi P. Beyond the use of food supplements: an empirical analysis in Italy. *Ital J Food Sci*. (2016) 28:15–24.

34. Boukid F, Klerks M, Pellegrini N, Fogliano V, Sanchez-Siles L, Roman S, et al. Current and emerging trends in cereal snack bars: implications for new product development. *Int J Food Sci Nutr*. (2022) 73:610–29. doi: 10.1080/09637486.2022.2042211

35. Kim H-Y. Statistical notes for clinical researchers: sample size calculation 2. Comparison of two independent proportions. *Restor Dent Endod*. (2016) 41:154–6. doi: 10.5395/rde.2016.41.2.154

36. Toldrá F. *Biochemistry of fermented meat. In food biochemistry and food processing*. US: John Wiley & Sons, Ltd (2012).

37. Godos J, Zappalà G, Bernardini S, Giambini I, Bes-Rastrollo M, Martinez-Gonzalez M. Adherence to the Mediterranean diet is inversely associated with metabolic syndrome occurrence: a meta-analysis of observational studies. *Int J Food Sci Nutr*. (2017) 68:138–48. doi: 10.1080/09637486.2016.1221900

38. Labayen Goñi I, Arenaza L, Medrano M, García N, Cadenas-Sanchez C, Ortega FB. Associations between the adherence to the Mediterranean diet and cardiorespiratory fitness with total and central obesity in preschool children: the PREFIT project. *Eur J Nutr*. (2018) 57:2975–83. doi: 10.1007/s00394-017-1571-3

39. Gardner CD, Vadiveloo MK, Petersen KS, Anderson CAM, Springfield S, van Horn L, et al. Popular dietary patterns: alignment with American Heart Association 2021 dietary guidance: a scientific statement from the American Heart Association. *Circulation*. (2023) 147:1715–30. doi: 10.1161/CIR.0000000000001146

40. Muscogiuri G, el Ghoch M, Colao A, Hassapidou M, Yumuk V, Busetto L, et al. European guidelines for obesity Management in Adults with a very low-calorie ketogenic diet: a systematic review and Meta-analysis. *Obes Facts*. (2021) 14:222–45. doi: 10.1159/000515381

41. Corciolani M. Navigating institutional complexity through emotion work: the case of Italian consumers adapting to a ketogenic diet. *J Bus Res*. (2023) 158:113657. doi: 10.1016/j.jbusres.2023.113657

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 03 July 2024
DOI 10.3389/fnut.2024.1418393

Check for updates

OPEN ACCESS

EDITED BY
Massimo Lucarini,
Research Centre for Food and Nutrition, Italy

REVIEWED BY
Rodrigo Meucci,
Federal University of Rio Grande, Brazil
Joeri Pen,
UZ Brussel, Belgium
Zugui Wu,
The Third Affiliated Hospital of Yunnan University of Chinese Medicine, China

*CORRESPONDENCE
Wei Jiang
✉ 13527580596@163.com
Zhenming Hu
✉ spinecenter@163.com

RECEIVED 16 April 2024
ACCEPTED 24 June 2024
PUBLISHED 03 July 2024

CITATION
Wang Y, Tang Y, Li Z, Jiang C, Jiang W and Hu Z (2024) Sugar-sweetened beverage intake and chronic low back pain. *Front. Nutr.* 11:1418393.
doi: 10.3389/fnut.2024.1418393



# Sugar-sweetened beverage intake and chronic low back pain

Yanting Wang[1,2], Yuchen Tang[1], Zhichun Li[3], Changhai Jiang[2], Wei Jiang[1]* and Zhenming Hu[1,4]*

[1]Department of Orthopedics, Orthopedic Laboratory of Chongqing Medical University, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2]Department of Orthopedics, The People's Hospital of Chuxiong Yi Autonomous Prefecture, Chuxiong, Yunnan, China, [3]Department of General Surgery, Caoxian People's Hospital, Heze, Shandong, China, [4]Department of Orthopedics, University-Town Hospital of Chongqing Medical University, Chongqing, China

**Background:** The consumption of sugar-sweetened beverages (SSBs) has become a major public health problem globally. However, no studies have specifically examined the relationship between SSB intake and chronic low back pain (CLBP). Therefore, the present study aimed to investigate the relationship between SSB intake and the risk of CLBP.

**Methods:** This cross-sectional study enrolled participants aged 20 to 69 from the National Health and Nutrition Examination Survey. CLBP was defined as persistent LBP for a consecutive three-month period. Furthermore, SSB intake was assessed and calculated based on dietary recall interviews. Moreover, survey-weighted logistic regression models were employed to evaluate the association between SSB intake and the risk of CLBP, while the restricted cubic spline (RCS) analysis was used to determine whether there were nonlinear associations between SSB intake and CLBP risk. In addition, subgroup analysis was performed using stratification and interaction analysis for all covariates.

**Results:** A total of 4,146 participants (mean age: 43.405 years) were enrolled in the final analysis. The results of survey-weighted logistic regression models showed that SSB consumption was significantly associated with an increased risk of CLBP among individuals aged 20 to 69 years. Moreover, the results of subgroup analysis and interaction analysis demonstrated that the association between SSB intake and the risk of CLBP was modified by smoking status and hypertension. Specifically, the SSB intake-associated CLBP risk was more pronounced among current smokers or individuals with hypertension.

**Conclusion:** Reduction of SSB consumption might contribute to the prevention of CLBP for individuals aged 20 to 69 years. Moreover, current smokers or individuals with hypertension should be more vigilant about the SSB intake-associated CLBP risk. Nevertheless, caution should be exercised when interpreting the results of this study, as further research is necessary to explore the association between SSB consumption and CLBP, given the limitations of the current study.

KEYWORDS

sugar-sweetened beverage, chronic low back pain, NHANES, smoking, hypertension

## Introduction

Low back pain (LBP) is a prevalent musculoskeletal disorder affecting a significant proportion of adults globally, with a prevalence ranging from 50 to 80% (1, 2). Chronic LBP (CLBP), characterized by pain persisting for more than 3 months and strongly associated with intervertebral disc degeneration (3, 4), is recognized as a major contributor to disability globally (5, 6), and this issue is exacerbated by the aging population and the growth of the population worldwide (7). Currently, there is a growing emphasis on the early prevention of CLBP due to the lack of effective therapeutic strategies. Moreover, cumulative evidence indicates that the pathogenesis of CLBP is complex and is associated with several risk factors, such as age, lifestyle factors, and dietary choices (8, 9). In addition, substantial evidence has implicated that diet and lifestyle interventions have beneficial effects on reducing the risk and improving the condition of CLBP (10, 11). Therefore, the exploration of risk factors for CLBP from the diet and lifestyle perspective has gained considerable attention in recent years and may provide theoretical guidance in the early prevention of CLBP.

Sugar-sweetened beverages (SSBs), including carbonated soft drinks, fruit drinks, and energy drinks, has been demonstrated to be leading sources of added sugars in the diet and to be associated with several adverse health outcomes, such as obesity, oral health, diabetes, and cardiovascular diseases (12–16). Therefore, the consumption of SSBs remains a major public health problem globally (17, 18), which also results in the formulation and implementation of interventions and policies, such as sugary drink warnings or SSB tax (19, 20). Previous evidence has suggested a potential link between high SSB consumption and musculoskeletal disorders, such as low bone mineral density and gout (21, 22). However, to the best of our knowledge, no studies have specifically examined the relationship between SSB intake and CLBP. In addition, it remains unknown whether there are potential factors that modify the association between SSB consumption and the risk of CLBP. Therefore, it is necessary to investigate and understand the relationship between SSB intake and CLBP further, which is crucial and may provide valuable insights into the role of dietary factors in the development and management of CLBP.

Based on the background above, the present study aimed to investigate the relationship between SSB intake and the risk of CLBP and to explore the potential factors that modified the relationship between SSB intake and CLBP, which may have important implications for public health policies, prevention strategies, and patient education regarding CLBP and SSB consumption.

## Materials and methods

### Study design and population

This cross-sectional study included participants from the National Health and Nutrition Examination Surveys (NHANES) 2009–2010, in which the data utilized in the present study is openly accessible on the NHANES website.[1] Participants who received the Inflammatory Arthritis Questionnaire, which was employed for CLBP assessment, were included in the present study. Moreover, the exclusion criteria for participants were listed as follows: (i) with missing data on SSB intake; (ii) with missing data on covariates. Furthermore, ethical approval for the NHANES was obtained from the ethics review board of the National Center for Health Statistics (23). All participants in the NHANES study were duly provided with and acknowledged informed consent (24). The present study conducted was a secondary analysis of deidentified, publicly available data, thus obviating the need for ethics approval. Additional comprehensive information was accessible on the NHANES website (25).

### CLBP assessment

CLBP, in which the definition was employed with reference to several previous studies (26, 27), was evaluated using the Inflammatory Arthritis Questionnaire [offering interview data pertaining to chronic back pain, Inflammatory Back Pain (IBP), and Spondyloarthritis (Spondyloarthritis or Spinal Arthritis)] (28, 29), with the study population consisting of a representative sample of United States adults aged 20 to 69 years. Moreover, all participants who received the Inflammatory Arthritis Questionnaire underwent the same assessments for CLBP, and a participant who was asked the question, "Had low back pain 3 months in a row?" met the criteria for CLBP if they reported experiencing persistent LBP for a consecutive three-month period. Detailed information on the Inflammatory Arthritis Questionnaire is available on the NHANES website (28, 29).

### SSB intake

SSB intake was evaluated through 24-h dietary recall interviews, which captured the consumption of various foods and beverages in the preceding 24 h. All reported food and beverage items were meticulously coded using the US Department of Agriculture (USDA) Food and Nutrient Database. Soft drinks, fruit drinks (not 100%), sports drinks, energy drinks, nutritional beverages, smoothies, grain drinks, carbonated water, and sweetened coffee and tea were considered the SSBs in the present study. The caloric content and nutrient composition of SSBs were determined by analyzing the reported quantities of food and beverages in conjunction with the nutrient data provided by the National Center for Health Statistics. Additional information regarding the methodology of dietary recall interviews can be accessed on the NHANES website (30).

### Covariates

Several demographic variables and variables considered as potential confounders of the relationship between SSB intakes and the risk of CLBP were included as the covariates in the subsequent analysis. Age, sex, race/ethnicity, education levels, body mass index (BMI), smoking status, drinking status, physical activity levels (mins/week, which were assessed by the Global physical activity questionnaire (GPAQ) (31) and included five aspects: vigorous work-related activity, moderate work-related activity, walking or bicycling for transportation, vigorous leisure-time physical activity, and moderate work-related activity), hypertension (diagnosed by doctors), diabetes (diagnosed by doctors), cancer (diagnosed by doctors),

1 https://www.cdc.gov/nchs/nhanes/index.htm

 10.3389/fnut.2024.1418393

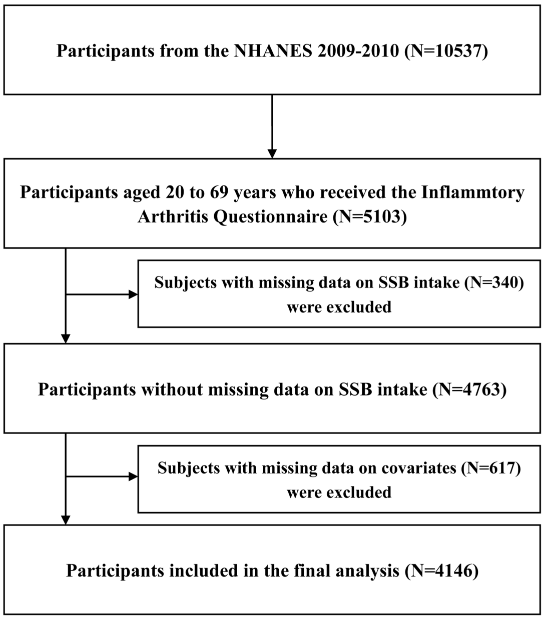



FIGURE 1
Flowchart of participants selection. CLBP, chronic low back pain; NHANES, National Health and Nutrition Examination Survey; SSB, sugar-sweetened beverage.

C-reactive protein (CRP), and total energy intake were selected as the covariates of the present study.

## Statistical analysis

Baseline characteristics of the study population were reported as means [standard errors (SEs)] for continuous variables and unweighted numbers (weighted proportions) for categorical variables, in which nationally representative estimates were calculated for all analyses by utilizing the recommended NHANES examinations sample weights (32). Furthermore, the differences between individuals with and without CLBP were assessed by survey-weighted linear regression models for continuous variables and survey-weighted Chi-square test for categorical variables. Moreover, the weighted binomial logistic regression models were employed to determine the association between SSB intake and the risk of CLBP and to calculate the odds ratios (ORs) and 95% confidence intervals (CIs), while restricted cubic spline (RCS) curves based on survey-weighted binomial logistic regression models were used to examine whether there were significant nonlinear associations between SSB intake and the risk of CLBP. In addition, subgroup analysis was performed using stratification and interaction analysis for all covariates mentioned above to determine whether there were potential factors that modified the association between SSB intake and the risk of CLBP. Statistical analyses were performed using R software version 4.2.1[2] and EmpowerStats version 4.2.[3] Two-sided $p$-values were utilized, with significance defined as $p < 0.05$.

2 https://cran.r-project.org/
3 http://www.empowerstats.com

# Results

## Study population selection

Overall, 10,537 participants from the NHANES 2009–2010 were included in this cross-sectional study, in which 5,103 participants aged 20–69 years received the Inflammatory Arthritis Questionnaire. Furthermore, individuals with incomplete data regarding SSB consumption ($N = 340$) or covariates ($N = 617$) were excluded from the analysis. Ultimately, a cohort of 4,146 participants was deemed suitable for inclusion in the final analysis. The selection process of the study population is visually represented in Figure 1.

## Baseline characteristics

Finally, 4,146 participants aged 20 to 69 years were included in the final analysis, and weighted samples of participants represent a population of 171,120,866. The mean age of the study population was 43.405 (0.382) years, and 50.062% were women. Furthermore, participants with CLBP tended to be older and show a higher prevalence of obesity, smoking, hypertension, diabetes, and cancer than those without CLBP. Moreover, the mean SSB intakes of the overall population, participants with CLBP, and those without CLBP were 120.017 (5.452) kcal/d, 149.249 (9.885) kcal/d, 114.911 (5.589) kcal/d, respectively, in which participants with CLBP showed significantly higher SSB intakes than those without CLBP. Other baseline characteristics of the study population are listed in Table 1.

## Association between SSB intake and CLBP

The results of weighted logistic regression models (Table 2) indicated that higher SSB intake (as a continuous variable) was associated with an increased risk of CLBP with or without adjustment for covariates. Moreover, when SSB intake was converted to a categorical variable (no SSB intake: 0 kcal/d, low SSB intake: 0–199 kcal/d, and high SSB intake: ≥200 kcal/d) according to the data distribution of SSB intake (Figure 2), participants with high SSB intake showed an elevated risk of CLBP compared with those with no SSB intake with or without adjusting for covariates. In addition, the results of RCS models (Figure 3) suggested that there were no significant nonlinear associations between SSB intake and the risk of CLBP with or without adjustment for covariates (P for nonlinear >0.05).

## Subgroup analysis

The results of subgroup analysis (Figure 4) demonstrated that higher SSB intake was associated with an increased risk of CLBP, which was observed in most of the subgroups with or without adjusting for covariates. Moreover, the results of interaction analysis suggested (Figure 4) that the association between SSB intake and the risk of CLBP were modified by smoking status and hypertension after adjusting for covariates (P for interaction <0.05). Furthermore, the results of weighted logistic regression analysis (Table 3) showed that current smokers, irrespective of the SSB intake, showed a significantly elevated risk of CLBP, and former smokers with high SSB intake

TABLE 1 Baseline characteristics.

| Characteristic | Total (*N* = 4,146)[a] | Without CLBP (*N* = 3,549)[a] | With CLBP (*N* = 597)[a] | *p*-value |
|---|---|---|---|---|
| *Age (years)* | 43.405 (0.382) | 42.963 (0.375) | 45.939 (0.765) | < 0.001 |
| *Sex* | | | | 0.644 |
| Men | 2,048 (49.938) | 1,771 (50.146) | 277 (48.745) | |
| Women | 2,098 (50.062) | 1,778 (49.854) | 320 (51.255) | |
| *Race/ethnicity* | | | | 0.070 |
| Non-Hispanic White | 1,876 (68.345) | 1,551 (67.350) | 325 (74.043) | |
| Non-Hispanic Black | 742 (11.218) | 651 (11.491) | 91 (9.656) | |
| Mexican American | 847 (8.847) | 745 (8.992) | 102 (8.018) | |
| Other races | 681 (11.590) | 602 (12.168) | 79 (8.283) | |
| *Education level* | | | | 0.027 |
| Under high school | 1,106 (17.044) | 937 (16.605) | 169 (19.557) | |
| High school or equivalent | 944 (22.000) | 793 (21.277) | 151 (26.142) | |
| Above high school | 2,096 (60.956) | 1,819 (62.118) | 277 (54.301) | |
| BM[b] | | | | 0.013 |
| Normal | 1,148 (30.308) | 1,023 (31.723) | 125 (22.206) | |
| Overweight | 1,375 (32.591) | 1,185 (32.399) | 190 (33.693) | |
| Obese | 1,623 (37.101) | 1,341 (35.878) | 282 (44.101) | |
| *Smoking status* | | | | < 0.001 |
| Never | 2,264 (55.392) | 2035 (57.570) | 229 (42.921) | |
| Former | 868 (22.781) | 709 (21.933) | 159 (27.635) | |
| Current | 1,014 (21.828) | 805 (20.497) | 209 (29.444) | |
| Drinking status[c] | | | | 0.406 |
| Never | 445 (8.571) | 393 (8.691) | 52 (7.879) | |
| Former | 537 (10.224) | 464 (10.012) | 73 (11.440) | |
| Current | 3,164 (81.205) | 2,692 (81.297) | 472 (80.682) | |
| *PA levels (mins/week)* | 688.890 (35.566) | 675.499 (34.302) | 765.556 (65.871) | 0.115 |
| *Hypertension* | | | | < 0.001 |
| Yes | 1,171 (25.093) | 921 (23.304) | 250 (35.339) | |
| No | 2,975 (74.907) | 2,628 (76.696) | 347 (64.661) | |
| *Diabetes* | | | | 0.009 |
| Yes | 400 (6.713) | 313 (6.064) | 87 (10.430) | |
| No | 3,746 (93.287) | 3,236 (93.936) | 510 (89.570) | |
| *Cancer* | | | | 0.005 |
| Yes | 258 (7.440) | 199 (6.760) | 59 (11.329) | |
| No | 3,888 (92.560) | 3,350 (93.240) | 538 (88.671) | |
| *CRP (mg/dL)* | 0.360 (0.017) | 0.352 (0.017) | 0.408 (0.032) | 0.055 |
| *Total energy intake (kcal/d)* | 2198.543 (19.625) | 2193.157 (19.816) | 2229.378 (58.626) | 0.557 |
| *SSB intake (kcal/d)* | 120.017 (5.452) | 114.911 (5.589) | 149.249 (9.885) | 0.002 |

[a]Unweighted number.
[b]Normal: <25 kg/m$^2$ ; Overweight: <30 but ≥ 25 kg/m$^2$; Obese: ≥30 kg/m$^2$.
[c]Never: participants who did not have at least 12 alcohol drinks in a lifetime; Former: participants who had at least 12 alcohol drinks in a lifetime but did not have at least 12 alcohol drinks for last 1 year; Current: participants who had at least 12 alcohol drinks in a lifetime and had at least 12 alcohol drinks for last 1 year. BMI, body mass index; CLBP, chronic low back pain; CRP, C-reactive protein; PA, physical activity; SSB, sugar-sweetened beverage.

showed a significantly increased risk of CLBP compared with never smokers with no SSB intake with or without adjustment for covariates. In addition, this study observed (Table 4) that only the hypertension group with high SSB intake showed a significantly elevated risk of CLBP compared with the non-hypertension group with no SSB intake after adjusting for all covariates.

10.3389/fnut.2024.1418393

TABLE 2 Association between SSB intake and the risk of CLBP.

| | Model 1[a] | | Model 2[b] | | Model 3[c] | |
|---|---|---|---|---|---|---|
| | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value |
| SSB intake (continuous variable) (Per 100 kcal/d increase) | 1.071 (1.035, 1.107) | <0.001 | 1.101 (1.059, 1.144) | <0.001 | 1.069 (1.022, 1.117) | 0.006 |
| SSB intake (categorical variable) | | | | | | |
| Group 1: 0 kcal/d | Ref (1) | – | Ref (1) | – | Ref (1) | – |
| Group 2: 1–199 kcal/d | 1.216 (0.763, 1.938) | 0.383 | 1.237 (0.766, 1.998) | 0.342 | 1.163 (0.716, 1.889) | 0.519 |
| Group 3: ≥200 kcal/d | 1.653 (1.232, 2.217) | 0.003 | 1.939 (1.432, 2.624) | <0.001 | 1.647 (1.163, 2.333) | 0.008 |
| P for trend | 0.005 | | 0.002 | | 0.018 | |

[a]Adjustment for no covariates were adjusted.
[b]Adjustment for age, sex, and race/ethnicity were adjusted.
[c]Adjustment for all covariates (including age, sex, race/ethnicity, education levels, BMI, smoking status, drinking status, PA levels, hypertension, diabetes, cancer, CRP, and total energy intake) were adjusted. BMI, body mass index; CI, confidence interval; CLBP, chronic low back pain; CRP, C-reactive protein; OR, odds ratio; PA, physical activity; SSB, sugar-sweetened beverage.

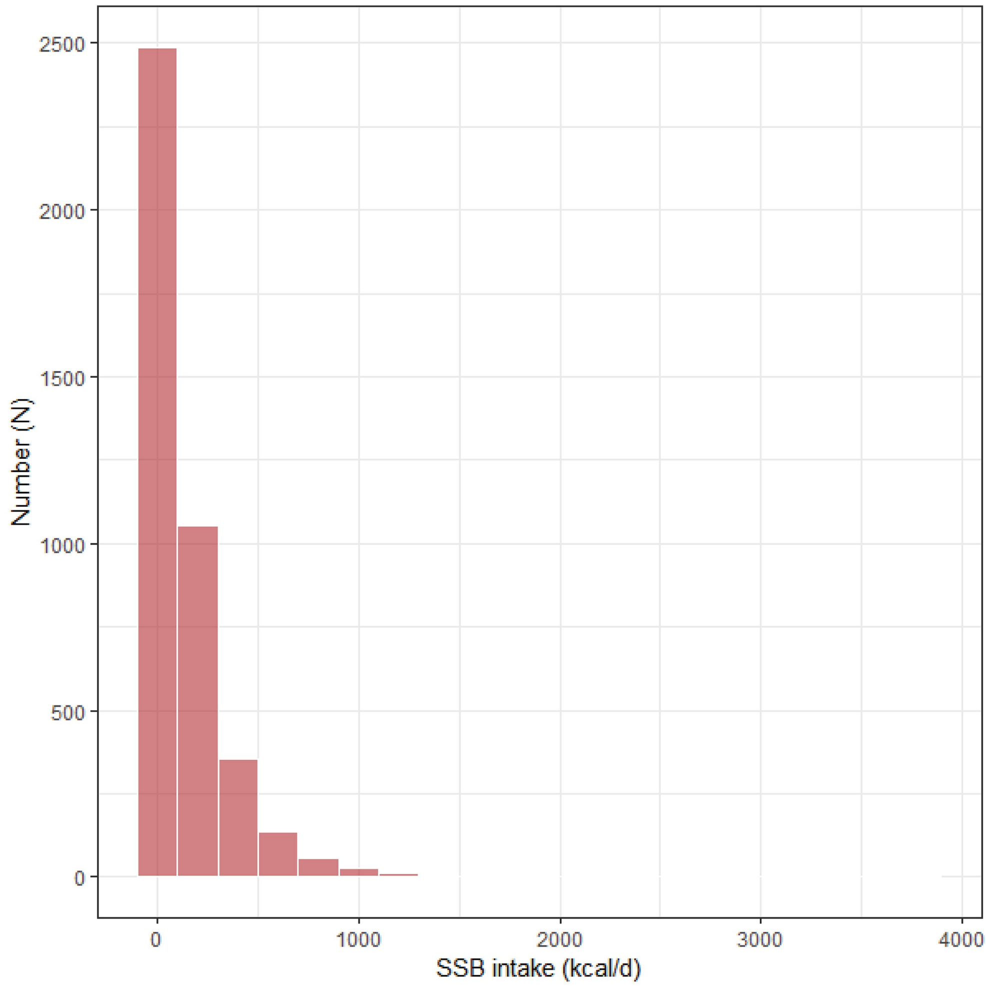



FIGURE 2
The data distribution of SSB intake. SSB, sugar-sweetened beverage.

## Discussion

Overall, this cross-sectional study observed that SSB consumption was significantly associated with an increased risk of CLBP among individuals aged 20 to 69 years. Moreover, we found that the association between SSB intake and the risk of CLBP was modified by smoking status and hypertension, in which the SSB intake-associated CLBP risk was more pronounced among current smokers or individuals with hypertension.

10.3389/fnut.2024.1418393

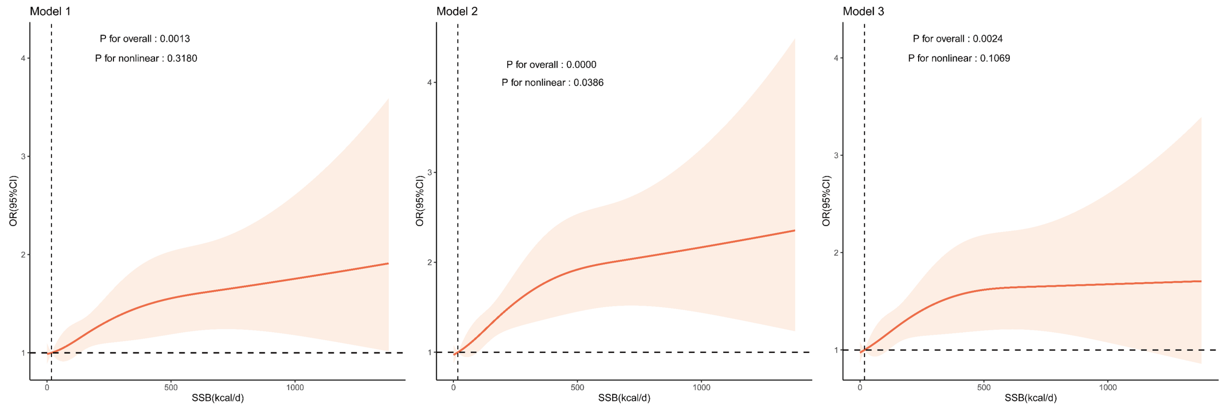
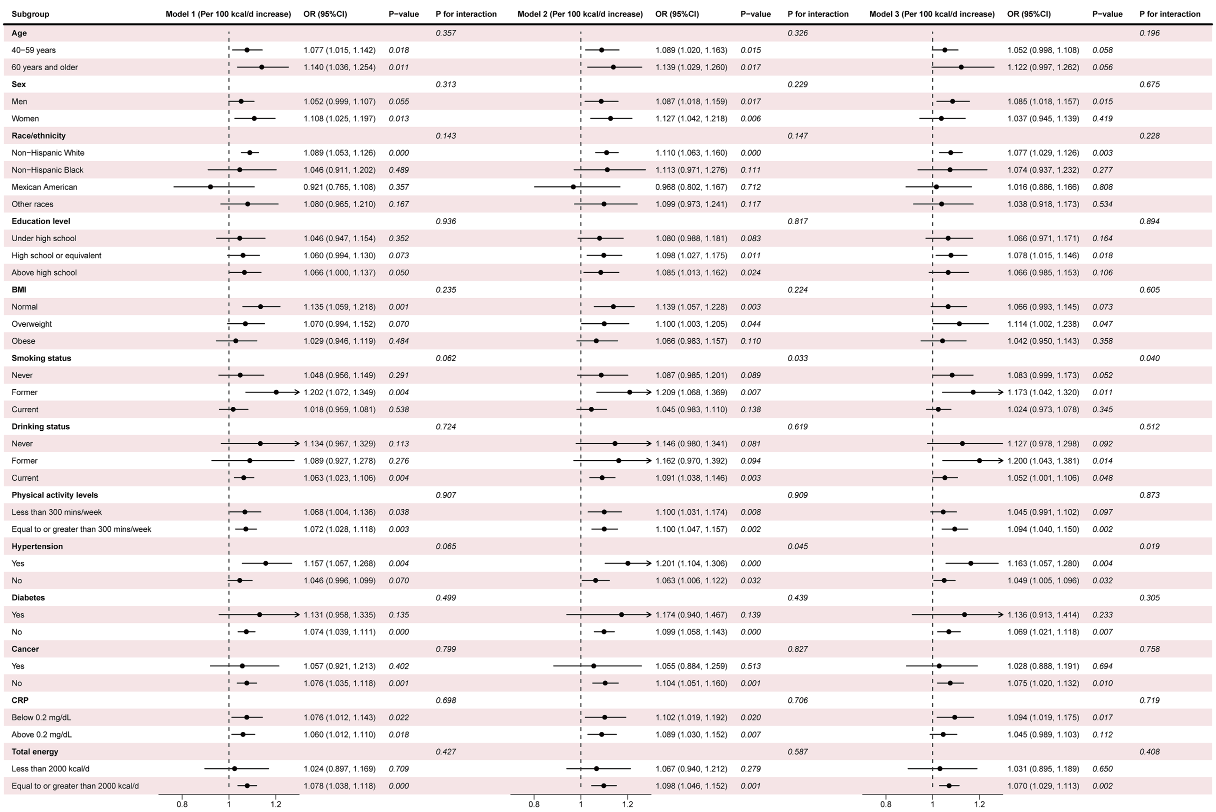



FIGURE 3
Relationship between SSB intake and the risk of CLBP. Model 1: adjustment for no covariates; Model 2: adjustment for age, sex, and race/ethnicity; Model 3: adjustment for all covariates. Data were fitted by a restricted cubic spline linear regression model, and the model was conducted with 4 knots at the 5th, 35th, 65th, 95th percentiles of SSB intake (reference is the median). Solid lines indicate OR values, and shadow shape indicates 95% CIs. CLBP, chronic low back pain; CI, confidence interval; OR, odds ratio; SSB, sugar-sweetened beverage.



| Subgroup | Model 1 OR (95%CI) | P-value | P for interaction | Model 2 OR (95%CI) | P-value | P for interaction | Model 3 OR (95%CI) | P-value | P for interaction |
|---|---|---|---|---|---|---|---|---|---|
| **Age** | | | 0.357 | | | 0.326 | | | 0.196 |
| 40–59 years | 1.077 (1.015, 1.142) | 0.018 | | 1.089 (1.020, 1.163) | 0.015 | | 1.052 (0.998, 1.108) | 0.058 | |
| 60 years and older | 1.140 (1.036, 1.254) | 0.011 | | 1.139 (1.029, 1.260) | 0.017 | | 1.122 (0.997, 1.262) | 0.056 | |
| **Sex** | | | 0.313 | | | 0.229 | | | 0.675 |
| Men | 1.052 (0.999, 1.107) | 0.055 | | 1.087 (1.018, 1.159) | 0.017 | | 1.085 (1.018, 1.157) | 0.015 | |
| Women | 1.108 (1.025, 1.197) | 0.013 | | 1.127 (1.042, 1.218) | 0.006 | | 1.037 (0.945, 1.139) | 0.419 | |
| **Race/ethnicity** | | | 0.143 | | | 0.147 | | | 0.228 |
| Non–Hispanic White | 1.089 (1.053, 1.126) | 0.000 | | 1.110 (1.063, 1.160) | 0.000 | | 1.077 (1.029, 1.126) | 0.003 | |
| Non–Hispanic Black | 1.046 (0.911, 1.202) | 0.489 | | 1.113 (0.971, 1.276) | 0.111 | | 1.074 (0.937, 1.232) | 0.277 | |
| Mexican American | 0.921 (0.765, 1.108) | 0.357 | | 0.968 (0.802, 1.167) | 0.712 | | 1.016 (0.886, 1.166) | 0.808 | |
| Other races | 1.080 (0.965, 1.210) | 0.167 | | 1.099 (0.973, 1.241) | 0.117 | | 1.038 (0.918, 1.173) | 0.534 | |
| **Education level** | | | 0.936 | | | 0.817 | | | 0.894 |
| Under high school | 1.046 (0.947, 1.154) | 0.352 | | 1.080 (0.988, 1.181) | 0.083 | | 1.066 (0.971, 1.171) | 0.164 | |
| High school or equivalent | 1.060 (0.994, 1.130) | 0.073 | | 1.098 (1.027, 1.175) | 0.011 | | 1.078 (1.015, 1.146) | 0.018 | |
| Above high school | 1.066 (1.000, 1.137) | 0.050 | | 1.085 (1.013, 1.162) | 0.024 | | 1.066 (0.985, 1.153) | 0.106 | |
| **BMI** | | | 0.235 | | | 0.224 | | | 0.605 |
| Normal | 1.135 (1.059, 1.218) | 0.001 | | 1.139 (1.057, 1.228) | 0.003 | | 1.066 (0.993, 1.145) | 0.073 | |
| Overweight | 1.070 (0.994, 1.152) | 0.070 | | 1.100 (1.003, 1.205) | 0.044 | | 1.114 (1.002, 1.238) | 0.047 | |
| Obese | 1.029 (0.946, 1.119) | 0.484 | | 1.066 (0.983, 1.157) | 0.110 | | 1.042 (0.950, 1.143) | 0.358 | |
| **Smoking status** | | | 0.062 | | | 0.033 | | | 0.040 |
| Never | 1.048 (0.956, 1.149) | 0.291 | | 1.087 (0.985, 1.201) | 0.089 | | 1.083 (0.999, 1.173) | 0.052 | |
| Former | 1.202 (1.072, 1.349) | 0.004 | | 1.209 (1.068, 1.369) | 0.007 | | 1.173 (1.042, 1.320) | 0.011 | |
| Current | 1.018 (0.959, 1.081) | 0.538 | | 1.045 (0.983, 1.110) | 0.138 | | 1.024 (0.973, 1.078) | 0.345 | |
| **Drinking status** | | | 0.724 | | | 0.619 | | | 0.512 |
| Never | 1.134 (0.967, 1.329) | 0.113 | | 1.146 (0.980, 1.341) | 0.081 | | 1.127 (0.978, 1.298) | 0.092 | |
| Former | 1.089 (0.927, 1.278) | 0.276 | | 1.162 (0.970, 1.392) | 0.094 | | 1.200 (1.043, 1.381) | 0.014 | |
| Current | 1.063 (1.023, 1.106) | 0.004 | | 1.091 (1.038, 1.146) | 0.003 | | 1.052 (1.001, 1.106) | 0.048 | |
| **Physical activity levels** | | | 0.907 | | | 0.909 | | | 0.873 |
| Less than 300 mins/week | 1.068 (1.004, 1.136) | 0.038 | | 1.100 (1.031, 1.174) | 0.008 | | 1.045 (0.991, 1.102) | 0.097 | |
| Equal to or greater than 300 mins/week | 1.072 (1.028, 1.118) | 0.003 | | 1.100 (1.047, 1.157) | 0.002 | | 1.094 (1.040, 1.150) | 0.002 | |
| **Hypertension** | | | 0.065 | | | 0.045 | | | 0.019 |
| Yes | 1.157 (1.057, 1.268) | 0.004 | | 1.201 (1.104, 1.306) | 0.000 | | 1.163 (1.057, 1.280) | 0.004 | |
| No | 1.046 (0.996, 1.099) | 0.070 | | 1.063 (1.006, 1.122) | 0.032 | | 1.049 (1.005, 1.096) | 0.032 | |
| **Diabetes** | | | 0.499 | | | 0.439 | | | 0.305 |
| Yes | 1.131 (0.958, 1.335) | 0.135 | | 1.174 (0.940, 1.467) | 0.139 | | 1.136 (0.913, 1.414) | 0.233 | |
| No | 1.074 (1.039, 1.111) | 0.000 | | 1.099 (1.058, 1.143) | 0.000 | | 1.069 (1.021, 1.118) | 0.007 | |
| **Cancer** | | | 0.799 | | | 0.827 | | | 0.758 |
| Yes | 1.057 (0.921, 1.213) | 0.402 | | 1.055 (0.884, 1.259) | 0.513 | | 1.028 (0.888, 1.191) | 0.694 | |
| No | 1.076 (1.035, 1.118) | 0.001 | | 1.104 (1.051, 1.160) | 0.001 | | 1.075 (1.020, 1.132) | 0.010 | |
| **CRP** | | | 0.698 | | | 0.706 | | | 0.719 |
| Below 0.2 mg/dL | 1.076 (1.012, 1.143) | 0.022 | | 1.102 (1.019, 1.192) | 0.020 | | 1.094 (1.019, 1.175) | 0.017 | |
| Above 0.2 mg/dL | 1.060 (1.012, 1.110) | 0.018 | | 1.089 (1.030, 1.152) | 0.007 | | 1.045 (0.989, 1.103) | 0.112 | |
| **Total energy** | | | 0.427 | | | 0.587 | | | 0.408 |
| Less than 2000 kcal/d | 1.024 (0.897, 1.169) | 0.709 | | 1.067 (0.940, 1.212) | 0.279 | | 1.031 (0.895, 1.189) | 0.650 | |
| Equal to or greater than 2000 kcal/d | 1.078 (1.038, 1.118) | 0.000 | | 1.098 (1.046, 1.152) | 0.001 | | 1.070 (1.029, 1.113) | 0.002 | |

FIGURE 4
Subgroup analysis and interaction testing. Model 1: adjustment for no covariates; Model 2: adjustment for age, sex, and race/ethnicity; Model 3: adjustment for all covariates. Covariates were not adjusted when stratified by their respective variables. 25(OH) D, 25-hydroxyvitamin D; BMD, bone mineral density; BMI, body mass index; CI, confidence interval; CRP, C-reactive protein; OR, odds ratio.

SSB consumption, which has been demonstrated to be associated with several adverse health outcomes (12–16), has become a major public health problem worldwide (17, 18). In the present study, we observed a significant association between the consumption of SSBs and an increased CLBP risk, the specific mechanisms of which are yet to be elucidated. However, we speculate that there are several possible causes of this phenomenon, including inflammatory, metabolic, nutritional,

TABLE 3 Association between SSB intake, smoking status, and the risk of CLBP.

| SSB intake | Smoking status | Model 1[a] | | Model 2[b] | | Model 3[c] | |
|---|---|---|---|---|---|---|---|
| | | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value |
| No SSB intake: 0 kcal/d | Never | Ref (1) | – | Ref (1) | – | Ref (1) | – |
| | Former | 1.671 (0.970, 2.879) | 0.061 | 1.479 (0.695, 3.144) | 0.198 | 1.511 (0.911, 2.506) | 0.103 |
| | Current | 2.211 (1.396, 3.501) | 0.004 | 2.251 (1.228, 4.126) | 0.024 | 2.252 (1.383, 3.666) | 0.003 |
| Low SSB intake: 1–199 kcal/d | Never | 1.264 (0.639, 2.501) | 0.451 | 1.264 (0.497, 3.212) | 0.483 | 1.202 (0.619, 2.337) | 0.565 |
| | Former | 1.962 (0.901, 4.270) | 0.081 | 1.750 (0.587, 5.220) | 0.201 | 1.585 (0.701, 3.588) | 0.249 |
| | Current | 2.482 (1.329, 4.636) | 0.010 | 2.632 (1.107, 6.260) | 0.038 | 2.668 (1.474, 4.830) | 0.003 |
| High SSB intake: ≥200 kcal/d | Never | 1.700 (0.991, 2.917) | 0.053 | 1.977 (0.940, 4.160) | 0.062 | 1.835 (1.121, 3.005) | 0.019 |
| | Former | 3.696 (2.026, 6.741) | 0.001 | 3.759 (1.774, 7.968) | 0.011 | 3.372 (1.912, 5.949) | <0.001 |
| | Current | 2.380 (1.302, 4.352) | 0.011 | 2.792 (1.299, 6.001) | 0.024 | 2.618 (1.456, 4.705) | 0.003 |

[a]Adjustment for no covariates.
[b]Adjustment for age, sex, and race/ethnicity.
[c]Adjustment for all covariates (including age, sex, race/ethnicity, education levels, BMI, drinking status, PA levels, hypertension, diabetes, cancer, CRP, and total energy intake). BMI, body mass index; CI, confidence interval; CLBP, chronic low back pain; CRP, C-reactive protein; OR, odds ratio; PA, physical activity; SSB, sugar-sweetened beverage.

TABLE 4 Association between SSB intake, hypertension, and the risk of CLBP.

| SSB intake | History of hypertension | Model 1[a] | | Model 2[b] | | Model 3[c] | |
|---|---|---|---|---|---|---|---|
| | | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value | OR (95%CI) | *p*-value |
| No SSB intake: 0 kcal/d | No | Ref (1) | – | Ref (1) | – | Ref (1) | – |
| | Yes | 1.806 (1.311, 2.489) | 0.002 | 1.604 (1.120, 2.298) | 0.018 | 1.333 (0.974, 1.824) | 0.070 |
| Low SSB intake: 1–199 kcal/d | No | 1.337 (0.719, 2.486) | 0.326 | 1.364 (0.702, 2.650) | 0.296 | 1.307 (0.706, 2.420) | 0.371 |
| | Yes | 1.711 (0.961, 3.048) | 0.065 | 1.519 (0.807, 2.861) | 0.157 | 1.236 (0.653, 2.338) | 0.492 |
| High SSB intake: ≥200 kcal/d | No | 1.432 (0.948, 2.163) | 0.082 | 1.608 (0.980, 2.639) | 0.057 | 1.347 (0.877, 2.068) | 0.160 |
| | Yes | 4.298 (2.522, 7.324) | <0.0001 | 4.304 (2.356, 7.862) | 0.001 | 3.411 (1.753, 6.637) | 0.001 |

[a]Adjustment for no covariates.
[b]Adjustment for age, sex, and race/ethnicity.
[c]Adjustment for all covariates (including age, sex, race/ethnicity, education levels, BMI, smoking status, drinking status, PA levels, diabetes, cancer, CRP, and total energy intake). BMI, body mass index; CI, confidence interval; CLBP, chronic low back pain; CRP, C-reactive protein; OR, odds ratio; PA, physical activity; SSB, sugar-sweetened beverage.

lifestyle, and psychological factors. For example, SSBs are known to have high levels of added sugars, which can lead to elevated inflammation levels in the body (33), which is believed to play a role in the development and persistence of pain, including CLBP (34–36). Furthermore, regular consumption of SSBs has been demonstrated to be associated with an elevated risk of weight gain, obesity, or diabetes (15, 37), which are also considered important risk factors for CLBP reported by numerous studies (8, 26, 38). Moreover, it is possible that individuals who consume higher amounts of SSBs might also have additional risk factors for CLBP, such as a sedentary lifestyle and higher stress levels (39, 40), which may be a possible explanation for the association between SSB consumption and an increased risk of CLBP. In addition, it should be noted that simple carbohydrates, such as fructose, have been demonstrated to have a direct nociceptive effect on pain sensation (41), which is also a probable cause for the association between high SSB consumption and the increased risk of CLBP. However, additional investigations are required to support our speculation due to the cross-sectional study design, which does not allow causal associations to be drawn.

Interestingly, this study observed that the association between SSB intake and the risk of CLBP was modified by smoking status and hypertension, in which the SSB intake-associated CLBP risk was more pronounced among current smokers or individuals with hypertension, suggesting that there might be a synergistic effect between SSB intake and smoking, as well as hypertension, in CLBP. On the one hand, both smoking and hypertension can contribute to elevated inflammation levels in the body (42, 43). SSBs, with their high sugar content, may further exacerbate inflammation levels (33). The synergistic effects of smoking, hypertension, and SSB consumption may lead to an even higher level of systemic inflammation, which has been demonstrated to be associated with an increased risk of CLBP (34–36). On the other hand, current smokers or individuals with hypertension may have other lifestyle factors that contribute to their increased risk of CLBP when combined with SSB consumption, such as poor dietary habits or a sedentary lifestyle, all of which can independently contribute to the development of CLBP (39, 40, 44). However, it should be noted that these potential reasons mentioned above are based on observations and correlations, and further research is needed to fully understand the underlying mechanisms and causality between SSB intake, smoking, hypertension, and CLBP.

The main findings of this study have implications for future clinical practice. To the best of our knowledge, this study is the first to investigate the association between SSB consumption and CLBP risk. Moreover, this study found a significant association between the consumption of SSBs and an increased risk of CLBP among individuals aged 20 to 69 years, which implies that SSB consumption may contribute to the development or progression of CLBP, while the reduction in SSB intake may serve to protect from CLBP. Furthermore, this study observed that SSB intake-associated CLBP risk was more pronounced among current smokers or individuals with hypertension. Therefore, these special populations need to be aware of the potential synergistic impact on CLBP risk. In addition, limiting SSB intake and addressing other risk factors, such as smoking and hypertension, may help reduce the burden of CLBP in the population.

This study is subject to certain limitations. Firstly, the cross-sectional study design utilized in this research precludes the establishment of a causal relationship between SSB intake and the risk of CLBP. Secondly, data on SSB intake and various covariates, including smoking status and history of hypertension, were obtained through dietary recall interviews or self-report questionnaires, potentially introducing reporting bias or recall bias. Thirdly, it should be noted that the participants in this study were drawn from the NHANES database, which represents the US population, suggesting the generalizability of the findings to populations in other countries or regions may be limited. Fourthly, the sample size of participants with CLBP was relatively small, which might influence the precision of estimation. Consequently, further research investigating the association between SSB consumption and the risk of CLBP is warranted to enhance the robustness of the evidence.

## Conclusion

SSB consumption was significantly associated with an elevated risk of CLBP among individuals aged 20 to 69 years, suggesting that the reduction in SSB intake might contribute to the prevention of CLBP. Moreover, the association between SSB intake and CLBP risk was modified by several lifestyles and diseases, including smoking and hypertension, suggesting such individuals should be more vigilant about the SSB intake-associated CLBP risk. However, the results from this study should be interpreted with caution, and additional studies are required in the future further to investigate the relationship between SSB consumption and CLBP, considering that there are several limitations of the present study.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: the datasets obtained and analyzed in this study are publicly available on the NHANES database (https://www.cdc.gov/nchs/nhanes/index.htm).

## Ethics statement

The NHANES was granted approval by the National Center for Health Statistics Ethics Review Board. All participants from the NHANES were duly provided with and acknowledged the informed consent. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

YW: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. YT: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. ZL: Investigation, Validation, Writing – review & editing. CJ: Validation, Writing – review & editing. WJ: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. ZH: Conceptualization, Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (No. 82002363) and Natural Science Foundation of Chongqing, China (No. cstc2020jcyj-msxmX0195).

## Acknowledgments

We thank the National Health and Nutrition Examination Survey participants and staff and the National Center for Health Statistics for their valuable contributions.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# References


1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low Back pain is and why we need to pay attention. *Lancet*. (2018) 391:2356–67. doi: 10.1016/s0140-6736(18)30480-x

2. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low Back pain. *Lancet*. (2021) 398:78–92. doi: 10.1016/s0140-6736(21)00733-9

3. Livshits G, Popham M, Malkin I, Sambrook PN, Macgregor AJ, Spector T, et al. Lumbar disc degeneration and genetic factors are the Main risk factors for low Back pain in women: the Uk twin spine study. *Ann Rheum Dis*. (2011) 70:1740–5. doi: 10.1136/ard.2010.137836

4. Global, Regional, and National Incidence. Prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. (2018) 392:1789–858. doi: 10.1016/s0140-6736(18)32279-7

5. Mokdad AH, Ballestros K, Echko M, Glenn S, Olsen HE, Mullany E, et al. The state of us health, 1990-2016: burden of diseases, injuries, and risk factors among us states. *JAMA*. (2018) 319:1444–72. doi: 10.1001/jama.2018.0158

6. Global, Regional, and National Incidence. Prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the global burden of disease study 2015. *Lancet*. (2016) 388:1545–602. doi: 10.1016/s0140-6736(16)31678-6

7. Clark S, Horton R. Low Back pain: a major global challenge. *Lancet*. (2018) 391:2302. doi: 10.1016/s0140-6736(18)30725-6

8. Wong CK, Mak RY, Kwok TS, Tsang JS, Leung MY, Funabashi M, et al. Prevalence, incidence, and factors associated with non-specific chronic low Back pain in community-dwelling older adults aged 60 years and older: a systematic review and Meta-analysis. *J Pain*. (2022) 23:509–34. doi: 10.1016/j.jpain.2021.07.012

9. Islam SMS, Uddin R, Das S, Ahmed SI, Zaman SB, Alif SM, et al. The burden of diseases and risk factors in Bangladesh, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet Glob Health*. (2023) 11:e1931–42. doi: 10.1016/s2214-109x(23)00432-1

10. Herrero P, Val P, Lapuente-Hernández D, Cuenca-Zaldívar JN, Calvo S, Gómez-Trullén EM. Effects of lifestyle interventions on the improvement of chronic non-specific low Back pain: A systematic review and network Meta-analysis. *Healthcare*. (2024) 12:505. doi: 10.3390/healthcare12050505

11. Liu S, Lv X, Deng X, Lai R, Du J, Wang C. Diet and risk of low Back pain: a Mendelian randomization analysis. *Eur Spine J*. (2024) 33:496–504. doi: 10.1007/s00586-023-07970-4

12. Malik VS, Hu FB. Sugar-sweetened beverages and Cardiometabolic health: an update of the evidence. *Nutrients*. (2019) 11:1840. doi: 10.3390/nu11081840

13. Malik VS, Hu FB. Fructose and Cardiometabolic health: what the evidence from sugar-sweetened beverages tells us. *J Am Coll Cardiol*. (2015) 66:1615–24. doi: 10.1016/j.jacc.2015.08.025

14. Valenzuela MJ, Waterhouse B, Aggarwal VR, Bloor K, Doran T. Effect of sugar-sweetened beverages on Oral health: a systematic review and Meta-analysis. *Eur J Pub Health*. (2021) 31:122–9. doi: 10.1093/eurpub/ckaa147

15. Malik VS, Hu FB. The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases. *Nat Rev Endocrinol*. (2022) 18:205–18. doi: 10.1038/s41574-021-00627-6

16. Malik VS, Popkin BM, Bray GA, Després JP, Willett WC, Hu FB. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a Meta-analysis. *Diabetes Care*. (2010) 33:2477–83. doi: 10.2337/dc10-1079

17. Singh GM, Micha R, Khatibzadeh S, Lim S, Ezzati M, Mozaffarian D. Estimated global, regional, and National Disease Burdens Related to sugar-sweetened beverage consumption in 2010. *Circulation*. (2015) 132:639–66. doi: 10.1161/circulationaha.114.010636

18. Popkin BM, Hawkes C. Sweetening of the global diet, particularly beverages: patterns, trends, and policy responses. *Lancet Diabetes Endocrinol*. (2016) 4:174–86. doi: 10.1016/s2213-8587(15)00419-2

19. Grummon AH, Hall MG. Sugary drink warnings: a Meta-analysis of experimental studies. *PLoS Med*. (2020) 17:e1003120. doi: 10.1371/journal.pmed.1003120

20. Andreyeva T, Marple K, Marinello S, Moore TE, Powell LM. Outcomes following taxation of sugar-sweetened beverages: a systematic review and Meta-analysis. *JAMA Netw Open*. (2022) 5:e2215276. doi: 10.1001/jamanetworkopen.2022.15276

21. Bragança M, Bogea EG, de Almeida Fonseca Viola PC, Dos Santos Vaz J, Confortin SC, AMB M, et al. High consumption of sugar-sweetened beverages is associated with low bone mineral density in young people: the Brazilian birth cohort consortium. *Nutrients*. (2023) 15:324. doi: 10.3390/nu15020324

22. Ebrahimpour-Koujan S, Saneei P, Larijani B, Esmaillzadeh A. Consumption of sugar sweetened beverages and dietary fructose in relation to risk of gout and hyperuricemia: a systematic review and Meta-analysis. *Crit Rev Food Sci Nutr*. (2020) 60:1–10. doi: 10.1080/10408398.2018.1503155

23. NCHS. NCHS Ethics Review Board (ERB) Approval (2024): Available at: https://www.cdc.gov/nchs/nhanes/irba98.htm

24. NHANES. Nhanes 2009–2010 brochures and consent documents (2024) Available at: https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Documents.aspx?BeginYear=2009

25. NCHS. National Health and nutrition examination surveys (2024). Available at: https://www.cdc.gov/nchs/nhanes/index.htm

26. Hassoon A, Bydon M, Kerezoudis P, Maloney PR, Rinaldo L, Yeh HC. Chronic low-Back pain in adult with diabetes: Nhanes 2009-2010. *J Diabetes Complicat*. (2017) 31:38–42. doi: 10.1016/j.jdiacomp.2016.10.025

27. Shmagel A, Ngo L, Ensrud K, Foley R. Prescription medication use among community-based U.S. adults with chronic low Back pain: a cross-sectional population based study. *J Pain*. (2018) 19:1104–12. doi: 10.1016/j.jpain.2018.04.004

28. NHANES. Inflammatory arthritis questionnaire (2009). Available at: https://wwwn.cdc.gov/nchs/data/nhanes/2009-2010/questionnaires/arq_f.pdf

29. NHANES. Sample person questionnaire hand cards (2009). Available at: https://wwwn.cdc.gov/nchs/data/nhanes/2009-2010/questionnaires/sp_handcards_0910.pdf

30. NHANES. Dietary recall (2024). Available at: https://wwwn.cdc.gov/nchs/data/nhanes/2009-2010/questionnaires/dietaryrecall_f.pdf.

31. WHO. Global Physical Activity Questionnaire (GPAQ) (2021). Available at: https://www.who.int/publications/m/item/global-physical-activity-questionnaire

32. Johnson CL, Paulose-Ram R, Ogden CL, Carroll MD, Kruszon-Moran D, Dohrmann SM, et al. National Health and Nutrition Examination Survey: Analytic guidelines, 1999–2010. National Center for Health Statistics *Vital Health Stat*. (2013) 2.

33. Schulze MB, Hoffmann K, Manson JE, Willett WC, Meigs JB, Weikert C, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *Am J Clin Nutr*. (2005) 82:675–84. doi: 10.1093/ajcn/82.3.675

34. Pinto EM, Neves JR, Laranjeira M, Reis J. The importance of inflammatory biomarkers in non-specific acute and chronic low Back pain: a systematic review. *Eur Spine J*. (2023) 32:3230–44. doi: 10.1007/s00586-023-07717-1

35. Scholz J, Woolf CJ. The neuropathic pain triad: neurons, immune cells and glia. *Nat Neurosci*. (2007) 10:1361–8. doi: 10.1038/nn1992

36. Baron R, Binder A, Attal N, Casale R, Dickenson AH, Treede RD. Neuropathic low Back pain in clinical practice. *Eur J Pain*. (2016) 20:861–73. doi: 10.1002/ejp.838

37. Qin P, Li Q, Zhao Y, Chen Q, Sun X, Liu Y, et al. Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose-response Meta-analysis of prospective cohort studies. *Eur J Epidemiol*. (2020) 35:655–71. doi: 10.1007/s10654-020-00655-y

38. Jacob L, Rathmann W, Koyanagi A, Haro JM, Kostev K. Association between type 2 diabetes and chronic low Back pain in general practices in Germany. *BMJ Open Diabetes Res Care*. (2021) 9:e002426. doi: 10.1136/bmjdrc-2021-002426

39. Hochheim M, Ramm P, Wunderlich M, Amelung V. Association between chronic low Back pain and regular exercise, sedentary behaviour and mental health before and during Covid-19 pandemic: insights from a large-scale cross-sectional study in Germany. *BMC Musculoskelet Disord*. (2022) 23:1–18. doi: 10.1186/s12891-022-05806-8

40. Kahere M, Ginindza T. The prevalence and risk factors of chronic low Back pain among adults in Kwazulu-Natal, South Africa: an observational cross-sectional hospital-based study. *BMC Musculoskelet Disord*. (2021) 22:955. doi: 10.1186/s12891-021-04790-9

41. García G, Gutiérrez-Lara EJ, Centurión D, Granados-Soto V, Murbartián J. Fructose-induced insulin resistance as a model of neuropathic pain in rats. *Neuroscience*. (2019) 404:233–45. doi: 10.1016/j.neuroscience.2019.01.063

42. Xiao L, Harrison DG. Inflammation in hypertension. *Can J Cardiol*. (2020) 36:635–47. doi: 10.1016/j.cjca.2020.01.013

43. Smith CJ, Perfetti TA, King JA. Perspectives on pulmonary inflammation and lung Cancer risk in cigarette smokers. *Inhal Toxicol*. (2006) 18:667–77. doi: 10.1080/08958370600742821

44. Feng C, Yao J, Xie Y, Yang F, Fan X. Association between different composite dietary antioxidant indexes and low Back pain in American women adults: a cross-sectional study from Nhanes. *BMC Public Health*. (2024) 24:147. doi: 10.1186/s12889-024-17649-0

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 06 January 2025
DOI 10.3389/fnut.2024.1503389

Check for updates

OPEN ACCESS

EDITED BY
Massimo Lucarini,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Sercan Karav,
Çanakkale Onsekiz Mart University, Türkiye
Didier G. Leibovici,
The University of Sheffield, United Kingdom
Lindung Parningotan Manik,
National Research and Innovation Agency (BRIN), Indonesia

*CORRESPONDENCE
Eva Valenčič
✉ eva.valencic@ijs.si

RECEIVED 28 September 2024
ACCEPTED 13 December 2024
PUBLISHED 06 January 2025

CITATION
Valenčič E, Beckett E, Bucher T, Collins CE and Koroušić Seljak B (2025) NutriBase – management system for the integration and interoperability of food- and nutrition-related data and knowledge. *Front. Nutr.* 11:1503389.
doi: 10.3389/fnut.2024.1503389



# NutriBase – management system for the integration and interoperability of food- and nutrition-related data and knowledge

Eva Valenčič[1,2,3,4]*, Emma Beckett[4,5,6], Tamara Bucher[4,5], Clare E. Collins[4,5] and Barbara Koroušić Seljak[1,2]

[1]Computer Systems Department, Jožef Stefan Institute, Ljubljana, Slovenia, [2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3]School of Health Sciences, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, NSW, Australia, [4]Food and Nutrition Research Program, Hunter Medical Research Institute, Newcastle, NSW, Australia, [5]School of Environmental and Life Sciences, College of Engineering, Science and Environment, University of Newcastle, Newcastle, NSW, Australia, [6]Department of Science, Nutrition Research Australia, Sydney, NSW, Australia

**Introduction:** Contemporary data and knowledge management and exploration are challenging due to regular releases, updates, and different types and formats. In the food and nutrition domain, solutions for integrating such data and knowledge with respect to the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles are still lacking.

**Methods:** To address this issue, we have developed a data and knowledge management system called NutriBase, which supports the compilation of a food composition database and its integration with evidence-based knowledge. This research is a novel contribution because it allows for the interconnection and complementation of food composition data with knowledge and takes what has been done in the past a step further by enabling the integration of knowledge. NutriBase focuses on two important challenges; data (semantic) harmonization by using the existing ontologies, and reducing missing data by semi-automatic data imputation made from conflating with existing databases.

**Results and discussion:** The developed web-based tool is highly modifiable and can be further customized to meet national or international requirements. It can help create and maintain the quality management system needed to assure data quality. Newly generated data and knowledge can continuously be added, as interoperability with other systems is enabled. The tool is intended for use by domain experts, food compilers, and researchers who can add and edit food-relevant data and knowledge. However, the tool is also accessible to food manufacturers, who can regularly update information about their products and thus give consumers access to current data. Moreover, the traceability of the data and knowledge provenance allows the compilation of a trustworthy management system. The system is designed to allow easy integration of data from different sources, which enables data borrowing and reduction of missing data. In this paper, the feasibility of NutriBase is demonstrated on Slovenian food-related data and knowledge, which is further linked with international resources. Outputs such as matched food components and food classifications have been integrated into semantic resources that are currently under development in various international projects.

KEYWORDS

database management system, food data compilation, food composition data, food composition database, knowledge base

# 1 Introduction

Food and nutrition-related data and knowledge (D&K) are essential for many research domains, including public health surveillance and promotion, dietary and health assessments, disease prevention, nutrition education, consumer protection, agriculture, food policy, and food labeling (1, 2). D&K, such as food composition data or dietary guidelines, are also necessary for stakeholders in the food industry, retail sector, non-government organisations, policymakers, and ultimately consumers. Consumers rely on D&K when making food and nutrition decisions, while policymakers use food and nutrition-related D&K to obtain accurate scientific evidence needed to design and promote strategies required to improve public health and overall well-being (3, 4).

However, D&K are complex, covering diverse areas such as food composition, food safety, food authenticity, and consumption. This paper focuses on food composition data (FCD) and knowledge for dietary assessment and advising. This is highly important for domain experts and policymakers, as well as consumers, including patients. While FCD contains detailed compositional, biochemical, and physiological data of foods (e.g., how much vitamin C apples contain), *knowledge* provides additional food-related information (e.g., what is the recommended intake of vitamin C). FCD and knowledge are compiled in various databases; however, their integration and interoperability are lacking (5). Improved integration would enable easier access the latest evidence-based D&K from different research areas within a single system.

Nowadays, FCD is compiled online in the form of a food composition database (FCDB). FCDBs are usually compiled at the national level but are often used internationally to conduct public health studies (2). Examples include multiple European FCDBs [available through the FoodEXplorer tool (6)], USDA's FoodData Central (7), FAO/INFOODS databases (8), Canadian FooDB (9), and others. In general, FCDBs contain data on traditional, ethnic, and local foods and dishes, with some combining generic and branded foods [e.g., Serbian (10)] and others maintaining separate databases for different food types [e.g., Dutch branded food database (11)]. In addition to institutional databases, numerous company-owned FCDBs also exist, such as the Edamam's food, grocery, and (restaurant) database composed using Natural Language Processing (NLP) techniques (12) and GS1 branded foods, and barcode databases maintained through the Global Data Synchronization Network (GDSN) (13).

There are two main challenges with existing FCDBs. Namely, data harmonization and missing data. First, FCBDs may contain data of different quality due to differences in data production methods (food sampling, analyses or estimation, (re)calculation, borrowing), data compilation (collection, aggregation, compilation, and dissemination), and data management. The challenge of data harmonization has been addressed by several networks of excellence. For example, the Food CEN standard (14), which defines requirements on the structure and semantics of food datasets and of interchange of food data. Another initiative, the ESFRI research infrastructure Metrofood (15), contributes to the development of aligned metrology services in the food domain. Moreover, when compiling a FCDB, guidelines and frameworks to assess the quality of data, datasets, and databases (16, 17) need to be acknowledged. Several frameworks also enable unified data classification and description, which need to be considered when harmonizing various FCDB (2, 18, 19). While these standards and frameworks facilitate the harmonization of food- and nutrition-related data, the problem of linking it with other data types (e.g., medical, environmental, and consumption-related) remains unresolved. The second challenge is related to missing data in FCDBs, which distorts data integrity. Analyzing all components of specific foods poses a significant financial burden for institutions; thus, no FCDB is complete, and updates are not done continuously. The challenge of missing FCD is being addressed in various ways, including borrowing data from other databases, performing tedious manual work, or using computer-supported methods for (semi-) automated data imputation (20, 21).

On the other hand, together with databases, knowledge bases (KBs) are also very important resources. By definition, a KB is an easily accessible online library of collected and organized information and documentation about certain topics (22). The important knowledge that should be included in food and nutrition KB should include, but not be limited to: standardized classification and description of coding systems [e.g., LanguaL (23), FoodEx2 (24), INFOODS (8)]; standardized value documentation (e.g., acquisition type, method type) (18); a chemical databases of molecular entities – ChEBI (25); retention and yield factors used to calculate the nutrient content of composite dishes or recipes (26); standardized household measurement units; national dietary reference values and dietary guidelines; physical activity standards; food components' bioavailability; food-drug interactions, and others.

As knowledge accumulates quickly, the creation and maintenance of a KB is tedious work, usually done manually by domain experts. However, semantic resources have complemented KBs and allowed interoperability of D&K from various research domains. Semantic resources like the ontologies [e.g., FoodOn (27), ISO-FOOD (28), FNS-Harmony (29), COMFOCUS (30)] or knowledge graphs [e.g., describing complex relationships between food and biomedical factors (31)] are being developed to formally describe knowledge as a set of concepts and the relationships between those concepts within a domain. To link FCD with semantic resources, FCD needs to be annotated with standardized metadata in machine-readable formats to enable connectivity of terms across different data sources.

Regardless of all research efforts, applicable KBs providing integrated knowledge on food and nutrition are still lacking. There are few KBs that focus on specific subdomains, such as FoodKG (32) for food recommendation based on diet-related knowledge or TasteAtlas (33), a world atlas of traditional dishes, local ingredients, and authentic restaurants.

Abbreviations: API, Application Programming Interface; D&K, Data and knowledge; DKBMS, Data- and knowledge base management system; FCD, Food composition data; FCDB, Food composition database; KB, Knowledge base; NLP, Natural Language Processing; KPI, Key performance indicators; MTBF, Mean time between failures.

The food and nutrition community has created many FCDBs as well as few KBs, but their integration and interoperability are currently missing. Even when limited just to the integration within FCDB, information is not harmonized because different coding systems, documentation or standards are used. Some examples of best practice using harmonized FCD are FoodEXplorer (6), FoodCASE (34), FoodData Central, Glycemic Index Research and GI News (35). Some of these tools even enable comparison of FCD from multiple countries. This is important as, with increasing globalization, the availability of international foods and dishes is increasing, and obtaining datasets of non-local foods is necessary. Having databases composed on a national level is important; however, for applied science, it would be useful if compilers could link and integrate not only FCD with each other but also FCDBs with KBs. This is something that we believe does not yet exist or is not publicly available in the food and nutrition domain. The importance of integration and interoperability was also highlighted in the recent paper by Durazzo et al. (36), which further emphasized the necessity of cooperation and D&K sharing between compilers. However, the connectivity among computer systems and/or online platforms is equally necessary.

In the current paper, we introduce a new database management system, called NutriBase, for integrating FCD from different databases with food- and nutrition-related knowledge. The integration is performed in a transparent way and enables, together with harmonization, a reduction in missing data. In Section 2, we explain how publicly available D&K resources, which (currently) represent the baseline of the NutriBase, were identified and collected. Next, we introduce NutriBase and describe its functionality. In Section 3, we describe the compilation process of the Slovenian FCDB and KB, identify issues, discuss possible solutions the system offers, and provide plans for future work. We conclude the paper in Section 4.

TABLE 1 FCDBs currently included in the NutriBase.

| Currently Imported FCDBs | | | | |
|---|---|---|---|---|
| Country code | No. of components | No. of food group levels‡ | No. of foods / dishes | Source file format |
| SI | 773* | 15 | 993 | .CSV/.XSL |
| | | 48 | | |
| | | 149 | | |
| FR | 60 | 10 | 2,807 | .CSV/.XSL |
| | | 58 | | |
| | | 83 | | |
| NL | 133 | 27 | 2,152 | .CSV/.XSL |
| DK | 197 | 18 | 1,186 | .CSV/.XSL |
| | | 127 | | |
| UK | 178 | 14 | 2,910 | .CSV/.XSL |
| | | 71 | | |
| | | 54 | | |
| AU | 249 | 22 | 1,534 | .CSV/.XSL |
| | | 97 | | |
| US | 235 | 28 | 7,793†, 210¨ | API |

*651 from EuroFIR Thesauri document and 122 subsequently added (own); ‡ = the top number is the highest level, the bottom number is the lowest (the most detailed) level (sub-level); † = SR Legacy Foods; ¨ = Foundation Foods.

# 2 Materials and methods

## 2.1 Data and knowledge collection

To demonstrate the feasibility of NutriBase, Slovenian FCD and both, national and international semantic resources were collected. Firstly, the analytical compositional data on generic foods from the Slovenian FCDB composed in 2006 and updated in 2012 (37) were imported. The recipes included in the Slovenian FCDB were imported separately, as they require different data handling, such as consideration of yield and retention factors, as well as standards for calculating recipes (38, 39). In addition, branded foods that can currently be purchased in Slovenia, are being uploaded through an application programming interface (API) from the Composition and Labeling Information System (CLAS) (40).

To complement the Slovenian FCDB for generic foods, six publicly available FCDBs (Table 1), together with associated metadata and documentation, were either downloaded or linked through an API in late 2020 or 2021. The imported FCDBs consisted of datasets in different formats, and not all of them adhered to the Food CEN standard (14). The imported metadata and documentation include various background information, such as explanations of data sources, procedures for data quality assurance, descriptions of foods and food group classification levels, and explanations of specific component descriptions, calculations and units used. Multiple foreign FCDBs needed to be imported because they contain different data. For example, FoodData Central (US in Table 2) in addition to FCD, provides also the data for household measurement units (e.g., tablespoon, cup, dash) which can be linked to generic foods. Moreover, different components are collected or analyzed across different FCDBs. For instance, some datasets contain data for total carbohydrates (digestible and indigestible, including dietary fiber), whereas others contain only data for available carbohydrates. From the currently imported FCDBs only three provide data for total carbohydrate, however all of them contain data for available carbohydrates and total dietary fiber, thus the total carbohydrates could be calculated. Lastly, relevant evidence-based food and nutrition knowledge was systematically reviewed and collected from publicly available national and international resources, and was further compiled into the NutriBase KB (Table 2).

The approaches and tools applied and described in the current paper can be used for D&K from any country. The Slovenian D&K are used as an example only. Unlimited publicly available FCDBs and/or KBs can be uploaded or linked via an API to create a new database, as long as they comply with the NutriBase requirements.

## 2.2 NutriBase - data- and knowledge base management system

NutriBase is designed to enable easy integration with other KBs and semantic resources conceptualizing the health, environmental, consumer behaviors, and food and nutrition domains in particular. This data- and knowledge base management system (DKBMS) has

TABLE 2 Resources included in the NutriBase KB.

| Scemantic resources | | | |
|---|---|---|---|
| Resource name/type and reference | Knowledge type | Description | Number of entities |
| Standardized classifications and description coding system | FoodEx2 classification | A food classification and description system developed by EFSA - includes different hierarchies and facets for different food safety domains. (e.g., A00KR#F27.A00KV$F27.A00LN $F27. A00LB$F27.A00LG; mixed leafy vegetables) | 4,445 |
| Standardized value documentation (11) | Component type | Component identifiers and descriptors (e.g., CHO; carbohydrate; use for total of those carbohydrates digested and absorbed in the intestine; total accessible carbohydrates include free sugars, polyols and dextrins, starch, and glycogen). | 660 (9 of these are for backward compatibility only) 56 classification identifiers (not used for new indexing) |
| | Unit | E.g., grams, millimoles, alpha- tocopherol equivalent, per cent. | 19 Additional 20 added (IU, g/kg body mass, etc.) |
| | Matrix unit | E.g., per 100 g of total food, per 100 mL food volume, per unit, per 100 g edible portion. | 20 matrixes |
| | Value type | E.g., arithmetic mean, best estimate, average, below limit of detection, trace. | 20 types |
| | Method type | Reporting if the value was analyzed, calculated or imputed (e.g., calculated as recipes, calculated from related food, analytical result). | 20 types |
| | Method indicator | Providing details for the analytical method or formulas used for calculation (e.g., chromatography, difference, ash calculated as sum of minerals). | 214 indicators |
| | Acquisition type | Describes the origin of the value (e.g., laboratory, food composition table, authoritative document). | 12 types |
| | Reference type | E.g., article in journal, file or database, product label, software. | 14 types |
| LanguaL thesaurus (10) | Cooking methods | E.g., griddled, cooked by microwave, deep fried. | 47 methods |
| FoodData Central at US Department of Agriculture (USDA) | Measurement and household units | E.g., tea spoon, slice, filet, cup, could be used for volume to weight conversions. | 115 (currently in use) out of 1923 |
| ChEBI - a chemical database and ontology of molecular entities, which is part of the Open biomedical ontologies at the EBI, and European ELIXIR infrastructure | Dictionary of molecular entities | Providing detailed data of chemical entities of biological interest (e.g., definitions, formulas, ontologies, chemical reactions, IUPAC names and identifiers) | 210 linked to added components |
| SciName Finder (26) | Search tool for scientific and common names of plants and animals | Providing precise identify plants and animals<br>Allows precise identification of plants and animals, and searching the information on scientific and common names provided by authoritative resources (and not from secondary sources) | More than 1,000,000 scientific and common names |
| Culinary groups [adapted from (18, 23)] | Culinary groups / subgroups related to retention and yield factors. | Providing the basics for obtaining nutrient content of foods by calculation methods (as recipe calculation), based on the amount of ingredients given in a recipe, nutrient composition of ingredients and factors that consider changes in nutrient content (retention factors), and weight (yield factors) during preparation. | 31 groups and subgroups related to yield factors, and 38 related to retention factors |
| Slovenian dietary reference values (DRVs) (27) based on the D-A-CH reference values adopted by the Ministry of Health of the Republic of Slovenia | DRVs | Reference values for energy and nutrient intake for children (at least 1-year old), adolescents, adults, elderly, pregnant women and nursing mothers. | 34 references for energy, macro- and micronutrients, for men and women (10 different age groups) |
| Latest dietary guidelines and recommendations | National and international dietary guidelines and recommendations | Relevant evidence-based guidelines and recommendations for different consumers (athletes, pregnant women and nursing mothers, healthy individuals from different age groups). | Currently defined for biomarkers (blood cholesterol and glucose) and endurance sports. |

*(Continued)*

TABLE 2 (Continued)

| Scemantic resources | | | |
|---|---|---|---|
| Resource name/type and reference | Knowledge type | Description | Number of entities |
| Physical activity related standards | Metabolic equivalent of task (METS) | E.g., basketball, swimming, mopping, walking, sitting. | 541 tasks |
| | Physical activity level (PAL) | E.g., sedentary or light activity lifestyle. | 5 levels per sex |

been implemented as a web-based tool (Figure 1) for food compilers to easily explore, compile and most importantly, link data from different FCDBs and KBs. The main goal of this process is achieving an optimal linking of D&K, which enables borrowing data respecting the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for data management (41), and reducing missing D&K.

### 2.2.1 FCDB compilation

To ensure a semi-automatic connectivity among different sources (FCDBs), standardized components and food groups matching had to be manually performed (Figure 2; Step 1 and 2). Since the composition of food depends on its geographical origin, it is important to also consider the data source and the data most closely related to local foods. Therefore, a pre-set priority list of data sources is integrated within the system and can be adapted if needed. For Slovenian example this means that European datasets are prioritized before non-EU datasets. This allows experts to semi-automatically compile datasets that are as complete as possible, while also transparently providing the source of specific data (e.g., component value). The pre-set priority list can easily be amended or set for different countries. Moreover, a comparison of a national dataset (in our case, Slovenian), with other, foreign datasets is also enabled. This feature allows borrowing specific data from other FCDBs. Together with food composition data, compilers can also check additional value information, such as value type and method type (if provided). Being able to check additional value information and standards, allows compilers to assess the quality of the data and select the most appropriate or accurate one. Additionally, during the FCDB compilation process, basic food information and metadata, such as generic and/or commercial names, allergens, ingredients, food origin, and food images, are also addressed and can be borrowed.

NutriBase presents an infrastructure that can be adapted for FCD from any country. However, to achieve an optimal linking of D&K and to ease and expedite FCDB compilation, various knowledge resources had to be considered.

### 2.2.2 Knowledge base compilation

In the NutriBase underlying thesaurus, knowledge about relevant food- and nutrition topics is collected and maintained. The KB, implemented within the DKBMS, is connected with all three steps of the workflow seen on Figure 2. Thus, all updates of the KB content will have an immediate impact on linked data in FCDB. That means whenever a new data or knowledge is published, it can easily be imported and linked to existing D&K or substituted for the latest findings. An important part of the implemented KB is food naming by using tags. It provides functionality for unique food naming and metadata annotation. While much work has already been conducted on unifying food description and classification, food naming is still an open issue. Therefore, we have implemented a new food-tagging approach to unify and standardize food naming within the FCDB. This is especially useful when different users are working on a FCDB, as it enables unambiguous communication between all users involved in the working process. In addition, together with using tags, setting rules for food naming has been proposed as another solution.

### 2.2.3 Usability of NutriBase

Lastly, the usability of the newly developed system was evaluated. We distributed the System Usability Scale (SUS) questionnaire among regular NutriBase users with different profile roles. The SUS tool is a reliable and validated tool for measuring the usability, which is frequently used by evaluators of mHealth services (42). It consists of a 10-item questionnaire with five response options for respondents (strongly agree to strongly disagree). The survey was completely anonymous and after collecting the responses, the participant's scores were carefully interpreted to produce a percentile ranking.

# 3 Results and discussion

## 3.1 The compilation of the Slovenian FCDB and KB

Throughout the entire compilation process (Figure 2), D&K were maintained in accordance with the FAIR principles. Managing D&K to ensure that the format of foreign FCDBs and KBs remains unchanged from the original sources has been a key requirement in NutriBase's development (Figure 3).

### 3.1.1 Components matching

To create and link the Slovenian database, the compilation process was initiated by components matching (Figure 2, Step 1). The Slovenian FCDB complies with the CEN Food standard (14), therefore the components specified with respect to the EuroFIR thesaurus for components (18) were manually matched with components from the foreign FCDBs (Figure 4 presents the user interface of this process). Although most of the foreign selected FCDBs also comply with the CEN Food standard, mismatched components (i.e., different names for the same components among different countries) were still present (examples are shown in Table 3).

Components were matched manually by domain experts to ensure a correct and unambiguous matching. Moreover, the result can be provided as an input to the FNS-Harmony ontology (43), which has been developed within the FNS-Cloud project to support interoperability of food- and nutrition-related data in the European Open Science Cloud (EOSC) and is available through the NCBO Bioportal. NutriBase could be integrated with FNS-Harmony, which reuses or incorporates several ontologies, including FoodOn (27). In

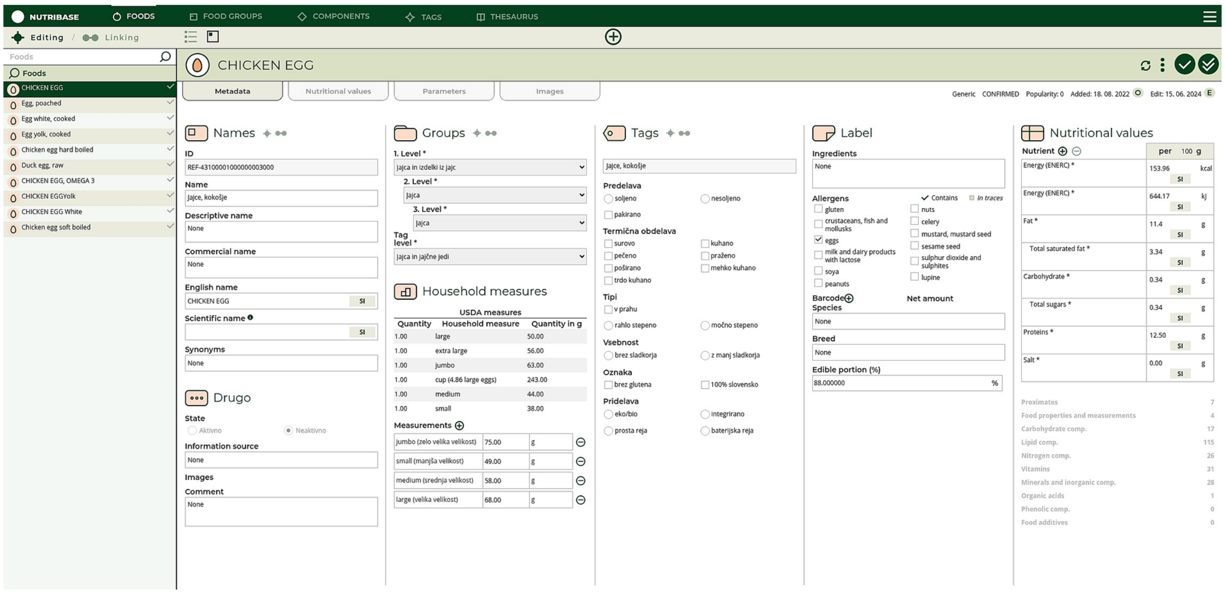



FIGURE 1
User interface of NutriBase.

this case, food compilers would not only be able to provide but also use new knowledge about semantic integration with other systems, such as GS1 GDSN (44).

### 3.1.2 Food groups matching

Firstly, food groups were designed based on the classification of foods used by relevant information systems in Slovenia, as well as the EuroFIR standard (18), which is intended for generic foods. Since Slovenian FCDB also includes branded foods, classification systems for these had to be considered as well. However, we found that different Slovenian institutions use different classification systems. This suggests that even within a single country, it might be necessary to follow and comply with several standards. For example, the Slovenian classification system, which is based on public procurement and is determined by law, or the Dunford classification system (45), specifically developed for branded foods. Currently, the Slovenian FCDB includes three hierarchical classification levels: 15 groups on the first, 48 groups on the second, and 160 on the third (and most detailed) level.

In addition to manually matching national food classification systems with one another, the food groups used in Slovenian FCDB were also matched with those used in the foreign FCDBs (Figure 2, Step 2). An example of a matched food group - *Fresh vegetables,* among FCDBs is presented in Table 4. The task of food groups matching was especially challenging, as different countries use different numbers of classification levels. For example, foods in France and the UK are classified into up to three levels, in Australia and Denmark into two levels, and in the Netherlands and USA into just one level. Moreover, the level of detail within food groups varies. As shown in Table 4, some countries group all vegetables together, while the others sub-classify them further (e.g., root vegetables, fruiting vegetables).

To ensure accurate food classification and assist users in using NutriBase, a feature was implemented allowing compilers to add examples of foods allocated to specific food group. This feature was found to be very useful, as it enables users to unambiguously select the correct food group. Additionally, manually matched food groups can also be provided as inputs into FNS-Harmony.

### 3.1.3 FCDB compilation

FCDB compilation process (Step 3 in Figure 2) began with manually checking and correcting a dataset of 14,064 entries for 443 generic foods analyzed by the Biotechnical Faculty of the University of Ljubljana in 2006 and 2012 (37). Together with the composition data, annotated metadata (e.g., value information) were also reviewed. Certain components were specifically checked to ensure compliance with the standards. For example, the differences between total available carbohydrates and total carbohydrates. This entire process aligns with the first 12 steps of the generic compilation process described by Westenbrink et al. (2), currently excluding Step 5 (attribution of quality index) and Step 11 (physical storage). The evaluation of Slovenian data quality (17) and the database quality evaluation, as suggested by the recently published FAO/INFOODS framework (16), are currently underway.

Next, the Slovenian name for each generic food was reviewed, and a scientific name (when appropriate), an English name, and synonyms were assigned based on the new food-tagging approach. To achieve this, tags were defined, and rules for their application were established within each food group. During this process, we found that similar foods might have different names. This can make searching for a specific food within the FCDB harder for compilers as well as for consumers accessing publicly available FCDBs. For example, the only difference between 'Baked eggplant with added cheese and tomato sauce' and 'Aubergine prepared in tomato sauce and cheese, frozen' is that one is baked and the other is frozen, but the names are very different. Therefore, using tags for food naming, helps unify the FCDB and simplifies searching for specific foods. Moreover, we ensured the naming is clear to all users, specifically for consumers accessing FCD

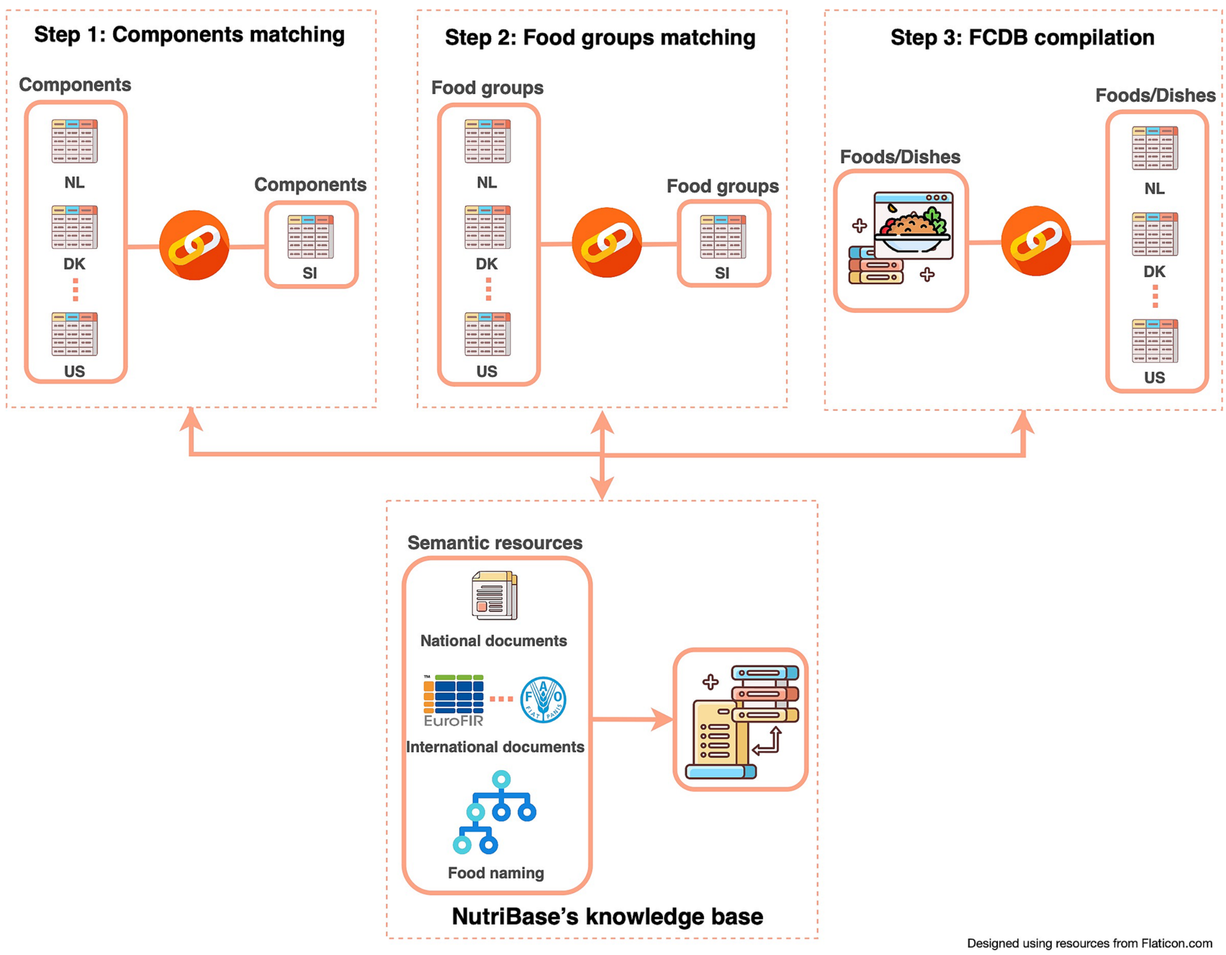



FIGURE 2
Flowchart of compilation process to link foods from different FCDBs.

(e.g., via a mobile app), who may find it challenging to understand the processing conditions of foods. For example, meat can be analyzed as raw (e.g., beef filet) or heat-treated (e.g., beef filet, grilled). However, experience shows, it is seen that consumers do not consider 'beef filet' as raw, but rather as ready-to-eat steak. Therefore, adding the 'raw' tag to raw meat seemed reasonable. On the other hand, it is clear to consumers that 'banana' is raw, and they do not expect this tag added to fresh fruits. Thus, the 'raw' tag is used in some food groups but not in others. In addition, the tag 'peeled' is used only when appropriate (e.g., 'apple, peeled', but not 'banana, peeled'). Currently, each food group at the third hierarchical level within the tool has an average of 15.4 tags.

Additionally, the initial Slovenian dataset of generic foods was manually linked with the same or similar food items from the selected foreign FCDBs. The linking was carried out by domain experts. First, the English names were compared, followed by a comparison of the main food components. In case the food composition was similar, the food items were linked together and the missing data were imputed from the foreign FCDBs. Table 5 presents an example of the number of imputed data for *Fresh vegetables* food group from a specific FCDB. As can be seen, only one value for total carbohydrates could be borrowed from US database, while the rest were taken from Slovenian FCDB. However, cystine values are missing in Slovenian FCDB, so they were borrowed from the Danish and US databases (the other FCDBs do not contain data for cystine). The NutriBase allows linking one food with multiple foods within one database or across multiple databases. For example, the Slovenian 'average white bread' can be linked with 'white baguette' and 'white loaf' from one FCDB, and with 'white bread' from the other FCDBs. The borrowed data will, however, be displayed based on the pre-set priority list of FCDBs. In our case, when a food item is linked with food item(s) from across different FCDBs, data from European datasets were prioritized before non-EU datasets. However, compilers can manually change the data source and select (borrow) non-EU data to be displayed if it is more appropriate. We found this approach to be very convenient, as it provides compilers with data most closely related to the local foods, but it still gives them freedom to select another data. Moreover, the manually matched foods present a valuable asset that can be used to construct a gold standard corpus, i.e., a corpus of text annotated with food entities required for NLP techniques, such as CafeteriaFCD (46).

Same as generic foods, the branded foods can also be linked with similar generic foods from either national FCDB or foreign FCDBs. In this case, the original FCD of a branded food is taken from the nutrition declaration table, while the FCD not provided on the nutrition declaration table (e.g., micronutrients) can be imputed from FCDBs and transparently marked as such. This is especially beneficial when

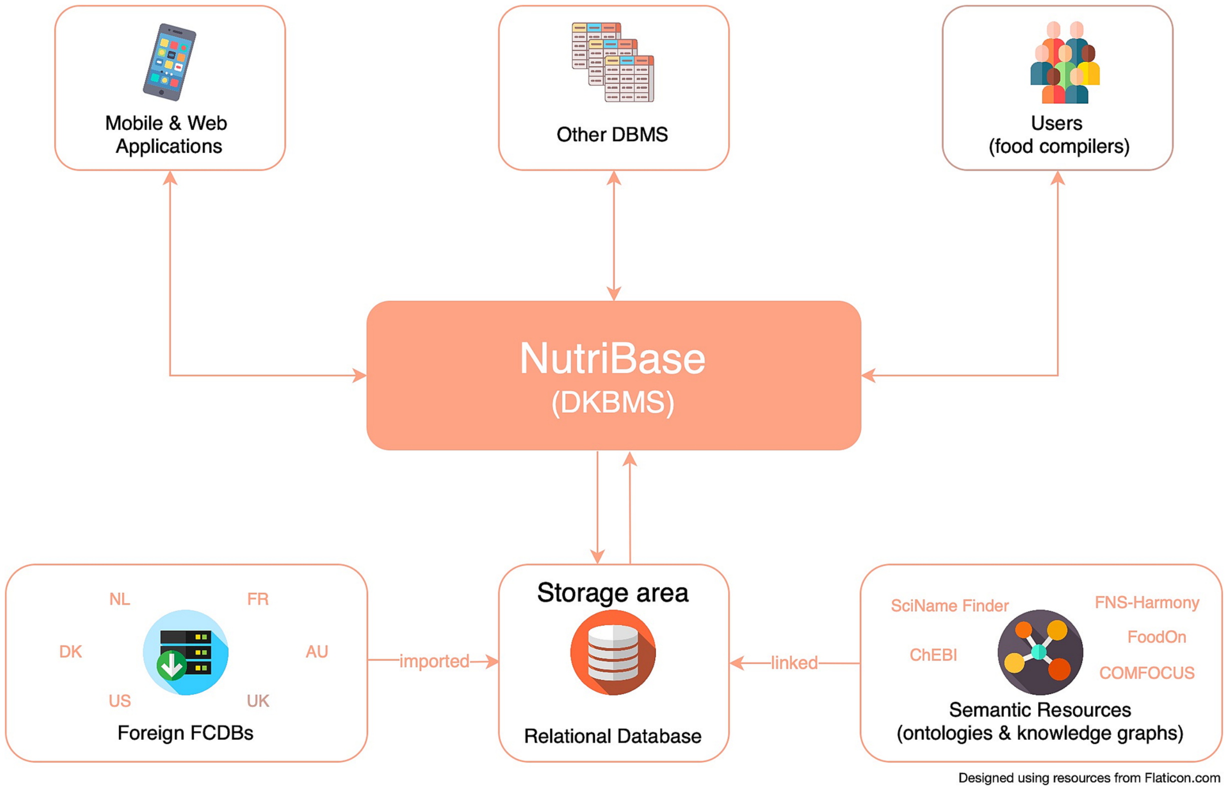



FIGURE 3
Overview of NutriBase structure.

collecting food consumption data for the national food consumption survey. As seen in the EU Menu project, consumers usually provide only the brand or production line of the food item when reporting food intake. For example, instead of reporting consumption of 'full fat milk', they reported a producer's name of such milk. Since the nutrition declaration table usually only provides the information of energy value and six other nutrients, the values of micronutrients are unknown. Thus, branded foods could be linked with generic foods to compose the complete dataset, which would provide the opportunity to more accurately assess food intake of individuals and overall population.

Finally, yet importantly, internationally accepted algorithms to avoid errors were selected and applied to produce aggregated data [e.g., recipe calculations) (Steps 14 and 15 according to Westenbrink et al. (2)]. In addition, the compiled and aggregated data within the NutriBase were verified [and corrected if needed) (Steps 16 and 17, according to Westenbrink et al. (2)] to prevent hazards related to data validation. The majority of the FCD validation has been done manually, however the tool automatically performs consistency checks for some metadata and components (e.g., content of specific component is not larger than 100 g (converted regardless of the unit), the sum of proximities is ≤105 g, value of saturated fatty acids is not larger than value of total fats, etc.). The validated data is then stored and disseminated [Steps 18 to 22, according to Westenbrink et al. (2)].

### 3.1.4 Knowledge base creation

Using semantic resources, a KB was created to support the optimal food compilation process, as well as for data quality assessment, traceability, calculations and validation. The KB implemented within the NutriBase is meant to be used by domain experts, as it collects the latest scientific evidence and documentation required for data management and data source management. The KB also consists of the reference list and it allows publication metadata to be imported in standardized formats (e.g., bib). These references can be further linked to specific data/information, which allows traceability of data and metadata. Moreover, the information can be edited or added to the existing KB and updated accordingly. For instance, units listed in the EuroFIR value documentation (18) can be supplemented or extended with other units (e.g., IU, ABV) to meet the compilers' needs, or they can be updated if changes are made to the existing EuroFIR standards.

### 3.1.5 Linking FCDB with knowledge

Linking FCD from different sources is important, and linking knowledge from various sources is equally crucial. Both types of linking can be performed in NutriBase; however, the system also enables the linking of FCD with knowledge. For instance, a specific component (e.g., vitamin C) can be linked with a relevant dietary recommendations, such as Slovenian DRVs (47). Therefore, within the tool, data (component; vitamin C) was interconnected and complemented with knowledge (dietary requirements for vitamin C), enabling access to combined information in one place. This approach takes what has been done in the past a step further by incorporating knowledge into the system, which can be especially useful for informing and educating consumers (e.g., via mobile apps). Instead of providing consumers or app users with just FCD, the incorporated knowledge can also be provided, which can deliver a more personalized approach. Our work is consistent with previous works (5, 27, 48), with the difference that NutriBase is a practical and applicable tool, whereas the previous works is theory based.

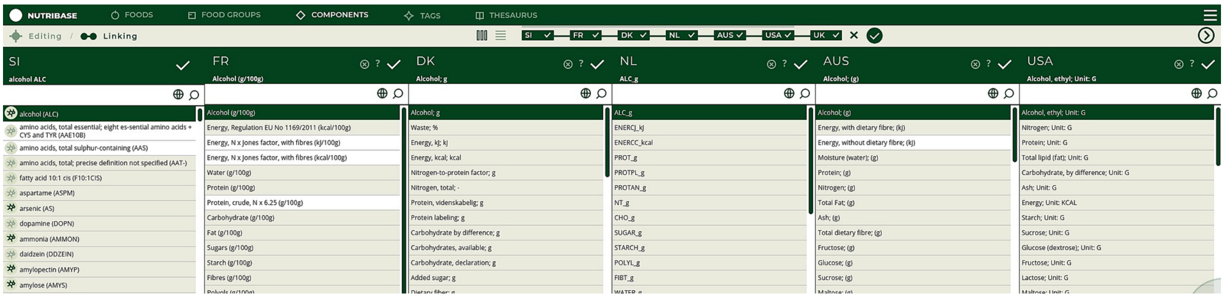



FIGURE 4
User interface of component matching process.

TABLE 3 An example of component matching of Slovenian components with components from foreign datasets.

| Component names among different FCDBs | | | | | | |
|---|---|---|---|---|---|---|
| SI | FR | NL | DK | UK | AU | US |
| Carbohydrate, total (CHOT) | / | / | Carbohydrate by difference; g | / | / | Carbohydrate, by difference; Unit: G |
| Carbohydrate (CHO) | Carbohydrate (g/100 g) | CHO g | Carbohydrates, available; g, Carbohydrate, declaration; g | Carbohydrate (g); CHO | Available carbohydrate, with sugar alcohols; (g) | Carbohydrate, by summation; Unit: G |
| Fiber, total dietary (FIBT) | Fibers (g/100 g) | FIBT_g | Dietary fiber; g | AOAC fiber (g); AOACFIB | Total dietary fiber; (g) | Total dietary fiber (AOAC 2011.25); Unit: G, Fiber, total dietary; Unit: G |
| Fat, total (FAT) | Fat (g/100 g) | FAT_g | Fat, total; g | Fat (g); FAT | Total Fat; (g) | Total lipid (fat); Unit: G |
| Fatty acids, total saturated (FASAT) | FA saturated (g/100 g) | FASAT_g | Sum saturated; g | Satd FA /100 g FA (g); SATFAC, Satd FA /100 g (g); SATFOD | Total saturated fatty acids;(%), Total saturated fatty acids; (g) | Fatty acids, total saturated; Unit: G |

### 3.1.6 Tool validation

The NutriBase and its functionalities were validated throughout the entire compilation process of the FCDB and KB. Seven experts who regularly use NutriBase evaluated it using the SUS tool, which is used for judging the perceived usability of systems. The SUS score was 78.9, which falls to 85$^{th}$ percentile and corresponds to grade A-. Moreover, six food compilers of different skills have performed various tasks (e.g., component matching, food linking) depending on their user profile role. For example, less skilled compilers have only edited D&K, whereas more experienced compilers performed more demanding tasks. Regardless of their skill level, all users agreed that the system is a helpful, easy-to-use tool when compiling a FCDB, especially because it collects all relevant and needed D&K in one place.

## 3.2 Strengths and limitation of DKBMS

While reviewing analytical data of generic foods from the past Slovenian FCDB and importing it into the DKBMS, some errors and gaps were identified and further discussed with compilers. The data was reviewed using spreadsheets, and it was found that errors were difficult to identify. However, when using NutriBase to review and edit the FCD, users agreed that it is a useful and reliable tool. Although spreadsheets are very popular when handling data, a similar finding was reported by Presser et al. (34).

To assess the quality of D&K, it is crucial to develop and maintain a quality management system (2). Currently available FCDBs contain data of varying quality, mainly due to the use of different resources and different methods of data acquisition. The metadata used to describe them, as well as the quantity of data differ among FCDBs. Therefore, compilers need to follow standardized guidelines, provide quality indexes for their original data, and further evaluate their FCDB. This will help domain experts select the best high-quality dataset and/or FCDB for their purposes, which can further be used to obtain accurate results in research, education, and in decision making for policy and programming (16). Not only is NutriBase a useful tool to help domain experts compare different datasets and therefore select the most appropriate one, it can also help national compilers to evaluate their own original data and metadata, and ensure the quality datasets. Moreover, an advantage of the system is also that food manufacturers can gain direct access, and add or edit food-related data of their products. In this way, important information about branded foods currently available in stores can be regularly updated and shared with consumers.

The usage of FCDBs may be significantly restricted due to the missing data (3). It has been proposed that it is better to include imputed data, transparently identified as such, than no data at all (3). However, data should only be borrowed or imputed among the same or similar foods. Several computational methods for missing data imputation within FCDBs have been previously researched (20, 21). All of them concluded that, in order to 'borrow' data, as many details as possible

TABLE 4 An example of matching one Slovenian food group with different foreign FCDBs.

| | Classification levels | | | |
|---|---|---|---|---|
| | | L1 | L2 | L3 |
| FCBD | SI | Vegetables | Vegetables, mushrooms and algae | Fresh vegetables |
| | FR | Fruits, vegetables, legumes and nuts | Vegetables | Vegetables, raw |
| | | | | Vegetables, cooked |
| | NL | Vegetables | / | / |
| | DK | Vegetables and vegetable products | Leaf and stem vegetables | / |
| | | | Root and tuber vegetables | |
| | | | "Fruit" vegetables | |
| | | | "Fruit" vegetables | |
| | UK | Vegetables | Vegetables, general | / |
| | AU | Vegetable products and dishes | Wild harvested vegetables, and vegetable dishes | / |
| | | | Cabbage, cauliflower and similar brassica vegetables | |
| | | | Carrot and similar root vegetables | |
| | | | Leaf and stalk | |
| | | | Tomato and tomato products | |
| | | | Other fruiting vegetables | |
| | | | Other vegetables and vegetable combinations | |
| | US | Vegetables and Vegetable Products | / | / |

L1, level 1 (the highest level in the hierarchy); L2, level 2; L3, level 3 (the lowest level in the hierarchy).

about the origin or source of the food are needed. In addition, when borrowing data, it is necessary to check whether the relevant values (e.g., nutrients) and metadata are similar. If the metadata or values deviate too much, the foods should not be linked, and a better match should be identified. Deviations may occur for various reasons, such as; different food origin, different analytical methods used or outdated data. The developed DKBMS may ease the process of comparing FCD among different datasets or resources, and help finding the best matches.

Although connecting data from just two FCDBs would be the easiest for compilers, it is not always feasible because different FCDBs contain different data. For example, all of the imported FCDBs contain data for the *total protein* content, but only three FCDBs provide data for specific amino acids. However, research suggests that emphasis should be given not only to the overall protein intake, but also to the specific amino acids [i.e., leucine in older adults, as it is proposed to prevent and treat sarcopenia (49)]. Thus, for experts to prepare dietary guidelines that focus also on specific amino acids and further disseminate them, FCDBs must first contain such information. Among the FCDBs currently imported into NutriBase, only the Danish, Australian, and American FCDBs provide data for leucine, for example. Currently, many imported FCDBs calculate the protein content of foods using a 6.25 nitrogen-to-protein conversion factor as the default factor. However, recent research suggests using specific conversion factors for different foods (50). The new factors and/or re-calculations of protein content can be updated when available and borrowed across FCDBs. Clearly, the DKBMS could also be used to identify globally missing data within the FCDBs.

Nowadays many web-based and mobile applications allow users to add or edit FCD without considering data standards. This may lead to imprecise data, which can further lead to incorrect dietary intake assessments. This is concerning because it raises the question: how can users be sure the data is of high quality? Hence, it is recommended that apps use FCD from approved and high-quality FCDBs, as these guarantee harmonized, scientifically collected, and reviewed data and information. Within the NutriBase, the data origin/source is clearly displayed, and traceability of it is enabled. Combining such trustworthy FCDB with all relevant KBs and sematic resources, can provide a baseline for other systems (e.g., mobile apps, web-based tools, online grocery stores), and it is an extension of what has been done in the past.

In addition, the created KB can be updated by adding and importing direct links to more relevant resources. Some examples of KBs and knowledge resources that could be added to the system are; the international network of food data systems - INFOODS [4], the Global Dietary Database (51), the chemical hazard database (52), different EFSA guidelines, standards and tools (53), etc. Uniting, linking and regularly updating all these resources, could present a baseline for experts and consumers by providing them with transparent, detailed and evidence-based food and nutrition D&K.

Despite the contributions of the current study, the limitations need acknowledgment. As already mentioned, some tasks had to be performed manually, which can be very tedious and usually requires the work of several people. Standardizing and harmonizing D&K among different research fields would allow us to avoid the manual work and expedite the process. In addition, currently only FoodEx2 coding system is implemented in the DKBMS. However, more coding systems could be imported to improve interoperability. Furthermore, although the tool's user interface is designed to be multilingual, it is currently available only in Slovenian, and not all parts of the tool have been translated into English yet. A complete translation of the tool would allow better distribution among different countries. Moreover, more expert users would need to use and test NutriBase to provide a more comprehensive evaluation. Lastly, while the development of the tool is based on research work,

TABLE 5 Number of data imputed from a specific FCDB for *Fresh vegetables* food group.

| Component | SI | FR | NL | DK | UK | AU | US |
|---|---|---|---|---|---|---|---|
| Carbohydrate, total (CHOT) | 42 | - | - | 0 | - | - | 1 |
| Carbohydrate (CHO) | 24 | 15 | 3 | 1 | 1 | 0 | 0 |
| Fiber, total dietary (FIBT) | 36 | 4 | 0 | 1 | 1 | 0 | 0 |
| Fat, total (FAT) | 42 | 0 | 0 | 0 | 1 | 0 | 0 |
| Fatty acids, total saturated (FASAT) | 28 | 12 | 1 | 0 | 1 | 0 | 2 |
| Fatty acids, total monounsaturated (FAMS) | 26 | 12 | - | 1 | 1 | 0 | 1 |
| Fatty acids, total polyunsaturated (FAPU) | 26 | 12 | 1 | 0 | 1 | 0 | 1 |
| Protein (PROT) | 42 | 0 | 0 | 0 | 1 | 0 | 0 |
| Energy, gross (ENERA) | 5 | 4 | 0 | 0 | 1 | 0 | 0 |
| Energy, total metabolizable (ENERC) | 33 | 1 | 5 | 0 | 0 | 0 | 5 |
| Water (WATER) | 41 | 1 | 0 | 0 | 1 | 0 | 0 |
| Ash (ASH) | 41 | 1 | 0 | 0 | - | 0 | 1 |
| Polyols (POLYL) | 0 | 10 | 22 | 0 | - | - | 0 |
| Alcohol (ALC) | 0 | 31 | 4 | 2 | 2 | 0 | 1 |
| Sodium (NA) | 39 | 2 | 0 | 0 | 1 | 0 | 0 |
| Salt (NACL) | 12 | 31 | - | - | - | - | - |
| Organic acids (OA) | 22 | 3 | 0 | 1 | - | - | - |
| Alanine (ALA) | 10 | - | - | 11 | - | 3 | 11 |
| Arginine (ARG) | 24 | - | - | 6 | - | 0 | 7 |
| Asparagine (ASN) | 0 | - | - | - | - | - | - |
| Cysteine (CYSTE) | 15 | - | - | - | - | - | 0 |
| Cystine (CYS) | 0 | - | - | 18 | - | - | 13 |
| Glutamic acid (GLU) | 10 | - | - | 11 | - | 3 | 11 |
| Glutamine (GLN) | 10 | - | - | - | - | - | 0 |
| Histidine (HIS) | 23 | - | - | 7 | - | 0 | 7 |
| Isoleucine (ILE) | 25 | - | - | 5 | - | 0 | 7 |
| Leucine (LEU) | 24 | - | - | 6 | - | 0 | 7 |
| Lysine (LYS) | 25 | - | - | 5 | - | 0 | 7 |
| Methionine (MET) | 25 | - | - | 5 | - | 0 | 7 |
| Phenylalanine (PHE) | 25 | - | - | 5 | - | 0 | 7 |
| Proline (PRO) | 10 | - | - | 11 | - | 3 | 11 |
| Serine (SER) | 9 | - | - | 12 | - | 3 | 11 |
| Taurine (TAU) | 0 | - | - | - | - | - | 0 |
| Threonine (THR) | 24 | - | - | 6 | - | 0 | 7 |
| Tryptophan (TRP) | 20 | - | - | 8 | - | 2 | 8 |
| Tyrosine (TYR) | 17 | - | - | 7 | - | 2 | 9 |
| Valine (VAL) | 24 | - | - | 6 | - | 0 | 7 |

- = the component is not present in the FCDB.

ongoing maintenance and upgrades will require additional and continuous financial support.

## 3.3 Future work

The development of NutriBase demonstrates the complexity of the food compilation process. It shows that many activities have to be performed to develop and maintain high-quality D&K, and to construct the semantic resources needed for the automation of specific steps. The results of the manually performed work presented in the current paper could serve as input for FNS-Harmony. Additionally, new computer-based methodologies to support our future work have been developed, and some solutions have already been implemented as openly available web services (e.g., through the FNS-Cloud catalog [36]). In order to speed up the compilation process, Ispirova et al. (54) developed the methodology for automatic identification of different names of the same foods or dishes (e.g., eggplant and aubergine).

To enable rapid upgrades of D&K, the tool will be integrated within existing or developing knowledge graphs [e.g., FoodKG on food recommendations, FooDB, knowledge graphs on food-disease and food-chemical relations (31, 55), and a knowledge graph on food consumer knowledge being under development within the COMFOCUS project (30)]. Since NutriBase is designed to integrate data with knowledge that is formalized with respect to standardized semantic resources, the connection with any healthcare information system compliant with the openEHR standard (56) is possible. Furthermore, for branded foods and recipes using branded foods as ingredients, the algorithm to calculate values for components that are not mandatory to be included on the nutrition declaration table, can be implemented by using the food matching web services developed within FNS Cloud (57).

Moreover, current FCDBs imported into NutriBase will be updated with the latest releases found, and additional FCDBs may be added. Complementing a FCDB with generic food images would also be beneficial; however, a database of standardized images for generic foods is currently lacking. Having such a database and linking it to FCDBs would facilitate food identification within the FCDBs and support research focusing on automated food image recognition (58). This could further assist in dietary intake assessments and portion size estimations, especially if measurement aids [e.g., (59)] are included.

## 4 Conclusion

The tool called NutriBase presented in the current paper is a comprehensive system that includes not only multiple FCDBs, but also KBs. Combining FCD with relevant knowledge is an extension of what has already been done in this research area. Moreover, all D&K imported are harmonized and compiled with respect to various well-established standards. NutriBase can help create and maintain the quality management system needed to ensure data quality. Merging quality management systems with data production and compilation management enhances the monitoring and assessment of FCDBs, thereby increasing their credibility among consumers, experts, policymakers, and other stakeholders. Additionally, using NutriBase reduces the time required to review FCD by enabling users to add, edit, link, and integrate data with knowledge, all in one place. Domain experts who evaluated and validated the tool would recommend using the system and believe that it is a very usable tool (SUS score 78.9). Moreover, NutriBase represents an important step in transparently borrowing imputed data, and therefore reducing missing data. Lastly, the system is highly modifiable and can be further customized to meet different requirements at the national or international level. Existing and newly generated D&K can be continuously added as long as they comply with standards, which would strengthen the tool even more.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: in the paper an infrastructure in a form of a new data- and knowledge base management system is presented, thus no dataset was generated. The management system is not yet publicly available. Requests to access these datasets should be directed to eva.valencic@ijs.si.

## Author contributions

EV: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. EB: Writing – review & editing. TB: Writing – review & editing. CC: Writing – review & editing. BKS: Conceptualization, Funding acquisition, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Ministry of Health of Slovenia [Grant Agreement No. C2711-19-185021] within the OPKP ('Upgrade of the Open platform for clinical nutrition with a mobile app') project, and by the Slovenian Research Agency (research core grant number P2-0098). Additionally, this work was undertaken within the Horizon Europe FishEUTrust project [Grant Agreement No. 101060712] and Horizon 2020 COMFOCUS project [Grant Agreement No. 101005259]. CC is supported by an NHMRC Leadership Research Fellowship (L3, APP2009340).

## Acknowledgments

The authors gratefully acknowledge the work of Matevž Ogrinc, Robert Modic, Andraž Simčič, Tina Kondić, Luka Šveigl, Jan Šuklje, Gregor Novak, Anemari Voršič, Gordana Ispirova and Peter Novak for their contribution to the development and design of the tool. We would also like to thank Erika Jesenko, Julija Repolusk, Jernej Ogrin, Kaja Kranjc and Emanuela Čerček Vilhar for their contribution to editing and supplementing the database. In addition, we would like to thank Blaž Ferjančič, Anja Bolha and Mojca Korošec, Tanja Pajk Žontar, Jasna Bertoncelj and Saša Piskernik from the Biotechnical Faculty of the University of Ljubljana for their contribution to the development, review and the update of Slovenian food composition database. Last but not least, we acknowledge scientific contribution of Tome Eftimov, Gjorgjina Cenikj, Ana Gjorgjevikj and Jan Drole.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Harrington RA, Adhikari V, Rayner M, Scarborough P. Nutrient composition databases in the age of big data: foodDB, a comprehensive, real-time database infrastructure. *BMJ Open*. (2019) 9:e026652. doi: 10.1136/bmjopen-2018-026652

2. Westenbrink S, Oseredczuk M, Castanheira I, Roe M. Food composition databases: the EuroFIR approach to develop tools to assure the quality of the data compilation process. *Food Chem*. (2009) 113:759–67. doi: 10.1016/j.foodchem.2008.05.112

3. Greenfield H, Southgate DA. Food composition data: production, management, and use. Rome: Food and Agriculture Organization (2003).

4. Elmadfa I, Meyer AL. Importance of food composition data to nutrition and public health. *Eur J Clin Nutr*. (2010) 64:S4–7. doi: 10.1038/ejcn.2010.202

5. Emara Y, Koroušić Seljak B, Gibney ER, Popovski G, Pravst I, Fantke P. Workflow for building interoperable food and nutrition security (FNS) data platforms. *Trends Food Sci Technol*. (2022) 123:310–21. doi: 10.1016/j.tifs.2022.03.022

6. Euro FIR. (2022). Euro FIR-Food EXplorer. Available at: https://www.eurofir.org/our-tools/foodexplorer/ (Accessed August 10, 2024).

7. U.S. Department of Agriculture. (2022). FoodData central, food composition database. Available at: . (https://fdc.nal.usda.gov)

8. Food and Agriculture Organization. (2022). INFOODS-international network of Food data systems. Available at: . (https://www.fao.org/infoods/en/)

9. Canadian Institutes of Health Research, Canada Foundation for Innovation, TMIC-the metabolomics innovation Centre. (2022). FooDB. Available at: . (https://foodb.ca)

10. Gurinović M, Milešević J, Kadvan A, Djekić-Ivanković M, Debeljak-Martačić J, Takić M, et al. Establishment and advances in the online Serbian food and recipe data base harmonized with EuroFIR™ standards. *Food Chem*. (2016) 193:30–8. doi: 10.1016/j.foodchem.2015.01.107

11. Westenbrink S, van der Vossen-Wijmenga W, Toxopeus I, Milder I, Ocké M. LEDA, the branded food database in the Netherlands: data challenges and opportunities. *J Food Compos Anal*. (2021) 102:104044. doi: 10.1016/j.jfca.2021.104044

12. EDAMAM. (2022). Food and grocery database. Available at: . (https://developer.edamam.com/food-database-api)

13. GS1 Standards. (2022). GS1. Available at: . (https://www.gs1.org/standards)

14. CEN. (2022). European standard. Food data–structure and interchange format. EN 16104: 2012. Available at: . (https://www.en-standard.eu/bs-en-16104-2012-food-data-structure-and-interchange-format/)

15. METROFOOD. (2022). Infrastructure for promoting metrology in Food and nutrition. Available at: . (https://www.metrofood.eu)

16. Charrondiere UR, Stadlmayr B, Grande F, Vincent A, Oseredczuk M, Sivakumaran S, et al. FAO/INFOODS evaluation framework to assess the quality of published food composition tables and databases - user guide. Rome, Italy: FAO (2023).

17. Oseredczuk M, Salvini S, Moller A, Roe M, Castanheira I, Colombani P, et al. (2013). Guidelines for quality index attribution to original data from scientific literature or reports for EuroFIR data interchange. Compilers' Toolbox™. Available at: http://toolbox.foodcomp.info/References/ValueDocumentation/EuroFIR_QE%20scirep%20guidelines_28March2013.pdf (Accessed September 15, 2024).

18. European Food Information Resource. (2022). The EuroFIR thesauri 2008 technical report D1.8.22. Available at: . (https://www.eurofir.org/our-resources/eurofir-thesauri/)

19. CEN/TC 387. (2022). Food data – Data structure. Available at: . (https://www.eurofir.org/food-information/how-are-fcdbs-made/quality-and-standards/common-standard/)

20. Ispirova G. Exploiting domain knowledge in predictive learning from Food and nutrition data. Ljubljana: G. Ispirova (2022).

21. Gjorshoska I, Eftimov T, Trajanov D. Missing value imputation in food composition data with denoising autoencoders. *J Food Compos Anal*. (2022) 112:104638. doi: 10.1016/j.jfca.2022.104638

22. ATLASSIAN. (2022). What is a knowledge base? Available at: https://www.atlassian.com/itsm/knowledge-management/what-is-a-knowledge-base (Accessed September 15, 2024).

23. Langual (2022) Langual-the international framework for food description. Available at: . (https://www.langual.org/default.asp)

24. European Food Safety Authority. The food classification and description system FoodEx2. Hoboken: Wiley (2015).

25. EMBL-EBI. (2022). ChEBI. Available at: https://www.ebi.ac.uk/chebi/ (Accessed September 17, 2024).

26. Machackova M, Giertlova A, Porubska J, Roe M, Ramos C, Finglas P. EuroFIR guideline on calculation of nutrient content of foods for food business operators. *Food Chem*. (2018) 238:35–41. doi: 10.1016/j.foodchem.2017.03.103

27. Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Sci Food*. (2018) 2:23. doi: 10.1038/s41538-018-0032-6

28. Eftimov T, Ispirova G, Potočnik D, Ogrinc N, Koroušić SB. ISO-FOOD ontology: a formal representation of the knowledge within the domain of isotopes for food science. *Food Chem*. (2019) 277:382–90. doi: 10.1016/j.foodchem.2018.10.118

29. FNS harmony. (2022). Application ontology for the H2020 FNS-cloud project. Available at: https://github.com/panovp/FNS-Harmony (Accessed August 12, 2024).

30. H2020 project. (2022). Community on Food consumer science (COMFOCUS). Available at: . (https://comfocus.eu)

31. Cenikj G, Strojnik L, Angelski R, Ogrinc N, Koroušić Seljak B, Eftimov T. From language models to large-scale food and biomedical knowledge graphs. *Sci Rep*. (2023) 13:7815. doi: 10.1038/s41598-023-34981-4

32. FoodKG. (2022). FoodKG: a semantics-driven knowledge graph for Food recommendation. Available at: . (https://foodkg.github.io)

33. TasteAtlas (2023). Travel global, eat local. Available at: . (https://www.tasteatlas.com)

34. Presser K, Weber D, Norrie M. FoodCASE: a system to manage food composition, consumption and TDS data. *Food Chem*. (2018) 238:166–72. doi: 10.1016/j.foodchem.2016.09.124

35. GI search. Glycemic index research and GI News. Camperdown: The University of Sydney (2022).

36. Durazzo A, Astley S, Kapsokefalou M, Costa HS, Mantur-Vierendeel A, Pijls L, et al. Food composition data and tools online and their use in research and policy: EuroFIR AISBL contribution in 2022. *Nutrients*. (2022) 14:4788. doi: 10.3390/nu14224788

37. Korošec M, Golob T, Bertoncelj J, Stibilj V, Seljak BK. The Slovenian food composition database. *Food Chem*. (2013) 140:495–9. doi: 10.1016/j.foodchem.2013.01.005

38. Food and Agriculture Organization (2022). FAO/INFOODS Recipes. Available at: https://www.fao.org/infoods/infoods/recipes/en/. (Accessed December 20, 2024).

39. Reinivuo H, Bell S, Ovaskainen ML. Harmonisation of recipe calculation procedures in European food composition databases. *J Food Compos Anal*. (2009) 22:410–3. doi: 10.1016/j.jfca.2009.04.003

40. Pravst I, Hribar M, Žmitek K, Blažica B, Koroušić Seljak B, Kušar A. Branded foods databases as a tool to support nutrition research and monitoring of the Food supply: insights from the Slovenian composition and labeling information system. *Front Nutr*. (2022) 8:8576. doi: 10.3389/fnut.2021.798576

41. GO FAIR. (2023). FAIR principles. Available at: https://www.go-fair.org/fair-principles/ (Accessed April 5, 2023).

42. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. *BMC Med Inform Decis Mak*. (2022) 22:22. doi: 10.1186/s12911-022-01764-2

43. FNS-H. (2023). FNS-harmony ontology. Available at: https://bioportal.bioontology.org/ontologies/FNS-H (Accessed September 26, 2024).

44. GSI. (2023). GSI. Available at: https://www.gs1.org/services/gdsn (Accessed December 27, 2022).

45. Dunford E, Webster J, Metzler AB, Czernichow S, Mhurchu CN, Wolmarans P, et al. International collaborative project to compare and monitor the nutritional composition of processed foods. *Eur J Prev Cardiol*. (2012) 19:1326–32. doi: 10.1177/1741826711425777

46. Ispirova G, Cenikj G, Ogrinc M, Valenčič E, Stojanov R, Korošec P, et al. CafeteriaFCD Corpus: Food consumption data annotated with regard to different Food semantic resources. *Food Secur*. (2022) 11:2648. doi: 10.3390/foods11172684

47. National Institute for public health. (2022). Reference values for energy and nutrient intake tabular recommendations for children (from 1 year of age), adolescents, adults, older adults, pregnant women and nursing mothers. Available at: https://www.

nijz.si/sites/www.nijz.si/files/uploaded/referencne_vrednosti_2020_3_2.pdf (Accessed September 22, 2024).

48. Zeb A, Soininen JP, Sozer N. Data harmonisation as a key to enable digitalisation of the food sector: a review. *Food Bioprod Process*. (2021) 127:360–70. doi: 10.1016/j.fbp.2021.02.005

49. Bauer J, Biolo G, Cederholm T, Cesari M, Cruz-Jentoft AJ, Morley JE, et al. Evidence-based recommendations for optimal dietary protein intake in older people: a position paper from the PROT-AGE study group. *J Am Med Dir Assoc*. (2013) 14:542–59. doi: 10.1016/j.jamda.2013.05.021

50. Mariotti F, Tomé D, Mirand PP. Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Crit Rev Food Sci Nutr*. (2008) 48:177–84. doi: 10.1080/10408390701279749

51. Global dietary database. (2022). Global dietary database. Available at: https://www.globaldietarydatabase.org (Accessed September 22, 2024).

52. EFSA. (2022). OpenFoodTox. Available at: https://www.efsa.europa.eu/en/data-report/chemical-hazards-database-openfoodtox (Accessed September 22, 2024).

53. EFSA. (2022). EFSA-European Food safety authority. Available at: https://www.efsa.europa.eu/en/publications (Accessed September 22, 2024).

54. Ispirova G, Eftimov T, Koroušić SB. Domain heuristic fusion of multi-word Embeddings for nutrient value prediction. *Mathematics*. (2021) 9:1941. doi: 10.3390/math9161941 (Accessed September 27, 2024).

55. Cenikj G, Seljak BK, Eftimov T. (2021). FoodChem: a food-chemical relation extraction model. In: 2021 IEEE symposium series on computational intelligence. IEEE: Orlando, FL. 1–8.

56. OpenEHR. (2024). The future of health & care is open. Available at: https://openehr.org (Accessed September 25, 2024).

57. Presser K., Kapela W., Astley S., Roe M., Finglas P. (2022). Catalogues for data sources and services. Available at: https://zenodo.org/record/6966070#.Y66RKS8w1Bx (Accessed September 15, 2024).

58. Mezgec S, Eftimov T, Bucher T, Koroušić SB. Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public Health Nutr*. (2019) 22:1193–202. doi: 10.1017/S1368980018000708

59. Bucher T, Weltert M, Rollo ME, Smith SP, Jia W, Collins CE, et al. The international food unit: a new measurement aid that can improve portion size estimation. *Int J Behav Nutr Phys Act*. (2017) 14:124. doi: 10.1186/s12966-017-0583-y

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 19 February 2025
DOI 10.3389/fnut.2025.1516521

Check for updates

OPEN ACCESS

EDITED BY
Massimo Lucarini,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Yuwei Yuan,
Zhejiang Academy of Agricultural Sciences, China
Oana Romina Botoran,
National Research and Development Institute for Cryogenic and Isotopic Technologies (ICSI), Romania

*CORRESPONDENCE
Nives Ogrinc
✉ nives.ogrinc@ijs.si

RECEIVED 24 October 2024
ACCEPTED 03 February 2025
PUBLISHED 19 February 2025

CITATION
Terro C, Modic R, Ogrinc M, Simčič A, Drole J, Eftimov T, Koroušić Seljak B and Ogrinc N (2025) IsoFoodTrack: a comprehensive database and management system based on stable isotope ratio analysis for combating food fraud.
*Front. Nutr.* 12:1516521.
doi: 10.3389/fnut.2025.1516521



# IsoFoodTrack: a comprehensive database and management system based on stable isotope ratio analysis for combating food fraud

Cathrine Terro[1,2], Robert Modic[3], Matevž Ogrinc[3], Andraž Simčič[3], Jan Drole[3], Tome Eftimov[2,3], Barbara Koroušić Seljak[2,3] and Nives Ogrinc[1,2]*

[1]Department of Environmental Sciences, Jožef Stefan Institute, Ljubljana, Slovenia, [2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3]Computer Systems, Jožef Stefan Institute, Ljubljana, Slovenia

The IsoFoodTrack database is a comprehensive, scalable, and flexible platform designed to manage isotopic and elemental composition data for a wide range of food commodities. It supports research in food authenticity and fraud detection by integrating isotopic data with rich metadata, including geographical, production, and methodological details. The database is built for scalability, allowing the addition of new commodities, analytical methods, and metadata fields, while ensuring interoperability with external databases through standardized formats and API integration. Based on the data collected in IsoFoodTrack using statistical, chemometric and machine learning approaches it has a capability to identify and classify the origin of food commodities. IsoFoodTrack also supports isotope mapping (isoscapes), providing spatially continuous predictions that enhance the detection of food fraud. Rigorous quality control measures ensure high data reliability, and the user-friendly web interface facilitates easy access and visualization. Openly accessible through platforms like National Center for Biomedical Ontology (NCBO) BioPortal, IsoFoodTrack is positioned for future expansion and integration of open-access data, making it a vital tool for researchers and regulatory agencies in ensuring food authenticity and traceability.

KEYWORDS

**database, stable isotope ratio analysis, elemental composition, food fraud, authenticity, interoperability**

## 1 Introduction

Food fraud, which refers to the economically motivated adulteration and mislabeling of food products, continues to be a major issue for food producers as well as consumers. Among the techniques available for detecting fraud, stable isotope fingerprinting is leading the way in establishing the authenticity and geographical origin of food products. This choice is based on the fact that the distribution of stable isotopes of carbon ($^{12}C$, $^{13}C$), nitrogen ($^{14}N$, $^{15}N$), sulfur ($^{32}S$,

10.3389/fnut.2025.1516521

$^{34}S$), hydrogen ($^{1}H$, $^{2}H$), and oxygen ($^{16}O$, $^{18}O$)[1] is influenced by fractionation processes linked to local climate, geology, and soil characteristics (1, 2). These processes result in varying rates of isotope transfer from natural sources such as water, soil, and the atmosphere to plant or animal tissues. For example, the isotope ratios in water ($^{2}H/^{1}H$ and $^{18}O/^{16}O$) provides critical information about local precipitation, surface water, and groundwater, influenced by factors like latitude, altitude, distance from the sea, precipitation levels, and evapotranspiration. The verification of regional origin becomes even more robust when isotope data are combined with elemental composition profiles (3). However, to determine authenticity, a suitable reference dataset of analyzed authentic products is required. This dataset should include samples representative of a wide range of geographical, seasonal, dietary, and production conditions. Authenticity is then assessed by comparing the values found in commercial samples with the limits estimated from the reference dataset, using a suitable statistical model to evaluate the best fit. These databases also need to be continuously curated and kept up to date, which is a considerable task, given the amount of variation that needs to be included.

A prime example of a well-established database is the European Wine DataBank, which the European Commission has maintained for over 20 years (4). However, while some databases are publicly available, many others are not freely shared due to intellectual property concerns and differences in sample pre-treatment methods. Nevertheless, the use of isotope databases is expanding; for example, the Stable Isotope Ratio Analysis (SIRA) database is already being applied to products like Parma ham, Grana Padano cheese, and Parmigiano Reggiano in Italy (5). Other examples include the pork origin database managed by the UK's Agriculture and Horticulture Development Board (AHDB), the egg database Kontrollierte Alternative Tierhaltungsformen (KAT), and asparagus databases in German food control labs.

To address the gaps in the availability and accessibility of such data, we have developed a comprehensive database and database management system (DBMS) called IsoFoodTrack.[2] This system provides extensive data on the stable isotopes of light elements and the elemental composition of authentic samples from various food commodities such as oils, milk and dairy products, meat, spices, truffles, seafood and vegetables. Furthermore, IsoFoodTrack is designed to be interoperable, allowing connection with other databases or centralized repositories. IsoFoodTrack represents a significant advancement over traditional food databases by prioritizing both accessibility and standardization. It incorporates open-access principles, ensuring that researchers from diverse regions can utilize its resources without significant barriers. Additionally, the database integrates standardized metadata protocols and harmonized data entry formats, which streamline cross-study comparisons and enhance reproducibility.

Additionally, there is a growing movement toward creating a centralized repository for isotopic data, as proposed by Pauli et al. (6), with the development of IsoBank. IsoBank aims to function similarly to GenBank in the field of genetics, serving as both an aggregator and repository of open-access isotope data. IsoBank is designed as a general-purpose repository for stable isotope data across all disciplines. It supports the storage and retrieval of isotope measurements irrespective of their context. It serves a broad research community, including fields like ecology, geology, archeology, and biology, among others. This resource will promote interdisciplinary research, facilitate data-sharing, and provide valuable educational opportunities by offering real-world isotopic data for students and researchers alike. On the other hand, IsoFoodTrack database serves as a specialized database aimed at practical applications in food traceability and authenticity verification, prioritizing functionality tailored to its specific use case. Its scope is narrower, targeting applications in food science, agriculture, and regulatory frameworks.

In this paper, we present the IsoFoodTrack database as the first effort to organize open-access stable isotope data for food research. It is organized in different sections including: database design, methods and technical aspects. Section 5 details the validation of IsoFoodTrack, demonstrating its practical application, while section 6 deals with database curation and availability. Finally, section 7 concludes the paper by discussing key achievements and contributions.

## 2 Database design

The design of the IsoFoodTrack database is crucial for ensuring the effective management of isotopic and elemental composition data for various food commodities. The database was developed with a focus on scalability, flexibility, and data integrity, enabling researchers to store, retrieve, and analyze stable isotope data in a structured and efficient manner. This section outlines the key aspects of the database design, including the data model, schema design, relationships between entities, and considerations for maintaining accuracy and performance.

### 2.1 Data model

The IsoFoodTrack database follows a relational database model, which is well-suited for managing structured data with clearly defined relationships between entities. The relational model enables efficient querying of data, as well as maintaining consistency and data integrity through the use of primary and foreign key constraints. This design ensures that all data points, including isotopic ratios, elemental compositions, and metadata about samples, are properly linked to their relevant entities.

The structure of IsoFoodTrack is presented in Figure 1.

The main entities in the database include:

- *Samples*: representing the physical food samples analyzed for isotope and elemental composition. This also includes the type of

1 Measurements of the stable isotope ratios of light elements are expressed in the $\delta$-notation in ‰ according to the equation: $\delta^{i}E = (R(^{i}E/^{j}E)_{sample}/R(^{i}E/^{j}E)_{standard}) - 1$, where *i* stands for the highest and *j* for the lowest atomic mass number of the element E (C, N, O, S), and R is the isotope ratio between the heavier and the lighter isotope of the element ($^{2}H/^{1}H$, $^{13}C/^{12}C$, $^{15}N/^{14}N$, $^{18}O/^{16}O$, $^{34}S/^{32}S$) in the *sample* or *standard*. The $\delta^{13}C$ values are expressed relative to V-PDB (Vienna-Pee Dee Belemnite) standard, $\delta^{15}N$ values relative to AIR, $\delta^{34}S$ values relative to V-CDT (Vienna-Canyon Diablo Troilite) standard and the $\delta^{2}H$ and $\delta^{18}O$ values relative to the VSMOW (Vienna-Standard Mean Ocean Water) standard.

2 http://isofoodtrack.ijs.si

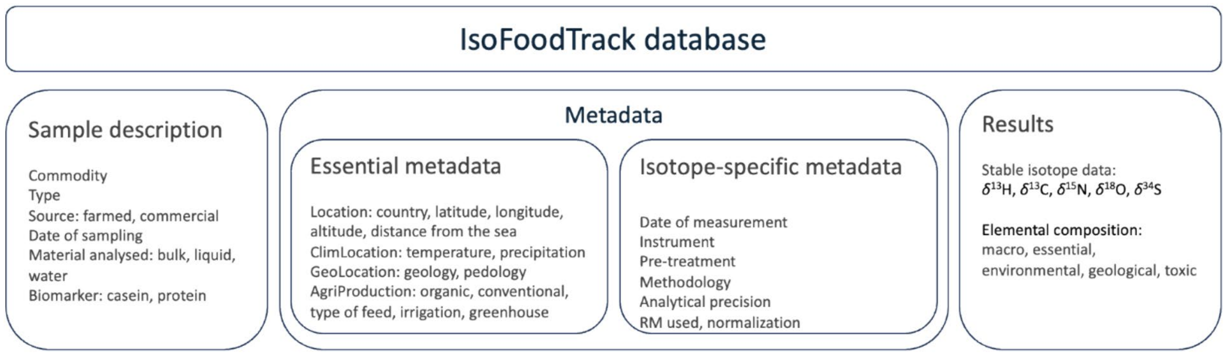



FIGURE 1
Structure of the IsoFoodTrack database.

sample, source of sample (authentic, commercial), date of sampling and compound analysed: bulk sample (freeze-dry or liquid), sample water, extracted components, fatty acid.

- *Metadata*: the metadata cover two categories: (i) essential metadata, describing every data record, and (ii) isotope-specific metadata. The essential metadata includes *geographical data:* storing information about the geographical location of each sample, including details such as latitude, longitude, altitude, distance from the sea. In addition, the data on yearly average amount of precipitation, average temperature of the location, geology and pedology are also included.

*Production data:* capturing details about the production and processing of the food samples, such as year of production, type of material (authentic, commercial), farming practices, seasonal information, and production methods (e.g., organic or conventional; if known).

The *isotope-specific metadata* includes reference materials used to normalize the data. Stable isotope data are produced in a wide range of research and commercial laboratories. Although the methods by which the majority of data, i.e., bulk carbon ($\delta^{13}$C) and nitrogen ($\delta^{15}$N) stable isotope values, are standardized, laboratories often use slightly different protocols and different laboratory reference materials to normalize data to internationally accepted scales. Other isotopes (e.g., $\delta^{2}$H and $\delta^{18}$O) have more fundamental issues associated with comparability of measurements (7). Hydrogen in an exchangeable position (e.g., when bound to oxygen in a hydroxyl group, as in proteinaceous material) will exchange with atmospheric water vapor, leading to potentially erroneous results unless controlled. Thus, to ensure data robustness, quality and user confidence, all pertinent analytical information for each piece of data is recorded. IsoFoodTrack metadata includes sample pretreatment methods (e.g., lipid extraction), reference materials used, type of normalization (one, two, multi-point normalization) and instrumentation used.

- *Isotope data*: storing detailed information about the isotopic composition of each sample, including ratios of isotopes such as $\delta^{2}$H, $\delta^{13}$C, $\delta^{15}$N, $\delta^{18}$O and $\delta^{34}$S (in ‰).
- *Elemental data*: capturing elemental concentrations for key elements: B, Na, Mg, Al, P, S, K, Ca, V, Cr, Mn, Fe, Co, Cu, Zn, As, Se, Rb, Sr., Mo, Cd, Cs, Ba, Hg, Pb.

## 2.2 Scalability and interoperability

The design of IsoFoodTrack anticipates the continuous expansion of the database as new samples are collected and analyzed. As such, the database architecture is scalable, with the ability to accommodate additional tables for new food commodities or analytical methods such as compound specific analysis.

# 3 Methods

The development of the IsoFoodTrack database involved several key stages, including the collection and preparation of authentic food samples, the analysis of isotopic and elemental compositions, the organization and management of data, and the validation of the database for practical applications in food authenticity testing. This section describes the methodology employed to create and curate the IsoFoodTrack database.

## 3.1 Sample collection and preparation

Sample collection represents a crucial step in the formation of the IsoFoodTrack database, and in order to ensure the accuracy of a food authenticity database, authentic samples must be used. Ideally, samples should be collected from primary producers by impartial collectors to ensure traceability and authenticity. In comparison, retail samples are less reliable due to possible contamination in the supply chain. When creating the database, sample size and variety are also important and should reflect natural variations, e.g., geography, breed, and climate. Additionally, the database should be validated for its intended use, and statistical analysis should be considered when determining sample size. The selection of reference data from the database is crucial and should be left to experts only (8, 9).

In our case, the sampling follows an appropriate protocol developed for various applications to mitigate potential biases caused by the overrepresentation of specific regions. This protocol is based on our prior experience and expert knowledge. To ensure a robust and representative dataset, samples were collected using the following criteria:

*Geographical diversity*: samples were sourced from a wide range of geographical locations, including different latitudes, altitudes, and climatic zones. This ensures that the database captures the natural variation in isotopic and elemental signatures that arise from environmental factors such as local precipitation, soil types, and temperature. In case of Slovenia, to account for natural variability, sampling was designed to reflect Slovenia's regionalization, which is divided into four distinct geographical regions: Dinaric, Mediterranean, Alpine, and Pannonian.

*Seasonal and temporal variation*: samples were collected over multiple growing seasons and harvest periods to account for seasonal changes in isotopic and elemental composition. These variations can be influenced by factors such as precipitation levels and temperature fluctuations throughout the year.

*Production methods*: both conventional and organic food production methods were represented in the sample set.

It is also useful to understand the production density of a foodstuff and the number of relevant authentic samples. For example, the Slovenian wine database, established in 1996, contains 25 authentic wine samples covering various geographical regions and varieties. The database is also included in EU Wine Databank. Another example is related to verification of correct labeling of selected fruits and vegetables on Slovenian market including strawberries, cherries, asparagous apples, kaki and garlic. Authentic samples are provided annually by the regional units of the Administration of the Republic of Slovenia for Food Safety, Veterinary, and Plant Protection from producers from various geographical production areas. This research began in 2018 and requires at least 30 authentic samples covering four geographical regions.

Once collected, each sample was carefully labeled with metadata, including information on the geographic origin, date of collection, food type, and any relevant production details. Samples were then prepared for isotopic and elemental analysis according to standardized operational protocols, ensuring consistency in the treatment of all samples.

Although IsoFoodTrack has been designed to cover only a limited area, such as Slovenia, it is crucial to recognize the limitations inherent in regions with sparse data availability on a global scale. Low-data regions may exhibit incomplete coverage, which can limit the robustness of isotopic analysis in those areas. To address this, IsoFoodTrack could adopt the following two strategies to enhance its global applicability: encouraging contributions from other researchers, initiatives that can fill data gaps and build regional datasets and leveraging artificial intelligence (machine learning) to predict isotopic baselines in low-data regions, taking into account appropriate uncertainty metrics.

## 3.2 Isotopic and elemental analysis

The analysis of stable isotope ratios and elemental composition was conducted using precise and well-established techniques. The two main analytical methods used were Isotope Ratio Mass Spectrometry (IRMS) and Inductively Coupled Plasma Mass Spectrometry (ICP-MS) for elemental profiling.

### 3.2.1 Isotopic analysis

There are only a few exceptions where no sample treatment is needed, such as determining the $\delta^{18}$O value of water in food and the $\delta^{2}$H, $\delta^{13}$C, and $\delta^{18}$O values of olive oil. However, in most cases, sample treatment is required since it permits the isolation of components that have a stronger geographical fingerprint than the bulk sample and with less interference. For example, samples containing lipids (e.g., meat, fish, milk, and cheese) are usually defatted because fat has a different C and H composition from the other food constituents, and therefore, its variable quantity can affect the overall isotopic signature. In the IsoFoodTrack samples, defatting was performed using a mixture of petroleum ether and diethyl ether (2:1 v/v). Before analysis, all fractions were freeze-dried and stored at room temperature.

Isotope ratios of hydrogen ($^{1}$H/$^{2}$H), carbon ($^{12}$C/$^{13}$C), nitrogen ($^{14}$N/$^{15}$N), oxygen ($^{16}$O/$^{18}$O), and sulfur ($^{32}$S/$^{34}$S) were measured using an isotope ratio mass spectrometer (IRMS) with different preparation systems. The solid and liquid samples are measured by elemental analyser coupled to IRMS (EA/IRMS), $^{1}$H/$^{2}$H and $^{16}$O/$^{18}$O in organic matrices with a TC/EA pyrolyser coupled to IRMS, while $^{1}$H/$^{2}$H and $^{16}$O/$^{18}$O in water in food samples is determined with MultiFlow Bio equilibration system connected to IRMS.

Validation of a database includes the data within it and its ability to complete the role for which it was designed. All data used to create the database must be validated, i.e., reliable, and all measurements must follow the protocol suggested by Skrzypek et al. (10). The validation tests highlighted in the manuscript were instrumental in assessing its reliability. These tests revealed occasional false positives and misclassifications during bivariate evaluations, particularly in food samples with isotopic compositions near boundary thresholds. For example, foods sourced from regions with similar climatic and environmental conditions exhibited overlapping isotopic signatures. To address these issues, enhanced multivariate analyses were implemented to improve classification accuracy, reducing false-positive rates to below 5% in most categories.

Further, if, upon re-analysis of the samples, data that are consistent with the initial "outlier" data are recorded, further investigation are undertaken to determine the underlying cause. Typically, outliers are due to particular and unique technological or geographical issues, such as a particular microclimate or technological choice. In this case, further investigations are carried out to understand if the outliers belong to another population of data or if they are just "outliers" falling in the percentage of error of the chosen confidence level (for example, 5% for 95% confidence level).

It is strongly recommended that laboratories are accredited to ISO17025 or can demonstrate that they have equivalent quality control systems. This is specifically required if we use the databank to verify the authenticity of commercial samples for food control purposes. According to the norm EN ISO/IEC 17025, the test result of an analytical measurement has to be stated with an estimate of its uncertainty, for example, when uncertainty affects compliance with an authenticity limit. Measurement uncertainty is usually reported in the reference methods (in the case of official and validated methods), or it can be estimated using different methods. Dunn et al. (11) recently published guidance for calculating measurement uncertainty of stable isotope ratio delta values. A further requirement of accreditation is that laboratory must participate in proficiency testing that comply with the ISO/IUPAC/AOAC International Harmonized Protocol for Proficiency

Testing of analytical laboratories. In our case this involves participation in the Food Analysis using Isotopic Techniques Proficiency Testing Scheme (FIT PTS), organized by Eurofins Scientific three times per year and includes samples from various food commodities.

### 3.2.2 Elemental analysis

The elemental composition of each sample was analyzed using Inductively Coupled Plasma Mass Spectrometry (ICP-MS), which is highly sensitive to trace elements. The elemental composition in our samples was performed on triple quadrupole inductively coupled plasma mass spectrometer, QQQ-ICP-MS (Agilent, USA). Prior to analysis, the samples were digested using appropriate chemical methods (e.g., acid digestion) to ensure the accurate quantification of elemental concentrations. Detection limits included in the table were calculated based on three standard deviations of blank measurements. A more detailed description of the optimization of the method for elemental analysis for fruits and vegetables can be found in (12). The elemental data were recorded as concentrations (mg/kg or ppm) for key elements that are relevant to distinguishing different geographical origins and production methods.

Both isotopic and elemental data are stored in the IsoFoodTrack database, alongside the associated sample metadata, to allow for comprehensive comparative analyses. In IsoFoodTrack, isotopic and elemental composition data are represented as single-point measurements.

## 4 Technical aspects

The construction of the database involves three main phases:

- Phase I: the database structure definition. The protocol for data preparation was evolved to minimize the labor involved in populating the database. The database also includes metadata that are important for the further evaluation of the data.
- Phase II: filling the database with data gained during the project evolution.
- Phase III: development of routines/queries for extracting data from the database.

The fundamental requirements of IsoFoodTrack included (i) the underlying database and (ii) the application layer (web application).

### 4.1 Underlying database

The following points were considered in the underlying database:

*Database platform*: the IsoFoodTrack database was implemented using PostgreSQL, an open-source relational database system known for its robustness, support for complex queries, and ability to handle large datasets. The specific technology was chosen for its reliability and synergy with other web technologies. PostgreSQL also enables the storage of semi-structured data when necessary.

*Data import and export*: bulk data import and export were handled using python scripts with PostgreSQL connectors. Data entry was streamlined by importing the isotopic and elemental data from excel file.

Underlying database schema is flexible to mitigate the need for further modification to accommodate growing metadata or analytical results requirements.

A detailed visualization of the database's structure is provided in the Supplementary Figure 1. This visualization includes an Entity-Relationship (ER) diagram that maps out the database's tables, the relationships between these tables, and the attributes that define each entity. The diagram identifies primary keys, foreign keys, and various constraints, providing a thorough overview of how data points are linked, referenced, and maintained across the entire IsoFoodTrack ecosystem.

### 4.2 Web application

The web application is a user-friendly interface for interacting and accessing the data from the IsoFoodTrack database. The IsoFoodTrack user interface was built using Django (a Python-based web framework), which allows for integration with the PostgreSQL database. The landing page (Figure 2) is designed to allow users to access the data quickly. All information is organized into categories, and when selecting a category (isotope, elemental composition), a tabular display is presented with the relevant data (Figure 3). The data of elements are grouped into macro, essential, environmental, geological, and toxic categories.

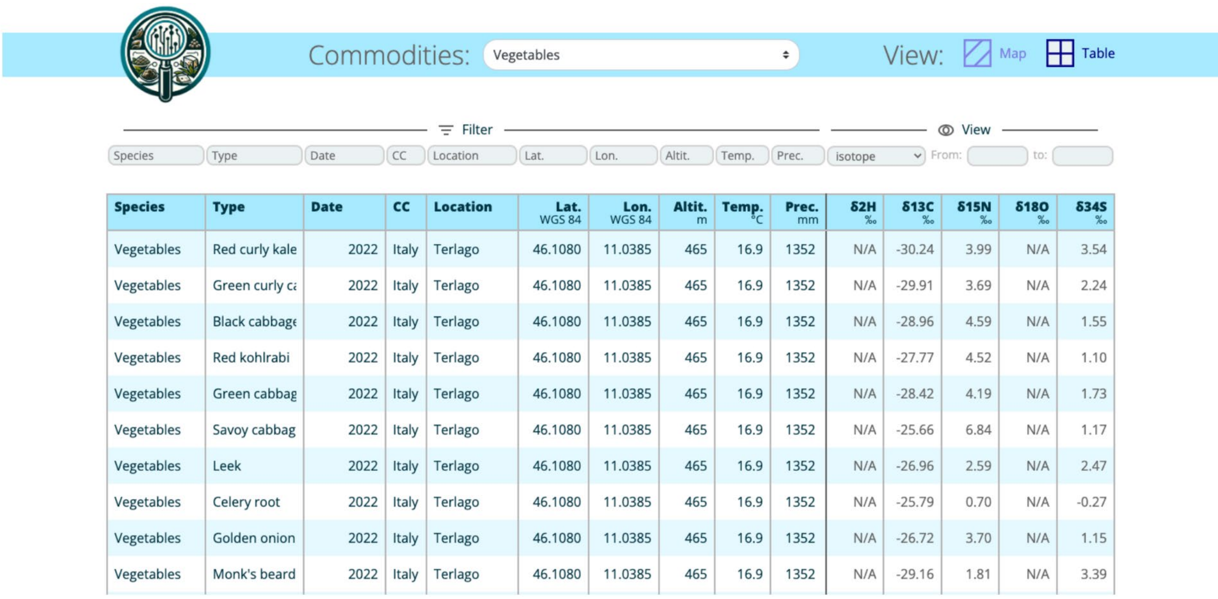
The following aspects were considered for the application:

- Security and Auditing (created, updated): The database is accessible to the public only via the web application. Once all data is published, access will be extended to interested users.
- Configuration or metadata and results: Key information, such as column names, data type, units, location data and categorization, are included in the database (Figure 3).
- Review process: Before any new results are added to the database, they are reviewed by the administrator and experts.
- Data Visualization: Django's ORM (Object-Relational Mapping) was used to query the database and display the results in a user-friendly format. Visualization of data was done using an interactive map with data points representing isotope measurements on different food items from various places around the world (Figure 4). Each data point is clickable and contains relevant information for easier comparison.

### 4.3 Interoperability

Leveraging the data interoperability within the IsoFoodTrack database, an API (Application Programming Interface) can be implemented to allow other systems to interact seamlessly with IsoFoodTrack. This permits external systems or researchers to query the database and retrieve data in standardized formats such as JSON or CSV. The use of standardized data formats and metadata conventions ensures compatibility, facilitating data exchange and cross-referencing with other isotope databases or centralized repositories, like IsoBank.


FIGURE 2
Interface of IsoFoodTrack database.




| Species | Type | Date | CC | Location | Lat. WGS 84 | Lon. WGS 84 | Altit. m | Temp. °C | Prec. mm | δ2H ‰ | δ13C ‰ | δ15N ‰ | δ18O ‰ | δ34S ‰ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vegetables | Red curly kale | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -30.24 | 3.99 | N/A | 3.54 |
| Vegetables | Green curly ca | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -29.91 | 3.69 | N/A | 2.24 |
| Vegetables | Black cabbage | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -28.96 | 4.59 | N/A | 1.55 |
| Vegetables | Red kohlrabi | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -27.77 | 4.52 | N/A | 1.10 |
| Vegetables | Green cabbag | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -28.42 | 4.19 | N/A | 1.73 |
| Vegetables | Savoy cabbag | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -25.66 | 6.84 | N/A | 1.17 |
| Vegetables | Leek | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -26.96 | 2.59 | N/A | 2.47 |
| Vegetables | Celery root | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -25.79 | 0.70 | N/A | -0.27 |
| Vegetables | Golden onion | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -26.72 | 3.70 | N/A | 1.15 |
| Vegetables | Monk's beard | 2022 | Italy | Terlago | 46.1080 | 11.0385 | 465 | 16.9 | 1352 | N/A | -29.16 | 1.81 | N/A | 3.39 |

FIGURE 3
The tabular display of a category.

 10.3389/fnut.2025.1516521

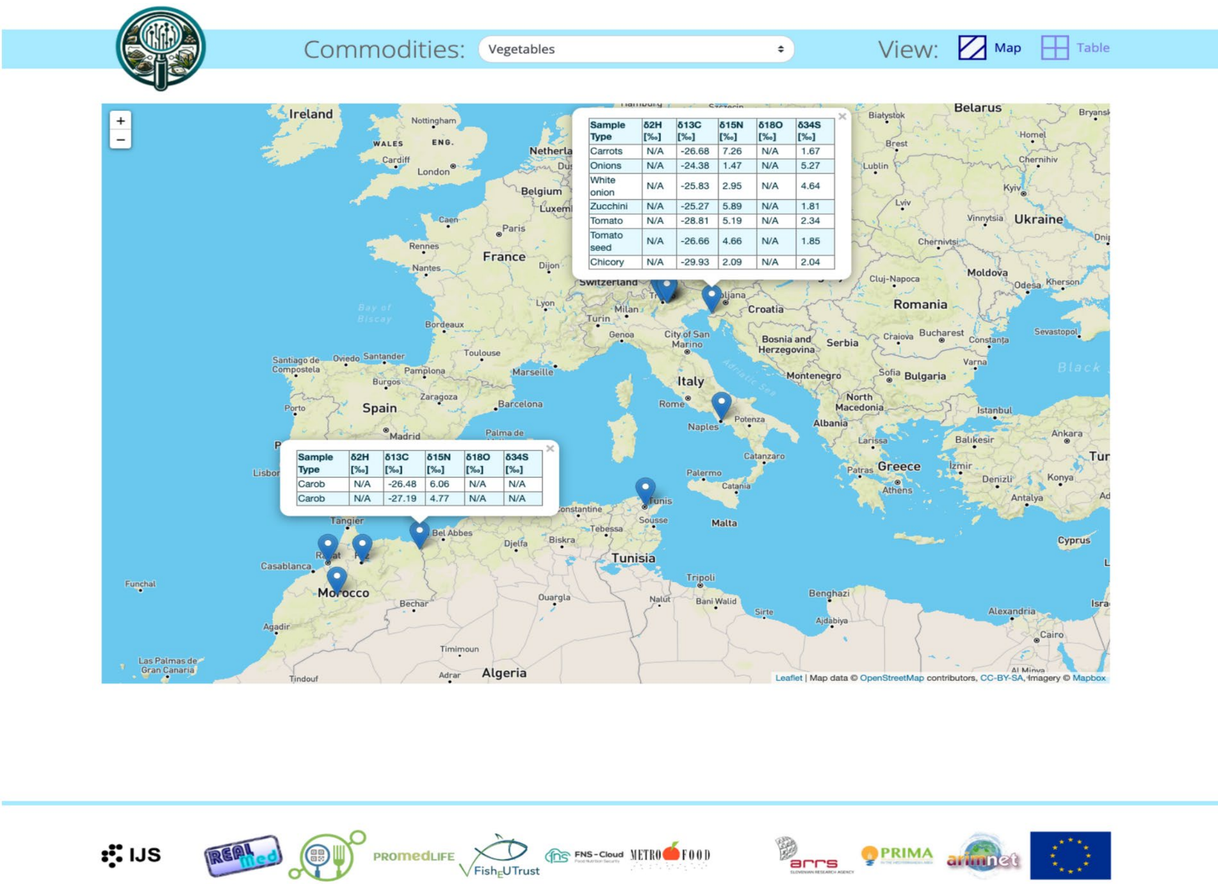



FIGURE 4
Interactive world map with the data.

Further, the metadata was selected from the ISO-FOOD ontology (13), which describes isotopic measurements with all of the necessary information required for future analysis. The ontology is linked with standard ontologies, such as Units of Measurement, Food, Nutrient and Bibliographic Ontologies.

From a higher-level perspective, the system framework and communication between the web application and its backend processes is presented in Figure 5.

Backend:

- PostgreSQL Database: the central storage of the system, containing the collected and structured data.
- Django (Data Processing): acts as the application layer for data processing and logic implementation, bridging the database and the next layer.
- Node.js Server: serves as the middleware or API server, facilitating communication between the backend and the frontend.

Frontend:

- Bootstrap: framework for IsoFoodTrack views.
- Leafletjs: API that serves map view.
- Jquery: a javascript library.

# 5 Application and validation

Validation of the IsoFoodTrack database was an essential step to ensure its practical application in food authentication and fraud detection. The validation process involved several stages:

## 5.1 Reference dataset validation

In order to reliably determine authenticity, the isotopic data of the test samples must be compared with the databank. The most straightforward and still the most recognized approach is that of univariate data evaluation, based on the arithmetic mean, median, standard deviation and authenticity limits considering the Student's t-distribution. These metrics enable users to understand the degree of heterogeneity within a given region and enhance the reliability of origin verification. A 95% confidence level is considered appropriate for commercial samples, which are produced in large batches and should have stable isotope values close to the mean values of the authentic materials. The test result can be clear, in terms of true or false, but also suspicious or unlikely, for example, when the reference databank is not robust enough to be considered reliable. The most efficient approach is to create yearly databases, particularly for vegetable and fruit commodities that exhibit significant variability in harvest and production from year to year.

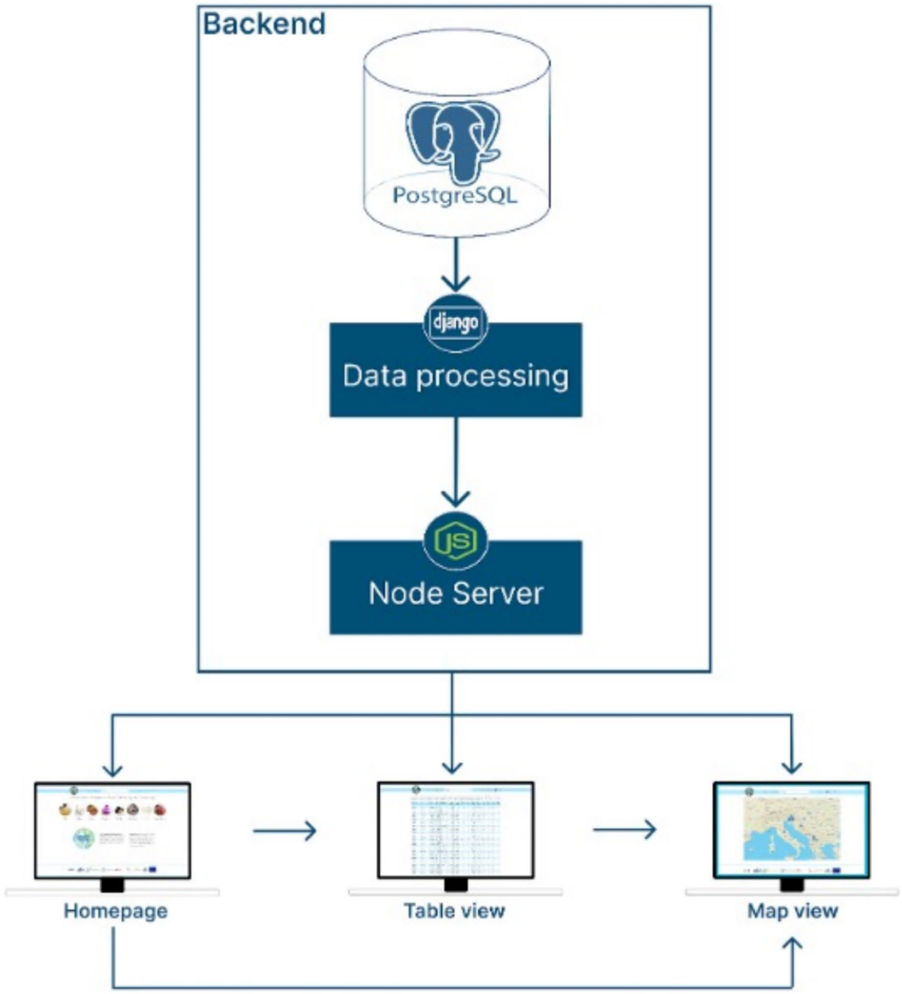



FIGURE 5
System architecture of the IsoFoodTrack framework.

Users should primarily rely on aggregated regional reference values for comparison to account for inherent variability within the dataset. In borderline cases, additional analyses such as examining secondary isotopes or incorporating external metadata (e.g., trace elements or supply chain information) are recommended. For samples classified as "suspicious," users are advised to conduct further investigations, including re-analysis or consultation with experts.

## 5.2 Statistical analysis

Chemometric methods or multivariate data analysis help separate information from noise, uncover hidden correlations, and visually represent them. There are three main approaches: (1) explorative analysis, (2) classification, and (3) calibration. The choice of method depends on the problem and experimental data (14). For example, principal component analysis (PCA) is commonly used initially for dimensionality reduction, highlighting the most representative features with minimal information loss and generating new variables called principal components (15). However, PCA does not consider group membership, so chemometric methods are used for classification and class modeling when focusing on product origin. Classification, synonymous with discriminant methods, assigns objects to predefined classes using techniques like linear discriminant analysis (LDA), k-nearest neighbors, partial least squares-discriminant analysis (PLS-DA), and artificial neural networks (ANN) (16).

Linear discriminant analysis, one of the simplest classifiers, requires a sufficient ratio (≥3) between samples and variables and struggles with highly collinear data common in chemistry (16). PLS-DA addresses these issues, creating a linear model statistically equivalent to LDA's solution but overcoming minimum sample-to-variable ratio and collinearity problems (16). Orthogonal partial least squares-discriminant analysis (OPLS-DA), a modification of PLS-DA, enhances interpretability by separating predictive variance from non-predictive variance (17). Model performance is evaluated by explained variation (R2X for PCA and R2Y for OPLS-DA) and predictive ability (Q2), with internal sevenfold cross-validation minimizing overfitting. OPLS-DA prediction performance is measured by sensitivity (true positives) and specificity (true negatives) (18). Discriminant markers are selected by Variable Importance in the Projection (VIP) values, with a threshold of one.

Class modeling, rather than discriminant analysis, is often recommended to confirm a sample's regional origin due to possible biases in one-class classification problems. Soft independent modeling of class analogy (SIMCA) is a standard method in chemometrics for such tasks (19, 20).

## 5.3 Cross-validation

A cross-validation approach was used to evaluate the robustness of the IsoFoodTrack database. This involved splitting the dataset into training and testing sets, where the training set was used to build a predictive model, and the testing set was used to evaluate the accuracy of the predictions. High predictive accuracy indicated the effectiveness of the database in identifying food fraud and verifying the geographical origin of samples.

## 5.4 Practical applicability

Finally, the practical application of the IsoFoodTrack database was demonstrated by analyzing a set of commercial food samples and comparing their isotopic and elemental profiles against the reference dataset. The results confirmed the ability of the database to detect discrepancies in geographical origin and production claims, thereby validating its utility as a tool for ensuring food authenticity. The Slovenian studies include the use of different chemometric approaches for verifying the geographical region of different commodities. For example, we investigated the possibility of determining the geographical origin of milk and dairy products. Using linear discriminant analysis, discrimination and specification of goat, cow and sheep milk and cheese was possible (21). Moreover, the existing database of authentic Slovenian cow milk was used to verify the correct assignment of regional provenance and declaration of origin. By applying discriminant analysis, the ability to discriminate between geographic regions was only possible when data were organized by year and season as a result of different feeding regimes. Based on the data, a discrimination model was developed to differentiate milk from European milk produced in Slovenia efficiently. Slovenian milk was statistically distinguishable from all other milk, where the most important parameters were $\delta^{18}$O, Sr., K and *Ca.* Commercial samples labeled as “Slovenian milk” were confirmed and classified as being authentic (22).

Despite the fact that the Slovenian truffles shared some similar characteristics with the samples originating from other countries, differences in the element concentrations suggest that respective truffle species may respond selectively to nutrients from a specific soil type under environmental and soil conditions. Cross-validation resulted in a 77% correct classification rate for determining the geographical origin and a 74% correct classification rate for discriminating between species. The critical parameters for geographical origin discriminations were Sr., Ba, V, Pb, Ni, Cr, Ba/Ca and Sr./Ca ratios, while from stable isotopes $\delta^{18}$O and $\delta^{13}$C values are the most important (23).

Discriminant and class-modeling methods have also been applied to assess the geographical classification and authentication of selected fruits and vegetables, including strawberries, cherries, apples, kaki, asparagus and garlic, using stable isotopes of light elements and elemental composition of samples harvested between 2018 and 2020. A good geographical classification of Slovenian and non-Slovenian strawberries was obtained despite different production years using discriminant approaches. Class models generated by data-driven soft independent modeling of class analogy (DD-SIMCA) had high sensitivity (96–97%) and good specificity (81–91%) on a yearly basis, while a more generalized model combining total yearly data gave a lower specificity (63%) (24). Of the 33 commercially available samples (test samples) with declared Slovenian origin, 39% were from outside of Slovenia. The specificity for garlic and asparagus was found to be higher compared to strawberries, indicating that the model for these two commodities is more robust for verifying the correct labeling.

These examples highlight the potential of isotopic and elemental analysis as reliable tools for food origin authentication while demonstrating that some commodities present more significant challenges compared to others.

In addition, the IsoFoodTrack database can be used to support advanced analytics based on statistical and explainable machine-learning approaches, enabling the development of discriminant models to differentiate selected food commodities based on species using elemental and stable isotope data. Machine learning (ML) is a branch of artificial intelligence (AI) that enables systems to learn and improve from experience without being explicitly programmed. The ML component accepts the food’s isotopic composition as input and predicts its geographical origin. Additionally, this approach offers explanations about which specific isotopes serve as indicators for that geographical origin, with the aim of increasing trust in AI-generated suggestions.

The metadata included in the database allow the user to enrich and complement a stable isotope reference database by a more novel approach, i.e., process-based modeling, such as isotopic mapping (isoscapes) (25). Isoscapes can be constructed to make predictive patterns and inform the likelihood of origin based on regional and localized characteristics. The basic concept of isoscapes is reflected in its name, derived from the words “isotope” and “landscape.” Isoscapes visualize the distribution of isotopic ratios (typically of light elements) in space, often using Geographic Information System (GIS) technology to incorporate these ratios into geographic maps. There are two primary methods for producing isoscapes: statistical and process modeling. In statistical modeling, various geostatistical approaches, such as inverse distance weighting and kriging, are used to model the expected isotopic composition of the material in question. These methods typically require extensive databases that densely cover the area of interest. Only a few national-scale isoscapes studies such as wine (26), milk (27), olive oil (28), and rice (29) have been published.

Conversely, process modeling involves obtaining variables with higher spatial density, such as meteorological or geological data, to model the isotopic composition based on the processes affecting isotopic fractionation. For example, food isoscapes are derived from the observation that food produced in a specific area often reflects the local climatic and geological characteristics. The advantage of process-based modeling over statistical modeling is that it requires a much smaller sample database and can be applied to areas with limited sampling. A good example of the spatial variability “GIS modeling Isoscapes” of oxygen and carbon stable isotope composition ($\delta^{13}$C, $\delta^{18}$O) of argan oil is also presented by Taous et al. (30). In order to make global, spatially continuous predictions for argan oil stable isotope ratios, the mechanistic models in ArcGIS software (ESRI Corporation ArcGIS 10.5) were implemented. The ordinary point kriging was used to spatially interpolate $\delta^{13}$C and $\delta^{18}$O values of argan oils from 25 individual samples collected at five independent regions.

These geospatial models – isoscapes may provide a cost-effective extension to the isotopic dataset approach.

## 6 Database curation and availability

The IsoFoodTrack database (see text footnote 2) curates isotope data from authentic food samples, primarily from Slovenia and other countries, as part of various projects. For instance, in the REALMED project, data includes Moroccan argan oil, Tunisian lamb, and truffles from Italy and Slovenia, while in FishEUTrust, data on sea bream from Malta, Portugal, and Spain is included.

Future upgrades will expand the database to cover additional commodities like honey, nuts, wheat, and cereals. In addition, we also intend to supplement our data with relevant, available open-access data from the literature. Open-access data will also supplement existing datasets such as FoodIntegrity, FNS-Cloud, and METROFOOD-RI, and norms for data-sharing, including embargo periods before public release, will be established. Additionally, it complies with the ISO-FOOD ontology (13), supporting semi-automated integration with data from other relevant sources, and is openly available through the NCBO BioPortal. The nomenclature of the elemental components complies with the CEN Standard of food data.

With new data and new methods for data analysis being integrated into IsoFoodTrack, both the tool and the ontology will enable interoperability with other platforms. The underlying concept of interlinking a dataset with semantic resources serves as an excellent example of open science e-infrastructures that will need to be developed in the future.

## 7 Conclusion

IsoFoodTrack represents one of the initial efforts to compile stable isotope and elemental data for food authenticity and fraud detection. Its robust, scalable, and flexible architecture makes it an invaluable resource for researchers, food control agencies, and global food authenticity initiatives. Here are the key achievements and contributions of IsoFoodTrack:

*Comprehensive data management*: The database integrates isotopic and elemental composition data with rich metadata, encompassing geographical, production, and methodological details.

*Flexibility and scalability*: IsoFoodTrack is designed for continuous expansion, allowing the inclusion of new food commodities, analytical methods, and metadata fields. This ensures the database remains adaptable to ongoing research and technological developments, making it future-proof.

*Interoperability and integration*: The use of standardized formats like JSON and CSV, along with the potential for API integration, enables IsoFoodTrack to exchange data seamlessly with other databases such as future IsoBank. This promotes collaboration and data sharing across global research initiatives.

*User-friendly access and data visualization*: IsoFoodTrack's web interface, developed using Django, provides an intuitive platform for users to access, query, and interact with the data. Features like interactive maps and tabular displays allow for easy visualization and comparison of data across different geographical locations and samples, enhancing user engagement and analytical capacity.

*Rigorous quality control*: The database implements a thorough validation process to ensure high-quality and reliable data. All samples undergo standardized treatment and analysis, with expert review ensuring that only accurate and validated data is included. This robust approach to quality control enhances the credibility of IsoFoodTrack in scientific research.

*Practical applicability*: Based on the data collected in IsoFoodTrack using statistical, chemometric and machine learning approaches the database has a capability to identify and classify the origin of food commodities. It also supports isoscapes, which visualize isotopic ratios across regions, providing spatially continuous predictions of geographical origins of food products. This novel approach adds depth to isotopic analysis and extends the database's predictive capabilities, particularly in regions with limited sampling.

*Future enhancements*: Future upgrades include expanding the database to cover more food commodities (e.g., honey, nuts, cereals) and integrating open-access data from literature. These enhancements ensure that IsoFoodTrack will continue to evolve as a critical resource for food authenticity research.

## Data availability statement

The raw data supporting the conclusions of this article are available from the corresponding author upon reasonable request.

## Author contributions

CT: Conceptualization, Data curation, Investigation, Writing – original draft. RM: Data curation, Methodology, Visualization, Writing – review & editing. MO: Data curation, Methodology, Visualization, Writing – review & editing. AS: Methodology, Visualization, Writing – review & editing. JD: Methodology, Writing – review & editing. TE: Methodology, Visualization, Writing – review & editing. BKS: Methodology, Supervision, Validation, Writing – review & editing. NO: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research work was financially supported by the Slovenian Research and Innovation Agency (ARIS) under programmes P1-0143 and P2-0098. Further, we acknowledge the financial assistance of the FoodTraNet Project funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodow-ska-Curie (Grant agreement no. 956265), Horizon Europe projects FishEUTrust (Grant agreement no. 101060712), METROFOOD-EPI (Grant agreement no. 101130162), and PROMEDLIFE project (PRIMA programme supported by the European Union, Grant agreement no. 2132).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2025.1516521/full#supplementary-material

## References

1. Danezis GP, Tsagkaris AS, Camin F, Brusic V, Georgiou CA. Food authentication: techniques, trends & emerging approaches. *TrAC Trends Anal Chem*. (2016) 85:123–32. doi: 10.1016/j.trac.2016.02.026

2. Kelly S, Heaton K, Hoogewerff J. Tracing the geographical origin of food: the application of multi-element and multi-isotope analysis. *Trends Food Sci Technol*. (2005) 16:555–67. doi: 10.1016/j.tifs.2005.08.008

3. Drivelos SA, Georgiou CA. Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC Trends Anal Chem*. (2012) 40:38–51. doi: 10.1016/j.trac.2012.08.003

4. Christoph N, Hermann A, Wachter H. 25 years authentication of wine with stable isotope analysis in the European Union–review and outlook. *BIO Web Conf*. (2015) 5:02020. doi: 10.1051/bioconf/20150502020

5. Camin F, Bontempo L, Perini M, Piasentier E. Stable isotope ratio analysis for assessing the authenticity of food of animal origin. *Compr Rev Food Sci Food Saf*. (2016) 15:868–77. doi: 10.1111/1541-4337.12219

6. Pauli JN, Newsome SD, Cook JA, Harrod C, Steffan SA, Baker CJ, et al. Why we need a centralized repository for isotopic data. *Proc Natl Acad Sci*. (2017) 114:2997–3001. doi: 10.1073/pnas.1701742114

7. Meier-Augenstein W, Hobson KA, Wassenaar LI. Critique: measuring hydrogen stable isotope abundance of proteins to infer origins of wildlife, food and people. *Bioanalysis*. (2013) 5:751–67. doi: 10.4155/bio.13.36

8. Camin F, Boner M, Bontempo L, Fauhl-Hassek C, Kelly SD, Riedl J, et al. Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends Food Sci Technol*. (2017) 61:176–87. doi: 10.1016/j.tifs.2016.12.007

9. Donarski J, Camin F, Fauhl-Hassek C, Posey R, Sudnik M. Sampling guidelines for building and curating food authenticity databases. *Trends Food Sci Technol*. (2019) 90:187–93. doi: 10.1016/j.tifs.2019.02.019

10. Skrzypek G, Allison CE, Böhlke JK, Bontempo L, Brewer P, Camin F, et al. Minimum requirements for publishing hydrogen, carbon, nitrogen, oxygen and sulfur stable-isotope delta results (IUPAC technical report). *Pure Appl Chem*. (2022) 94:1249–55. doi: 10.1515/pac-2021-1108

11. Dunn PJH, Hai L, Malinovsky D, Goenaga-Infante H. Simple spreadsheet templates for the determination of the measurement uncertainty of stable isotope ratio delta values. *Rapid Commun Mass Spectrom*. (2015) 29:2184–6. doi: 10.1002/rcm.7376

12. Potočnik D, Hudobivnik MJ, Mazej D, Ogrinc N. Optimization of the sample preparation method for determination of multi-elemental composition in fruit samples by ICP-MS analysis. *Measur Sensors*. (2021) 18:100292. doi: 10.1016/j.measen.2021.100292

13. Eftimov T, Ispirova G, Potočnik D, Ogrinc N, Koroušić Seljak B. ISO-FOOD ontology: a formal representation of the knowledge within the domain of isotopes for food science. *Food Chem*. (2019) 277:382–90. doi: 10.1016/j.foodchem.2018.10.118

14. Granato D, Putnik P, Kovačević DB, Santos JS, Calado V, Rocha RS, et al. Trends in Chemometrics: food authentication, microbiology, and effects of processing. *Compr Rev Food Sci Food Saf*. (2018) 17:663–77. doi: 10.1111/1541-4337.12341

15. Covaciu FD, Moldovan Z, Dehelean AA, Magdas DA, Feher IC, Puscas R, et al. Determination of pesticides, elements, and stable isotopes in strawberries. *Anal Lett*. (2016) 49:2560–72. doi: 10.1080/00032719.2016.1140175

16. Marini F. Classification methods in Chemometrics. *Curr Anal Chem*. (2009) 6:72–9. doi: 10.2174/157341110790069592

17. Rongai D, Sabatini N, Del Coco L, Perri E, Del Re P, Simone N, et al. $^{1}$H NMR and multivariate analysis for geographic characterization of commercial extra virgin olive oil: a possible correlation with climate data. *Food Secur*. (2017) 6:96. doi: 10.3390/foods6110096

18. Fiamegos Y, Dumitrascu C, Papoci S, de la Calle MB. Authentication of PDO paprika powder (Pimentón de la Vera) by multivariate analysis of the elemental fingerprint determined by ED-XRF. A feasibility study. *Food Control*. (2021) 120:107496. doi: 10.1016/j.foodcont.2020.107496

19. Oliveri P. Class-modelling in food analytical chemistry: development, sampling, optimization and validation issues–a tutorial. *Anal Chim Acta*. (2017) 982:9–19. doi: 10.1016/j.aca.2017.05.013

20. Rodionova OY, Titova AV, Pomerantsev AL. Discriminant analysis is an inappropriate method of authentication. *TrAC Trends Anal Chem*. (2016) 78:17–22. doi: 10.1016/j.trac.2016.01.010

21. Nečemer M, Potočnik D, Ogrinc N. Discrimination between Slovenian cow, goat and sheep milk and cheese according to geographical origin using a combination of elemental content and stable isotope data. *J Food Compos Anal*. (2016) 52:16–23. doi: 10.1016/j.jfca.2016.07.002

22. Potočnik D, Nečemer M, Perišić I, Jagodic M, Mazej D, Camin F, et al. Geographical verification of Slovenian milk using stable isotope ratio, multi-element and multivariate modelling approaches. *Food Chem*. (2020) 326:126958. doi: 10.1016/j.foodchem.2020.126958

23. Hamzić Gregorčič S, Strojnik L, Potočnik D, Vogel-Mikuš K, Jagodic M, Camin F, et al. Can we discover truffle's true identity? *Molecules*. (2020) 25:2217. doi: 10.3390/molecules25092217

24. Strojnik L, Potočnik D, Hudobivnik MJ, Mazej D, Japelj B, Škrk N, et al. Geographical identification of strawberries based on stable isotope ratio and multi-elemental analysis coupled with multivariate statistical analysis: a Slovenian case study. *Food Chem*. (2022) 381:132204. doi: 10.1016/j.foodchem.2022.132204

25. Bowen GJ. Isoscapes: spatial pattern in isotopic biogeochemistry. *Annu Rev Earth Planet Sci*. (2010) 38:161–87. doi: 10.1146/annurev-earth-040809-152429

26. West J, Ehlerinwqger J, Cerling T. Geography and vintage predicted by a novel GIS model of wine $\delta^{18}O$. *J Agric Food Chem*. (2007) 55:7075–83. doi: 10.1021/jf071211r

27. Ehtesham E, Baisden WT, Keller ED, Hayman AR, Van Hale R, Frew RD. Correlation between precipitation and geographical location of the δ2H values of the fatty acids in milk and bulk milk powder. *Geochim Cosmochim Acta*. (2013) 111:105–16. doi: 10.1016/j.gca.2012.10.026

28. Chiocchini F, Portarena S, Ciolfi M, Brugnoli E, Lauteri M. Isoscapes of carbon and oxygen stable isotope compositions in tracing authenticity and geographical origin of Italian extra-virgin olive oils. *Food Chem*. (2016) 202:291–301. doi: 10.1016/j.foodchem.2016.01.146

29. Sheng M, Zhang W, Nie J, Li C, Zhu A-X, Hu H, et al. Predicting isoscapes based on an environmental similarity model for the geographical origin of Chinese rice. *Food Chem*. (2022) 397:133744. doi: 10.1016/j.foodchem.2022.133744

30. Taous F, Amenzou N, Marah H, Maia R, Maguas C, Bahmad L, et al. Stable isotope ratio analysis as a new tool to trace the geographical origin of Argan oils in Morocco. *Forensic Chem*. (2020) 17:100198. doi: 10.1016/j.forc.2019.100198

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 18 March 2025
DOI 10.3389/fnut.2025.1527771

Check for updates

OPEN ACCESS

EDITED BY
Massimo Lucarini,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Lingqiong Meng,
University of Arizona, United States
Rositsa Chamova,
Medical University of Varna, Bulgaria

*CORRESPONDENCE
Chuan Xiang
✉ chuanxiang@sxmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 13 November 2024
ACCEPTED 05 March 2025
PUBLISHED 18 March 2025

CITATION
Wang X, Wang M, Guo Z and Xiang C (2025) Association between plain water intake and the risk of osteoporosis among middle-aged and elderly people in the United States: a cross-sectional study.
*Front. Nutr.* 12:1527771.
doi: 10.3389/fnut.2025.1527771



# Association between plain water intake and the risk of osteoporosis among middle-aged and elderly people in the United States: a cross-sectional study

Xudong Wang[1†], Meng Wang[2†], Zijian Guo[1] and Chuan Xiang[1*]

[1]Department of Orthopedics, The Second Hospital of Shanxi Medical University, Taiyuan, Shanxi, China, [2]Academy of Medical Sciences, Shanxi Medical University, Taiyuan, Shanxi, China

**Background:** The connection between plain water intake (PWI) and osteoporosis risk is still unclear. The investigation aimed to identify the relationship between PWI and osteoporosis risk in middle-aged and elderly individuals in the United States (US).

**Methods:** This cross-sectional study was conducted among participants aged 50 years and older in the following waves of the National Health and Nutrition Examination Survey (NHANES): 2007–2008, 2009–2010, 2013–2014, and 2017–2018. The relationship between PWI and osteoporosis risk was examined by multivariable logistic regression models, accompanied by subgroup analyses and interaction tests. Smooth curve fitting and threshold effect analysis were utilized.

**Results:** The present investigation included 6,686 participants. In accordance with the fully adjusted model, individuals in the highest PWI tertile had a significantly reduced risk of osteoporosis in contrast to those in the lowest tertile [odds ratio (OR) = 0.62; 95% confidence interval (CI): 0.49–0.77; *P* for trend<0.001]. After adjusting for all covariates, a higher PWI was linked to a decreased risk of osteoporosis (OR = 0.92; 95% CI: 0.86–0.98; $p = 0.008$). No significant interactions were detected in the subgroup analyses for age, gender, race, body mass index, diabetic history, hypertension status, smoking history, consumption of prednisone or cortisone, or moderate or strenuous activity (all *P* for interaction>0.05). Smooth curve fitting and threshold effect analysis revealed that when PWI was less than 1,220 mL/day, there was a significant negative connection between PWI and osteoporosis risk (OR = 0.79; 95% CI: 0.70–0.89; $p < 0.001$); nevertheless that association was not significant when PWI was greater than 1,220 mL/day (OR = 1.06; 95% CI: 0.95–1.17; $p = 0.288$).

**Conclusion:** The outcomes of our investigation indicated that among middle-aged and older US adults, a higher PWI was connected with a moderately reduced osteoporosis risk. Managing PWI might reduce the osteoporosis risk.

KEYWORDS

**plain water intake, osteoporosis, middle-aged and elderly people, cross-sectional study, National Health and Nutrition Examination Survey**

## Introduction

Osteoporosis is a systemic skeletal disorder defined by diminished bone mineral density (BMD) and microarchitectural degradation of bone tissue (1–3). With the advanced aging of the population, osteoporosis has emerged as the most prevalent bone metabolic disorder (4). Almost 14.1 million individuals aged 50 and older suffer from osteoporosis in the United States (US), and the incidence rate exhibits a steady increase (5–7). Osteoporosis can result in higher fragility of the bone and an elevated fracture risk, which impacts almost all skeletal sites due to the systemic nature of the disease (1, 3, 8, 9). Traditionally, hip and vertebral fractures have been regarded as prototypical osteoporotic fractures (1). However, a far greater incidence of osteoporotic fractures has been observed at all other sites (i.e., excluding the hip and vertebrae) (10). The consequences of osteoporotic fractures include serious complications, reduced quality of life, elevated disabilities, and raised death rates (11). Moreover, osteoporosis and its associated fractures impose an enormous financial burden on patients, their families, and society (12–14). Therefore, preventing osteoporosis is vital.

Diet has a vital role in modifying the risk of osteoporosis and contributing to its prevention (15). Water is an important nutrient in the diet and is connected with several physiological functions, such as metabolism, modulation of body temperature, transportation of nutrients, and elimination of waste products (16–18). There are various sources of water consumed in daily life, including tea, coffee, sugar-sweetened beverages, and plain water. The source of water consumed is important for bone health. Huang et al. discovered that consuming tea provided a protective effect against osteoporosis, especially in women and middle-aged adults (19). Xu et al. (20) reported that regular moderate consumption of coffee may provide protection against osteoporosis between older US adults and those in middle age. Notably, a systematic review and meta-analysis comprising 26 publications exhibited that the intake of beverages that were sweetened by sugar was negatively related to BMD in adults (21). Nevertheless, few investigations have focused on the connection between plain water intake (PWI) and osteoporosis risk.

Therefore, this cross-sectional study was performed to determine the link between PWI and osteoporosis risk among older US adults and those in middle age.

## Materials and methods

### Study population

The nutritional and health status of the US people were evaluated utilizing the National Health and Nutrition Examination Survey (NHANES), which is a large-scale cross-sectional study executed by the National Center for Health Statistics (NCHS). Information on diet, demographics, questionnaires, examinations, and laboratories has been published every 2 years. The Institutional Review Board of the NCHS authorized the entire program, and every individual signed an informed consent form.

All of the information in this investigation was retrieved from the following NHANES cycles: 2007–2008, 2009–2010, 2013–2014, and 2017–2018. Because it was only during these cycles that femur dual-energy X-ray absorptiometry (DXA) data and information about vitamin D intake and dietary supplements were recorded. The criteria for participant inclusion in our investigation were as follows: (1) complete PWI data; (2) availability of femoral BMD data; and (3) aged 50 years and older. The criteria of exclusion were established as follows: (1) missing PWI data; (2) missing femur DXA data; (3) age younger than 50 years; and (4) missing data on other covariates. Firstly, data for 40,115 participants were chosen from the following NHANES cycles: 2007–2008, 2009–2010, 2013–2014, and 2017–2018. Subsequently, 9,658 participants were excluded owing to absent PWI information. Participants with missing information on femur DXA ($N$ = 14,556) and those with age less than 50 years ($N$ = 7,886) or with lost information on other covariates ($N$ = 1,329) were also excluded. Ultimately, the present study comprised 6,686 participants (Figure 1).

### Measurement of PWI

PWI is known as the overall amount of water consumed over a 24-h timeframe, including bottled water, ordinary tap water, spring water, and water obtained from the consumption of fountains or water coolers. The 24-h PWI of each participant was collected via face-to-face interviews, and the information was subsequently collected via telephone interviews 3–10 days later. The present investigation utilized the mean of two recordings for statistical analysis to determine the long-term average PWI of the population in the US.

### Definition of osteoporosis

The BMD of the femoral region, as measured by DXA, was used to evaluate whether an individual was diagnosed with osteoporosis. Depending on the classification standards of the World Health Organization, osteoporosis was diagnosed when BMD measurements in any femoral region were greater than 2.5 standard deviations (SDs) below those of the reference group (young adults) (22). The current study examined the femoral BMD at the whole femur, neck of the femur, trochanter, and intertrochanter sites. The diagnostic thresholds were 0.68 g/cm$^2$ for the whole femur, 0.59 g/cm$^2$ for the femur neck, 0.49 g/cm$^2$ for the trochanter, and 0.78 g/cm$^2$ for the intertrochanter (23).

### Other covariates

Based on prior research and clinical experience, we collected data on covariates that may influence the connection between PWI and osteoporosis risk. The selected covariates were obtained from demographic, examination, questionnaire, laboratory and dietary data. The covariates extracted from demographic data included age, gender, race, level of education, marital status, and family poverty–income ratio (PIR). Body mass index (BMI) data were extracted from the examination information. The factors obtained from the questionnaire data consisted of diabetic history, hypertension status, thyroid disease, smoking history, consumption of prednisone or cortisone, milk product consumption, engagement in moderate or strenuous activity, and fracture. The covariates extracted from laboratory data comprised serum vitamin D, total calcium, alanine aminotransferase, asparate aminotransferase, creatinine, and uric acid. The covariates obtained from the dietary data included alcohol consumption, tea consumption,

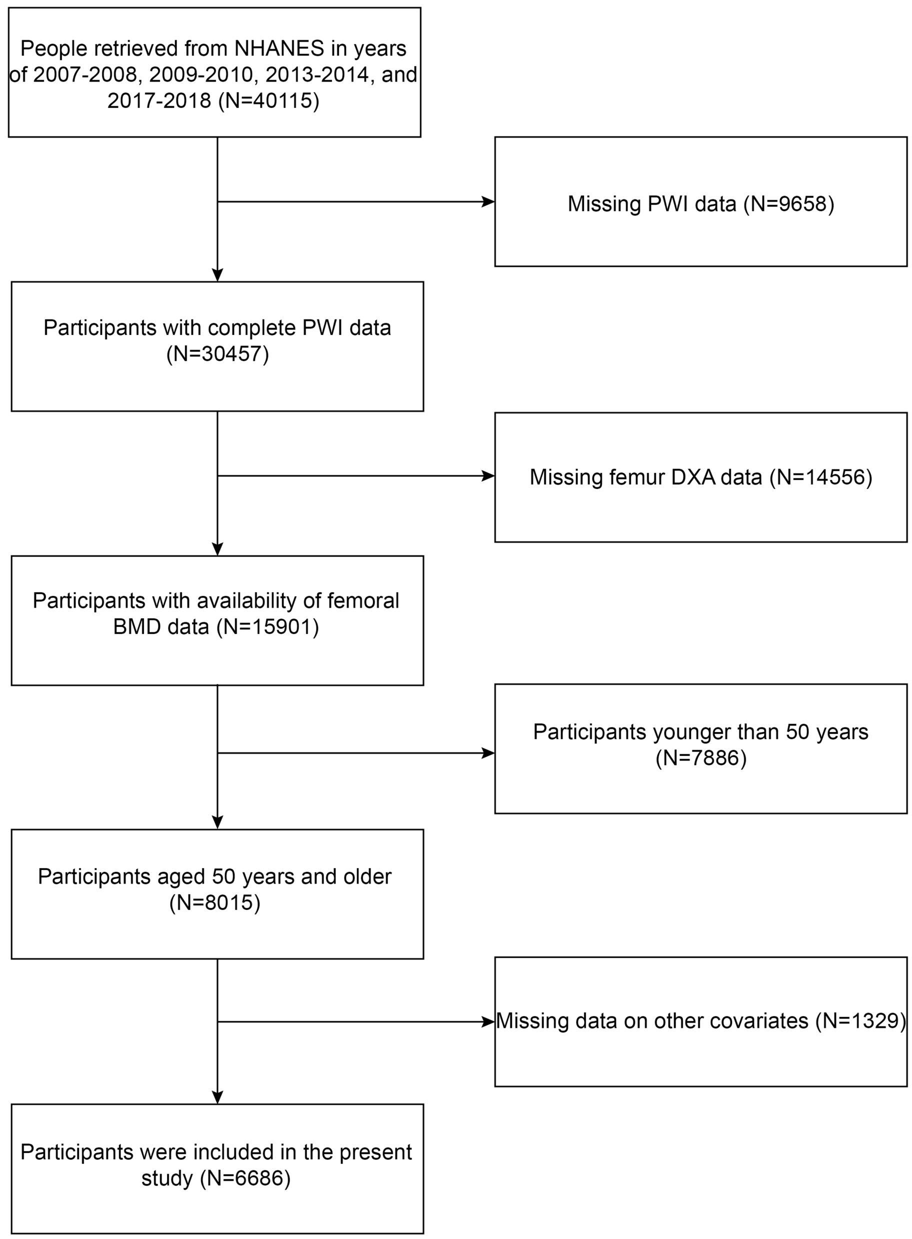



FIGURE 1
Flow chart of the inclusion and exclusion criteria.

vitamin D supplementation, calcium supplementation, vitamin D intake, calcium intake, caffeine intake, energy intake, protein intake, and other liquid intake. Smoking history was ascertained by inquiring if individuals had consumed a minimum of 100 cigarettes throughout their lives. Moderate or strenuous activity was characterized as a minimum of ten continuous minutes of sports, fitness, or recreational activities that resulted in a minor or significant elevation in heart rate or breathing over the preceding 30 days or during a typical week.

## Statistical analysis

The R Version 3.4.3 (The R Foundation, http://www.R-project.org) and Empower (X&Y Solutions, Inc., Boston, MA, United States) programs were utilized to perform the statistical analysis. We employed proportions to provide a summary of categorical data and means ± SDs to characterize continuous variables. We utilized a chi-square test for categorical variables and a Student's t-test for continuous variables in order to assess variations among patients classified by either the existence or absence of osteoporosis. The correlation between PWI and osteoporosis risk was examined by employing multivariable logistic regression models. Model 1 was unadjusted for covariates; Model 2 underwent adjustment for gender, age, and race; and Model 3 underwent adjustment for all covariates, encompassing age, gender, race, level of education, marital status, PIR, BMI, diabetic history, hypertension status, thyroid disease, smoking history, consumption of prednisone or cortisone, moderate or strenuous activity, fracture, milk product consumption, alcohol consumption, tea consumption, vitamin D supplementation, calcium supplementation, vitamin D intake, calcium intake, caffeine intake, serum vitamin D, total calcium, alanine aminotransferase, asparate aminotransferase, creatinine, uric acid, energy intake, protein intake, and other liquid intake. To strengthen the data analysis, we utilized each 500 mL/day PWI as a unit and classified PWI into three groups according to tertiles. The dependability of the regression analysis outcomes was improved utilizing a trend test. Furthermore, we conducted subgroup analyses and interaction tests for particular variables, including age, gender, race, BMI, diabetic history, hypertension status, smoking history, consumption of prednisone or cortisone, or moderate or strenuous activity, to explore heterogeneity across subgroups. Smooth curve fitting and threshold effect analysis were utilized to explore possible nonlinear connections between PWI and osteoporosis risk. A $p$-value below 0.05 was deemed statistically significant.

# Results

## Baseline features of the enrolled participants

Table 1 presents the baseline features of the recruited participants. There were 734 individuals in the osteoporosis group and 5,952 individuals in the nonosteoporosis group. Individuals without osteoporosis had a greater PWI (every 500 mL/day) than did those with osteoporosis (1.80 ± 1.63 vs. 1.49 ± 1.51, $p < 0.001$). Moreover, there were significant between-group variations in gender, race, age, level of education, PIR, marital status, BMI, thyroid disease, smoking history, consumption of prednisone or cortisone, moderate or strenuous activity, fracture, alcohol consumption, vitamin D supplementation, calcium supplementation, calcium intake, caffeine intake, serum vitamin D, alanine aminotransferase, creatinine, uric acid, energy intake, protein intake, and other liquid intake ($p < 0.05$). Nevertheless, no significant variation was detected in diabetic history, hypertension status, milk product consumption, tea consumption, vitamin D intake, total calcium, and asparate aminotransferase between both groups ($p > 0.05$).

## Associations between PWI and the risk of osteoporosis

The associations between PWI and osteoporosis risk are shown in Table 2. PWI was altered from a continuous variable into a categorical variable according to tertiles. According to Model 1 (not adjusted for covariates), participants in the greatest PWI tertile group had a 44% reduced risk of osteoporosis in contrast to those in the group of the lowest PWI tertile [odds ratio (OR) = 0.56; 95% confidence interval (CI): 0.47–0.68; $P$ for trend<0.001]. Similarly, participants in the group of the greatest PWI tertile had a significantly lower risk of osteoporosis in contrast to those in the group of the lowest PWI tertile, as shown by Model 2 (adjusted for the main covariates; OR = 0.54; 95% CI: 0.44–0.66; $P$ for trend<0.001) and Model 3 (adjusted for all covariates; OR = 0.62; 95% CI: 0.49–0.77; $P$ for trend<0.001).

## Subgroup analyses

It was observed that after adjusting for all covariates, a higher PWI was linked to a decreased risk of osteoporosis (OR = 0.92; 95% CI: 0.86–0.98; $p = 0.008$, Table 3). Subgroup analyses were performed to examine whether this relationship varied across the different characteristics of the participants. No significant interactions were identified in the subgroup analyses for age, gender, race, BMI, diabetic history, hypertension status, smoking history, consumption of prednisone or cortisone, or moderate or strenuous activity (all $P$ for interaction>0.05).

## Smooth curve fitting and threshold effect analysis

Smooth curve fitting demonstrated a nonlinear relationship between PWI and osteoporosis risk (Figure 2). The outcomes of the threshold effect analysis are displayed in Table 4. The inflection point, which was identified utilizing a two-piecewise linear regression model, was 2.44 (1,220 mL/day). When PWI was less than 1,220 mL/day, there was a significant negative connection between PWI and osteoporosis risk (OR = 0.79; 95% CI: 0.70–0.89; $p < 0.001$), indicating that osteoporosis risk decreased by 21% for every 500 mL/day increase in PWI. However, that association was not significant when PWI was greater than 1,220 mL/day (OR = 1.06; 95% CI: 0.95–1.17; $p = 0.288$), demonstrating that increasing PWI beyond 1,220 mL/day did not further significantly reduce the risk of osteoporosis.

# Discussion

In this cross-sectional study with 6,686 participants, the three distinct models indicated that individuals in the greatest PWI tertile had a significantly decreased risk of osteoporosis in contrast to those in the lowest tertile. After adjusting for all covariates, a higher PWI was linked to a decreased risk of osteoporosis. Subgroup analyses exhibited that this trend remained consistent across different population settings. Furthermore, smooth curve fitting and threshold effect analysis indicated that when PWI was less than 1,220 mL/day, a greater PWI was connected with a diminished osteoporosis risk, although an increase in PWI did not further significantly reduce osteoporosis risk when PWI was more than 1,220 mL/day. Notably, 4,953 participants (74.08% of the sample) reported a PWI of less than 1,220 mL/day. Many older US adults and those in middle age may ignore the importance of PWI. Therefore, our investigation is of high significance in the field of public health.

Dietary nutrients are essential for life and serve as the foundation for numerous metabolic processes. A diet rich in balanced nutrients is acknowledged as a preventive measure against osteoporosis, and the impact of nutrition on bone health has garnered growing interest (24). Many dietary nutrients, especially dietary micronutrients, including calcium, vitamin C, iron, potassium, magnesium, phosphorus, and vitamin D, may be strongly associated with osteoporosis (25–32). For example, Lee et al. (33) reported that bone mass might be improved by increasing calcium consumption and maintaining a high dietary calcium/phosphorus ratio. Liu et al. (27) reported that moderate rises in iron consumption were linked with a diminished osteoporosis risk

TABLE 1 Baseline features of the enrolled participants.

| Feature | Nonosteoporosis | Osteoporosis | *P*-value |
|---|---|---|---|
| No. of participants | 5,952 | 734 | |
| Age, *n* (%) | | | <0.001 |
| < 65 | 3,342 (56.15%) | 208 (28.34%) | |
| ≥ 65 | 2,610 (43.85%) | 526 (71.66%) | |
| Gender, *n* (%) | | | <0.001 |
| Male | 3,253 (54.65%) | 158 (21.53%) | |
| Post-menopausal female | 2,406 (40.42%) | 546 (74.39%) | |
| Non-menopausal female | 293 (4.92%) | 30 (4.09%) | |
| Race, *n* (%) | | | <0.001 |
| Hispanic | 1,327 (22.30%) | 118 (16.08%) | |
| Non-Hispanic White | 2,984 (50.13%) | 490 (66.76%) | |
| Non-Hispanic Black | 1,200 (20.16%) | 61 (8.31%) | |
| Other | 441 (7.41%) | 65 (8.86%) | |
| Level of education, *n* (%) | | | 0.003 |
| Less than high school | 1,433 (24.08%) | 192 (26.16%) | |
| High school | 1,405 (23.61%) | 205 (27.93%) | |
| More than high school | 3,114 (52.32%) | 337 (45.91%) | |
| Marital status, *n* (%) | | | <0.001 |
| Married/Living with partner | 3,817 (64.13%) | 345 (47.00%) | |
| Widowed/Divorced/Separated | 1750 (29.40%) | 349 (47.55%) | |
| Never married | 385 (6.47%) | 40 (5.45%) | |
| PIR, *n* (%) | | | <0.001 |
| ≤ 1 | 909 (15.27%) | 128 (17.44%) | |
| 1–3 | 2,467 (41.45%) | 370 (50.41%) | |
| > 3 | 2,576 (43.28%) | 236 (32.15%) | |
| BMI, n (%) | | | <0.001 |
| < 30 | 3,594 (60.38%) | 606 (82.56%) | |
| ≥ 30 | 2,358 (39.62%) | 128 (17.44%) | |
| Diabetic history, *n* (%) | | | 0.058 |
| Yes | 1,151 (19.34%) | 125 (17.03%) | |
| No | 4,572 (76.81%) | 590 (80.38%) | |
| Borderline | 229 (3.85%) | 19 (2.59%) | |
| Hypertension status, *n* (%) | | | 0.645 |
| Yes | 3,182 (53.46%) | 399 (54.36%) | |
| No | 2,770 (46.54%) | 335 (45.64%) | |
| Thyroid disease, *n* (%) | | | <0.001 |
| Yes | 844 (14.18%) | 166 (22.62%) | |
| No | 5,108 (85.82%) | 568 (77.38%) | |
| Smoking history, *n* (%) | | | 0.041 |
| Yes | 3,035 (50.99%) | 345 (47.00%) | |
| No | 2,917 (49.01%) | 389 (53.00%) | |
| Consumption of prednisone or cortisone, *n* (%) | | | 0.009 |
| Yes | 355 (5.96%) | 62 (8.45%) | |

*(Continued)*

TABLE 1 (Continued)

| Feature | Nonosteoporosis | Osteoporosis | *P*-value |
|---|---|---|---|
| No | 5,597 (94.04%) | 672 (91.55%) | |
| Moderate or strenuous activity, *n* (%) | | | <0.001 |
| Yes | 2,662 (44.72%) | 250 (34.06%) | |
| No | 3,290 (55.28%) | 484 (65.94%) | |
| Fracture, *n* (%) | | | <0.001 |
| Yes | 698 (11.73%) | 167 (22.75%) | |
| No | 5,254 (88.27%) | 567 (77.25%) | |
| Milk product consumption, *n* (%) | | | 0.250 |
| Yes | 4,856 (81.59%) | 586 (79.84%) | |
| No | 1,096 (18.41%) | 148 (20.16%) | |
| Alcohol consumption, *n* (%) | | | <0.001 |
| Yes | 1747 (29.35%) | 159 (21.66%) | |
| No | 4,205 (70.65%) | 575 (78.34%) | |
| Tea consumption, *n* (%) | | | 0.256 |
| Yes | 2,310 (38.81%) | 269 (36.65%) | |
| No | 3,642 (61.19%) | 465 (63.35%) | |
| Vitamin D supplementation, *n* (%) | | | 0.003 |
| Yes | 2,736 (45.97%) | 380 (51.77%) | |
| No | 3,216 (54.03%) | 354 (48.23%) | |
| Calcium supplementation, *n* (%) | | | <0.001 |
| Yes | 2,830 (47.55%) | 397 (54.09%) | |
| No | 3,122 (52.45%) | 337 (45.91%) | |
| Vitamin D intake (μg/day, mean ± SDs) | 4.68 ± 4.40 | 4.61 ± 4.45 | 0.748 |
| Calcium intake (mg/day, mean ± SDs) | 868.49 ± 451.09 | 812.56 ± 432.28 | <0.001 |
| Caffeine intake, (mg/day, mean ± SDs) | 164.08 ± 182.88 | 138.75 ± 152.92 | <0.001 |
| Serum vitamin D (nmol/L, mean ± SDs) | 70.79 ± 28.81 | 76.17 ± 33.67 | <0.001 |
| Total calcium (mg/dL, mean ± SDs) | 9.42 ± 0.38 | 9.43 ± 0.41 | 0.341 |
| Alanine aminotransferase (U/L, mean ± SDs) | 24.25 ± 18.95 | 20.19 ± 11.35 | <0.001 |
| Asparate aminotransferase (U/L, mean ± SDs) | 25.45 ± 13.83 | 25.15 ± 14.48 | 0.576 |
| Creatinine (mg/dL, mean ± SDs) | 0.97 ± 0.46 | 0.96 ± 0.57 | <0.001 |
| Uric acid (mg/dL, mean ± SDs) | 5.69 ± 1.41 | 5.25 ± 1.49 | <0.001 |
| Energy intake (kcal/day, mean ± SDs) | 1922.26 ± 749.85 | 1690.95 ± 680.17 | <0.001 |
| Protein intake (g/day, mean ± SDs) | 76.47 ± 32.33 | 65.29 ± 28.03 | <0.001 |
| PWI (every 500 mL/day, mean ± SDs) | 1.80 ± 1.63 | 1.49 ± 1.51 | <0.001 |
| Other liquid intake (ml/day, mean ± SDs) | 1746.25 ± 802.85 | 1533.24 ± 668.69 | <0.001 |

PWI, plain water intake; PIR, family poverty–income ratio; BMI, body mass index; SDs, standard deviations.

TABLE 2 Associations between PWI and osteoporosis risk.

| | Model 1 | Model 2 | Model3 |
|---|---|---|---|
| | (OR, 95% CI, *P*-value) | (OR, 95% CI, *P*-value) | (OR, 95% CI, *P*-value) |
| PWI (categorical) | | | |
| Tertile 1 (≤ 414.50 mL/day) | Reference | Reference | Reference |
| Tertile 2 (414.50–1024.06 mL/day) | 0.73 (0.61, 0.87) <0.001 | 0.66 (0.54, 0.80) <0.001 | 0.70 (0.57, 0.86) <0.001 |
| Tertile 3 (> 1024.06 mL/day) | 0.56 (0.47, 0.68) <0.001 | 0.54 (0.44, 0.66) <0.001 | 0.62 (0.49, 0.77) <0.001 |
| *P* for trend | <0.001 | <0.001 | <0.001 |

OR, odds ratio; CI, confidence interval; PWI, plain water intake.

in females. In addition, protein is necessary to maintain bone health and lower daily protein consumption may be connected with a greater risk of osteoporosis (34–36). A cross-sectional study of 4,707 participants revealed that elderly people and those in middle age in the US have an increased risk of osteoporosis when their daily dietary protein consumption is reduced (37). Importantly, fatty acids are one of the components of fat, and consuming fatty acids could be good for bone health (38–40). Fang et al. (38) reported that saturated, monounsaturated, and polyunsaturated fatty acids intake was positively connected with overall BMD among people aged 20 to 59 years. Notably, a cross-sectional study of 4,447 participants revealed that increased carbohydrate consumption was linked to decreased BMD (41). Recently, dietary fiber has been found to have potential benefits for bone health (42–44). Zhang et al. (44) conducted a cross-sectional study with 2,829 individuals and revealed that postmenopausal females with a dietary ratio of carbohydrate/fiber greater than 17.09 have a greater osteoporosis risk, while increased dietary fiber consumption is connected with a diminished risk of osteoporosis. Water is the richest nutrient in the diet, and plain water is the most affordable and accessible source of water consumed in daily life. Nevertheless, the link between PWI and osteoporosis risk has rarely been investigated.

Adequate PWI is crucial for proper body function (45, 46). Prior investigations have discovered the connection between PWI and a variety of diseases or metabolic disorders (47–53). For instance, Li et al. (47) conducted a cross-sectional study of 5,882 individuals and reported that PWI was inversely linked to the risk of periodontitis in those in middle age and older adults in the US. Another cross-sectional study of 16,434 individuals demonstrated that a greater PWI was independently linked with less afresh diagnosed nonalcoholic fatty liver disease in men but not in females (48). Furthermore, Pan et al. (49) performed a 5-year cohort study of 3,200 participants and discovered that a PWI over 4 cups a day was connected with a decreased risk of developing new-onset overweight for people with normal body weight. Most importantly, Lee et al. (50) drew conclusions from a cross-sectional study of 112,250 participants that there was a significant connection between lower PWI and increased risk of self-reported depression or suicidality. This research revealed that, among older US adults and those in middle age, a greater PWI was connected with a moderately diminished risk of osteoporosis. Managing PWI may decrease the osteoporosis risk.

To explain the connection between PWI and osteoporosis risk, we propose several potential mechanisms. Initially, a greater PWI was related to healthier dietary patterns described by greater intake of vegetables, fruits, and dairy products with low and reduced fat (54, 55). Thus, plain water is considered a possible dietary component that could improve dietary micronutrient profiles (56). Dietary micronutrients, including iron, phosphorus, magnesium, calcium, vitamin C, potassium, and vitamin D, may be closely related to osteoporosis. Therefore, a greater PWI may protect bone health through healthier dietary patterns associated with moderately increased intake of certain dietary micronutrients. Second, people with greater PWI were more likely to reduce their sugar-sweetened beverages intake (57). A high intake of sugar-sweetened beverages may reduce BMD (58). As a result, PWI may enhance bone health by decreasing sugar-sweetened beverages consumption. In addition, there were variations in the gut microbiota between people who drank more water and those who consumed less water (59). The gut microbiota can participate in preserving bone balance and protecting against osteoporosis development (60). Consequently, greater PWI may help individuals maintain their bone health through changes in the gut microbiota. Finally, increased daily PWI was shown to decrease the blood urea nitrogen concentration and inhibit the decrease in the estimated glomerular filtration rate (61). The osteoporosis risk was elevated in individuals with a decreased

**TABLE 3** Subgroup analyses between PWI (every 500 mL/day) and osteoporosis risk.

| Feature | Osteoporosis | | *P* for interaction |
|---|---|---|---|
| | *N* | Odds ratio (95% confidence interval), *P*-value | |
| Total | 6,686 | 0.92 (0.86, 0.98) 0.008 | |
| Age | | | 0.531 |
| < 65 | 3,550 | 0.94 (0.85, 1.03) 0.179 | |
| ≥ 65 | 3,136 | 0.90 (0.83, 0.98) 0.013 | |
| Gender | | | 0.773 |
| Male | 3,411 | 0.90 (0.79, 1.02) 0.111 | |
| Post-menopausal female | 2,952 | 0.93 (0.86, 1.00) 0.041 | |
| Non-menopausal female | 323 | 0.82 (0.56, 1.19) 0.298 | |
| Race | | | 0.349 |
| Hispanic | 1,445 | 0.95 (0.82, 1.10) 0.469 | |
| Non-Hispanic White | 3,474 | 0.90 (0.83, 0.98) 0.014 | |
| Non-Hispanic Black | 1,261 | 0.75 (0.57, 0.98) 0.035 | |
| Other | 506 | 0.99 (0.82, 1.20) 0.925 | |
| Body mass index | | | 0.170 |
| < 30 | 4,200 | 0.94 (0.87, 1.01) 0.080 | |
| ≥ 30 | 2,486 | 0.85 (0.74, 0.97) 0.016 | |
| Diabetic history | | | 0.911 |
| Yes | 1,276 | 0.91 (0.79, 1.05) 0.212 | |
| No | 5,162 | 0.92 (0.86, 0.99) 0.019 | |
| Borderline | 248 | 0.83 (0.52, 1.32) 0.430 | |
| Hypertension status | | | 0.171 |
| Yes | 3,581 | 0.95 (0.88, 1.04) 0.261 | |
| No | 3,105 | 0.88 (0.80, 0.96) 0.006 | |
| Smoking history | | | 0.052 |
| Yes | 3,380 | 0.97 (0.89, 1.05) 0.444 | |
| No | 3,306 | 0.86 (0.78, 0.94) 0.001 | |
| Consumption of prednisone or cortisone | | | 0.983 |
| Yes | 417 | 0.91 (0.73, 1.15) 0.434 | |
| No | 6,269 | 0.91 (0.85, 0.97) 0.006 | |
| Moderate or strenuous activity | | | 0.313 |
| Yes | 2,912 | 0.95 (0.86, 1.05) 0.361 | |
| No | 3,774 | 0.89 (0.83, 0.97) 0.007 | |

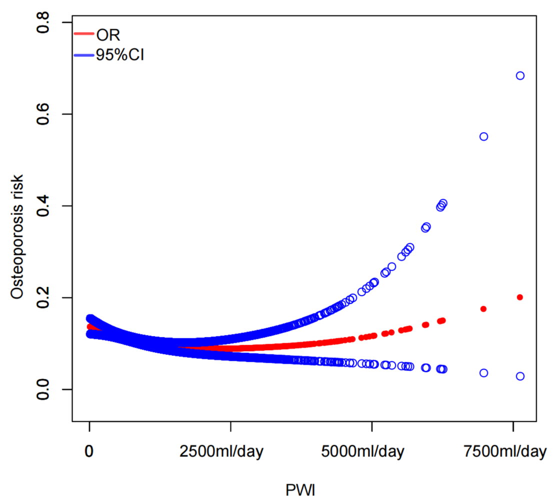



FIGURE 2
Smooth curve fitting illustrated a nonlinear association between PWI and the risk of osteoporosis.

**TABLE 4 Threshold effect analysis of PWI (every 500 mL/day) on osteoporosis utilizing a two-piecewise linear regression model.**

| Outcome | Osteoporosis (OR, 95% CI, *P*-value) |
|---|---|
| Fitting by standard linear model | 0.92 (0.86, 0.98) 0.008 |
| Fitting by two-piecewise linear model | |
| Inflection point | 2.44 (1,220 mL/day) |
| < 2.44 (*N* = 4,953) | 0.79 (0.70, 0.89) <0.001 |
| > 2.44 (*N* = 1733) | 1.06 (0.95, 1.17) 0.288 |
| Logarithmic likelihood ratio test *P*-value | 0.002 |

OR, odds ratio; CI, confidence interval.

estimated glomerular filtration rate (62). Therefore, a greater PWI may help preserve bone health by inhibiting a reduction in the estimated glomerular filtration rate. Notably, these mechanisms are speculative, and we intend to conduct more research in the future to verify the underlying mechanism(s).

In our study, the ORs between PWI and risk of osteoporosis for all subgroups were < 1. Notably, the association between PWI and osteoporosis risk was significant ($p < 0.05$) in certain subgroups such as post-menopausal female, Non-Hispanic White, or Non-Hispanic Black, whereas it was not significant ($p > 0.05$) in male, non-menopausal female, or any other races. However, larger *p*-values should not be interpreted as indicating no association or no effect: absence of evidence is not evidence of absence (63, 64). In addition, Hodzic-Santor et al. (65) reported that studies with smaller samples are more likely to have larger *p*-values, and studies with larger samples are more likely to have smaller *p*-values. The small number of participants in certain subgroups is a possible reason why the association was not significant.

Our investigation has multiple strengths. First, this study is the first investigation of the correlation between PWI and osteoporosis risk in elderly individuals and those who are middle age in the US. Second, we utilized nationally representative data, which greatly increased the sample size. Third, to ensure the reliability of our outcomes, we adjusted for confounders as much as possible. Finally, we enhanced the robustness of the data analysis by treating each 500 mL/day PWI as a unit and dividing participants into three PWI tertile groups. However, several limitations exist in our study. First, participants could be from different parts of the US, where the chemical composition of the soil, and therefore the water varies. Second, due to data source restrictions, we failed to additionally validate the outcomes using additional NHANES cycles. Third, the 24-h PWI of each participant was determined based on interviews, which may have led to recall bias. Finally, owing to the cross-sectional nature of this study, a causative association between PWI and osteoporosis risk could not be determined. Additional prospective and experimental investigation is necessary in the future to validate the causal link between PWI and osteoporosis risk and to elucidate the underlying processes.

## Conclusion

The findings of our study suggested that among middle-aged and elderly people in the US, a greater PWI was connected with a moderately lower osteoporosis risk. Managing PWI might diminish osteoporosis risk.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: National Health and Nutrition Examination Surveys database (https://www.cdc.gov/nchs/nhanes/).

## Ethics statement

The studies involving humans were approved by Research Ethics Reviewer Board of the National Center for Health Statistics. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

XW: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. MW: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. ZG: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. CX: Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the National Natural Science Foundation of China (No. 81972075) and the Central Guided Local Science and Technology Development Funds (No. YDZJSX20231A062).

## Acknowledgments

We thank the participants of the NHANES databases.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Compston JE, McClung MR, Leslie WD. Osteoporosis. *Lancet*. (2019) 393:364–76. doi: 10.1016/S0140-6736(18)32112-3

2. Lupsa BC, Insogna K. Bone health and osteoporosis. *Endocrinol Metab Clin N Am*. (2015) 44:517–30. doi: 10.1016/j.ecl.2015.05.002

3. Sun A, Hu J, Wang S, Yin F, Liu Z. Association of the visceral adiposity index with femur bone mineral density and osteoporosis among the U.S. older adults from NHANES 2005-2020: a cross-sectional study. *Front Endocrinol (Lausanne)*. (2023) 14:1231527. doi: 10.3389/fendo.2023.1231527

4. Wang G, Fang ZB, Liu DL, Chu SF, Li HL, Zhao HX. Association between caffeine intake and lumbar spine bone mineral density in adults aged 20-49: a cross-sectional study. *Front Endocrinol (Lausanne)*. (2022) 13:1008275. doi: 10.3389/fendo.2022.1008275

5. Sarafrazi N, Wambogo EA, Shepherd JA. Osteoporosis or low bone mass in older adults: United States, 2017-2018. *NCHS Data Brief*. (2021):1–8. doi: 10.15620/cdc:103477

6. Wright NC, Looker AC, Saag KG, Curtis JR, Delzell ES, Randall S, et al. The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine. *J Bone Miner Res*. (2014) 29:2520–6. doi: 10.1002/jbmr.2269

7. Chen K, Wang T, Tong X, Song Y, Hong J, Sun Y, et al. Osteoporosis is associated with depression among older adults: a nationwide population-based study in the USA from 2005 to 2020. *Public Health*. (2024) 226:27–31. doi: 10.1016/j.puhe.2023.10.022

8. Hak DJ. The biology of fracture healing in osteoporosis and in the presence of anti-osteoporotic drugs. *Injury*. (2018) 49:1461–5. doi: 10.1016/j.injury.2018.04.016

9. Stone KL, Seeley DG, Lui LY, Cauley JA, Ensrud K, Browner WS, et al. BMD at multiple sites and risk of fracture of multiple types: long-term results from the study of osteoporotic fractures. *J Bone Miner Res*. (2003) 18:1947–54. doi: 10.1359/jbmr.2003.18.11.1947

10. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res*. (2007) 22:465–75. doi: 10.1359/jbmr.061113

11. Cauley JA. Public health impact of osteoporosis. *J Gerontol A Biol Sci Med Sci*. (2013) 68:1243–51. doi: 10.1093/gerona/glt093

12. Qu B, Ma Y, Yan M, Wu HH, Fan L, Liao DF, et al. The economic burden of fracture patients with osteoporosis in western China. *Osteoporos Int*. (2014) 25:1853–60. doi: 10.1007/s00198-014-2699-0

13. Dempster DW. Osteoporosis and the burden of osteoporosis-related fractures. *Am J Manag Care*. (2011) 17:S164–9.

14. Rashki Kemmak A, Rezapour A, Jahangiri R, Nikjoo S, Farabi H, Soleimanpour S. Economic burden of osteoporosis in the world: a systematic review. *Med J Islam Repub Iran*. (2020) 34:154. doi: 10.34171/mjiri.34.154

15. Albertazzi P, Coupland K. Polyunsaturated fatty acids. Is there a role in postmenopausal osteoporosis prevention? *Maturitas*. (2002) 42:13–22. doi: 10.1016/S0378-5122(02)00022-1

16. Jéquier E, Constant F. Water as an essential nutrient: the physiological basis of hydration. *Eur J Clin Nutr*. (2010) 64:115–23. doi: 10.1038/ejcn.2009.111

17. Sawka MN, Cheuvront SN, Carter R 3rd. Human water needs. *Nutr Rev*. (2005) 63:S30–9. doi: 10.1111/j.1753-4887.2005.tb00152.x

18. Popkin BM, D'Anci KE, Rosenberg IH. Water, hydration, and health. *Nutr Rev*. (2010) 68:439–58. doi: 10.1111/j.1753-4887.2010.00304.x

19. Huang YP, Chen LS, Feng SH, Liang YS, Pan SL. Tea consumption and the risks of osteoporosis and hip fracture: a population-based longitudinal follow-up study. *Osteoporos Int*. (2023) 34:101–9. doi: 10.1007/s00198-022-06569-7

20. Xu J, Zhai T. Coffee drinking and the odds of osteopenia and osteoporosis in middle-aged and older Americans: a cross-sectional study in NHANES 2005-2014. *Calcif Tissue Int*. (2024) 114:348–59. doi: 10.1007/s00223-024-01184-6

21. Ahn H, Park YK. Sugar-sweetened beverage consumption and bone health: a systematic review and meta-analysis. *Nutr J*. (2021) 20:41. doi: 10.1186/s12937-021-00698-1

22. Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. WHO study group. *Osteoporos Int*. (1994) 4:368–81. doi: 10.1007/BF01622200

23. Looker AC, Orwoll ES, Johnston CC Jr, Lindsay RL, Wahner HW, Dunn WL, et al. Prevalence of low femoral bone density in older U.S. adults from NHANES III. *J Bone Miner Res*. (1997) 12:1761–8. doi: 10.1359/jbmr.1997.12.11.1761

24. Rizzoli R, Biver E, Brennan-Speranza TC. Nutritional intake and bone health. *Lancet Diabetes Endocrinol*. (2021) 9:606–21. doi: 10.1016/S2213-8587(21)00119-4

25. Al-Daghri NM, Hussain SD, Alnaami AM, Aljohani N, Sabico S. Dietary calcium intake and osteoporosis risk in Arab adults. *Nutrients*. (2023) 15:15. doi: 10.3390/nu15132829

26. Zhang YW, Song PR, Wang SC, Liu H, Shi ZM, Su JC. Diets intervene osteoporosis via gut-bone axis. *Gut Microbes*. (2024) 16:2295432. doi: 10.1080/19490976.2023.2295432

27. Liu X, An J. Dietary iron intake and its impact on osteopenia/osteoporosis. *BMC Endocr Disord*. (2023) 23:154. doi: 10.1186/s12902-023-01389-0

28. Singh W, Kushwaha P. Potassium: a frontier in osteoporosis. *Horm Metab Res*. (2024) 56:329–40. doi: 10.1055/a-2254-8533

29. Ha J, Kim SA, Lim K, Shin S. The association of potassium intake with bone mineral density and the prevalence of osteoporosis among older Korean adults. *Nutr Res Pract*. (2020) 14:55–61. doi: 10.4162/nrp.2020.14.1.55

30. He B, Xia L, Zhao J, Yin L, Zhang M, Quan Z, et al. Causal effect of serum magnesium on osteoporosis and Cardiometabolic diseases. *Front Nutr*. (2021) 8:738000. doi: 10.3389/fnut.2021.738000

31. Brzezińska O, Łukasik Z, Makowska J, Walczak K. Role of vitamin C in osteoporosis development and treatment-a literature review. *Nutrients*. (2020) 12:12. doi: 10.3390/nu12082394

32. Polzonetti V, Pucciarelli S, Vincenzetti S, Polidori P. Dietary intake of vitamin D from dairy products reduces the risk of osteoporosis. *Nutrients*. (2020) 12:1743. doi: 10.3390/nu12061743

33. Lee KJ, Kim KS, Kim HN, Seo JA, Song SW. Association between dietary calcium and phosphorus intakes, dietary calcium/phosphorus ratio and bone mass in the Korean population. *Nutr J*. (2014) 13:114. doi: 10.1186/1475-2891-13-114

34. Kędzia G, Woźniak M, Samborski W, Grygiel-Górniak B. Impact of dietary protein on osteoporosis development. *Nutrients*. (2023) 15:15. doi: 10.3390/nu15214581

35. Shams-White MM, Chung M, Du M, Fu Z, Insogna KL, Karlsen MC, et al. Dietary protein and bone health: a systematic review and meta-analysis from the National Osteoporosis Foundation. *Am J Clin Nutr*. (2017) 105:1528–43. doi: 10.3945/ajcn.116.145110

36. Durosier-Izart C, Biver E, Merminod F, van Rietbergen B, Chevalley T, Herrmann FR, et al. Peripheral skeleton bone strength is positively correlated with total and dairy protein intakes in healthy postmenopausal women. *Am J Clin Nutr*. (2017) 105:513–25. doi: 10.3945/ajcn.116.134676

37. Zhang YW, Cao MM, Li YJ, Dai GC, Lu PP, Zhang M, et al. Dietary protein intake in relation to the risk of osteoporosis in middle-aged and older individuals: a cross-sectional study. *J Nutr Health Aging*. (2022) 26:252–8. doi: 10.1007/s12603-022-1748-1

38. Fang ZB, Wang GX, Cai GZ, Zhang PX, Liu DL, Chu SF, et al. Association between fatty acids intake and bone mineral density in adults aged 20-59: NHANES 2011-2018. *Front Nutr*. (2023) 10:1033195. doi: 10.3389/fnut.2023.1033195

39. Järvinen R, Tuppurainen M, Erkkilä AT, Penttinen P, Kärkkäinen M, Salovaara K, et al. Associations of dietary polyunsaturated fatty acids with bone mineral density in elderly women. *Eur J Clin Nutr*. (2012) 66:496–503. doi: 10.1038/ejcn.2011.188

40. Nawata K, Yamauchi M, Takaoka S, Yamaguchi T, Sugimoto T. Association of n-3 polyunsaturated fatty acid intake with bone mineral density in postmenopausal women. *Calcif Tissue Int*. (2013) 93:147–54. doi: 10.1007/s00223-013-9743-5

41. Gao S, Qian X, Huang S, Deng W, Li Z, Hu Y. Association between macronutrients intake distribution and bone mineral density. *Clin Nutr*. (2022) 41:1689–96. doi: 10.1016/j.clnu.2022.05.019

42. Weaver CM. Diet, gut microbiome, and bone health. *Curr Osteoporos Rep*. (2015) 13:125–30. doi: 10.1007/s11914-015-0257-0

43. Scholz-Ahrens KE, Schrezenmeir J. Inulin and oligofructose and mineral metabolism: the evidence from animal trials. *J Nutr*. (2007) 137:2513s–23s. doi: 10.1093/jn/137.11.2513S

44. Zhang L, Zhao L, Xiao X, Zhang X, He L, Zhang Q. Association of dietary carbohydrate and fiber ratio with postmenopausal bone mineral density and prevalence of osteoporosis: a cross-sectional study. *PLoS One*. (2024) 19:e0297332. doi: 10.1371/journal.pone.0297332

45. Li S, Xiao X, Zhang X. Association between plain water intake and risk of hypertension: longitudinal analyses from the China health and nutrition survey. *Front Public Health*. (2023) 11:1280653. doi: 10.3389/fpubh.2023.1280653

46. Johnson EC. Adams WM: water intake, body water regulation and health. *Nutrients*. (2020) 12:12. doi: 10.3390/nu12030702

47. Li X, Wang L, Yang L, Liu X, Liu H, Mu Y. The association between plain water intake and periodontitis in the population aged over 45: a cross-sectional study based on NHANES 2009-2014. *BMC Oral Health*. (2024) 24:27. doi: 10.1186/s12903-023-03809-y

48. Wang X, Lin S, Gan S, Gu Y, Yang Y, Zhang Q, et al. Higher plain water intake is related to lower newly diagnosed nonalcoholic fatty liver disease risk: a population-based study. *Eur J Clin Nutr*. (2021) 75:1801–8. doi: 10.1038/s41430-021-00891-9

49. Pan XB, Wang HJ, Zhang B, Liu YL, Qi SF, Tian QB. Plain water intake and association with the risk of overweight in the Chinese adult population: China health and nutrition survey 2006-2011. *J Epidemiol*. (2020) 30:128–35. doi: 10.2188/jea.JE20180223

50. Lee JW, Kim Y. Association of plain water intake with self-reported depression and suicidality among Korean adolescents. *Epidemiol Health*. (2024) 46:e2024019. doi: 10.4178/epih.e2024019

51. Haghighatdoost F, Feizi A, Esmaillzadeh A, Rashidi-Pourfard N, Keshteli AH, Roohafza H, et al. Drinking plain water is associated with decreased risk of depression and anxiety in adults: results from a large cross-sectional study. *World J Psychiatry*. (2018) 8:88–96. doi: 10.5498/wjp.v8.i3.88

52. Carroll HA, Davis MG, Papadaki A. Higher plain water intake is associated with lower type 2 diabetes risk: a cross-sectional study in humans. *Nutr Res*. (2015) 35:865–72. doi: 10.1016/j.nutres.2015.06.015

53. Wang JS, Chiang HY, Chen HL, Flores M, Navas-Acien A, Kuo CC. Association of water intake and hydration status with risk of kidney stone formation based on NHANES 2009-2012 cycles. *Public Health Nutr*. (2022) 25:2403–14. doi: 10.1017/S1368980022001033

54. Popkin BM, Barclay DV, Nielsen SJ. Water and food consumption patterns of U.S. adults from 1999 to 2001. *Obes Res*. (2005) 13:2146–52. doi: 10.1038/oby.2005.266

55. Duffey KJ, Popkin BM. Adults with healthier dietary patterns have healthier beverage patterns. *J Nutr*. (2006) 136:2901–7. doi: 10.1093/jn/136.11.2901

56. Yang M, Chun OK. Consumptions of plain water, moisture in foods and beverages, and total water in relation to dietary micronutrient intakes and serum nutrient profiles among US adults. *Public Health Nutr*. (2015) 18:1180–6. doi: 10.1017/S136898001400007X

57. Dibay Moghadam S, Krieger JW, Louden DKN. A systematic review of the effectiveness of promoting water intake to reduce sugar-sweetened beverage consumption. *Obes Sci Pract*. (2020) 6:229–46. doi: 10.1002/osp4.397

58. Bragança M, Bogea EG, De Almeida Fonseca Viola PC, Dos Santos Vaz J, Confortin SC, AMB M, et al. High consumption of sugar-sweetened beverages is associated with low bone mineral density in young people: the Brazilian birth cohort consortium. *Nutrients*. (2023) 15:324. doi: 10.3390/nu15020324

59. Vanhaecke T, Bretin O, Poirel M, Tap J. Drinking water source and intake are associated with distinct gut microbiota signatures in US and UK populations. *J Nutr*. (2022) 152:171–82. doi: 10.1093/jn/nxab312

60. Ibrahim I, Syamala S, Ayariga JA, Xu J, Robertson BK, Meenakshisundaram S, et al. Modulatory effect of gut microbiota on the gut-brain, gut-bone axes, and the impact of cannabinoids. *Meta*. (2022) 12:1247. doi: 10.3390/metabo12121247

61. Nakamura Y, Watanabe H, Tanaka A, Yasui M, Nishihira J, Murayama N. Effect of increased daily water intake and hydration on health in Japanese adults. *Nutrients*. (2020) 12:12. doi: 10.3390/nu12041191

62. Tung CW, Hsu YC, Shih YH, Chang PJ, Lin CL. Dipstick proteinuria and reduced estimated glomerular filtration rate as independent risk factors for osteoporosis. *Am J Med Sci*. (2018) 355:434–41. doi: 10.1016/j.amjms.2017.12.011

63. Mansournia MA, Nazemipour M. Recommendations for accurate reporting in medical research statistics. *Lancet*. (2024) 403:611–2. doi: 10.1016/S0140-6736(24)00139-9

64. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. (1995) 311:485. doi: 10.1136/bmj.311.7003.485

65. Hodzic-Santor E, Fralick M. Treating patients, not P values. *NEJM Evid*. (2025) 4:EVIDe2400405. doi: 10.1056/EVIDe2400405

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 19 March 2025
DOI 10.3389/fnut.2025.1552367

Check for updates

OPEN ACCESS

EDITED BY
Alessandra Durazzo,
Council for Agricultural Research and Economics, Italy

REVIEWED BY
Mojca Korošec,
University of Ljubljana, Slovenia
Megan Kocher,
University of Minnesota Twin Cities, United States

*CORRESPONDENCE
Selena Ahmed
✉ selena.ahmed@heart.org

†These authors have contributed equally to this work and share first authorship

RECEIVED 27 December 2024
ACCEPTED 21 February 2025
PUBLISHED 19 March 2025

CITATION
Brinkley S, Gallo-Franco JJ, Vázquez-Manjarrez N, Chaura J, Quartey NKA, Toulabi SB, Odenkirk MT, Jermendi E, Laporte M-A, Lutterodt HE, Annan RA, Barboza M, Amare E, Srichamnong W, Jaramillo-Botero A, Kennedy G, Bertoldo J, Prenni JE, Rajasekharan M, de la Parra J and Ahmed S (2025) The state of food composition databases: data attributes and FAIR data harmonization in the era of digital innovation.
*Front. Nutr.* 12:1552367.
doi: 10.3389/fnut.2025.1552367



# The state of food composition databases: data attributes and FAIR data harmonization in the era of digital innovation

Sarah Brinkley[1†], Jenny J. Gallo-Franco[2†], Natalia Vázquez-Manjarrez[3], Juliana Chaura[4], Naa K. A. Quartey[5], Sahar B. Toulabi[6], Melanie T. Odenkirk[6], Eva Jermendi[7], Marie-Angélique Laporte[8], Herman E. Lutterodt[5], Reginald A. Annan[5], Mariana Barboza[9], Endale Amare[10], Warangkana Srichamnong[11], Andres Jaramillo-Botero[4,12], Gina Kennedy[1], Jaclyn Bertoldo[13], Jessica E. Prenni[6], Maya Rajasekharan[2], John de la Parra[14] and Selena Ahmed[13]*

[1]The Periodic Table of Food Initiative, Bioversity International, CGIAR, Rome, Italy, [2]The Periodic Table of Food Initiative, The International Center for Tropical Agriculture (CIAT), CGIAR, Cali, Colombia, [3]Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [4]iOMICAS Research Institute, Pontificia Universidad Javeriana Cali, Cali, Colombia, [5]Department of Biochemistry and Biotechnology and Department of Food Science and Technology, College of Science Kwame Nkrumah University of Science and Technology - KNUST, Kumasi, Ghana, [6]Department of Horticulture and Landscape Architecture, Colorado State University, Fort Collins, CO, United States, [7]Division of Human Nutrition and Health, Wageningen University and Research, Wageningen, Netherlands, [8]The Periodic Table of Food Initiative, Bioversity International, Montpellier, France, [9]Genome Center and Innovation Institute for Food and Health, University of California Davis, Davis, CA, United States, [10]Nutrition, Environmental Health and Non-Communicable Diseases Research Directorate, Ethiopian Public Health Institute, Addis Ababa, Ethiopia, [11]Institute of Nutrition, Mahidol University, Salaya, Thailand, [12]Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, United States, [13]The Periodic Table of Food Initiative, American Heart Association, Dallas, TX, United States, [14]The Rockefeller Foundation, New York, NY, United States

**Introduction:** Food composition databases (FCDBs) are essential resources for characterizing, documenting, and advancing scientific understanding of food quality across the entire spectrum of edible biodiversity. This knowledge supports a wide range of applications with societal impact spanning the global food system. To maximize the utility of food composition data, FCDBs must adhere to criteria such as validated analytical methods, high-resolution metadata, and FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable). However, complexity and variability in food data pose significant challenges to meeting these standards.

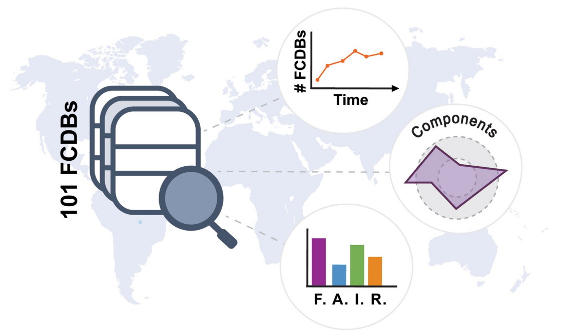
**Methods:** In this study, we conducted an integrative review of 35 data attributes across 101 FCDBs from 110 countries. The data attributes were categorized into three groups: general database information, foods and components, and FAIRness.

**Results:** Our findings reveal evaluated databases show substantial variability in scope and content, with the number of foods and components ranging from few to thousands. FCDBs with the highest numbers of food samples (≥1,102) and components (≥244) tend to rely on secondary data sourced from scientific articles or other FCDBs. In contrast, databases with fewer food samples and components predominantly feature primary analytical data generated in-house. Notably, only one-third of FCDBs reported data on more than 100 food components. FCDBs were infrequently updated, with web-based interfaces being updated more frequently than static tables. When assessed for FAIR compliance, all FCDBs met the criteria for Findability. However, aggregated scores for Accessibility, Interoperability, and Reusability for the reviewed FCDBs were 30, 69, and 43%, respectively.

**Discussion:** These scores reflect limitations in inadequate metadata, lack of scientific naming, and unclear data reuse notices. Notably, these results are associated with country economic classification, as databases from high-income countries showed greater inclusion of primary data, web-based interfaces, more regular updates, and strong adherence to FAIR principles. Our integrative review presents the current state of FCDBs highlighting emerging opportunities and recommendations. By fostering a deeper understanding of food composition, diverse stakeholders across food systems will be better equipped to address societal challenges, leveraging data-driven solutions to support human and planetary health.

KEYWORDS

**food composition database, food composition data management, food composition data, food quality, metadata, food components, FAIR data, nutritional database**




GRAPHICAL ABSTRACT

## 1 Introduction

Food composition data are essential for informing solutions to today's human and planetary health challenges including loss of biodiversity, food insecurity, and diet-related chronic disease (1–3). Food composition databases (FCDBs) are foundational tools across sectors, including agriculture, food science, nutrition, public health, and policymaking, supporting crop breeding, product development, nutritional assessments, and public health initiatives. By advancing the availability and accessibility of FCDBs it is possible to promote evidence-based solutions to harness the power of food to foster well-being and sustainable practices across food systems. Contributors and curators of FCDBs perform a critical role in providing access to and enabling the use of reliable, high-quality data on food and food composition (4–6). To maximize the utility of data from FCDBs for diverse applications, these databases should meet three key criteria: (i) the utilization of validated methods and computational approaches to ensure data accuracy and consistency (1) (ii) the inclusion of detailed, high-resolution metadata that provide essential context about the source, preparation, and analysis of foods (7), and (iii) adherence to FAIR Guiding Principles for data management and stewardship, making data Findable, Accessible, Interoperable, and Reusable (FAIR), to facilitate integration, sharing, and practical application across sectors (8).

Given the importance of FCDBs, conducting an integrative review of their current state was timely. Here, we evaluated FCDBs spanning multiple countries worldwide based on 35 data attributes, emphasizing the range of foods and components included, data harmonization, and adherence to FAIR data governance principles. We also present an overview of the inception and historical evolution of these databases, highlighting their role in enabling researchers to monitor trends in food crop variation, particularly in response to climate change and biodiversity loss. Additionally, we examined how the data are presented and accessed from the end user's perspective, including researchers, policymakers, and food systems stakeholders, ensuring accessibility and usability across sectors. Finally, we assessed the compliance of the analyzed databases with the FAIR data criteria—Findable, Accessible, Interoperable, and Reusable. The results of this evaluation provide a detailed snapshot of the current state of FCDBs and identify key opportunities to enhance their functionality as dynamic and integrative resources. These enhancements can foster cross-sectoral collaboration and drive innovative solutions. Strengthening FCDBs will unlock their potential to support global efforts in preserving food biodiversity, addressing nutrition insecurity, and mitigating diet-related chronic diseases through evidence-based strategies.

## 2 Background

In their first iterations in the 1800s, food composition data were compiled in Food Composition Tables (FCT) that strictly focused on proximate composition (e.g., carbohydrates, fat, protein, moisture, and ash) of a limited cross-section of foods in a "typical" (i.e., Western-leaning) diet (9, 10). In contrast, modern FCDBs are characterized by a high degree of heterogeneity encompassing a diverse range of data sources, analytical methods, nomenclature and terminologies, food types, data processing methods, data formats, and overall relevance to various audiences (11). This diversity reflects advancements in analytical technologies, which have expanded the scope of nutritional data to include foodomics-level insights such as the thousands of specialized metabolites in foods including bioactive polyphenols, sterols, terpenes, and carotenoids (1, 12, 13). However, despite these advancements, foodomics data remain underrepresented in FCDBs, and the relationship of the thousands of specialized metabolites to adequate nutrition and health is still being established.

Efforts to address the variability and gaps in FCDBs have been ongoing for decades. European compilers at the International Network of Food Data Systems (INFOODS) administered by the Food and

Agriculture Organization (FAO) of the United Nations, the EU Network of Excellence: European Food Information Resource (EuroFIR), and others have long recognized the need for and led efforts on the harmonization of food composition data (14–16). Since their inception, these international efforts have advanced food composition database management by promoting the inclusion of mandatory metadata thesauri, standardized analytical methods (e.g., AOAC), food and food component nomenclature, and methods of conversion (17).

Despite these advances, national FCDBs, which track the nutrient composition of foods based on dietary intake patterns at the national level, often reflect regional biases. For instance, the United States Department of Agriculture's FoodData Central (FDC) (18) widely recognized as the gold standard in food composition data, is a crucial resource for aggregating food data which shapes the U.S. national nutrition guidelines and associated food and nutrition policies (10). However, with a federal mandate to survey the nation's most widely consumed foods, FDC may still have sparse coverage of foods found in regionally distinct diets (19). For example, Lozano et al. (20), report 97 commonly consumed foods of Hawaii, like taro-based poi or pohole (i.e., fiddlehead fern or *Diplazium esculentum*) are not represented in FDC's Food and Nutrient Database for Dietary Studies (FNDDS). This paucity of food representation leaves nutrition professionals to rely on closely related food analogs which may result in dietary assessment error disproportionately impacting the health outcomes of the populations who depend on these foods (20, 21).

While efforts to increase the edible biodiversity represented in global food composition databases exist (15), there is still a panoply of edible species yet to be characterized (22–24). To overcome the disparity of underrepresentation, national FCDBs must be enriched with data on regionally distinct staples and less utilized, culturally significant foods, for example, edible insects like house cricket (*Acheta domesticus*) and dung beetle (*Paragymnopleurus aethiops*) in Thailand, African palm weevil (*Rhynchophorus phoenicis*) in Ghana (25–27), and *Amaranthus* spp. endemic throughout sub-Saharan Africa and the Americas (28). The characterization of traditional foods, like amaranth or nopal (*Opuntia ficus-indica*) from Mexico and other regions, will allow for further safeguarding of traditional knowledge while integrating nutrient-dense ingredients with high potential for reducing noncommunicable disease (28, 29) into regional and global nutrition frameworks. Inclusion of these foods is crucial for accurately reflecting true biodiversity and addressing food security challenges. For instance, the moriche palm (*Mauritia flexuosa*) serves as a vital resource in Colombia, providing not only nutritious fruits rich in vitamins A and E for traditional dishes but also providing materials for crafts and construction, thus supporting local economies and cultural practices (30). Expanding the characterization of the world's edible biodiversity will not only reduce assessment error and improve cross-cultural relevance (31), greater understanding of chemodiversity will also inspire a cornucopia of future foods (23).

Secondary data (i.e., food composition data from another FCDB, peer-reviewed manuscript, or another external source) may also lead to data homogenization or inaccurate representations of the local food supply. Due to resource constraints, national FCDBs often rely on the primary data generated by the USDA or other literature-reported primary food composition data. Primary data refer to food composition data derived from in-house, laboratory analysis, which is generated specifically for the purpose of compiling the FCDB. Databases may recycle primary data directly, use methods of conversion, or publish an amalgamation of both primary and secondary data. While the use of secondary data facilitates faster data compilation there are often challenges in harmonizing analytical methodologies, conversion factors, and other technical aspects related to data processing and reporting (11, 32). Additionally, the nutrient content of some foods can vary significantly between countries and regions due to factors such as genetics (i.e., cultivars, variety, or breed), environment (i.e., climate, soil, geography, and biotic and abiotic factors), and agricultural management, not to mention postharvest and processing factors (10, 33).

Building on these advancements, modern FCDBs are presented with an opportunity to adopt better data governance and stewardship principles. International quality standards such as the FAIR Data Guiding Principles (34) promote the inclusion of metadata and ontologies (35) to make food composition data more discoverable, shareable, usable, and citable. Originally, FAIR Data Principles were created to increase the exchange of scholarly data products (34), but by extension, the utility of FAIRness for food composition data management and stewardship facilitates the sharing of knowledge on foods and food composition (36). FCDBs with harmonized food composition data support international research and policy-making, including addressing cross-border nutritional challenges and promoting a transnational understanding of the world's food supply with increasingly interconnected food systems (37). With this unified approach, there is potential for a greater comparative understanding of nutritional resources globally, yet data harmonization should not result in data homogenization.

# 3 Materials and methods

## 3.1 Assessing the landscape of food composition databases

We conducted an integrative review of food composition databases. The systematic approach of the integrative review followed a rigorous systematic literature review methodology but integrated diverse sources of research (i.e., disparate food composition databases) (38, 39). We began our investigation with a broad research question aimed at assessing the current landscape of food composition databases globally: "What are the gaps and opportunities in food composition analysis and data collection in an era of digital innovation"? We then conducted multiple searches using Google Search and Google Scholar in private browser tabs using the keywords "nutritional database" OR "food composition database". The search process took place between April and December 2023 with all queries performed in English. The locations where the searches were conducted were globally distributed between Europe, North America, and South America. All the results were reviewed. A minimum of two researchers independently conducted each search for food composition databases.

The search results were carefully reviewed to identify national and international FCDBs, foodomics databases, and other nutritional databases for inclusion in our integrative review. The search uncovered resources dating back to the 1950s, which informed our decision to include databases from 1950 to 2024. Our search additionally revealed food composition database repositories (i.e., collections of food composition databases) from authoritative sources essential for research and policymaking. In addition to individual FCDBs, we also included resources previously compiled by such authoritative sources including FAO/INFOODS (40), EuroFIR (14), Danish Food Informatics (16), and the World Nutrient Databases for Dietary Studies (WNDDS) (41). The integration of FCDBs from these open-access repositories enhanced the accuracy and scope of this integrative review, enabling more robust analyses and comparisons.

After the systematic search, teams of reviewers were assembled to conduct quality appraisal and data extraction steps. A set of 35 data attributes was established (Figure 1) to characterize the identified FCDBs. Multidisciplinary teams of 16 researchers from eight countries, with expertise in food science, nutrition, public health, agricultural sciences, analytical chemistry, biology, biotechnology, bioinformatics, and computer science were assembled to define FCDBs characteristics and extract data on the 35 data attributes used in the review. The transdisciplinary researchers were from multiple geographies and cultural-linguistic backgrounds, and thus, they reviewed databases in Dutch, English, French, Hungarian, Italian, and Portuguese in their original language. For databases available in languages other than these languages, researchers used Google Translate to translate the necessary information for data extraction. All the FCDBs were distributed in random order for data extraction by the transdisciplinary research teams. Teams of two reviewers conducted a thorough and independent review and quality appraisal of each food composition database to decide upon inclusion or exclusion. Finally, a third round of revisions was carried out to resolve any discrepancies identified by the first two reviewers and to ensure consistency and data quality across the entire dataset. A smaller group of researchers was selected to conduct a third and final review.

The 35 attributes were categorized into three groups: general information about the database (23 attributes), food and nutrient data (8 attributes), and FAIR Data Principles (4 attributes evaluated against 13 criteria).

The general database attributes included: Database name, Acronym, Link, Reference, Related repository, Funding source/ Governance, Country of creation, Creation Country Economy Classification (WBA), Countries included, Economy classification of countries included (WBA), Creation date, Last update, Reported languages, Data source type, Database interface, Export availability, Access fee, Foodomics data availability, Experimental data inclusion, Search interface, Data source, Overall objective, and General description. For definitions of the data attributes, refer to the Readme file in Annex 1 of the Supplementary materials. For the economic classification of countries, we applied the World Bank economic classification system (42). To simplify interpretations, we grouped countries classified as High-Income together with those categorized as Upper-Middle-Income, and countries classified as Low-Income together with those categorized as Lower-Middle-Income.

The food and nutrient data attributes included: the number of food samples (i.e., total number of food samples including diverse food types such as multi-ingredient foods), food groups, number of components, data type, proximate composition, minerals and trace elements, vitamins, and specific compounds. Due to the lack of standardization in food-specific metadata, we evaluated food coverage across 13 predefined food groups, based on the methodology of Jarvis et al. (23). These groups included: (i) algae, (ii) mushrooms, (iii) herbs and spices, (iv) oily plants, (v) beverages, (vi) nuts and seeds, (vii) processed foods, (viii) beans and pulses, (ix) fruits, (x) terrestrial animal products, (xi) vegetables, (xii) aquatic animal products, and (xiii) cereals and grains. Notably, processed foods were an additional category not included in Jarvis et al. (23). Processed foods are defined as any foods that are not in a raw or minimally processed state.

Criteria were established to determine if a database should be included or excluded from this integrative review: (i) problems accessing the database, (ii) the absence of food composition data, (iii)

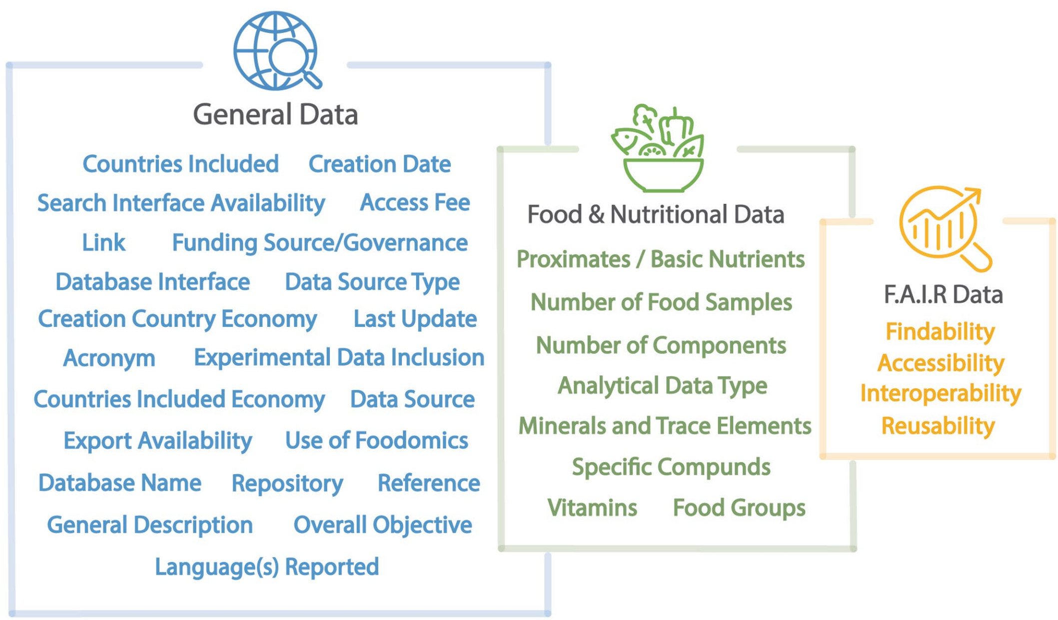



FIGURE 1
The data attributes used to catalog the food composition database characteristics. For the full descriptions of the 35 attributes used and the data collected, refer to Annex 1 and Supplementary Tables S1–S5.

the presence of only one metabolite unrelated to its presence in foods, and (iv) repositories or lists of databases.

## 3.2 Data stewardship and FAIR data guiding principles

To assess data stewardship best practices, all FCDBs underwent an evaluation of FAIR Data Guiding Principles, which included assessments of data Findability, Accessibility, Interoperability, and Reusability. Emphasis was placed on machine readability, as described by Wilkinson et al. (34), due to the increasing scale of big data, the advent of artificial intelligence, and the need for computational support in research. FCDBs that met the individual criteria used to assess FAIR Principles were assigned a 1. Any misalignment of FCDBs or ambiguous agreements with criteria on specific FAIR data assessments were assigned a 0 (Supplementary Tables S3, S4).

The initial step in adhering to the FAIR Data Guiding Principles involves locating the data. The findable criteria were defined as the database being easily discoverable by both humans and machine-learning interfaces. To determine if the database is findable, it must possess a globally unique, persistent identifier such as a Uniform Resource Locator (URL) or a Digital Object Identifier (DOI) (see Supplementary Table 3). In some cases, the search engines or the global databases provided broken URLs for the FCDBs, requiring extensive searches (i.e., Archive.org) to find the updated links (43).

To evaluate data accessibility, four criteria were assessed. First, FCDBs were classified as publicly accessible or available in a controlled manner for users with appropriate permissions, either for free or for a fee. Although databases were not penalized for requiring a fee, open-access was considered a positive factor in data accessibility. Second, we examined whether the database allowed data downloads, and if so, whether the output was provided in a machine-readable format (e.g., CSV, XML, JSON, and MySQL). For this integrative review, PDFs were considered non-machine readable. Finally, the presence of an application programming interface (API) was assessed as the final metric for accessibility (see Supplementary Table 3).

The interoperability of a database was assessed based on the adoption of standardized protocols and formats that enable both humans and machines to retrieve and interpret the data effectively. To evaluate interoperability, we examined whether the database employed a formal, accessible, shared, and widely applicable language for representing knowledge related to: (i) food groups, (ii) scientific names, (iii) nutritional components, and (iv) analysis methods for primary and secondary data types (see Supplementary Table 3).

Lastly, databases were considered reusable if they met the following criteria: (i) inclusion of a clear and accessible data usage license, (ii) association with detailed provenance information, and (iii) provision of metadata that comprehensively describes the context in which the data were generated, as outlined by Wilkinson et al. (34), (see Supplementary Table 3).

## 3.3 Meta-analysis and statistics

Metadata were harmonized for statistical analysis with RStudio Version 4.3.2 (R Studio, Boston, MA, United States). Statistical analysis and data visualization were performed using the following packages: R World Maps, ggplot2, maps. Code is available on GitHub at: https://github.com/scbrinkley/ptfi-fs.

# 4 Results

## 4.1 Identification and inclusion of databases

A total of 117 FCDBs were compiled from web resources. A set of inclusion criteria was applied to each database to establish if it should be included in this integrative review. These criteria included: accessibility and availability of nutritional data, inclusion of comprehensive information about food components, and/or inclusion in a national or international effort to evaluate food composition. From the initial list of FCDBs, we excluded 17 databases because they did not meet the inclusion criteria (Supplementary Table 6). Specifically, these 17 FCDBs were excluded because the webpage was not available, they contained only single food components, they only contained data on food flavor, and/or they solely provided lists of other food composition databases. We arrived at a final count of 101 FCDBs (Figure 2).

## 4.2 Global overview of food composition databases

An inventory of 101 FCDBs was evaluated. Each database was characterized using 22 general database attributes, eight food and nutritional data attributes, and four criteria for FAIRness evaluation (Annex 1; Supplementary Tables S1–S5). Out of the 101 databases assessed, 73 (72%) FCDBs provide nutritional data for foods typically consumed within a single country, focusing primarily on local foods. In contrast, 28 (28%) of the databases compile nutritional data from food collected and consumed across multiple countries, often involving regions or neighboring countries. Among these international databases, 16 explicitly list the names of all countries contributing to the food data, whereas the remaining 12 adopt a broader international scope without specifying the countries included (Figure 3).

A note on database governance, 68 (67%) FCDBs were funded and managed by national governments, 24 (24%) FCDBs were transnational or international efforts, 16 (16%) FCBDs were managed by or associated with public universities, and 8 (8%) FCDBs were nonprofit initiatives. Among the nonprofit organizations of note, the Alliance of Bioversity-CIAT and International Food Policy Research Institute (IFPR), both CGIAR institutions, have supported the formation of several FCDBs including *The Periodic Table of Food Initiative* (PTFI), Biodiversity for Food and Nutrition (B4FN), and HarvestPlus' *A Food Composition Table for Central and Eastern Uganda* (44, 45, 80).

The FCDBs were further stratified based on the origin of the data used. Eleven databases reported exclusively primary data, which consists of analytical data generated directly by the database entities; 42 relied only on secondary data, gathered from different secondary sources such as scientific articles or other FCDBs; 44 used a mix of primary and secondary data (Figure 3). Notably, countries hosting more than one database typically included both primary and secondary data. Likewise, databases from countries in Africa, Central

10.3389/fnut.2025.1552367

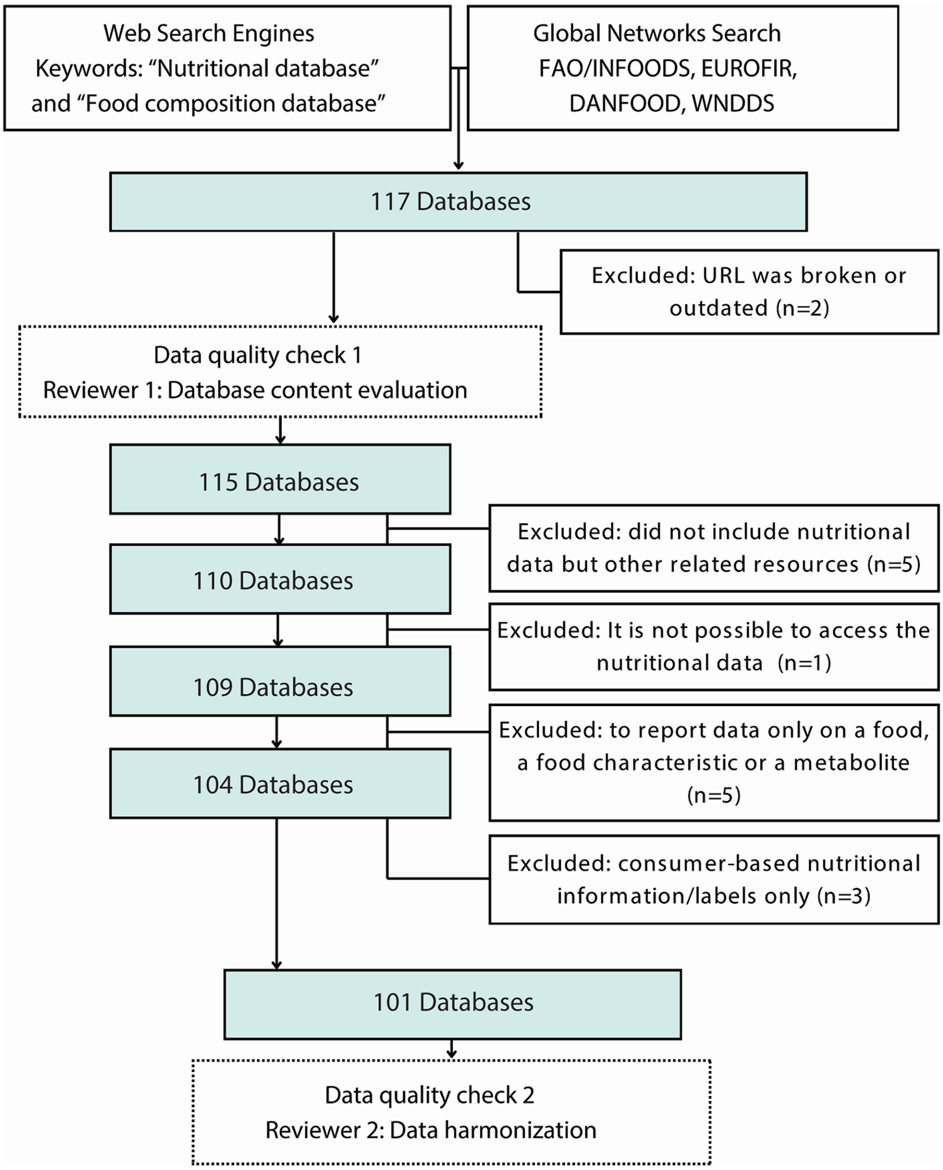
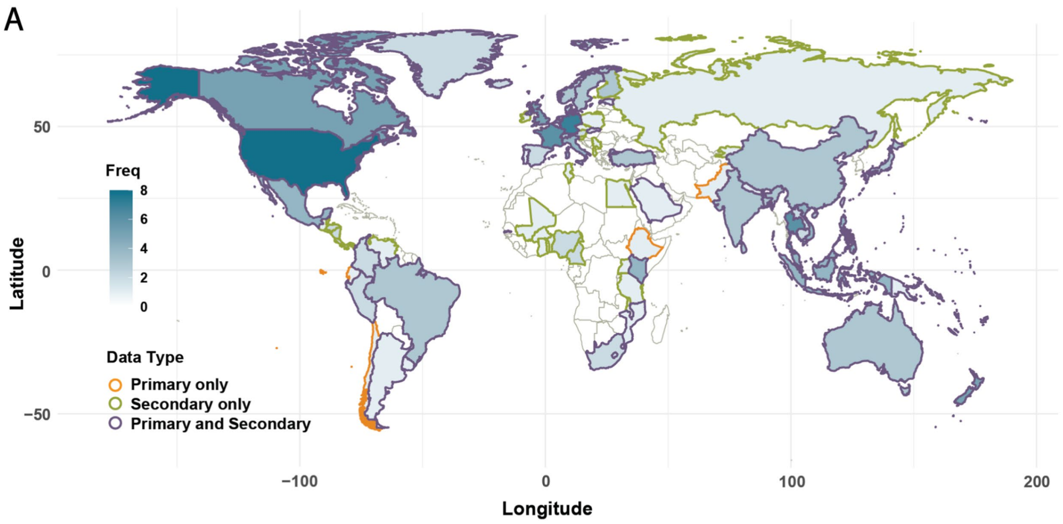



FIGURE 2
Food composition databases search, filtering, and quality appraisal protocol. All the databases were submitted to two data quality check controls to be considered in this integrative review.

America, and Eastern Europe predominantly contain secondary data (Figure 3).

We further analyzed the relationship between the data source (primary or secondary) and three key attributes: date of creation, number of food samples, and number of compounds per food across the 101 FCDBs. The oldest database in this analysis, the *Standard Tables of Food Composition in Japan* (46), dates back to 1950 In contrast, the most recent ones include the *Albanian Food Composition Table* (47) and *The PTFI* (PTFI Research Hub – Research Community and Resources for the Periodic Table of Food Initiative) published in 2022 and 2024, respectively.

The number of food samples listed across these databases varied significantly, ranging from a single food type in the *Bovine Milk Proteome Database* (48) and *The Milk Composition Database* (49) to 65,993 in the *L'observatoire de l'alimentation* database (*OQALI*), with an average of 2,523 food samples. Notably, 90% of the databases reported fewer than 4,000 food samples. Our analysis revealed that databases solely based on secondary data had the highest average number of food samples (range: 1 to 65,993; average: 3,614). Excluding outliers such as the *L'observatoire de l'alimentation* database (50), *Food and Nutrient Database* (51), and *The European Food Safety Authority Food Composition Data* (52), which report 65,993, 19,500, and 16,500 food samples respectively, the adjusted average for secondary data databases drops to 1,390. Mixed databases (primary and secondary data) ranged from 1 to 15,000 foods, averaging 1,988. Excluding outliers databases *FoodData Central* (10) and *German Nutrient Database* (53), which report 13,682 and 15,000 food samples respectively, the average number of food samples falls to 1,400. Databases with only primary data reported between 16 and 1,892 foods, averaging




B

| North America | Central America | South America | Africa | Europe | Asia | Pacific |
|---|---|---|---|---|---|---|
| 15 | 2 | 11 | 18 | 33 | 22 | 6 |

FIGURE 3
Global distribution of food composition data. **(A)** World map showing the frequency of countries' food composition data representation across databases. *Note: the blue color gradient (0–8+) indicates the frequency of inclusion reported per country. Darker blue represents higher frequencies (8+), while lighter blue and white indicate lower frequencies (0–1). Countries are outlined in orange if they only report primary data, green if they only report secondary data, and purple if they report both primary and secondary data. **(B)** Summary of food composition data reported by geographic region.

488 (Figure 4A; outliers were excluded from graphical representation).

The count of compounds reported per database also showed considerable variation, from six in the *European Database of Carotenoids* (54) and *The Proximate Composition of New Zealand Marine Finfish and Shellfish* (55) to 70,926 in *FooDB* (56), with an average of 1,223 compounds. However, 90% of databases reported fewer than 550 components. According to our results, databases including primary data averaged the highest number of measured compounds (range: 6 to 24,721; average: 2313), but removing the outlier database *The PTFI,* which includes 24,721 compounds, reduces the average to 73. Mixed data databases varied from 15 to 70,926 compounds, averaging 1,756, but excluding the outlier FooDB drops the average to 147. Secondary databases ranged from 6 to 10,642 compounds, averaging 419, and removing the outlier *Bovine Milk Protein Database* with 10,642 compounds, adjusted the average to 181 (outlier databases were excluded from graphical representation - Figure 4A).

Analyzed FCDBs were formatted in different interfaces, 48 were web interfaces (48%), 45 were static tables (44%), and 8 included both web interfaces and static tables (8%). Originally, databases were primarily static tables. However, in the early 2000s, there was a clear shift toward web-based interfaces (Figure 4B). Interestingly, despite the growing popularity of DBs with web interfaces, table formats have continued to be a prevalent method for presenting nutritional information. Based on the World Bank economic classification of countries (42), 77% of FCDBs were created by High-Income countries, while 23% were created by Low-Income countries. As expected, FCDBs developed by High-Income countries primarily incorporate a web-based interface, while most of the FCDBs developed by Low-Income Countries rely on static tables (Figures 4C,D).

Although the number of table and web-based FCDBs has increased over time, their update frequencies show considerable variation (Figure 4F). Of all databases analyzed, 38 (39%) have never been updated. Among the remaining 59 databases, 11 (11%) have not received updates in the last decade. The update frequency also differs by database format; among table-based databases, 28 (61%) have never been updated and 7 (15%) were last updated over a decade ago. In contrast, among the 51 databases with web interfaces, only 10 (20%) have never been updated, with the majority (69%) updated in the last 5 years (Figure 4F). Additionally, the update frequency seems influenced by the economic classification of the country of creation of the database, with databases from High-Income countries generally showing more recent updates (Figure 4E).

## 4.3 Food and nutritional coverage across FCDBs

Supplementary Table S2 summarizes the findings from the assessment of FCDBs, focusing on the food samples and components measured in each database, including both nutritional and bioactive compounds. We first examined the inclusion of commonly reported macro- and micronutrients, including proximate composition, minerals, and vitamins (Supplementary Table S7; Figures 5A–C). Among the evaluated FCDBs, 95 contained data on proximates or basic nutrients, 94 included minerals, and 91 covered vitamins. In

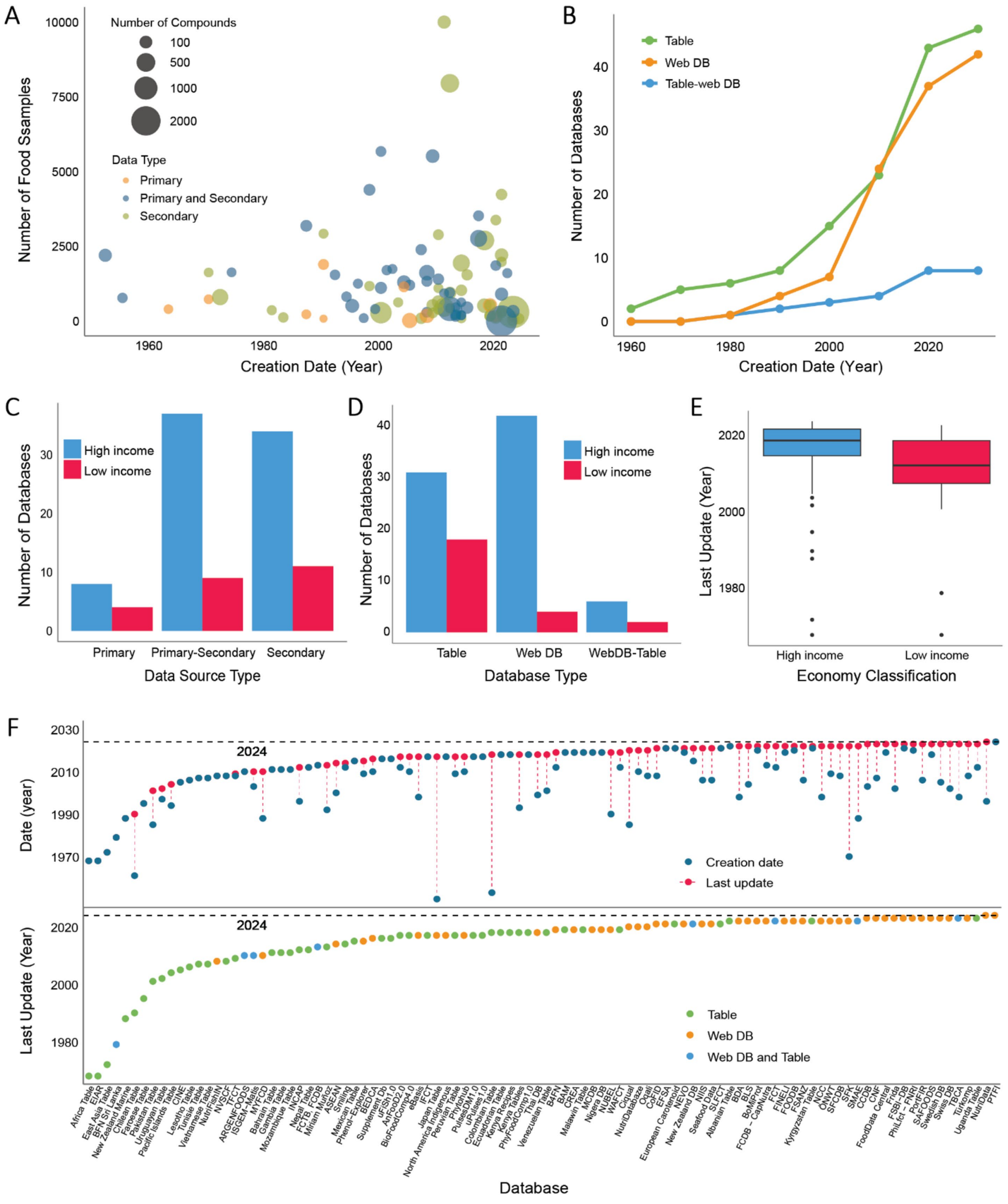



FIGURE 4
Overview of food composition databases (FCDBs) over time. **(A)** Bubble plot showing the relationship between the number of food samples covered in each database, the number of components reported, the type of data used (primary, secondary, or both), and the creation date. Eight outlier databases were excluded for better visualization. The size of each bubble represents the number of molecular components. **(B)** Line plot depicting the growth in the number of FCDBs created from 1950 to 2024 (year of review). Different colors represent the type of database interface: Table (green), Web interface (purple), and both Table and Web interface (blue). **(C)** Bar plot showing the number of FCDBs created by High-Income and Low-Income countries, categorized by database interface type (Table, Web, or both). **(D)** Bar plot representing the number of databases created by High-Income and Low-Income countries, categorized by data source type: Primary, Secondary, or both (Primary and Secondary). **(E)** Box plot comparing the distribution of the last update year for FCDBs created by High-Income and Low-Income countries. Countries grouped as High-income also include countries considered Upper-Middle income, and the category of Low-Income countries also includes countries classified as Lower-Middle income (42). **(F)** Upper graph: A timeline illustrating the creation dates (blue dots) and most recent updates (red dots) for the 97 FCDBs analyzed. Lower graph: A timeline displaying the last update dates for the 97 FCDBs, categorized by database interface: Table (green), Web interface (purple), or both Table and Web interface (blue).

contrast, only four specialized, compound-specific databases, such as *Phytohub* (57) and the *Bioactive Substances in Food Information Systems database* (58), focused exclusively on plant-based bioactive compounds.

Across all databases, a total of seven proximates were reported: water, ash, energy, fiber (i.e., crude and/ or dietary fiber), carbohydrates, fat, and protein. Among these, protein and fat were the most frequently included, reported in 90 and 91% of databases, respectively (Figure 5A). A total of 43 minerals were identified, but only 19 were reported in at least 10 databases (Figure 5B). Likewise, 18 vitamins were identified, with 12 of them being reported in at least 10 databases (Figure 5C; Supplementary Table S7). Specific bioactive compounds were also assessed across all databases. Notably, only 15% of the databases did not report any specific bioactive compounds beyond proximates, vitamins, and minerals. Of the 85% of databases reporting bioactives, the main groups of compounds were identified and reported as either compound class (e.g., fatty acids, amino acids, polyphenols, etc.) or specific compounds (e.g., cholesterol, tryptophan, beta-carotene, etc.; Supplementary Table S7). However, due to the diversity of compounds and nomenclature used, further comparison to evaluate the coverage of compounds across databases was not possible.

The landscape of edible biodiversity reported in food composition databases is extensive. However, in many cases, critical food-specific metadata and/or standardized food coding are absent. This lack of harmonization complicates the comparison of foods across databases beyond broad, culinary food group classifications. To address this limitation, we used 13 predefined food groups to assess and compare food coverage across all databases (Figure 5D). Our analysis revealed that *Aquatic animal products* and *Cereals and grains* were the most common food groups, present in 89 and 88% of FCDBs, respectively. *Fruits and vegetables* were reported in 86% of databases, followed by *Terrestrial animals* reported in 85% of databases. Notably, infrequent coverage was observed for *Mushrooms* and *Algae*, which were included in only 45 and 25% of databases, respectively. Furthermore, the number of food groups represented in the FCDBs was analyzed (Supplementary Figure 1) revealing that 60% of databases included 10 or more of the 13 food groups. In contrast, 9% of databases reported only one single food group.

## 4.4 FAIRness of FCDBs

Following the criteria outlined in this manuscript for evaluating the FAIRness of global nutritional databases (Supplementary Table S3),

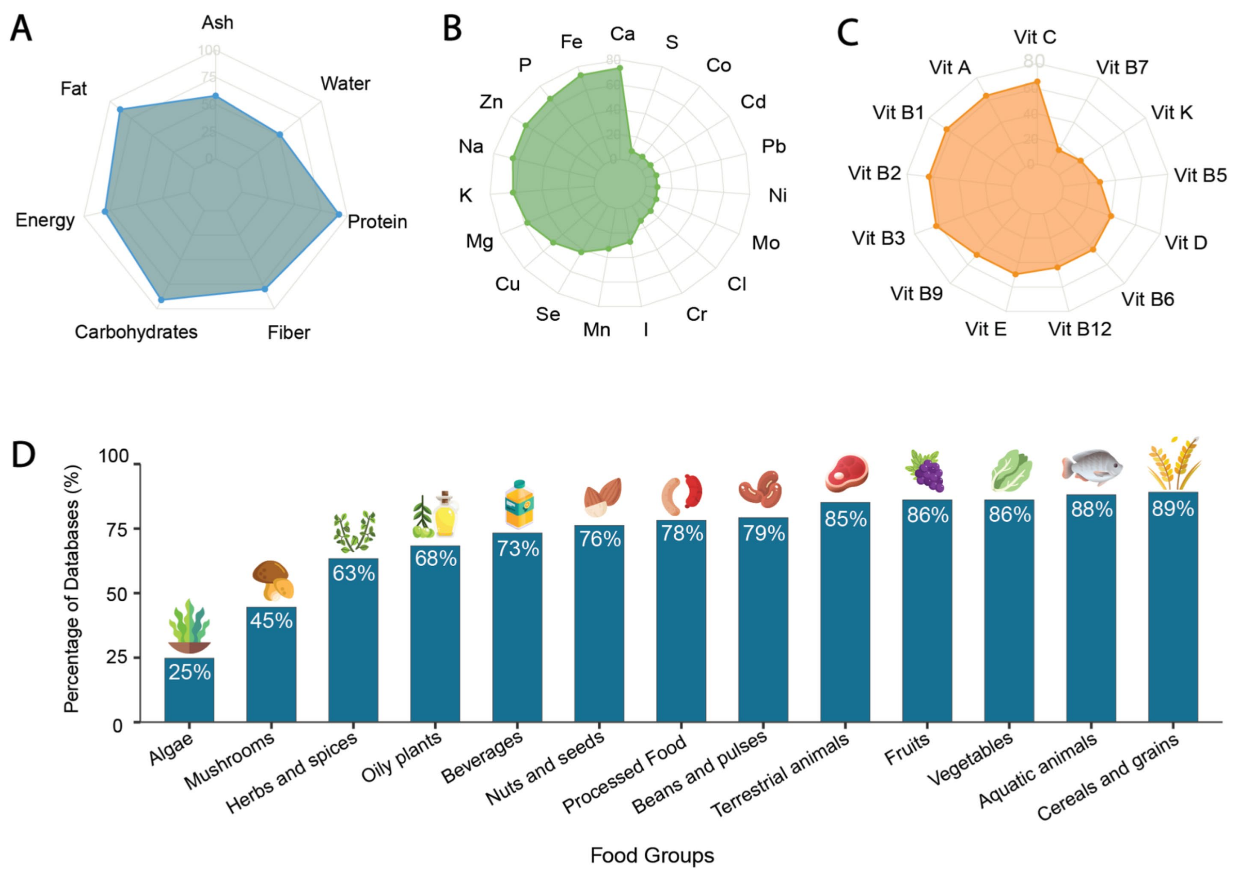



FIGURE 5
Coverage of food composition across food databases. **(A)** Report of proximate composition in Food Composition Databases (FCDBs). Chart axes represent different nutritional components each scale from 0 to 100. The plotted values indicate the number of databases that include data on each respective compound **(B)** Number of databases documenting minerals reported in at least 10 databases assessed in this study. **(C)** Number of databases documenting Vitamins reported in at least 10 databases assessed in this study **(D)** Presence of 13 food groups across the evaluated food composition databases, presented as the percentage of databases that include each food group.

we established the percentage of databases that adhere to FAIR principles of being Findable, Accessible, Interoperable, and Reusable (Figure 6) (34). Based on our findings, every database included in this study met the findable criteria, meaning that each one is assigned a persistent identifier. This identifier, which can be a Uniform Resource Locator (URL) or a Digital Object Identifier (DOI), ensures that the databases can be accurately identified by both human users and computers. However, two databases were excluded in the first step because their URLs were broken (i.e., access yielded a 404 error: page not found).

Our analysis revealed that 100% of the databases adhere to the criteria of being publicly accessible through either open or paid access. However, a closer examination of the access modes reveals limitations in access. Specifically, 10% of the databases did not have download availability, thereby making it impossible to interact with the raw data. Among the databases that did allow downloads, only 64% support the downloading of machine-readable files, underscoring a limitation in the versatility of data formats provided. Additionally, the option to download data via an API is scarce with only 32% of databases offering this path for automated data retrieval. Collectively, only 30% of databases met all accessibility criteria evaluated.

Evaluation of the interoperability of FCDBs showed that food classification systems were consistently reported across all databases. Food components metadata were included in 96% of the databases, and metadata for methods used in secondary databases was present in 91% of cases. The areas with the least compliance were the inclusion of scientific names for the foods analyzed, which was found in 80% of FCDBs, and the inclusion of metadata related to the analytical methods used in primary databases, which was present in 80% of the databases. Overall, 69% of the FCDBs met all the criteria for interoperability.

The assessment of data reusability revealed high compliance in reporting data provenance (90%) and providing descriptive general metadata (81%). However, the most commonly missing criterion was the lack of a licensing statement regarding the use of the data, which was present in only 49% of the databases. Overall, only 43% of FCDBs met all the criteria for data reusability.

Overall, the analysis of the FAIR principles indicates that, although several FAIR criteria are being implemented by national and international FCDBs, significant gaps in adoption remain. Notably, only 17 databases (17%) satisfied all 13 FAIR criteria evaluated in this integrative review. All of these FCDBs are accessible through web interfaces and are associated with High-Income countries (Table 1).

It is important to note that while we evaluated the FAIRness of FCDBs, we did not assess the accuracy of food composition data. The FAIRness criteria for interoperability only indicated if analytical methods—used in the collection of food composition data—were reported. We did not evaluate the validity of the analytical methods used to generate primary data. These factors are crucial for assessing the reliability and accuracy of nutritional data but fall outside the scope of this review.

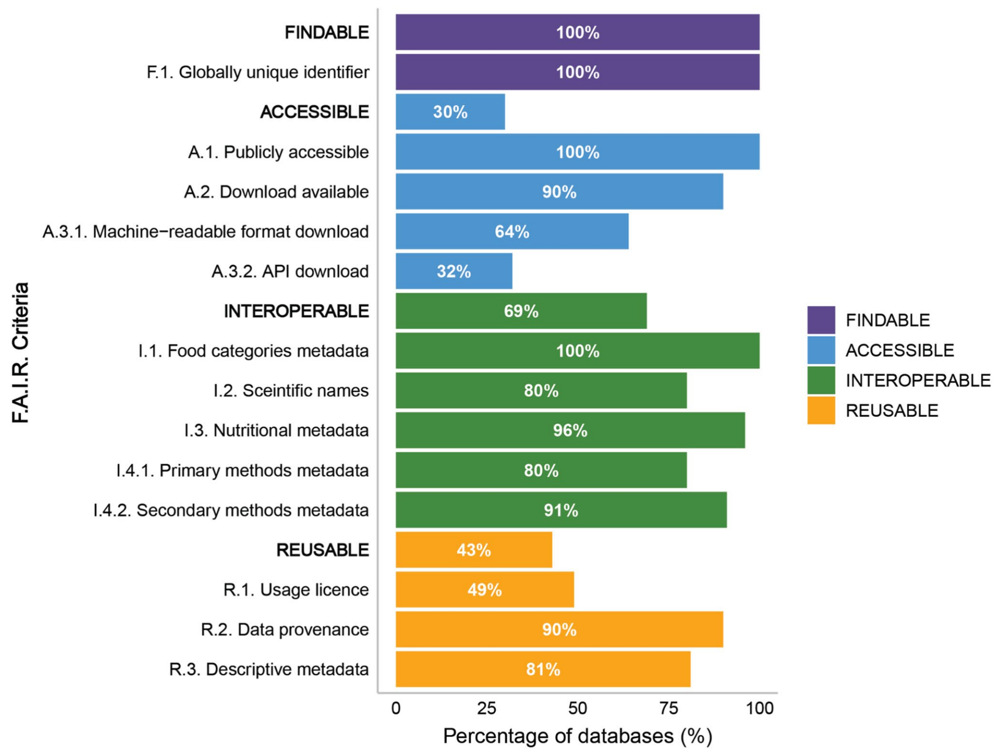



FIGURE 6
FAIR Data Principles criteria for Food Composition Databases (FCDBs). Bar graph illustrating the percentage of databases meeting the criteria for each FAIR principles Findable (purple), Accessible (blue), Interoperable (green), and Reusable (orange). For each category, the first item in uppercase represents the overall compliance percentage of databases with all the listed criteria under that principle, followed by the individual compliance percentages for specific criteria within the category. Detailed information on the criteria used to evaluate the FAIR data principles is provided in Supplementary Table S3.

TABLE 1 Food composition databases that met all the FAIR criteria analyzed in this integrative review.

| Database | Creation date | Last update | Creation country | Database type |
|---|---|---|---|---|
| ANSES-CIQUAL | 1985 | 2020 | France* | Web DB |
| Canadian Nutrient File (CNF) | 2007 | 2023 | Canada* | Web DB |
| Composition of Foods Integrated Dataset (CoFID) | 2008 | 2021 | United Kingdom* | Web DB |
| Czech Food Composition Database (NutriDatabaze) | 2010 | 2020 | Czech Republic* | Web DB |
| Dutch Food Composition Database (NEVO) | 2019 | 2021 | Netherlands* | Web DB |
| Finnish National Food Composition Database (FINELI) | 2019 | 2022 | Finland* | Web DB |
| Frida Food Data | 2002 | 2023 | Denmark* | Web DB |
| Food Composition Database for Epidemiological Studies in Italy (BDA) | 1998 | 2022 | Italy* | Web DB |
| FOODB | 2020 | 2022 | Canada* | Web DB |
| Food Data Central | 2019 | 2023 | United States* | Web DB |
| The Periodic Table of Food Initiative (PTFI) | 2024 | 2024 | United States*, International | Web DB |
| New Zealand Food Composition Database | 2015 | 2021 | New Zealand* | Web DB; Table |
| Spanish Food Composition Database (BEDCA) | 2010 | 2016 | Spain* | Web DB |
| Tabelle di Composizione Degli Alimenti (CREA) | 2019 | 2019 | Italy* | Web DB |
| The Norwegian Food Composition Table (FCT) | 2012 | 2022 | Norway* | Web DB; Table |
| The Swiss Food Composition Database | 2002 | 2023 | Switzerland* | Web DB |
| Turkish Food Composition Database (Türkomp) | 2008 | 2023 | Turkey* | Web DB |

*High-Income Economy based on World Bank Data (42).

## 5 Discussion

In the era of digital innovation, food data quality and utility are key. Yet our integrative review of 101 FCDBs revealed that global efforts often have inadequate coverage of both foods and food components in the world's food supply. We found a skewed geographic distribution, with North America, Europe, and Asia having the highest representation with more than 15 databases per continent. Even where FCDBs are plentiful there is still an opportunity to improve the coverage of both foods and food components. Our search revealed, on average, FCBDs contained 2,523 food samples and 1,206 food components; however, only 38 components (i.e., proximate composition, select minerals and vitamins) were found to be common among at least 10 databases. To fill in gaps in both the number of foods and components, FCDBs often recycle secondary data from existing databases; we found 85% of databases used at least some secondary data. Food composition and the composition of diets can evolve over time due to environmental, economic, and social dynamics. Yet, 39% of FCDBs have yet to be updated, speaking to the opportunity for regions to renew their understanding of their own food supply. However, we recognize that countries' capacity to update their food composition databases is very much dependent upon economic status. For example, only 12 databases that meet the FAIR criteria are web-based FCDBs, and they are maintained by high and upper-middle-income economies. Overall, we recommend global FCDBs work toward expanded coverage of foods and components, unified methods of analysis, and enhanced metadata and FAIR data adherence, all to improve scope and harmonization across FCBDs globally. Through our assessment of the profile of food composition databases, we found emergent opportunities to improve the quality and usefulness of FCDB content and propose the following key recommendations for improvement:

### 5.1 Emergent opportunities

(1) Geographic distribution of food composition databases

Through this integrative review of the state of FCDBs, we found that many countries around the world do not produce or maintain a national FCDB containing nutritional information about their locally consumed foods. Further, where national databases exist, we found an absence of primary analytical data, particularly in regions like Central America, East Asia, and across the continent of Africa. The discrepancy in data availability is often correlated with economic classification; typically, high-income countries not only include more primary data but also frequently update their databases and utilize web interfaces. The disparity between high and low-income countries' capacity to generate data on their own food composition data has downstream implications for dietary guidelines, food and agriculture policymaking, and ultimately human and planetary health outcomes. On the continent of Africa, the absence of national FCDBs with primary data further complicates efforts to devise optimal dietary improvement strategies to combat the prevalence of malnutrition and chronic and hidden hunger (18, 59). In Africa, Southeast Asia, and beyond, opportunities exist to apply sustainable food-based approaches to address micronutrient deficiencies through biodiverse dietary recommendations powered by high-quality, primary food composition data (60–62).

(2) Prevalence of primary versus secondary data

The reliance on secondary data, where primary data are unavailable, poses considerable risk. Secondary data do not accurately reflect the current local food supply in coverage of foods nor in food components. Primary data compiled into secondary databases from other geographies should not be used as a one-to-one swap (37), especially in the absence of metadata and FAIR data standards of interoperability. Further, the use of secondary data often leads to instances where data do not reflect recent advances in analytical methods, plant breeding or agricultural advancements, or changes in food processing methods (10, 63, 64). Data inaccuracies are also propagated when FCDB data are used to describe composite meals. Ingredient substitutions with foods bearing similar common names or with foods grown in different geographies often present a multitude of confounders that lead to dietary assessment errors in human nutrition studies (20). The information obtained from current dietary assessment tools carries an inherent bias that is rooted in their retrospective nature. This bias is further amplified by the inaccurate compositional analyses of the habitual diets of individuals which encumbers the understanding of robust diet-health associations (65).

Additionally, in the era of digital innovation, AI tools using large language models trained on recycled secondary data will undoubtedly result in misleading conclusions termed artificial hallucinations (11, 66), particularly when certain geographies are overrepresented in the data. Knowing that only 15% of databases were solely powered by primary food composition data strongly points to the need for democratized tools to generate primary data to support AI applications.

(3) Number of food components and coverage

Since the 1990s, food biomolecular diversity among databases has increased slightly, but most databases are still limited to under 100 food components measured. Our findings demonstrate that FCDBs report on average 1,206 food components (ranging from 6 to 70,926). However, 90% of databases reported fewer than 550 components with known bioactivity. Moreover, the reality is that only 38 components were found to be common among 10 databases with proximate composition, minerals, and vitamins dominating the landscape. We observed a slight trend deviating from these few components since the 1990s, but there is much work to be done to further uncover the nutritional dark matter of food, particularly when the chemical complexity in our diet ranges from an estimated 26,000 to 49,000 distinct biomolecules (67).

Two noteworthy outliers stood out among the rest in addressing this unknown chemical space. The Bovine Milk Protein Database and FooDB with 10,642 compounds and 70,926 compounds, respectively, are unique in the sheer number of compounds reported although secondary data was included. The PTFI was notable for reporting a large number of compounds (i.e., 24,721 compounds) generated as high-confidence, primary data. This is a welcome development in the FCDB space since, through our review, comparability of specific compounds beyond proximate composition, nutrients, and vitamins was an insurmountable limitation. Biomolecular diversity stands out as a major limitation of data comparability.

These challenges highlight the urgent need for improved standardization and reproducibility in generating primary data. Beyond the most commonly measured nutrients, there are limited globally accepted standardized methods to evaluate food's diversity of bioactive molecules. Additionally, analytical limitations can arise from the need for diverse instrumentation tailored to each type of biomolecule. The complexity of the food matrix drives accessibility challenges of costly, time-consuming, and low-throughput extraction and analytical methods (1). To address these challenges, international food composition databases and ontologies are emerging, establishing improved data standards, particularly for food components.

(4) Number of food samples and coverage

Our search revealed, on average, FCBDs contained just 2,576 food samples underscoring the opportunity to improve food coverage in databases globally. We identified some outlier databases such as OQALI which includes a large number of food samples derived from secondary data. By contrast, the Malaysian Food Composition Database, reported only self-generated, primary data on 1,892 foods. Overall, databases with primary data average only 488 food samples. Food samples in our review are defined as food items. Food items include whole, raw foods, and minimally processed foods, but also multi-ingredient meals and processed and packaged food items, etc. A large count of food samples is not necessarily indicative of edible species diversity, and in most cases, food species-specific metadata were missing to make that determination.

On the topic of food coverage, only 20% of the databases reviewed encompassed all 13 culinary food groups. Yet, we additionally found the food group categorization often lacked standardization. We found ambiguity in classification of certain food classes like mushrooms and insects, with very few databases including metadata to support more accurate food group classification. The best examples of food-specific metadata appear in EuroFIR FoodEXplorer, FAO/INFOODS, and USDA FDC databases.

By our estimation, INFOODs and USDA FDC (i.e., Foundation Foods and Standard Reference Legacy) combined report approximately 119,922 food samples but only 767 diverse species (18, 23, 40). By contrast, the PTFI, when leading a search to determine gaps in species coverage, created an inspiring list of 1,650 species of high priority in need of biomolecular exploration (23). In terms of both the bio- and chemodiversity of food, these initiatives represent a small fraction (i.e., ~5%) of the estimated 35,000 edible plant (22, 24, 68), animal, insect (25, 69), and fungal (70, 71) species worldwide. This leaves 95% of named edible species yet to be explored.

(5) Frequency of update

In addition to improving the breadth of food and nutritional data, measuring food composition consistently and over time can provide a basis for identifying drivers of food quality, such as genetic variation, agricultural practices, climate, food processing and preparation, and consumer preferences (4, 11). Yet, 59 out of the 101 databases and tables we surveyed have not published updated food composition data in the last decade. The update frequency among databases was better than data tables, with a majority (69%) of web-based FCDBs having been updated in the last 5 years. However, the opportunity for more relevant food composition data remains, particularly to keep pace with a rapidly changing climate.

(6) Adherence to FAIR data management and stewardship

Harmonizing food composition data is foundational in ensuring consistency, accuracy, interoperability, and traceability across various food composition databases and sources. Currently, there is no standard for assessing the data quality of FCDBs (8), but the *FAIR Guiding Principles for scientific data management and stewardship provide an initial framework* to understand how food data might be structured and utilized within these FCBDs. The FAIR principles also point to a significant need for enhancing the homogenization and comparability of FCDBs. Several challenges related to the findability and accessibility of FCDBs were identified such as broken URLs. Notably, some of the databases surveyed in this integrative review were embedded in food composition data indexes that act as repositories for FCDBs. These embedded data sources were not independently findable through a web search and could only be located by visually scanning food composition data index web pages. Although most databases are publicly accessible, the formats available often hinder effective interaction with the data, as many only provide PDF-based food tables and web-based interfaces lack APIs to facilitate data exchange.

Interoperability is still a critical challenge according to our results underscoring the need for clearer descriptions of analytical methods and scientific nomenclature for food components, which are crucial for enhancing data reliability and comparability. The analytical methodologies used in populating these databases often vary by country or even institution, as does the naming of foods and nutritional components making the comparisons between databases challenging (60). While there has been widespread adoption of FAO/INFOODS tag names and EuroFIR thesauri for food components, facilitating some standardization in language across databases, there remain significant gaps in the standardization of these components across other databases globally (40, 72).

Noteworthy, most FCDBs adequately describe their in-house metadata, which not only supports the potential for reusability but also facilitates deeper analysis and broader application in diverse research contexts. Efforts to harmonize procedures for better data comparability and interchange, such as the FCDBs of FAO/INFOODS and EuroFIR FoodEXplorer and independent ontologies like the Food Ontology (FoodON) have aimed to address these challenges (35). However, wider adoption among other FCDBs has been slow, hindering the effectiveness of data interoperability (11). Furthermore, the description of usage rights attached to the data and metadata often remain vague, which can significantly restrict the potential for reusing and sharing data across studies and applications.

The FAIRness of FCDBs is crucial for analyzing and comparing data from different databases considering the criteria used in each country, the diverse objectives pursued by each project, and the intended users (11, 73). Studies from the late 1990s suggest that nutrient intake estimations from the same diet can vary by as much as 20–45%, depending on the database used, owing to systematic and random errors that include discrepancies in naming, terminology, and calculation methods (74). It is worth noting that the FAIRness of the analyzed databases is highly related to the income classification of the country that developed them. This underscores the need for greater support, resources, and guidelines to ensure more consistent and accurate comparisons across databases globally.

## 6 Recommendations

The depth of our understanding of food composition significantly influences our ability to develop sustainable diets and improve nutritional outcomes. Recognizing the various challenges involved in the collection and dissemination of quality, standardized, and well-organized food composition data, credit is given to the significant efforts that have been made. These efforts complement existing strategies aimed at enhancing dietary quality. Overall, we recommend the following actions for advancing the utility of food composition data for diverse types of users. We recommend efforts be made in five key areas:

(1) Broaden database content and frequency of update through globally coordinated and place-based approaches

To accurately profile the vast diversity of modern diets and the global edible biodiversity more broadly, it is essential to expand the range of foods and components included in food composition databases, including those that reflect diverse cultures, agricultural practices, and geographies. Ensuring regular updates is crucial for maintaining the accuracy and relevance of the data, particularly in a changing climate and with changes in land use, agriculture, and food systems. Advancements in technology such as foodomics approaches further warrant the continuous update of food composition databases as new methodologies are developed to more precisely profile food composition. While resources remain a constraint in profiling a wide range of foods and biomolecules in FCDBs as well as their update frequency, a globally coordinated approach among countries would support economies of scale and enable countries to benefit from learnings globally. Such a globally coordinated approach should include place-based efforts representative of local food systems including underutilized crops as well as novel crops currently under development. Further, web-based platforms, known for their flexibility and ease of access, can significantly facilitate these updates and allow for monitoring of shifts in food component data over time.

(2) Harmonize data and enhance quality using standardized approaches and metadata

To achieve a comprehensive and cohesive approach to food data collection, we recommend a minimum, globally agreed upon, set of food components generated using standardized methodologies for inclusion in FCDBs to enhance interoperability globally and provide evidence on the world's food supply at scale. Standardized methods, including foodomics approaches (1, 75) to comprehensively profile food components using both targeted methods and untargeted methods (i.e., techniques to measure unknown compounds with relative quantitation), is essential to expand our understanding of the vast, unknown "dark matter" of food. Complementing these untargeted analytical techniques with relative quantification of compounds of importance for human and planetary health can further enhance the functionality of these data. Yet even more fundamentally, to add context to food data, we recommend the inclusion of metadata. Accurate descriptions of analytical methods through the use of metadata and data dictionaries, including nomenclature for foods and components, will be a hallmark of this next era of digital innovation.

(3) Incorporate FAIR principles and ethical governance and stewardship

Incorporating Findable, Accessible, Interoperable, and Reusable (FAIR) principles within FCDBs will enhance their utility. This involves improving the findability and accessibility of FCDBs by maintaining functional links and incorporating APIs and machine-readable formats, which simplifies data integration and usage. Furthermore, implementing clear and standardized usage licenses is essential to promote the reusability of FCDB data across various studies and applications, ensuring that data usage rights are well-defined and communicated. Ultimately, leveraging existing ontologies and building out new food systems-focused thesauri will enhance the interoperability of food composition data in this new era of digital innovation (76, 77).

Beyond data-centric FAIR principles, there is an emerging awareness of data governance, stewardship, and ethics globally. Because food composition data and associated metadata are effectively digital sequence information (DSI), food composition repositories should likewise be tasked with adhering to access and benefits sharing modalities governing the use of other data derived from genetic resources. In light of the outcome of negotiations at the United Nations Conference of the Parties Convention on Biological Diversity (CDB COP16), databases compilers, and by extension FCDBs, have a call to action to infuse FAIR, CARE, and TRUST principles into their data governance policies. CARE, or the *CARE Principles for Indigenous Data Governance,* is an acronym meaning Collective Benefit, Authority to Control, Responsibility, and Ethics (78). CARE principles promote data sovereignty ensuring responsible data collection, accreditation, and equitable data reuse. Whereas TRUST is an acronym for Transparency, Responsibility, User-focus, Sustainability, and Technology (79). TRUST principles foster sustainable governance of digital repositories by promoting reliable and secure infrastructure over the long term. Integration of the complementary principles FAIR, CARE, and TRUST will encourage database managers to honor both people and purpose in the stewardship of food data. Food composition databases potentially contain vast amounts of digital Indigenous data and traditional knowledge, and therefore, have a responsibility to steward these datasets ensuring that the data is safeguarded from historical inequities.

(4) Strengthening capacity for generating and applying food composition data across food systems

While food composition data have historically been utilized for nutritional assessments and by nutritionists and dieticians, there are increasing opportunities to apply food composition data across food systems by diverse stakeholders including farmers, producers, crop breeders, and agricultural researchers, but also food scientists, food manufacturers, chefs, consumers, and other diverse users. For example, for these diverse users to know how to apply food data, there is a need for capacity strengthening. In addition, as novel approaches for generating food data such as foodomics provide emerging opportunities, there is a need to provide capacity strengthening to scientists globally on utilizing these novel technologies. Capacity strengthening in the form of technology transfer should not only bolster the technical aspects of FCDBs but also enhance their applicability and use in diverse cultural contexts.

(5) Prioritize investment to develop and maintaining FCDBs

High-quality FCDBs require notable resources. There is a pronounced need for increased investment to support the creation of accurate, accessible, and culturally relevant FCDBs, particularly in countries with limited resources and capacities. As food data are beneficials for those across food and health systems, there is a need to prioritize investment in FCDB infrastructure for diverse users across food and health sectors. Likewise, such efforts should be equitable globally. In an increasingly globalized food system, it is essential for high-income countries to support FCDB efforts in low-and middle-income countries from where they often procure food.

## 7 Conclusion

Food composition data are essential for informing solutions and decision making to today's human and planetary health crises, including biodiversity loss, food insecurity, and diet-related chronic diseases. Despite the critical role of FCDBs, this integrative review reveals a significant opportunity to improve the coverage, structure, and comparability of data on food components. Many FCDBs include data on only a few foods and components, with a small subset consistently reported across databases. In addition, there is a high level of reliance on secondary data and a widespread use of static tables for representing the data. These challenges underscore major gaps in the availability of robust and updated nutritional data, limiting the relevance of these databases in specific cultural and geographic contexts.

Data stewardship guidelines, like the FAIR data principles, demonstrate areas for improvement and progress. While all FCDBs meet some criteria, such as Findability, only a few fully adhere to all FAIR principles, with a clear need to improve machine-readability and data reusability. Notably, high-income countries are frequent adopters of web-based interfaces, frequently updated platforms, and FAIR principles compared to middle-low-income countries. Encouragingly, some efforts have arisen to address these challenges, resulting in several international food composition databases with improved data standards, especially for food components. FCDBs like FoodData Central and FoodDB have set high standards for data quality and breadth, respectively, while newer projects, such as the PTFI, contribute innovative analytical approaches, meta-data and data harmonization.

To overcome the current challenges in FCDBs, we recommend: (i) broadening the coverage of foods and bioactive compounds included in FCDBs to better represent global dietary diversity, (ii) establishing standardized methods for data generation, curation, and reporting to enhance interoperability, (iii) comprehensively implementing FAIR principles, including higher resolution metadata, to improve data accessibility and usability, and (iv) increasing investments in capacity building and technological infrastructure, particularly in resource-limited regions. Strengthening FCDBs through these strategies will significantly enhance their utility for policymakers, researchers, and practitioners. This advancement will support the development of evidence-based nutritional profiling and guidelines, foster biodiversity conservation, and contribute to more sustainable diets and equitable food systems that promote human and planetary health.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

## Author contributions

SB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. JG-F: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. NV-M: Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review & editing. JC: Data curation, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review & editing. NQ: Writing – review & editing, Data curation, Investigation, Methodology, Validation. ST: Data curation, Investigation, Methodology, Writing – review & editing. MO: Formal analysis, Methodology, Visualization, Writing – review & editing. EJ: Data curation, Investigation, Validation, Writing – review & editing. M-AL: Methodology, Supervision, Writing – review & editing. HL: Data curation, Methodology, Supervision, Writing – review & editing. RA: Data curation, Methodology, Supervision, Writing – review & editing. MB: Data curation, Writing – review & editing. EA: Data curation, Methodology, Writing – review & editing. WS: Data curation, Methodology, Supervision, Writing – review & editing. AJ-B: Supervision, Writing – review & editing. GK: Writing – review & editing. JB: Data curation, Formal analysis, Investigation, Writing – review & editing. JEP: Funding acquisition, Supervision, Writing – review & editing. MR: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. JP: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. SA: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. We are grateful for the support of The Periodic Table of Food Initiative (PTFI) by The Rockefeller Foundation, The Foundation for Food and Agriculture Research, Seerave Foundation, Fourfold Foundation, Atria Health Collaborative, and the Bill & Melinda Gates Foundation. The content, findings and conclusions presented are those of the authors and do not necessarily reflect the official views, positions or policies of the funders or the institutions with which the funders or the authors are affiliated.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2025.1552367/full#supplementary-material

## References

1. Ahmed S, de la Parra J, Elouafi I, German B, Jarvis A, Lal V, et al. Foodomics: a data-driven approach to revolutionize nutrition and sustainable diets. *Front Nutr*. (2022) 9:874312. doi: 10.3389/fnut.2022.874312

2. Smith P, Arneth A, Barnes DKA, Ichii K, Marquet PA, Popp A, et al. How do we best synergize climate mitigation actions to co-benefit biodiversity? *Glob Chang Biol*. (2022) 28:2555–77. doi: 10.1111/gcb.16056

3. Watts N, Amann M, Arnell N, Ayeb-Karlsson S, Beagley J, Belesova K, et al. The 2020 report of the lancet countdown on health and climate change: responding to converging crises. *Lancet*. (2021) 397:129–70. doi: 10.1016/S0140-6736(20)32290-X

4. Delgado A., Issaoui M., Vieira M. C., Carvalho I. S.De, Fardet A. (2021). Food composition databases: does it matter to human health? *Nutrients*, 13:816. doi: 10.3390/nu13082816

5. Mingay E, Hart M, Yoong S, Hure A. Why we eat the way we do: a call to consider food culture in public health initiatives. *Int J Environ Res Public Health*. (2021) 18:11967. doi: 10.3390/ijerph182211967

6. Muscaritoli M. The impact of nutrients on mental health and well-being: insights from the literature. *Front Nutr*. (2021) 8:656290. doi: 10.3389/fnut.2021.656290

7. McKillop K, Harnly J, Pehrsson PR, Fukagawa NK, Finley J. FoodData central, USDA's updated approach to food composition data systems. *Curr Dev Nutr*. (2021) 5:596. doi: 10.1093/cdn/nzab044_027

8. Li Z, Forester S, Jennings-Dobbs E, Heber D. Perspective: a comprehensive evaluation of data quality in nutrient databases. *Adv Nutr*. (2023) 14:379–91. doi: 10.1016/j.advnut.2023.02.005

9. Atwater WO, Woods CD, Bryant AP. The chemical composition of American food materials. United States: U. S. Government Printing Office (1906).

10. Fukagawa NK, McKillop K, Pehrsson PR, Moshfegh A, Harnly J, Finley J. USDA's FoodData central: what is it and why is it needed today? *Am J Clin Nutr*. (2022) 115:619–24. doi: 10.1093/ajcn/nqab397

11. Ferraz de Arruda H, Aleta A, Moreno Y. Food composition databases in the era of big data: vegetable oils as a case study. *Front Nutr*. (2023) 9:1052934. doi: 10.3389/fnut.2022.1052934

12. Capozzi F, Bordoni A. Foodomics: a new comprehensive approach to food and nutrition. *Genes Nutr*. (2013) 8:1–4. doi: 10.1007/s12263-012-0310-x

13. Rodríguez-Carrasco Y. Foodomics: current and future perspectives in food analysis. *Food Secur*. (2022) 11:1238. doi: 10.3390/foods11091238

14. Castanheira I, André C, Oseredczuk M, Ireland J, Owen L, Robb P, et al. Improving data quality in food composition databanks: a EuroFIR contribution. *Accred Qual Assur*. (2007) 12:117–25. doi: 10.1007/s00769-006-0225-6

15. Charrondière R, Stadlmayr B, Rittenschober D, Mouille B, Nilsson E, Medhammar E, et al. FAO/INFOODS food composition database for biodiversity. *Food Chem*. (2013) 140:408–12. doi: 10.1016/j.foodchem.2012.08.049

16. Møller A. *Danish food Compositics home page*. Danish Food Informatics. (2010). Available online at: https://www.danfood.info/links.asp.

17. Westenbrink S, Presser K, Roe M, Ireland J, Finglas P. Documentation of aggregated/compiled values in food composition databases; EuroFIR default to improve harmonization. *J Food Compos Anal*. (2021) 101:103968. doi: 10.1016/j.jfca.2021.103968

18. USDA ARS. *FoodData Central*. (2019). Available online at: https://fdc.nal.usda.gov/.

19. Haytowitz DB, Pehrsson PR. USDA's National Food and nutrient analysis program (NFNAP) produces high-quality data for USDA food composition databases: two decades of collaboration. *Food Chem*. (2018) 238:134–8. doi: 10.1016/j.foodchem.2016.11.082

20. Lozano CP, Aveiro KH, Wilkens LR, Martin CK, Apolzan JW, Novotny R. A food and nutrient database for dietary studies (FNDDS) tailored to Hawaii may reduce dietary assessment error. *Curr Dev Nutr*. (2024) 8:103157. doi: 10.1016/j.cdnut.2024.103157

21. Martinez-Victoria E. Ingesta de energía y nutrientes; armonización de las bases de datos de composición de alimentos. *Nutr Hosp*. (2015) 3:168–76. doi: 10.3305/nh.2015.31.sup3.8764

22. Diazgranados M., Allkin B., Black N., Cámara-Leret R., Canteiro C., Carretero J., et al. *World checklist of useful plant species*. Produced by the Royal Botanic Gardens, Kew. Knowledge Network for Biocomplexity; KNB Data Repository. (2020).

23. Jarvis A, Gallo-Franco J, Portilla J, German B, Debouck D, Rajasekharan M, et al. Periodic table of food initiative for generating biomolecular knowledge of edible biodiversity. *Nat Food*. (2024) 5:189–93. doi: 10.1038/s43016-024-00941-y

24. Ulian T, Diazgranados M, Pironon S, Padulosi S, Liu U, Davies L, et al. Unlocking plant resources to support food security and promote sustainable agriculture. *Plants People Planet*. (2020) 2:421–45. doi: 10.1002/ppp3.10145

25. Jongema Y. *Worldwide list of recorded edible insects. List of edible insects of the world (April 1, 2017)*. (2017). Available online at: https://www.wur.nl/en/research-results/chair-groups/plant-sciences/laboratory-of-entomology/edible-insects/worldwide-species-list.htm.

26. Krongdang S, Phokasem P, Venkatachalam K, Charoenphun N. Edible insects in Thailand: an overview of status, properties, processing, and utilization in the food industry. *Food Secur*. (2023) 12:162. doi: 10.3390/foods12112162

27. Smith MR, Stull VJ, Patz JA, Myers SS. Nutritional and environmental benefits of increasing insect consumption in Africa and Asia. *Environ Res Lett*. (2021) 16:065001. doi: 10.1088/1748-9326/abf06c

28. Alegbejo JO. Nutritional value and utilization of *Amaranthus* (*Amaranthus* spp.) – a review. Bayero. *J Pure Appl Sci*. (2013) 6:136. doi: 10.4314/bajopas.v6i1.27

29. López-Romero P, Pichardo-Ontiveros E, Avila-Nava A, Vázquez-Manjarrez N, Tovar AR, Pedraza-Chaverri J, et al. The effect of Nopal (Opuntia Ficus Indica) on postprandial blood glucose, Incretins, and antioxidant activity in Mexican patients with type 2 diabetes after consumption of two different composition breakfasts. *J Acad Nutr Diet*. (2014) 114:1811–8. doi: 10.1016/j.jand.2014.06.352

30. Torres-Mora MA, Rubio-Cruz MA, Trujillo-González JM. Aproximación de la importancia socio-cultural de la palmera Moriche (*Mauritia fleuxuosa* L. F.) en la comunidad indígena Wacoyo (Sikuani) en el municipio de Puerto Gaitán, Colombia. *Orinoquia*. (2015) 19:231–6. doi: 10.22579/20112629.337

31. Herforth AW, Ballard T, Rzepa A. Development of the diet quality questionnaire for measurement of dietary diversity and other diet quality indicators. *Curr Dev Nutr*. (2024) 8:103798. doi: 10.1016/j.cdnut.2024.103798

32. Kapsokefalou M, Roe M, Turrini A, Costa HS, Martinez-Victoria E, Marletta L, et al. Food composition at present: new challenges. *Nutrients*. (2019) 11:1714. doi: 10.3390/nu11081714

33. Ahmed S, Brinkley S, Smith E, Sela A, Theisen M, Thibodeau C, et al. Climate change and coffee quality: systematic review on the effects of environmental and management variation on secondary metabolites and sensory attributes of *Coffea arabica* and *Coffea canephora*. *Front Plant Sci*. (2021) 12:13. doi: 10.3389/fpls.2021.708013

34. Wilkinson MD, Dumontier M, Appleton G, Axton M, Baak A, Blomberg N, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. (2016) 3:160018. doi: 10.1038/sdata.2016.18

35. Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ Sci. Food*. (2018) 2:23. doi: 10.1038/s41538-018-0032-6

36. Brinkley S., Franco J. G., Laporte M. A., Rajasekharan M., Chien C.-M., de la Parra J., et al. *The periodic table of food: Metadata harmonization from point-of-production to plate*. ASA, CSSA, SSSA International Annual Meeting. (2023). Available online at: https://scisoc.confex.com/scisoc/2023am/meetingapp.cgi/Paper/151437.

37. Marconi S, Durazzo A, Camilli E, Lisciani S, Gabrielli P, Aguzzi A, et al. Food composition databases: considerations about complex food matrices. *Food Secur*. (2018) 7:2. doi: 10.3390/foods7010002

38. Moher D, Liberati A, Tetzlaff J, Altman DGPRISMA Group. Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement. *PLoS Med*. (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097

39. Toronto CE. Overview of the integrative review. A step-by-step guide to conducting an integrative review. Cham: Springer, pp. 1–9. (2020).

40. INFOODS. *FAO/INFOODS Databases*. (2019). Available online at: https://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases/en/.

41. World Nutrient Databases for Dietary Studies (WNDDS). *Agriculture and Food Systems Institute*. (n.d.). Available online at: https://foodsystems.org/resources/wndds/ (Accessed July 3, 2024).

42. World Bank Country and Lending Groups. (2025). Available online at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

43. Wayback Machine. (n.d.). Available online at: https://web.archive.org/ (Accessed November 13, 2024).

44. Biodiversity for Food and Nutrition. *Species database: biodiversity for food and nutrition*. (2019). Available online at: http://www.b4fn.org/resources/species-database/.

45. PTFI. *Research hub – Research community and resources for the periodic table of food initiative*. (n.d.). Available online at: https://ptfi.versobio.com/ (September 14, 2024).

46. Ministry of Education, Culture, Sports, Science and Technology (MEXT) -Japan. *MEXT: STANDARD TABLES OF FOOD COMPOSITION IN JAPAN - 2015—(seventh revised version)*. (2015). Available online at: https://www.mext.go.jp/en/policy/science_technology/policy/title01/detail01/1374030.htm.

47. Hoxha L., Giertlová A., Bartošová L., Kongoli R., Lika E., Koto R., et al. *First edition of the Albanian food composition tables*. The 14th European Nutrition Conference FENS 2023, p. 269. (2024).

48. Das A, Giri K, Behera RN, Maity S, Ambatipudi K. BoMiProt 2.0: an update of the bovine milk protein database. *J Proteome*. (2022) 267:104696. doi: 10.1016/j.jprot.2022.104696

49. Foroutan A, Guo AC, Vazquez-Fresno R, Lipfert M, Zhang L, Zheng J, et al. The chemical composition of commercial Cow's Milk. *J Agric Food Chem*. (2019) 67:4897–914. doi: 10.1021/acs.jafc.9b00204

50. OQALI. *French food observatory*. (n.d.). Available online at: https://www.oqali.fr/en/home/

51. Nutrition Coordinating Center (NCC), University of Minnesota. *NCC Food and Nutrient Database*. (2024). Available online at: http://www.ncc.umn.edu/ndsr-database-page/.

52. European Food Safety Authority (EFSA). *Food composition data|EFSA*. (2021). Available online at: https://www.efsa.europa.eu/en/data-report/food-composition-data.

53. German Nutrient Database–Bundeslebensmittelschlüssel. *Bundeslebensmittelschlüssel*. (n.d.). Available online at: https://www.blsdb.de/.

54. Dias MG, Borge GIA, Kljak K, Mandić AI, Mapelli-Brahm P, Olmedilla-Alonso B, et al. European database of carotenoid levels in foods. Factors affecting carotenoid content. *Foods*. (2021) 10:912. doi: 10.3390/foods10050912

55. Vlieg P. *Proximate composition of New Zealand marine finfish and shellfish*. Biotechnology Division. (1988). Available online at: https://www.fao.org/4/ae581e/ae581e00.htm

56. FooDB. *FooDB*. (n.d.). Available online at: https://foodb.ca/.

57. Bento da Silva A., Giacomoni F., Fillatre Y., Rothwell J. A., Knox C., Roman E., et al. *PhytoHub V1.4: A new release for the online database dedicated to food phytochemicals and their human metabolites*. HAL Open Science. The 1st International Conference on Food Bioactives and Health, Norwich, United Kingdom. (2016).

58. Plumb J, Pigat S, Bompola F, Cushen M, Pinchen H, Nørby E, et al. eBASIS (bioactive substances in food information systems) and bioactive intakes: major updates of the bioactive compound composition and beneficial bioeffects database and the development of a probabilistic model to assess intakes in Europe. *Nutrients*. (2017) 9:320. doi: 10.3390/nu9040320

59. Ene-Obong H, Schönfeldt HC, Campaore E, Kimani A, Mwaisaka R, Vincent A, et al. Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. *Proc Nutr Soc*. (2019) 78:496–505. doi: 10.1017/S0029665118002926

60. Al-Balushi B, Al-Balushi R, Waly M, Al-Attabi Z. Global status of food composition database: a short review. *Int J Nutr Pharmacol Neurol Dis*. (2023) 13:240–2. doi: 10.4103/ijnpnd.ijnpnd_48_23

61. Hulshof P, Doets E, Seyha S, Bunthang T, Vonglokham M, Kounnavong S, et al. Food composition tables in Southeast Asia: the contribution of the SMILING project. *Matern Child Health J*. (2019) 23:46–54. doi: 10.1007/s10995-018-2528-8

62. IRD. *Sustainable micronutrient interventions to control deficiencies and improve nutritional status and general health in Asia (SMILING) initiative*. (2020). Available online at: https://nutrition-smiling.eu/.

63. Ispirova G, Eftimov T, Seljak BK. Evaluating missing value imputation methods for food composition databases. *Food Chem Toxicol*. (2020) 141:111368. doi: 10.1016/j.fct.2020.111368

64. Van Puyvelde H, Perez-Cornago A, Casagrande C, Nicolas G, Versele V, Skeie G, et al. Comparing calculated nutrient intakes using different food composition databases: results from the European prospective investigation into Cancer and nutrition (EPIC) cohort. *Nutrients*. (2020) 12:2906. doi: 10.3390/nu12102906

65. Cuparencu C, Bulmuş-Tüccar T, Stanstrup J, La Barbera G, Roager HM, Dragsted LO. Towards nutrition with precision: unlocking biomarkers as dietary assessment tools. *Nat Metab*. (2024) 6:1438–53. doi: 10.1038/s42255-024-01067-y

66. Bailey RL, MacFarlane AJ, Field MS, Tagkopoulos I, Baranzini SE, Edwards KM, et al. Artificial intelligence in food and nutrition evidence: the challenges and opportunities. *PNAS Nexus*. (2024) 3:461. doi: 10.1093/pnasnexus/pgae461

67. Barabási A-L, Menichetti G, Loscalzo J. The unmapped chemical complexity of our diet. *Nat Food*. (2020) 1:33–7. doi: 10.1038/s43016-019-0005-1

68. Marrelli M, Statti G, Conforti F. A review of biologically active natural products from Mediterranean wild edible plants: benefits in the treatment of obesity and its related disorders. *Molecules*. (2020) 25:649. doi: 10.3390/molecules25030649

69. Van Itterbeeck J, Pelozuelo L. How many edible insect species are there? A not so simple question. *Diversity*. (2022) 14:143. doi: 10.3390/d14020143

70. Li H, Tian Y, Menolli N Jr, Ye L, Karunarathna SC, Perez-Moreno J, et al. Reviewing the world's edible mushroom species: a new evidence-based classification system. *Compr Rev Food Sci Food Saf*. (2021) 20:1982–2014. doi: 10.1111/1541-4337.12708

71. Pérez-Moreno J, Mortimer PE, Xu J, Karunarathna SC, Li H, Pérez-Moreno J, et al. Global perspectives on the ecological, cultural and socioeconomic relevance of wild edible fungi. *Stud Fungi*. (2021) 6:408–24. doi: 10.5943/sif/6/1/31

72. European Food Information Resources (EuroFIR). *EuroFIR food information*. (2009). Available online at: https://www.eurofir.org/our-tools/.

73. Slimani N, Deharveng G, Unwin I, Southgate DAT, Vignat J, Skeie G, et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr*. (2007) 61:1037–56. doi: 10.1038/sj.ejcn.1602679

74. Charrondière UR, Vignat J, Riboli E. Comparable nutrient intake across countries is only possible through standardization of existing food composition tables (FCT). Nutrition and Lifestyle: Opportunities for Cancer Prevention. Lyon: IARC Press (2002). 45–48.

75. Álvarez-Rivera G., Valdés A., León C., Cifuentes A. (2021). *Foodomics – fundamentals, state of the art and future trends*. Foodomics.

76. Valenčič E, Beckett EL, Bucher T, Collins CE, Korousic Seljak B. NutriBase - management system for the integration and interoperability of food-and nutrition-related data and knowledge. *Front Nutr*. (2024) 11:389. doi: 10.3389/fnut.2024.1503389

77. Youn J, Li F, Simmons G, Kim S, Tagkopoulos I. FoodAtlas: automated knowledge extraction of food and chemicals from literature. *Comput Biol Med*. (2024) 181:109072. doi: 10.1016/j.compbiomed.2024.109072

78. Carroll S, Garba I, Figueroa-Rodriguez O, Holbrook J, Lovett R, Materechera S, et al. The CARE principles for indigenous data governance. *Data Sci J*. (2020) 19:43. doi: 10.5334/dsj-2020-043

79. Lin D, Crabtree J, Dillo I, Downs RR, Edmunds R, Giaretta D, et al. The TRUST principles for digital repositories. *Sci Data*. (2020) 7:144. doi: 10.1038/s41597-020-0486-7

80. Hotz C., Abdelrahman L., Sison C., Moursi M., Loechl C. *A food composition table for central and eastern Uganda*. (2012). Available online at: https://assets.publishing.service.gov.uk/media/57a08a90e5274a27b200066d/Tech_Mono_9_Web_0.pdf.

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 06 June 2025
DOI 10.3389/fnut.2025.1519401

Check for updates

OPEN ACCESS

EDITED BY
Lorenza Mistura,
Council for Agricultural Research and Agricultural Economy Analysis (CREA), Italy

REVIEWED BY
Ezequiel Pinto,
University of Algarve, Portugal
Sandra Crispim,
Federal University of Paraná, Brazil

*CORRESPONDENCE
Laura A. Bardon
✉ laura.bardon@quadram.ac.uk

RECEIVED 29 October 2024
ACCEPTED 16 May 2025
PUBLISHED 06 June 2025

CITATION
Bardon LA, Bennett G, Weech M, Hwang F, Kelly EFA, Lovegrove JA, Panov P, Astley S, Finglas P and Gibney ER (2025) The development and evaluation of a quality assessment framework for reuse of dietary intake data: an FNS-Cloud study.
*Front. Nutr.* 12:1519401.
doi: 10.3389/fnut.2025.1519401



# The development and evaluation of a quality assessment framework for reuse of dietary intake data: an FNS-Cloud study

Laura A. Bardon[1,2,3]*, Grace Bennett[1,2,3], Michelle Weech[4,5], Faustina Hwang[6], Eve F. A. Kelly[4,5], Julie A. Lovegrove[4,5], Panče Panov[7], Siân Astley[8], Paul Finglas[1,8] and Eileen R. Gibney[2,3]

[1]Food and Nutrition National Bioscience Research Infrastructure, Quadram Institute Bioscience, Norwich, United Kingdom, [2]Institute of Food and Health, University College Dublin (UCD), Dublin, Ireland, [3]School of Agriculture and Food Science, University College Dublin, Dublin, Ireland, [4]Department of Food and Nutritional Science, Hugh Sinclair Unit of Human Nutrition, Institute of Food, Nutrition and Health, University of Reading, Reading, United Kingdom, [5]Department of Food and Nutritional Science, Hugh Sinclair Unit of Human Nutrition, Institute for Cardiovascular and Metabolic Research, University of Reading, Reading, United Kingdom, [6]Biomedical Engineering Section, School of Biological Sciences, University of Reading, Reading, United Kingdom, [7]Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia, [8]EuroFIR AISBL, Brussels, Belgium

A key aim of the FNS-Cloud project (grant agreement no. 863059) was to overcome fragmentation within food, nutrition and health data through development of tools and services facilitating matching and merging of data to promote increased reuse. However, in an era of increasing data reuse, it is imperative that the scientific quality of data analysis is maintained. Whilst it is true that many datasets *can* be reused, questions remain regarding whether they *should* be, thus, there is a need to support researchers making such a decision. This paper describes the development and evaluation of the FNS-Cloud data quality assessment tool for dietary intake datasets. Markers of quality were identified from the literature for dietary intake, lifestyle, demographic, anthropometric, and consumer behavior data at all levels of data generation (data collection, underlying data sources used, dataset management and data analysis). These markers informed the development of a quality assessment framework, which comprised of decision trees and feedback messages relating to each quality parameter. These fed into a report provided to the researcher on completion of the assessment, with considerations to support them in deciding whether the dataset is appropriate for reuse. This quality assessment framework was transformed into an online tool and a user evaluation study undertaken. Participants recruited from three centres ($N$ = 13) were observed and interviewed while using the tool to assess the quality of a dataset they were familiar with. Participants positively rated the assessment format and feedback messages in helping them assess the quality of a dataset. Several participants quoted the tool as being potentially useful in training students and inexperienced researchers in the use of secondary datasets. This quality assessment tool, deployed within FNS-Cloud, is openly accessible to users as one of the first steps in identifying datasets suitable for use in their specific analyses. It is intended to support researchers in their decision-making process of whether previously collected datasets under consideration for reuse are fit their new intended research purposes. While it has been developed and evaluated, further testing and refinement of this resource would improve its applicability to a broader range of users.

KEYWORDS

data reuse, dietary intake data, quality assessment, tool development, lifestyle data, demographic data

## Introduction

Within the field of nutrition research, there is a wealth of existing dietary intake datasets that have been collected within national, regional or targeted sub-population group studies. Fewer studies exist that have been collected across multiple countries or regions. The few pan-European nutritional studies that do exist, including the Food4Me study (1), Feel4Diabetes study (2), EPIC (3) and the Seven Countries study (4), enable deeper analyses to be completed including country-to-country comparisons. Although these analyses are invaluable in nutrition research, they are costly representing a loss of scientific opportunities and waste of time and financial resources. Data reuse and merging of existing datasets can help achieve insights without the same time and expenses but strategies to effectively reuse data need to be considered.

Numerous methodological approaches to the collection and analysis of data exist, making it challenging to merge or compare datasets (5). In more recent years, initiatives such as EUMenu by EFSA have sought to create harmonised data collection approaches across countries facilitating comparison of different datasets or merging of datasets for combined analysis (6). Furthermore, initiatives including FAO/WHO GIFT (7) and the Global Dietary Database (8) are examples of how datasets can be harmonised and accessed for effective reuse by the community. Currently, large amounts of (often publicly funded) money are used to generate big datasets that are usually not exploited for reuse, despite this increasingly becoming a requirement for funding bodies.

FAIR principles (findable, accessible, interoperable and reusable) were established as a guide to support maximal benefits from data, tools, services and algorithms (9). Applying FAIR principles to data is mutually beneficial for both scientific research and society. Recognising this, the European Commission (EC) has established an expert group that aims to turn the concept into reality to open up science and research (10), through the European Open Science Cloud (EOSC), which federates existing European research infrastructures and aims to realise a web of FAIR data and related services, making more data interoperable and machine actionable (11). Making data FAIR is an increasingly important requirement of European funding requirements and is likely to be mandatory in future (with some exceptions), enabling existing datasets to be accessed and reused. These principles were applied in Food Nutrition Security Cloud (FNS-Cloud) (grant agreement no. 863059) underpinning data reuse (9). FNS-Cloud aimed to improve access to datasets, tools and services in the domains of food, nutrition and health, making access more equable across Europe enhancing research capacity through defragmentation of food, nutrition, and security data and the development of tools and services to facilitate matching and merging of data to promote increased data reuse (12).

In this era of increasing data reuse, when using existing, open datasets to answer new research questions, it is important that researchers understand and consider the quality and provenance of data before being reused (13, 14). Challenges exist around dietary intake data due to the variety of methods for collection available, approaches to describe/quantify portion sizes, and underlying composition tables used to generate mean daily intakes; these should be adequately considered before reusing dietary intake data. Several dietary assessment methods exist to collect dietary data at food group or individual food item levels, including food frequency questionnaires or 24-h dietary recalls and food diaries (15, 16). Depending on the method chosen, portion size of foods can be quantified (using actual weights) or estimated (including using portion size pictures, household measures, photographic food atlases or by applying average intakes). There are many food composition datasets available. These can be national composition tables, such as the Composition of Foods Integrated Database (CoFID) for the United Kingdom, or databases for larger regions, such as the EFSA database for Europe (17). Selection of a composition dataset, which is appropriate for the population examined, is essential to ensure the accuracy of resulting data. These challenges, among others, impact the accuracy of resulting data and how it can be used. Dietary intake data has a range of uses including development of food based dietary guidelines, assessment of nutrient deficiencies in populations, and examination of dietary, meal patterns, and food choice in a given population or subgroup (18, 19).

Although development of a prototype Cloud infrastructure through the FNS-Cloud project represents an advancement, and a new direction for food and nutrition science, it is important that data are reused appropriately, to ensure the quality of resulting scientific outputs remain high. When considering the quality of specific datasets, it is important to note this must be in the context of an individual research question. Each user must assess whether the datasets they have selected are appropriate for their research question. This relies on scientific integrity among researchers and appropriate knowledge of datasets prior to reuse. Whilst the onus is, and should remain, on researchers to ensure outputs are based on sound science, there is also a need to support researchers in the decision-making process of whether a dataset is appropriate for their purpose. Within FNS-Cloud, a quality assessment tool acts as a guide for data users to assess whether a dietary intake dataset is suitable to answer their research question, facilitating an informed final decision by the data user. Thus, the aim of this work was to develop and evaluate a quality assessment framework for FNS-Cloud to support researchers in their decision-making process around data reuse, specifically if datasets under consideration are *fit for their intended purpose*.

## Methods

Development of this framework followed processes for developing any quality assessment tool, as described by Whiting et al. (20). This approach consisted of three stages, initial steps (defining scope, identification of parameters of quality), tool development (development of dietary intake dataset quality assessment decision trees, output from decision trees, testing of framework design, transformation of quality assessment framework into an online tool, evaluation of quality assessment tool and contents) and dissemination. An overview of the actions taken within this body of work is summarised in Table 1.

TABLE 1 Overview of the process of developing the quality assessment framework.

| Stage | Quality assessment framework for dietary intake, consumer behavior, lifestyle and demographic data for FNS-Cloud |
|---|---|
| **Stage 1: Initial steps** | |
| 1.1 Identify need | A tool to support the reuse of existing dietary intake datasets |
| 1.2 Obtain funding | This work was conducted within Food Nutrition Security Cloud (FNS-Cloud) (grant agreement no. 863059) |
| 1.3 Assemble team | Larger group FNS-Cloud Consortium (*n* = 35 partner institutions)<br>Working group (*n* = 15 researchers across 7 partner institutions) |
| 1.4 Manage project | Core group (UCD, *n* = 3 researchers) |
| 1.5 Define scope | Appropriateness of reuse of existing dietary intake datasets<br>Domain based flowcharts<br>Questions with defined answer options and personalised messages with considerations |
| **Stage 2: Tool development** | |
| 2.1. Generate items | Targeted literature review of data domains focusing on data collection, data handling, use of underlying data sources, data uses and analysis<br>Summarised parameters of quality<br>Formation of trees |
| 2.2. Agree items | Virtual face-to-face meeting |
| 2.3. Produce first draft | Core group |
| 2.4. Pilot and refine | (1) Paper based feedback from consortium members on main data domain (dietary intake data)<br>(2) Application of paper-based version of the form on *n* = 19 datasets across 2 example research questions<br>(3) External feedback through evaluation activity in 3 centres |
| **Stage 3: Dissemination** | |
| 3.1 Publication | Planned peer review publication of tool development process<br>Entry of tool into FNS Cloud catalogues |
| 3.2 Website | Integration into FNS-Cloud (https://catalogues.fns.foodcase-services.com/catalogues) |
| 3.3 Uptake | Presentation to FNS-Cloud consortium |
| 3.4 Translations | None planned thus far |

## Defining scope

As described, the aim was to develop a quality assessment framework and user-friendly tool to support selection of dietary intake data for reuse, thereby ensuring the quality of outputs is maintained when exploiting data in future research. The core domain of interest for this framework was dietary intake data, however additional FNS-Cloud data domains—including demographic, anthropometric, lifestyle and consumer behavior data—were also included as they are often collected in conjunction with dietary intake data for context. Inclusion of multiple data domains expands the types of research questions that can be answered and, therefore, maximises the scope of food and nutrition data that might be included. For example, links can be made between lifestyle, diet and the development of health conditions; and the food environment can impact consumer behavior and subsequent dietary intake. Other complementary data collected generally encompasses lifestyle, physical activity, and measures of consumer behavior such as purchase, preparation, and consumption.

## Identification of parameters of quality

Firstly, to identify parameters of quality, targeted searches of peer-reviewed literature (including PubMed Central and Web of Science) were performed for each of the domains (dietary intake, lifestyle, anthropometric, demographic, consumer behavior). Searches focused on where quality can be affected during data generation, namely during collection (method of collection chosen, validation, period of collection, days of week, training of data collectors), selection of underlying data sources (portion size quantification, composition databases), how raw data were handled (identification of under/over-reporters, systems used for coding foods), and uses and analysis of data (whether analysing data based on nutrients, foods or food groups). From this review, individual parameters of quality were identified; these were reviewed by researchers with expert knowledge where additional or overlooked parameters of quality were identified.

## Development of dietary intake dataset quality assessment decision tree

Once the parameters were defined, assessment was developed in the form of decision trees. An overview of the structure of the decision trees is presented in Supplementary Figures 1A–C. Decision trees have been used previously in healthcare to support clinical decision making (21, 22) and also in the delivery of personalised nutrition advice (23). The parameters of quality were transformed into questions with structured categorical answer options, e.g., yes, no, do not know. Follow up questions were developed, where necessary, forming branches within the decision tree. Individual branches were developed for each data domain with separate branches also created at different levels where quality can be impacted in the generation of dietary intake data. Each branch concluded with delivery of a personalised message based on the answers selected. The personalised messages give information on the parameter in question, describing why the parameter is important and how it might influence quality of the dataset based on the answer(s) selected.

## Output from decision trees

Following completion of the quality assessment, a user is presented with a personalised feedback report compiling all messages that were produced. The content of these messages varies depending on the relevance to parameters in question but provide the user with considerations to support their decision on whether to use the dataset to answer their research question. Key findings from the literature review of quality parameters for the data domains informed the content of these feedback messages.

## Testing of framework design

A prototype decision tree framework consisting of decision trees illustrated in a powerpoint format was developed and tested in two phases. First, internal testing was conducted at a face-to-face workshop during an FNS-Cloud consortium meeting, attended by ~30 food and health researchers from across the FNS-Cloud partner institutions ($n = 35$ institutions across 12 EU member states, Serbia and Switzerland) in Sardinia, Italy in June 2022, whereby the structure of the dietary intake data branches (questions, response options, and messages) were presented and feedback collected. The consortium comprised a diverse group of nutrition researchers, IT professionals, software developers, and communications specialists. Participating consortium members were asked to review the framework contents and evaluate whether any parameters of quality were missing; the appropriateness of the questions and responses; and, whether the messages were useful for the researcher. Their feedback was used to modify the prototype and develop complementary data branches of the framework. Once fully developed, the paper-based framework was used in case studies by two independent researchers (LAB, MW) to manually assess the quality of existing datasets that had previously been identified to answer two (example) research questions [$N = 11$ datasets assessed for research question 1: "*What are the factors influencing dietary patterns and adherence to sustainable healthy eating guidelines across Europe?*" (LAB) and $N = 8$ datasets assessed for research question 2: *"Does diet quality and dietary intake differ across key adult life stages, and are these influenced by demographic factors, such as European region and sex?"* (MW)]. Researchers answered each question in the framework and created a table of feedback responses the tool generated for each dataset. An informed decision regarding whether each dataset was suitable for reuse to answer the research question was made considering the feedback messages received.

## Transformation of quality assessment framework into online tool

Following feedback and testing, the revised decision trees were transformed from paper-based format into conditional expressions (IF/THEN statements) and a prototype of the online dietary intake data quality assessment tool produced (Figures 1A–C). Then followed an iterative refinement process between two researchers from the core development group (LAB and GB) who identified issues, bugs and glitches in the prototype and the technical team[1] who solved the identified problems. Example research questions were formulated and used to test the accuracy of the workflow. Suggestions for improving the usability of the tool were also shared with the technical team.

## Critical evaluation of the tool and its contents

Following development of the online tool, wider evaluation was conducted among a group of participants with prior research experience in analysing dietary intake data. These evaluations were performed either virtually using Microsoft Teams or in person and were conducted from August 2023 until January 2024 at three centres: University College Dublin (UCD), Ireland (researchers GB, LAB); University of Reading (UoR), United Kingdom (researchers FH, MW); and Quadram Institute Bioscience (QIB), Norwich, United Kingdom (researcher LAB). The researchers undertook targeted recruitment of people with dietary intake domain knowledge within their departments so that they could evaluate the appropriateness of the tool contents and feedback messages. Ethical approval was granted by the UoR research ethics committee (number: 32/2023) and informed written consent was obtained from each participant before interviews were conducted.

Participants were asked to use the tool to perform a quality assessment on a dietary intake dataset that they had prior experience of using. Participants were provided with the URL to the tool and login details, after which they completed independently. Researchers observed them to determine how users navigated the tool and collected comments from participants as they were using the tool. Following completion, these participants were interviewed using an indicative interview script co-developed and agreed previously by all researchers. This guide included a list of questions to ask the participants to assess critical elements of the tool, quality assessment, and feedback messages. All interviews (virtual and in-person) were recorded, with the consent of participants, for later analysis. Basic demographic information including participant sex, career stage, years of experience with dietary intake data, and, self-rated experience and knowledge of dietary intake data were captured through multiple choice questions. Finally, participants were shown and asked a number of open-ended and Likert-item questions to (1) evaluate clarity of the tool's purpose and whether users could navigate the tool easily, (2) verify the tool's contents and assess the relevance and clarity of questions, and (3) gather overall feedback on the tool and its future usability.

1 www.scalefocus.com

To analyse user evaluations of the tool, automated transcripts of the interview videos were generated using Microsoft Teams and subsequently verified using the recordings. One researcher (LAB) reviewed all interview recordings and transcripts multiple times to extract content as well as observational data from interviewer notes. Data were collated for each participant individually, under section headings used during the assessment (for example specific response and reaction to tool content and use of question hints). An inductive analytical approach was applied whereby key phrases discussed by participants were identified (24). An inductive analytical approach allows the content of data to inform emerging patterns. Similar statements of feedback were collated and assigned category labels to determine the frequency with which certain opinions were mentioned by participants. Researchers applied a code frequency approach to determine key themes. All identified categories of feedback were divided into overarching themes of positive elements of the tool or elements requiring future consideration. Participant feedback on the assessment, feedback messages, and overall tool experience was reviewed by two researchers (LAB and GB).

# Results

## Quality framework development

An overview of the parameters of quality identified within each data domain is presented in Supplementary Table 1. $N = 25$ parameters of quality were identified with the majority (60%) being within the dietary intake data domain. Individual decision trees were developed for demographic, dietary intake, consumer behavior, anthropometric and lifestyle data. The dietary intake decision trees contained four levels: "data collection," "data handling/dataset management," "underlying data sources applied," and "uses and analysis." Lifestyle

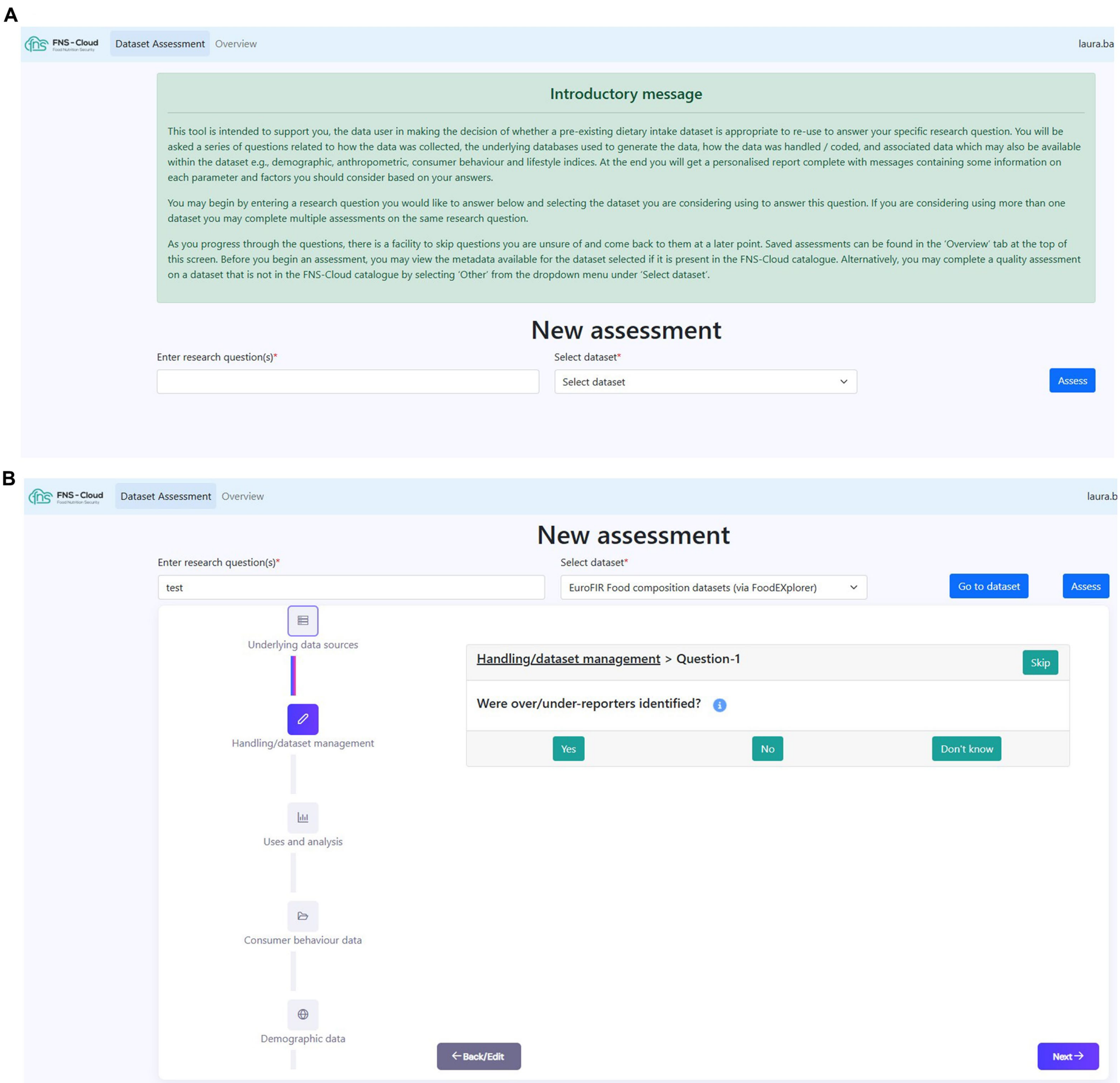



FIGURE 1 (Continued)

C

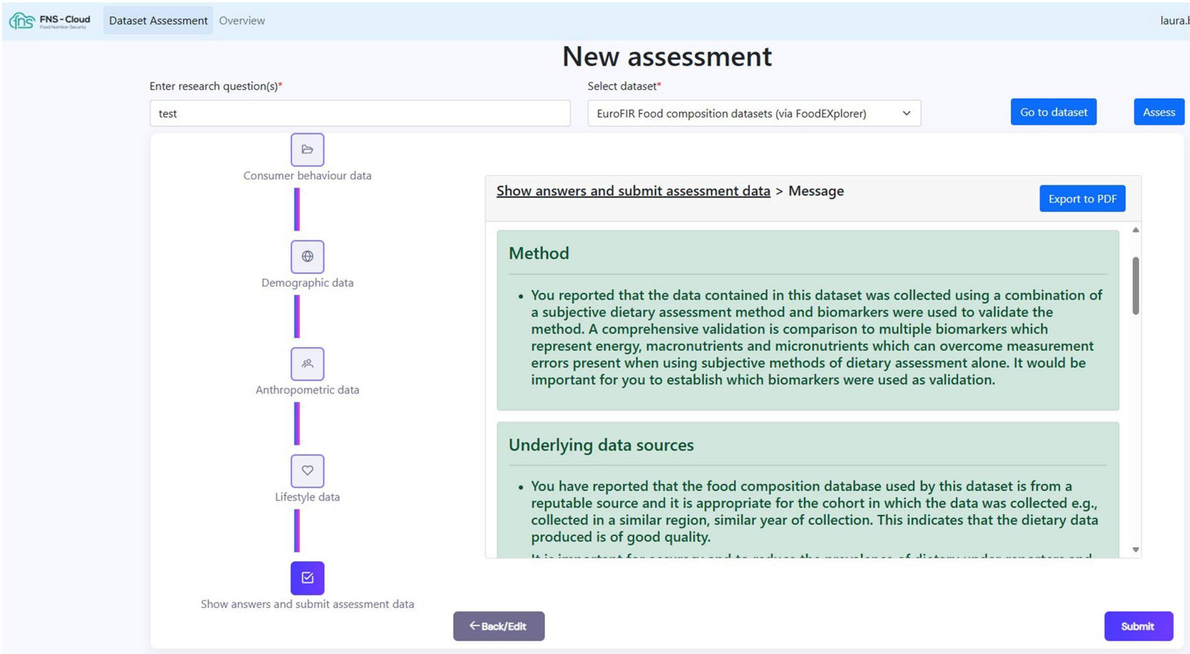



FIGURE 1
Snapshots of developed quality assessment tool. **(A)** Tool introductory page. **(B)** Dataset assessment flow. **(C)** Personalised feedback report.

data was divided into "data collection" and "data handling/dataset management," and consisted of $n = 3$ branches, $n = 4$ distinct questions and $n = 5$ distinct personalised messages. The remaining three data domains (consumer behavior, demographic and anthropometric data) had one branch each with a total of $n = 4$, $n = 5$ and $n = 3$ distinct questions, and $n = 8$, $n = 4$ and $n = 5$ distinct messages, respectively. All feedback messages developed for the decision trees are presented in Supplementary Table 2. Under uses and analysis, a decision was made not to create a decision tree asking about parameters of quality due to the wide range of analytical possibilities. Instead, a generic message was produced describing considerations when using and analysing dietary intake data. Dietary intake data was predominant with $n = 7$ branches, $n = 24$ distinct questions and $n = 37$ distinct personalised messages. All domains were divided into levels where quality can be affected during generation of data. Table 2 provides a breakdown of the branches within each data domain, and the numbers of distinct questions and messages developed for each.

Following creation of the online prototype of the tool (Figures 1A–C), modifications included addition of an introductory message, describing the purpose of the tool for users; user ability to skip questions; user ability to save completed assessments; pop-up help icons to further explain certain terminology within the questions; and, ability to download personalised feedback report after the assessment.

## Evaluation of the tool and contents

A total of $n = 13$ participants ($n = 5$ UCD, $n = 5$ UoR, $n = 3$ QIB) completed the evaluation; the average interview time was 1 h and 9 min and ten participants completed the evaluation virtually via Microsoft Teams. Responses to structured demographic questions are shown in Table 3. Most respondents were female (77%), had been working with dietary intake data for at least four years (69%), and considered themselves to be very experienced with dietary intake data (62%). Participants were from a range of career stages but almost half (46%) were postgraduate students.

Overall, participants rated individual aspects of the tool positively (Figure 2). All participants rated the assessment format as either "somewhat easy" or "very easy" to navigate ($n = 13$, 100%), and the majority felt the information in the personalised feedback report was "somewhat useful" ($n = 6$, 46%) or "very useful" ($n = 6$, 46%) in helping decide if a dataset was appropriate to reuse for their purpose. When rating the messages, the majority rated the contents as "somewhat" or "very appropriate" ($n = 11$, 85%), length as "about right" ($n = 10$, 77%), and the clarity as "clear" or "very clear" ($n = 12$, 92%). All except two participants "agreed" or "strongly agreed" that they would use the tool in their research. The majority of participants ($n = 8$, 62%) rated the user friendliness of the tool as "excellent".

Feedback from the user evaluation was categorized into positive aspects and facets that needed future consideration (Table 4). In general, participants were positive about the tool and its contents. Some believed they would use this tool for future research ($n = 10$, 77%), primarily with datasets they have not collected themselves ($n = 2$, 15%) as this would help identify the strengths and weaknesses of a dataset. Even those who did not feel the tool would be useful for their work, did speak about benefits for students or inexperienced researchers ($n = 5$, 38%). The information asked for during the assessment, especially within the dietary intake data section, was deemed relevant for measuring quality of data and included some information that is often poorly considered in dietary research.

TABLE 2 Overview of numbers of branches, questions and messages within the tool.

| Data domain | Branches | Branches (*n*) | Distinct questions (*n*) | Distinct messages (*n*) |
|---|---|---|---|---|
| Dietary intake | Methods | 7 | 12 | 16 |
| | Underlying data sources A | | 8 | 14 |
| | Underlying data sources B | | | |
| | Underlying data sources C | | | |
| | Underlying data sources D | | | |
| | Handling of data A | | 4 | 7 |
| | Handling of data B | | | |
| Consumer behavior | Methods | 1 | 4 | 6 |
| Demographic | Methods | 1 | 5 | 8 |
| Anthropometric | Methods | 1 | 3 | 4 |
| Lifestyle | Methods A | 3 | 4 | 5 |
| | Methods B | | | |
| | Methods C | | | |

TABLE 3 Demographic characteristics of evaluation study participants.

| | *N* (%) |
|---|---|
| Female sex | 10 (76.9) |
| **Education/career stage** | |
| Postgraduate student | 6 (46.2) |
| Postdoctoral researcher | 3 (23.1) |
| Researcher <5 years | 0 |
| Researcher 5–9 years | 2 (15.4) |
| Researcher >10y years | 2 (15.4) |
| **Years experience with dietary intake data** | |
| <1 year | 0 |
| 1–3 years | 4 (30.8) |
| 4–6 years | 4 (30.8) |
| >6 years | 5 (38.5) |
| **Self-rated experience with dietary intake data** | |
| Moderately experienced | 5 (38.5) |
| Very experienced | 8 (61.5) |
| Extremely experienced | 0 |
| **Self-rated knowledge of dietary intake data quality** | |
| Moderately knowledgeable | 7 (53.8) |
| Very knowledgeable | 5 (38.5) |
| Extremely knowledgeable | 1 (7.7) |

However, some elements of the tool were described as text heavy, in particular the tool introduction page and the feedback messages for lifestyle and demographic data domains. In addition, feedback messages were not always deemed useful for specific research purposes nor were they based on specific responses provided during the assessment, being described as overly generic. Some specific improvements such as altering the wording of some questions within the assessment as well as specific technical and functional improvements were suggested by participants (Supplementary Table 3).

## Discussion

This paper presents the development and user evaluation of a novel quality assessment tool for dietary intake data designed for use in nutrition research. The tool was designed to assess appropriateness of existing dietary intake datasets for reuse in addressing new research questions. User evaluation was undertaken to understand potential applicability and functionality of the tool. The tool was intended for use within nutrition research with the user evaluation identifying inexperienced nutrition researchers and students as ideal users.

As research questions around nutrition are increasingly focused on food security, sustainable diets, and the interplay of diets with health and environmental consequences, effective nutrition research increasingly requires data from multiple disciplines. In the absence of largescale multiple country databases with data from many areas, there is a greater need for merging datasets for secondary uses. Data reuse and exploitation for new aims presents many opportunities to improve the pace of research and increase capacity to answer more complex problems facing society. However, as part of researcher integrity, user communities have a duty to ensure scientific quality is not compromised. The development of tools and frameworks are an important part of this transition to facilitate data reuse and ensure that researchers are adequately supported. This tool was designed to act as a support for the researcher, but responsibility still lies with the researcher to ensure they adequately understand the dataset in question before deciding to use it. Furthermore, supporting data reuse underpinned by FAIR principles are priorities for European Open Science Cloud (*EOSC*) (11). To ensure these are successfully implemented in the health and life sciences communities, there is a need to upskill researchers and to engage data curators. This was emphasized in the user evaluation, where over half the participants suggested tool assessment would be quicker and feedback possibly more accurate if data owner(s)/provider(s) completed the assessment.

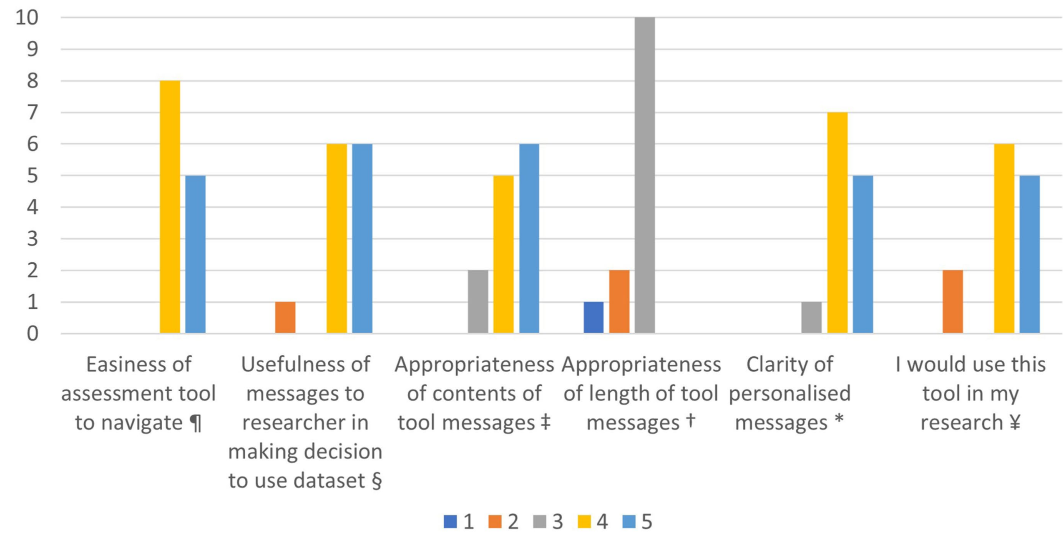



FIGURE 2
Participant self-ratings of aspects of the tool. ¥ 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree. * 1 = very unclear, 2 = unclear, 3 = neutral, 4 = clear, 5 = very clear. † 1 = way too long, 2 = too long, 3 = about right, 4 = too short, 5 = way too short. ‡ 1 = not at all appropriate, 2 = somewhat inappropriate, 3 = neutral, 4 = somewhat appropriate, 5 = very appropriate. § 1 = being useless, 2 = somewhat useless, 3 = neutral, 4 = somewhat useful, 5 = very useful. ¶ 1 = very difficult, 2 = somewhat difficult, 3 = neither difficult nor easy, 4 = somewhat easy, 5 = very easy.

Study participants spoke about the importance of supporting users to assess the appropriateness of reusing dietary intake data. While most felt the duration of the assessment was appropriate, some were concerned about the time it might take new users to complete, who were not familiar with the selected dataset. These participants believed that, ideally, the owner/provider of the data should complete the quality assessment of their dataset, as they would have greater knowledge about the methodologies used. This would revise the scope of the tool, whereby potential users are presented with a report about the strengths and weaknesses of the dataset, and under which circumstances the dataset might be appropriate to be used in. Additional aspects of data quality such as questions about the size and age range of the population, representativeness of the sample, measurement of anthropometrics in fasted vs. unfasted participants, and seasonal variation in intakes were deemed missing from the tool, both in the assessment and feedback report, which many participants expressed as important when assessing quality of food and nutrition data. Guidelines on dietary assessment have been developed in a similar way, highlighting the importance of an open and reiterative process, refining the contents following a series of expert panel reviews (25).

Whilst there are several quality assessment initiatives and tools that have been developed for the food and nutrition domain (26) such as Nutritools (25, 27), Quisper,[2] and DAPA[3], to the best of the authors' knowledge, this is the first attempt to assess the quality of previously collected data for reuse. Like other quality assessment frameworks, the design of this tool is not intended to definitively advise the user whether datasets are suitable to answer research questions; rather, the tool supports decision making through personalised messages containing additional quality and 'fitness for use' factors they may not have previously considered. Quality assessment frameworks are not designed to recommend a single best approach. Instead, they provide a systematic approach to ascertain whether a certain element is fit for the intended purpose and provide suggestions on how to approach different situations (20). Within the space of medical research, several frameworks have been developed in an attempt to systematically assess the quality of health records for reuse (28–31). Some of these frameworks have since been expanded and tailored for specific areas of research, such as heart failure biomarkers to promote identification of appropriate quality studies for reuse in this field (32). The work presented in this paper takes a more generalized approach, as the framework can be applied to all types of dietary intake data but goes beyond previous frameworks as it has been transformed into an online open access tool that can be easily accessed and used by all.

## Future work

This paper presents the development of a quality assessment framework for assessing dietary intake datasets for reuse and its transformation into a first iteration tool. Although a user evaluation study showed the tool was broadly accepted and a particular value was seen in training inexperienced researchers and students in thinking about data quality, the tool would benefit from further development to optimize the user experience. The tool could be further developed to be formally included in nutrition sciences curricula as a training resource for students. Several participants cited the desire for a definitive rating of the datasets quality thus there is a need to make the purpose of the tool, to support the researcher's own decision making, clearer. Participant feedback has highlighted revisions that would be useful to include in a

2 https://quisper.eu/
3 https://www.measurement-toolkit.org/

TABLE 4 User feedback from evaluation study.

| Aspect | Positive aspects identified | Aspects in need of future consideration |
|---|---|---|
| Tool | Tool purpose was well understood (*n* = 8, 62%).<br>Useful for inexperienced users of nutrition data (*n* = 8, 62%) or during study design phase (*n* = 4, 31%). | Introductory message – text heavy (*n* = 4, 31%).<br>Consider how research question is incorporated into assessment (*n* = 2, 15%).<br>Ideally data owner would complete assessment; challenging and time consuming for users unfamiliar with the data (*n* = 8, 62%).<br>Technical improvements such as a side panel listing questions to display progress (*n* = 4, 31%). |
| Assessment | Overall questions were deemed as important and relevant to assessing data quality (*n* = 12, 92%).<br>Hints associated with each question were appreciated and used throughout assessment (*n* = 10, 77%). | Phrasing could be improved for certain questions/data elements that all users may not be familiar with, e.g., food coding systems (*n* = 6, 46%).<br>A greater number of response options or the option to select multiple responses would be useful (*n* = 8, 62%).<br>Some additional questions were suggested, listed in Supplementary material (*n* = 5, 38%). |
| Feedback report | Dietary intake and anthropometric data domain reports were clear, easily understood and examples were appropriate (*n* = 7, 54%).<br>Information included in dietary intake feedback report was deemed relevant to quality of nutrition data (*n* = 5, 38%). | Feedback provided was overly generic. Messaging could be tailored to the specific dataset/research question provided (*n* = 5, 38%).<br>Lifestyle and demographic messages were repetitive (*n* = 8, 62%).<br>Report sections were quite long and wordy (*n* = 4, 31%).<br>Consider visual presentation of information (*n* = 2, 15%).<br>Would like definitive indication of usability or good/bad quality rating (*n* = 2, 15%). |

*n*=: indicates the number of participants who discussed these sentiments in their assessment of their tool.

next version of the software, mainly around the need to condense text on the introductory page and in certain feedback messages as well as addition of further response options and hint icons (user support). The current text heavy version of the tool may be unappealing and a barrier to use for some users who deem it too time consuming. In order to improve the tool's uptake, some participants suggested visualisation of results or generating a summary table of "key messages." Large amounts of text could mean that users less experienced with dietary data may misinterpret or become overwhelmed by the information provided. Furthermore, the evaluation study described in this paper included only thirteen participants, predominantly postgraduate students, who were recruited from 3 research centres across the United Kingdom and Ireland. This may limit the generalizability of our findings to other groups outside of these locations thus a broader evaluation study including more diverse participants from other geographical locations would be important so that it could be used more broadly. Evaluation of the tool by a wider variety of intended users (research, clinical, non-nutrition disciplines) alongside a wider range of experience levels may identify additional improvements which could be made to the tools content and clarity. It is intended to be a living tool that can be further developed and potentially expanded over time. Participants in the user evaluation suggested tailoring some questions and or responses to the research question provided. Addressing these elements could be the focus of any future iterations. Within this first iteration, one data type (dietary intake data) was chosen and the most commonly associated sub-data types (demographic, lifestyle, anthropometric, consumer behavior) added. This list is not exhaustive and future versions could be expanded to include further data types.

## Strengths and limitations

There are many strengths to the tool. Quality parameters were identified through a combination of literature searches and knowledge from domain experts. The tool utilizes a standardized framework that asks consistent questions and covers all areas where quality might be affected during data generation—from data collection, data handling processes, use of underlying data sources, through to how the data are intended to be used and analysed. To the authors' knowledge, no such tool currently exists for dietary intake datasets thus it addresses an important gap. Further, the tool has potential to enhance research capacity through supporting researchers to address new research questions by exploiting existing data.

Some limitations must also be acknowledged. Although the design and tool have been evaluated by test users, there could still be relevant parameters that have not been identified or included. Almost half of the participants in the user evaluation were postgraduate students and there was a lack of participants with extensive experience which may have impacted the findings. User evaluation interviews were conducted by three separate sets of researchers across three different centres as opposed to a single researcher. To minimize differences emerging as a result of this, researchers co-developed a single interview script that was followed for all interviews.

## Conclusion

The tool presented here can support users assessing the suitability of dietary intake datasets for reuse. Although not designed to definitively inform a user whether a dataset is appropriate for their purpose, the use of personalised feedback messages provides users with important considerations to support decision-making. In particular, evaluation of the tool suggested that students and early career researchers might benefit most and the tool could have benefits as a training resource to develop their thinking. The tool is openly available from the FNS-Cloud platform.[4] Future work could expand this framework to incorporate further data types.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by University of Reading research ethics committee (number: 32/2023). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

LB: Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. GB: Formal analysis, Methodology, Visualization, Writing – review & editing. MW: Formal analysis, Methodology, Writing – review & editing. FH: Formal analysis, Methodology, Writing – review & editing. EK: Formal analysis, Methodology, Writing – review & editing. JL: Methodology, Writing – review & editing. PP: Writing – review & editing. SA: Writing – review & editing. PF: Writing – review & editing. EG: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Food Nutrition Security Cloud (FNS-Cloud) has received funding from the European Union's Horizon 2020 Research and Innovation programme (H2020-EU.3.2.2.3. — a sustainable and competitive agri-food industry) under grant agreement No. 863059 – www.fns-cloud.eu. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was part funded by the BBSRC Core Capability Grant BB/CCG2260/1 and its constituent project BBS/E/QU/23NB0006 (Food & Nutrition National Bioscience Research Infrastructure).

## Acknowledgments

The authors would like to thank the Scalefocus team who completed the technical transformation of the framework into the electronic tool, the participants who took part in the evaluation study, and the wider FNS-Cloud consortium members who provided feedback on the initial tool design.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2025.1519401/full#supplementary-material

4 https://aai.fnscloud.eu/realms/fns-cloud/protocol/openid-connect/auth?client_id=sf-da-public-stag&redirect_uri=https%3A%2F%2Fdataset-assessment.scalefocus.dev%2F&state=a9d16e40-eef1-4036-ae84-1d5eb7c70e0d&response_mode=fragment&response_type=code&scope=openid&nonce=b1a5e35c-6259-4394-8375-7c10b19f7b0f

## References

1. Celis-Morales C, Livingstone KM, Marsaux CF, Forster H, O'Donovan CB, Woolhead C, et al. Design and baseline characteristics of the Food4Me study: a web-based randomised controlled trial of personalised nutrition in seven European countries. *Genes Nutr*. (2015) 10:450. doi: 10.1007/s12263-014-0450-2

2. Manios Y, Androutsos O, Lambrinou CP, Cardon G, Lindstrom J, Annemans L, et al. A school-and community-based intervention to promote healthy lifestyle and prevent type 2 diabetes in vulnerable families across Europe: design and implementation of the Feel4Diabetes-study. *Public Health Nutr*. (2018) 21:3281–90. doi: 10.1017/S1368980018002136

3. Riboli E, Kaaks R. The EPIC project: rationale and study design. European prospective investigation into Cancer and nutrition. *Int J Epidemiol*. (1997) 26 Suppl 1:S6–S14. doi: 10.1093/ije/26.suppl_1.s6

4. Keys A, Menotti A, Aravanis C, Blackburn H, Djordevič BS, Buzina R, et al. The seven countries study: 2,289 deaths in 15 years. *Prev Med*. (1984) 13:141–54. doi: 10.1016/0091-7435(84)90047-1

5. Bailey RL. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. *Curr Opin Biotechnol*. (2021) 70:91–6. doi: 10.1016/j.copbio.2021.02.007

6. European Food Safety A. Guidance on the EU menu methodology. *EFSA J*. (2014) 12:3944. doi: 10.2903/j.efsa.2014.3944

7. Leclercq C, Allemand P, Balcerzak A, Branca F, Sousa RF, Lartey A, et al. FAO/WHO GIFT (global individual Food consumption data tool): a global repository for

harmonised individual quantitative food consumption studies. *Proc Nutr Soc*. (2019) 78:484–95. doi: 10.1017/S0029665119000491

8. Miller V, Singh GM, Onopa J, Reedy J, Shi P, Zhang J, et al. Global dietary database 2017: data availability and gaps on 54 major foods, beverages and nutrients among 5.6 million children and adults from 1220 surveys worldwide. *BMJ Glob Health*. (2021) 6:e003585. doi: 10.1136/bmjgh-2020-003585

9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*. (2016) 3:160018. doi: 10.1038/sdata.2016.18

10. European Commission: Directorate-General for R, Innovation. Turning FAIR into reality – Final report and action plan from the European Commission expert group on FAIR data. Luxembourg: Publications Office (2018).

11. European Commission (2020). European Open Science cloud. Availabe online at: https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud (accessed 22nd August 2024)

12. FNS-Cloud (2019) Welcome to FNS-cloud. Availabe online at: https://www.fns-cloud.eu/cookies-policy/overview-new/

13. Prat NAA. Augmented analytics. *Bus Inf Syst Eng*. (2019) 61:375–80. doi: 10.1007/s12599-019-00589-0

14. Zhang R, Indulska M, Sadiq S. Discovering data quality problems. *Bus Inf Syst Eng*. (2019) 61:575–93. doi: 10.1007/s12599-019-00608-0

15. Welch A. Dietary intake measurement: methodology In: En cyclopedia of human nutrition. *Third edition* ed: Academic Press (2013)

16. Shim J-S, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. *Epidemiology Health*. (2014) 36:e2014009–9. doi: 10.4178/epih/e2014009

17. Roe MA, Bell S, Oseredczuk M, Christensen T, Westenbrink S, Pakkala H, et al. Updated food composition database for nutrient intake. *EFSA Support Public*. (2013) 10:355E. doi: 10.2903/sp.efsa.2013.EN-355

18. Roberge AG, Sevigny J, Seoane N, Richard L. Dietary intake data: usefulness and limitations. *Prog Food Nutr Sci*. (1984) 8:27–42.

19. Institute of Medicine (US) Subcommittee on Interpretation and Uses of Dietary Reference Intakes. DRI dietary reference intakes: Applications in dietary assessment. Washington (DC): National Academies Press (US) (2000).

20. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. *Syst Rev*. (2017) 6:204. doi: 10.1186/s13643-017-0604-6

21. Yin Foo R, Guppy M, Johnston LM. Intelligence assessments for children with cerebral palsy: a systematic review. *Dev Med Child Neurol*. (2013) 55:911–8. doi: 10.1111/dmcn.12157

22. Keikes L, Kos M, Verbeek XAAM, Van Vegchel T, Nagtegaal, ID, Lahaye M, et al. Conversion of a colorectal cancer guideline into clinical decision trees with assessment of validity. *Int J Qual Health Care*. (2021) 33:1–8. doi: 10.1093/intqhc/mzab051

23. Forster H, Walsh MC, O'Donovan CB, Woolhead C, McGirr C, Daly EJ, et al. A dietary feedback system for the delivery of consistent personalized dietary advice in the web-based multicenter Food4Me study. *J Med Internet Res*. (2016) 18:e150. doi: 10.2196/jmir.5620

24. Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs*. (2008) 62:107–15. doi: 10.1111/j.1365-2648.2007.04569.x

25. Cade JE, Warthon-Medina M, Albar S, Alwan NA, Ness A, Roe M, et al. DIET@NET: best practice guidelines for dietary assessment in health research. *BMC Med*. (2017) 15:202. doi: 10.1186/s12916-017-0962-x

26. Dao MC, Subar AF, Warthon-Medina M, Cade JE, Burrows T, Golley RK, et al. Dietary assessment toolkits: an overview. *Public Health Nutr*. (2019) 22:404–18. doi: 10.1017/S1368980018002951

27. Warthon-Medina M, Hooson J, Hancock N, Alwan NA, Ness A, Wark PA, et al. Development of Nutritools, an interactive dietary assessment tools website, for use in health research. *Lancet*. (2017) 390:S94. doi: 10.1016/S0140-6736(17)33029-5

28. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)*. (2016) 4:1244. doi: 10.13063/2327-9214.1244

29. Ozonze O, Scott PJ, Hopgood AA. Automating electronic health record data quality assessment. *J Med Syst*. (2023) 47:23. doi: 10.1007/s10916-022-01892-2

30. Razzaghi H, Greenberg J, Bailey LC. Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. *Learn Health Syst*. (2022) 6:e10264. doi: 10.1002/lrh2.10264

31. Tute E, Ganapathy N, Wulff A. A data driven learning approach for the assessment of data quality. *BMC Med Inform Decis Mak*. (2021) 21:302. doi: 10.1186/s12911-021-01656-x

32. Lee K, Weiskopf N, Pathak J. A framework for data quality assessment in clinical research datasets. *AMIA Annu Symp Proc*. (2017) 2017:1080–9.

frontiers | Frontiers in Nutrition

TYPE Original Research
PUBLISHED 18 June 2025
DOI 10.3389/fnut.2025.1568089



OPEN ACCESS

EDITED BY
Igor Pravst,
Institute of Nutrition, Slovenia

REVIEWED BY
Panchali Moitra,
SNDT Women's University, India
Rodica Siminiuc,
Technical University of Moldova, Moldova

*CORRESPONDENCE
Mariana Vieira dos Santos Kraemer
✉ marianavskraemer@gmail.com

RECEIVED 28 January 2025
ACCEPTED 02 June 2025
PUBLISHED 18 June 2025

CITATION
Kraemer MVS, Fernandes AC, Chaddad MCC, Scapin T, Barros BV, Milano E, Padovan M, Uggioni PL, Bernardo GL, Khandpur N, Ares G and Proença RPC (2025) Assessing nutritional composition and ingredients of packaged foods in Brazil: an in-store census method for creating a comprehensive food label database.
*Front. Nutr.* 12:1568089.
doi: 10.3389/fnut.2025.1568089



# Assessing nutritional composition and ingredients of packaged foods in Brazil: an in-store census method for creating a comprehensive food label database

Mariana Vieira dos Santos Kraemer[1]*, Ana Carolina Fernandes[1], Maria Cecília Cury Chaddad[2], Tailane Scapin[1,3], Beatriz Vasconcellos de Barros[1], Elisa Milano[1], Marina Padovan[1], Paula Lazzarin Uggioni[1], Greyce Luci Bernardo[1], Neha Khandpur[4], Gastón Ares[5] and Rossana Pacheco da Costa Proença[1]

[1]Nutrition in Foodservice Research Centre, Nutrition Postgraduate Program, Federal University of Santa Catarina, Florianópolis, Brazil, [2]Movimento Põe no Rótulo, São Paulo, Brazil, [3]Global Centre for Preventive Health and Nutrition (GLOBE), Institute for Health Transformation, School of Health and Social Development, Deakin University, Burwood, VIC, Australia, [4]Division of Human Nutrition and Health, Wageningen University, Wageningen, Netherlands, [5]Sensometrics and Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Montevideo, Uruguay

**Background:** The consumption of ultra-processed packaged foods has surged worldwide with important health implications. It is pertinent to study the composition of packaged foods through information provided on labels. However, there is limited methodological discussion in the field. This study aimed at discussing methodological evolution and challenges of in-store census methods for assessing the composition of packaged foods, and characterizing a Brazilian food label database.

**Methods:** The first Brazilian food label database reported in the scientific literature, based on data of in-store census method, was created in 2010 by the Nutrition in Foodservice Research Centre (NUPPRE at the Federal University of Santa Catarina). The in-store census method involves collecting primary data directly from the labels of packaged foods available for sale through retail food outlets. In 2020, the in-store census was carried out in partnership with the FoodSwitch Program. The NUPPRE/FoodSwitch Brazil 2020 database was developed in four steps: pre-data collection, data collection, data tabulation, and database construction and processing. The database was characterized by calculating the prevalence of foods per food group and foods that declared mandatory nutrition and health information on food labels according to Brazilian regulation.

**Results:** The nutritional profile and ingredients of packaged foods was obtained from four food label censuses (2010, 2011, 2013 and 2020), supporting the Brazilian government on food labeling regulations and public policies. The experience prompted reflections about the methodological aspects of food label studies, and enabled improvements to the research process, such as a more accurate data collection, the inclusion of all packaged foods and beverages available for sale in the supermarket and the inclusion of more variables to the analysis. It is noteworthy the relevance of building nationwide food labeling databases. However, there are important challenges regarding the costs and efforts needed to maintain and update the data, especially in continental countries such as Brazil. The NUPPRE/FoodSwitch Brazil 2020 database consists of 7,828 packaged foods, 94% of the sampled brands sold nationwide. Most foods presented the mandatory information according to Brazilian regulation.

**Conclusion:** This study proposed a series of methodological procedures to be carefully considered, designed, and executed during planning, data collection, data tabulation, and database processing. Greater rigor and detail are needed in the methods section of scientific articles, to aid replication.

KEYWORDS

**nutrition labeling, ingredient list, nutrition information, ultra-processed foods, methods, supermarkets**

## 1 Background

Over the past decade, the consumption of processed and ultra-processed foods has surged worldwide and these products have partially replaced fresh unprocessed foods with important public health implications (1). In particular, consumption of ultra-processed foods has been associated with the development of several diseases, such as cardiovascular diseases, type 2 diabetes, obesity, numerous types of cancers, mental disorders, as well as with an increased risk of all-cause mortality (2, 3). The proposed mechanisms to explain these associations are related to the processing-altered composition of ultra-processed products: high content of sugars, saturated fat, trans fat, and/or sodium, and the presence of food additives, as well as neoformed compounds, and contamination with contact materials (3–5).

In this context, monitoring the composition of the processed and ultra-processed products available in the marketplace is relevant to inform the design, implementation, monitoring, and evaluation of public health interventions aimed at improving diet quality. Food composition data are also relevant for clinical and epidemiological research, particularly in relation to the health effects of ultra-processed foods (6).

Food composition data are available on food labels as manufacturers are mandated to include the list of ingredients and nutrient declarations in most countries (7). Therefore, food labels can be used to monitor the composition of processed and ultra-processed products (8), to the extent that they are reported transparently, accurately and consistently by food manufacturers. Any informational limitation of the nutritional labels may also be reflected in the databases constructed from them. The main advantage of this approach is that large databases can be compiled in short-time frames at relative low cost compared to chemical analyses. Food labels are also the key source of information for consumers, enabling them to access detailed information about the nutritional and ingredient composition of packaged foods (6).

Databases containing food labeling information have been developed worldwide, which differ in data collection procedures and data quality control, among other aspects. Some publicly available databases (e.g., Open Food Facts) rely on crowdsourcing, where food label photographs are submitted by users (9). Other labeling databases contain information provided directly by food manufacturers. Examples include the USDA's *FoodData Central* (10) and the Mintel database (11). However, the completeness and accuracy of the data may not always be guaranteed (6, 12). In addition, some of these databases are behind paywalls (e.g., Mintel), which may limit their accessibility.

Label information can also be collected from the websites of food manufacturers and retailers (13). However, these websites usually include limited information. A recent study reported that food composition information was frequently unavailable on supermarket websites in Australia, such as allergen information, the nutrition facts panel, and the ingredients list (13). Food composition databases have also been created by researchers by collecting primary data directly from the labels of packaged foods available for sale through retail food outlets (12, 14, 15). Data collection is carried out in person at supermarkets or other food outlets, by collecting information from the labels of all the available products to generate a comprehensive database. In the current study, this strategy is referred to as the in-store census method.

Databases have been generated using in-store census worldwide. Examples include the FoodSwitch database from Australia (16, 17), the Food Label Information Program (FLIP), from Canada (18), and the Composition and Labeling Information System (CLAS), from Slovenia (6). In Brazil, the Nutrition in Foodservice Research Centre (NUPPRE) at the Federal University of Santa Catarina (UFSC) has been performing in-store census of food labels since 2010.

Despite the importance of monitoring food composition using label information, the great majority of the articles published with food labeling data do not discuss the methodological aspects related to the construction of the database, which may directly impact the quality of the data. Few studies have been published describing in detail the procedures for collecting, tabulating, and constructing databases from food label information (6, 12, 14, 15). In view of these gaps, this study aimed at: (i) discussing methodological evolution and challenges of in-store census methods for assessing the composition of packaged foods, and (ii) describing and characterizing the NUPPRE/FoodSwitch Brazil 2020 food label database.

### 1.1 The Brazilian food labeling regulatory framework

The Brazilian Health Regulatory Agency determines the criteria for the declaration of food labeling information through several regulations. Brazilian Resolution No. 259/2002 establishes general rules for the labeling of packaged foods, requiring the declaration of mandatory information on food labels, such as the ingredient list. The only exception to this requirement applies to foods composed of a single ingredient, such as sugar, coffee, and salt, which do not need to comply with those requirements (19).

Another legal instrument, Resolution No. 360/2003, addresses the nutrition and health aspects of labeling, and made nutrition labeling mandatory for all packaged foods. Nutrition labeling includes information presented on the nutrition facts panel (mandatory) and nutrition claims (voluntary). The following items must be declared on the nutrition facts panel, accompanied by their respective quantities per serving: energy value (kcal and kJ), carbohydrates (g), proteins (g), total fat (g), saturated fat (g), trans fat (g), dietary fiber (g), and sodium (mg). The declaration of vitamins and other minerals is optional if the product contains 5% or more of the recommended daily intake per serving, whereas it should be mandatory if the front of the package contains any claim about these nutrients (20). Additionally, this resolution did not apply to the following foods: alcoholic beverages; food additives and processing aids; spices; mineral waters; vinegars; salt; coffee, yerba mate, tea, and other herbs without additional ingredients; and fresh, chilled, and frozen meats, fruits, and vegetables. Therefore, data on these foods were not collected in 2010, 2011 and 2013.

Three other regulations concerning the declaration of nutrition and health information related to gluten, allergens, and lactose were in force at the time of data collection. All packaged foods with labels must contain the warning "contains gluten" or "does not contain gluten," as appropriate (21). Furthermore, all packaged foods with labels must declare the presence of the following allergens: wheat, rye, barley, oat and oat hybrids, crustaceans, eggs, fish, peanut, soybean, milk from any species of mammalian animals, almond, hazelnut, cashew nut, Brazil nut, macadamia nut, pecan nut, pistachio, pine nut, chestnut, and natural latex (22). Finally, regarding lactose, the applicable resolution requires the declaration of lactose presence for all packaged foods with labels that contain more than 100 mg of lactose per 100 g or 100 mL (23).

It should be noted that all these regulations were in force by the time of 2010, 2011, 2013 and 2020 data collections, therefore, being applied uniformly across the entire studied periods. Most of these mandatory requirements are still in effect, though they are now regulated by a recently approved resolution (24), which consolidated the general labeling legislation. Also, a new regulation regarding nutrition labeling was approved in October 2020 (25), significantly altering the regulatory framework. However, the impacts of this change are not within the scope of this investigation, given that the regulation was not in force by the time of the data collections.

## 1.2 Historical and methodological evolution of the in-store food label data collection in Brazil: an overview of the NUPPRE/UFSC census method for creating comprehensive food labeling databases

The first Brazilian food label database reported in the scientific literature, based on data collected using an in-store census method, was created in 2010 by NUPPRE/UFSC (26). Over time, the methodological aspects were refined to enhance the validity and reliability of the database. Table 1 presents the main methodological aspects of the label census method developed by NUPPRE/UFSC (2010, 2011, 2013) and NUPPRE/FoodSwitch Brazil 2020.

The first data collection, in 2010, took place in the context of a study on the declaration of trans fat, serving sizes, and household measures on packaged food labels. Food label data were gathered at a medium-sized supermarket store belonging to a large Brazilian chain. This supermarket chain remains one of the leading chains in Brazil, according to the revenue ranking published by the Brazilian Association of Supermarkets (27).

The data were recorded on a paper-based form. Information was manually copied from labels at the supermarket by trained data collectors. The sample included all foods likely to contain trans fat that were available at the supermarket at the time of data collection. Variations of the same type of food product (different flavors and packaging sizes) were counted as distinct items, as it was observed that products often had different characteristics and compositions depending on the size and type of packaging. Foods outside the scope of nutritional labeling legislation (20) were not included in data collection. Additionally, foods intended for infants and young children were also excluded, as they were regulated by specific legislation (28). Subsequently, the data were transcribed into Microsoft Excel spreadsheets for analysis. The following information was collected: trade name, brand, manufacturer, country of origin, product type (e.g., cookies, milk drink, and chocolate), flavor, price, package weight, presence/absence and order of declaration of trans fat or ingredients likely to contain trans fat in the ingredients list, presence of nutrition information on trans fat on the nutrition facts panel, serving size, household measure, and claims related to the absence of trans fat.

In 2011, a new data collection was carried out at a large supermarket store of the same chain chosen in the previous year. Methodological procedures for data collection and tabulation were also the same but focused on the analysis of sodium declaration in packaged foods. The sample included foods that could contain sodium, were covered by the applicable nutrition labeling legislation (20), and were available for sale at the supermarket. The following information was collected from product labels: trade name, brand, manufacturer, country of origin, product type, flavor, price, package weight, presence/absence and order of declaration of added sodium (salt and sodium-containing food additives) in the ingredients list, nutrition information on foods containing added sodium, sodium declaration on the nutrition facts panel, serving size, household measure, sodium-related claims, and claims targeted at children.

In 2013, NUPPRE/UFSC conducted a third in-store census of food labels. Data collection was performed at the same supermarket store sampled in 2011. The data were recorded in-store by trained data collectors using tablets and an electronic form (EpiCollect Plus®), based on the previously used paper-based form. For this data collection, all food products also had their packages photographed. Subsequently, the information collected on the electronic form was exported to Microsoft Excel. The photos were used to extract data from the ingredients lists of each product, and this information was tabulated in Microsoft Excel. The sample included all foods available for sale at the supermarket and covered by applicable legislation (20). The information collected from product labels included trade name, brand, manufacturer, country of origin, product type, flavor, price, package weight, regulated nutrition claims (29), nutrition facts panel information (serving size, household

TABLE 1 Description of the four in-store census methods (2010, 2011, 2013, and 2020) used to develop comprehensive food labeling databases in Brazil.

| Methodological aspect | In-store label census | | | |
|---|---|---|---|---|
| | NUPPRE Brazil 2010 (*N* = 2,327) | NUPPRE Brazil 2011 (*N* = 4,286) | NUPPRE Brazil 2013 (*N* = 5,620) | NUPPRE/ FoodSwitch Brazil 2020 (*N* = 7,828) |
| Study site | | | | |
| Large supermarket chain[§] | X | X | X | X |
| Medium-size store | X | | | |
| Large-size store | | X | X | X |
| Trained data collectors | X | X | X | X |
| Data collection instrument | | | | |
| Paper form | X | X | | |
| Electronic form | | | X | |
| Smartphone app with a photo capture feature | | | | X |
| Information collected from food labels | | | | |
| Product identification | X | X | X | X |
| Trans fat information (quantity and claims) | X | | | |
| Sodium information (quantity and claims) | | X | | |
| All nutrient information for all nutrients available (quantity and claims) | | | X | X |
| All claims related to trans fat | X | | | |
| All claims related to sodium | | X | | |
| All claims presented in the packaged (nutrition claims, health claims and others) | | | X | X |
| Trans fat ingredients | X | | | |
| Sodium-containing ingredients (salt and sodium-based ingredients) | | X | | |
| All ingredients (food ingredients and food additives) | | | X | X |
| Diet/light claims | | X | | X |
| Marketing strategies targeting children | | X | X | X |
| Transgenic (GMO) symbol | | | X | X |
| Additional information displayed on the label | | | X | X |
| Inclusion criteria | | | | |
| Foods that may contain trans fat | X | | | |
| Foods that may contain sodium | | X | | |
| Foods within the scope of Brazilian nutritional labeling legislation** (except foods intended for infants and young children) | X | X | X | |
| All packaged foods and beverages available for sale at the supermarket | | | | X |
| Data collection procedures | | | | |
| Data collection begins after supermarket authorization | X | X | X | X |
| Paper forms | X | X | | |
| Electronic forms | | | X | |

*(Continued)*

TABLE 1 (Continued)

| Methodological aspect | In-store label census | | | |
|---|---|---|---|---|
| | NUPPRE Brazil 2010 (*N* = 2,327) | NUPPRE Brazil 2011 (*N* = 4,286) | NUPPRE Brazil 2013 (*N* = 5,620) | NUPPRE/ FoodSwitch Brazil 2020 (*N* = 7,828) |
| All sides of the package of all packaged foods available for sale at the supermarket by the time of data collection were photographed | | | X | X |
| Data tabulation | | | | |
| Transcription of collected information into a Microsoft Excel spreadsheet | X | X | | |
| Transcription of ingredients lists into a Microsoft Excel spreadsheet | | | X | |
| Information collected in the electronic form transferred directly to Microsoft Excel spreadsheets | | | X | |
| Transcription of collected information into a monitoring database | | | | X |
| Database processing | | | | |
| Exclusion of duplicated products | X | X | X | X |
| Data transferred from the monitoring database (FoodSwitch program) to Microsoft Excel | | | | X |
| Quality control: verification of tabulated data in 10% of the database | X | X | X | |
| Quality control: verification of tabulated data in 20% of the database and in each study based on variables of interest | | | | X |
| Focus of data analysis | | | | |
| Trans fat | X | | X | X |
| Trans fat substitutes | | | X | X |
| Serving sizes and household measures | X | | | |
| Sodium | | X | | |
| Foods targeted at children or consumed by children | | X | X | X |
| Added sugars | | | X | X |
| Added sugars in foods targeted at children | | | | X |
| Free sugars from fruits | | | | X |
| Sweeteners (food additives) | | | X | X |
| Homemade, traditional, and similar claims | | | X | X |
| Whole grain claims | | | X | |
| Claims of functional and health properties | | | | X |
| Claims of functional and health properties in foods targeted at children | | | | X |
| Genetically modified organisms | | | X | |
| Vitamins and minerals in foods targeted at children | | | X | |
| Food additives | | | | X |
| Food additives in foods targeted at children | | | | X |

[$]According to the revenue ranking annually published by the Brazilian Association of Supermarkets *Brazilian Health Regulatory Agency (20); **Brazilian Health Regulatory Agency (30).

measure, total energy value, carbohydrates, proteins, total fat, saturated fat, trans fat, fiber, sodium, vitamins, and minerals), and ingredients list. As in the two previous data collections, product variations (different flavors and package sizes) were considered distinct items, and infant and children's foods were excluded.

For the first time, in 2020, data collection was carried out in partnership with FoodSwitch. The sampled supermarket store was the same site as 2011 and 2013 data collections. A database was created using data on the nutritional composition of all packaged foods and beverages available for sale at the time of data collection.

Information on the nutritional profile and ingredients of packaged foods was obtained from the four food label censuses conducted by our research group. Such data were utilized for various analyses related to the nutritional and ingredient composition of packaged foods, including cross-sectional and longitudinal analyses, comparisons, and data monitoring. The studies were carried out in the form of postdoctoral fellowships, doctoral theses, master's dissertations, scientific initiation projects, and undergraduate capstone projects with Pan American Health Organization (PAHO), Brazilian Health Regulatory Agency (ANVISA) and Brazilian national agencies for research support (CAPES and CNPq) grants (Supplementary Table S1). Additionally, our research group has been supporting ANVISA and the Brazilian Ministry of Health in important actions, as recognition and result of the work developed over the four NUPPRE in-store censuses (2010, 2011, 2013, and 2020), assisting the development and reformulation of public policies in food, nutrition, and health. In this perspective, efforts have been made toward reforming national legislation regarding general and nutritional food labeling, eliminating trans fat, defining the appropriate use for the term "whole" in cereal- and pseudocereal-based foods, establishing quality standards for oils and fats, and engaging in discussions about food additives, sweeteners, and added sugars.

With the experience gathered after every data collection, it was possible to improve the methodological procedures. The main methodological differences between the four censuses (2010, 2011, 2013, 2020) refer to (i) the type of foods included in data collection, (ii) the information retrieved from food labels, and (iii) data collection instruments and techniques.

With each new data collection, efforts were made to expand the range of packaged foods analyzed. In the first two data collections, only foods possibly containing trans fat or sodium were analyzed. In the third data collection, the aim was to gather data from all packaged foods covered by Brazilian nutrition labeling legislation. In 2020, all labeled foods and beverages marketed by the supermarket were censused, including those outside the scope of nutritional labeling regulations, such as alcoholic beverages, mineral waters, vinegars, salt, coffee, meats, fresh fruits and vegetables, and other items described above. Thus, data were collected from all labeled foods available for sale in the supermarket at the time of data collection.

From 2013 onwards data collection no longer occurred using paper forms, but in a mixed manner: an electronic form was filled, and photographs were taken of food products. In 2020, data collection was carried out entirely by taking photographs of food labels. This method helped reduce the risk of bias related to errors in data collection and allowed further expanding the amount of information collected. Thus, in addition to the data tabulated in 2013, in 2020, it was possible to gather more information available in food labeling, such as allergens, gluten, and lactose.

Regarding claims (nutrition, health, and others), those related to trans fat were included in the 2010 database; the claims related to salt, and sodium were included in the 2011 database; and the nutrition claims as regulated by Brazilian legislation (29) were included in the 2013 database. In 2020, although front-of-pack claims were available in the photos taken during data collection, this information was not tabulated and, consequently, was not included in the database. This decision was taken because of the wide diversity of claims displayed on the front of the package by manufacturers. Future data tabulation by researchers with expertise in the subject could enhance precision and accuracy, reducing the potential for errors. For instance, the tabulation of claims linked to marketing strategies targeting children was handled by researchers working on the theme, allowing for the application of more specific and accurate criteria to identify such claims.

Concerning data collection instruments, the use of an electronic form in 2013 made it possible to export part of the information directly to a Microsoft Excel spreadsheet; thus, only the ingredients list was transcribed. In 2020, an app was used to take photographs of food labels, increasing the speed of data collection. This method helped reduce the risk of bias related to errors in data collection.

All collections were carried out in the same supermarket chain, which is among the largest supermarket chains in Brazil according to annual revenue (27). In the first census, a smaller store was selected due to the limited technical capacity of the team and the pioneering nature of the study, ensuring quality and continuity in data collection. As the team's technical skills and experience enhanced, the second data collection was conducted in a larger store. According to data from the supermarket chain, the store chosen for the 2011, 2013, and 2020 censuses is the largest with regard to size, number of products, and number of brands. Given the large store size, it was possible to collect information from a greater diversity of products and brands.

## 2 A step-by-step approach to building a food label composition database: description of the NUPPRE/FoodSwitch 2020 in-store census

The development of the NUPPRE/FoodSwitch Brazil 2020 database comprised four distinct and consecutive steps: pre-data collection procedures, data collection, data tabulation, and database construction and processing. Each step was carried out in stages, as depicted in Figure 1.

### 2.1 Pre-data collection procedures

#### 2.1.1 Study site selection

The study was conducted during November 2020 in a large supermarket outlet in Brazil. The outlet was chosen intentionally, to enable data collection with the available financial and human resources. In addition, the outlet was selected because it belonged to one of the 15 largest supermarket chains in Brazil according to annual revenue (27) and was the largest supermarket outlet in the Brazilian state where the study was conducted (Santa Catarina). The supermarket manager provided consent for data collection.

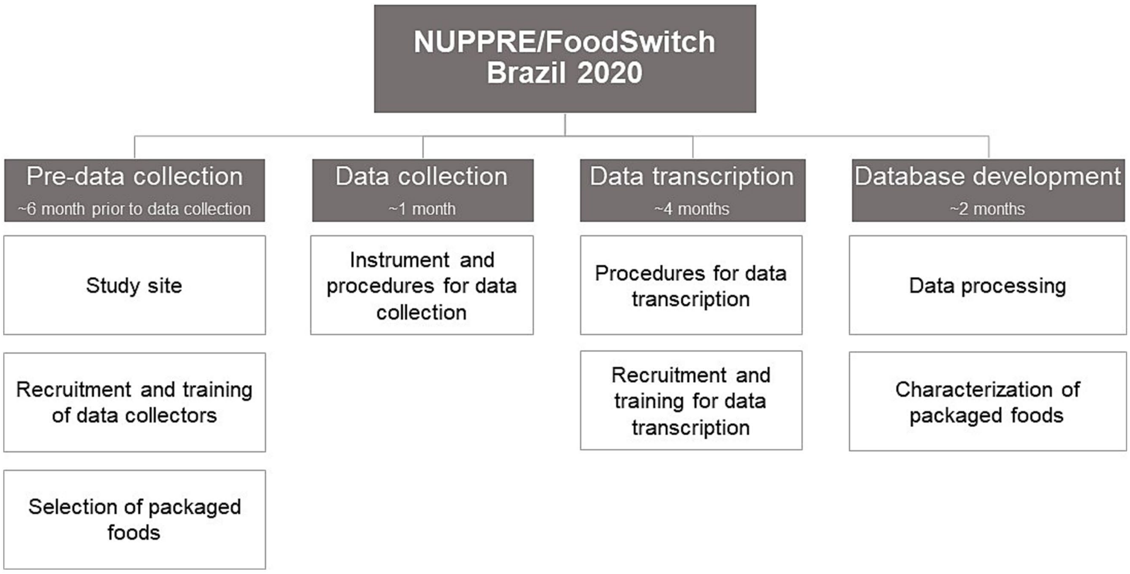



FIGURE 1
Stages of development of the NUPPRE/FoodSwitch Brazil 2020 database.

### 2.1.2 Recruitment and training of data collectors

Ten data collectors were recruited among graduate and undergraduate students in Nutrition & Dietetics at UFSC. All data collectors received theoretical-practical training, offered in English by researchers from the Australian FoodSwitch program. The training session was conducted virtually and covered the configuration and operation of the data collection instrument, as well as practical and technical aspects of data collection in supermarkets. The research coordinators offered a reinforcement training session in Portuguese, the native language of data collectors. Additionally, the data collectors received a document outlining the data collection protocol. The protocol included a detailed explanation of how to use the data collection instrument, encompassing operations such as filling in identification data, scanning barcodes, taking and saving photos, and completing the data collection session.

A pre-test of the data collection process was carried out to identify potential errors in operating the instrument and to ensure that photos of food packaging were taken correctly. In the pre-test, data collectors were instructed to use the data collection instrument to take photos of six distinct food labels, following the procedures taught during training. Then, the Australian FoodSwitch team provided feedback on the quality and legibility of photographs and collected information.

### 2.1.3 Inclusion criteria

Information was collected from the labels of all packaged foods available for sale at the time of data collection that met the following inclusion criteria:

i All foods that are marketed and packaged in the absence of the customer and ready for consumer purchase, according to Resolution No. 259/2002 (19), which specifies labeling requirements for packaged foods.

ii Specific foods for infants and young children, as defined by Law No. 11265/2006. These products include formulas for infants, follow-on formulas for infants and young children; fluid, powdered, and modified milk products and similar products of plant origin; transition and cereal-based foods indicated for infants and/or young children; foods or beverages, whether milk-based or not, suitable for infants and young children (28).

iii Alcoholic beverages and mineral waters.

Foods with different barcode numbers were treated as distinct products. Therefore, all variations of a food product (different flavors and package sizes) were sampled, as products can have distinct characteristics and composition depending on packaging size and type. Unpackaged fresh fruits, vegetables, meats, breads and bakery products sold without a label were not surveyed.

## 2.2 Data collection

Labels were photographed using a mobile phone application developed by The George Institute's FoodSwitch program, Australia (16, 17). The application was adapted for collecting data from Brazilian food labels on iOS and Android smartphones. The app enabled data collectors to scan the barcodes of each packaged food for identification and then take photos of the information displayed on the labels. Figure 2 shows the information retrieved from photographs, presented in the order in which they were recorded.

If necessary, more than one photo could be taken to capture all required information. For example, if it was not possible to fit the entire nutrition facts panel into a single photo, collectors could take as many photos as needed. Once all the information had been retrieved, the data collection for that product was completed, and the next item was surveyed by scanning the barcode.

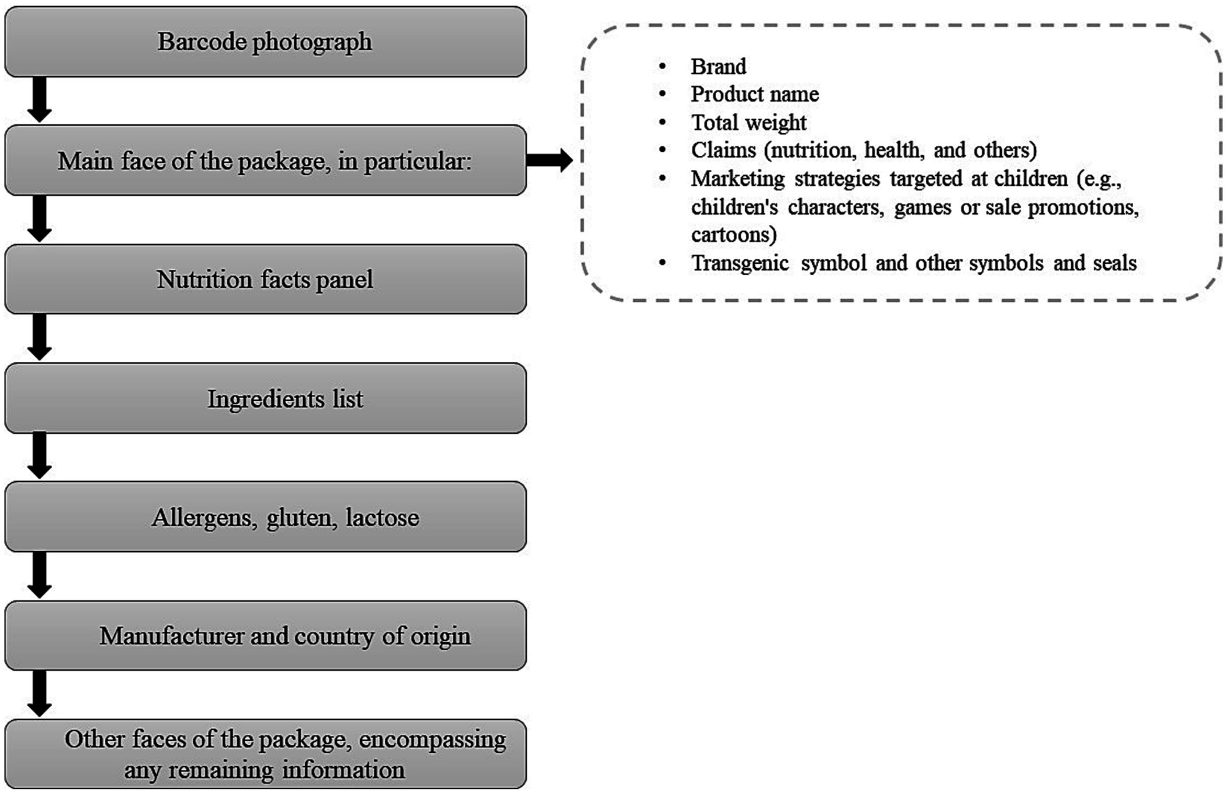



FIGURE 2
Information retrieved from photographs taken during data collection for the creation of the NUPPRE/FoodSwitch Brazil 2020 database, in the order in which they were recorded.

As in the previous years (2010, 2011, and 2013), a coordinator was present during data collection in 2020 and assisted data collectors in case of difficulties. On each day, the coordinator informed the data collectors which supermarket sector would be surveyed and instructed that data from all food products available in that sector should preferably be collected on the same day. All foods available for sale at the time of data collection were surveyed.

At the end of data collection, the photographs were uploaded to the FoodSwitch system. These photos were later made available for data tabulation.

## 2.3 Data tabulation

### 2.3.1 Data tabulation procedures

Tabulation consisted of transcribing the information contained in food label photos to a monitoring database developed by FoodSwitch (The George Institute's Food and Beverage Information Content Management System—FBI CMS). The monitoring database is an online platform where photos are organized by food products, based on their barcode numbers. Next to each photo, there are fields for transcribing the details shown in the images.

The following information was tabulated by the researchers, in Portuguese: manufacturer, brand, product name, total package weight, and nutrition facts panel information (serving size, household measure, macro- and micronutrient contents, ingredient list, list of allergens, contains/does not contain gluten, alcohol content).

### 2.3.2 Recruitment and training of data tabulators

As for data collection, 10 data tabulators were recruited among graduate and undergraduate students in nutrition at UFSC. All data tabulators participated in a 2 h training session, provided virtually in English by researchers from the Australian FoodSwitch program. The training session addressed practical questions about the monitoring database, including which information should be entered in each field. The research coordinator offered a reinforcement training session in Portuguese, the native language of data tabulators. This reinforcement session lasted about 1 h and was aimed at clarifying doubts and highlighting key points of data tabulation. Additionally, data tabulators received a step-by-step protocol. The protocol provided a detailed explanation of how to use the monitoring database and indicated where each piece of information should be entered.

## 2.4 Database development and processing

After tabulation, the data were exported to a Microsoft Excel® spreadsheet, creating the NUPPRE/FoodSwitch Brazil 2020 database. The database was made available on a remote desktop, with individual access granted to each researcher.

In a preliminary treatment step, two different researchers reviewed the database for duplicate products, which were identified in the Excel spreadsheet by their barcode numbers. When the same food was tagged as a duplicate by the two researchers, one of the duplicate entries was excluded from the database. Of note, the only criterion for excluding food products from the database was the presence of duplicate entries. A lack of nutrition, ingredient, or health-related information was not adopted as an exclusion criterion.

The next step involved the food products classification into groups and subgroups, according to the Brazilian nutrition labeling legislation in effect at the time (20). Additional groups were created for foods not covered by this classification, namely Baby and infant foods; Mineral waters; Non-sugar sweeteners, colorings, flavorings, raising agents, and yeasts; Tea, herbs, and coffee; Vinegar and salt; and Supplements.

Quality control has been carried out in 20% of the database, as well as in all studies conducted by NUPPRE researchers using this database, focusing on specific variables of interest. For instance, a study analyzing trans fat content in packaged foods would verify fat information for a portion of the foods entered in the database, compared to the information collected manually or through the photo, proposing corrections as needed.

### 2.5 Database characterization

The NUPPRE/FoodSwitch Brazil 2020 database was characterized by calculating the number of food products per food group, stratified as defined by the applicable nutrition labeling legislation (30).

The declaration of mandatory nutrition and health information on food labels was assessed based on the following criteria: (i) presence of the nutrition facts panel (20), (ii) presence of the ingredients list (19), (iii) presence of allergen information (22), (iv) presence of the "contains/does not contain gluten" warning (21), and (v) presence of lactose information (23).

## 3 Characterization of foods composing the NUPPRE/FoodSwitch Brazil 2020 database

The NUPPRE/FoodSwitch Brazil 2020 database includes 7,828 packaged products of 1,035 different brands, 94% of which were sold nationwide. Table 2 shows the frequencies of packaged foods and the mandatory components of food labels related to nutrition and health information, stratified by food group. Most food items belong to the non-alcoholic beverages, sweets, and confectionery group (25%). Sweet biscuits, chocolates, non-alcoholic beverages, and savory snacks are the most frequent foods in this group, corresponding to 50% of the items in the group. The second most prevalent group was cereals, vegetables, and tubers (14%), in particular salted biscuits, breads, and pasta (42%).

Most food items presented the mandatory information on the presence of gluten (98%), ingredient list (92%), nutrition facts panel (81%), and allergens (59%) (Table 2). It was notable that all items of the baby and infant food group (which includes infant formulas and infant cereals) had labels containing an ingredient list, nutrition facts panel, and gluten information. Information on lactose was the least frequent on the food labels of all groups.

It was found that most foods complied with national regulations regarding mandatory health and nutrition information. As expected, information on the presence of lactose was the least frequent, as it applies to a smaller universe of foods, that is, only those with more than 100 mL of lactose in 100 g or 100 mL of food (23). Gluten must be declared on all packaged foods and beverages (21). The absence of other mandatory items on food labels does not necessarily constitute non-compliance with Brazilian legislation, as there are exceptions to their presence on labels. For example, the presence of allergens must be declared in foods that contain these substances or are at risk for unintentional contamination. Thus, foods that do not contain allergens and do not pose a risk of contamination are not required to include allergen statements on the label. Similarly, there are specific rules for lactose, ingredients lists, and nutrition facts panels, as explained in the Methods section. Considering the public health relevance of allergens and lactose declaration on food labeling, it is important to develop studies aiming to analyze the labeling of these components in order to monitor compliance with specific regulations and assess whether accurate information is being provided to consumers.

## 4 Challenges and lessons learned during the application of the in-store census method

The experience gained over the four data collections by NUPPRE/UFSC prompted reflections about the methodological aspects of food label studies, which are still incipient in scientific literature.

During the planning of this study, we highlight the importance of previously establishing clear criteria on which foods would compose the database and what information should be retrieved from food labels. Factors such as financial resources, human resources, available time, and technical capacity to conduct the studies and work with the data need to be considered. Another important factor is defining where data are to be collected. The supermarket, or other points of sale, should be chosen based on the objectives of the study, establishing clear criteria and seeking the place that best suits the context. It is important to conduct data collections that cover the greatest diversity of foods, brands, and manufacturers possible, thereby addressing data on products that are available to a greater number of people.

Collecting data from food labels through photographs is faster and more accurate compared to paper or electronic forms, when the aim is to collect all information available on food labels. A smartphone application that takes photographs and sends them to data clouds according to the respective barcode number was used here and in previous research conducted in other countries (16–18, 31). Automated tools that enhance the speed of data collection also contribute to avoiding errors generally caused by data collectors. It is important to adequately manage in-store data collection by organizing the team and conducting prior training so that photographs are taken correctly, and products are not missed. For errors and unforeseen events to be avoided during data collection, it is important to carry out a pre-test of the collection instrument, as well as a pilot test of data collection. These procedures allow improving the use of instruments and collection techniques.

In all data collections carried out by NUPPRE, foods with varying packaging sizes were considered as different products. This criterion was adopted because, since the first collection, it was observed that some products have different characteristics and compositions according to packaging size and type, in addition to having different barcode numbers. This criterion is considered a relevant methodological measure, ensuring that all foods sold at the time of

TABLE 2 Characteristics of food items composing the NUPPRE/FoodSwitch Brazil 2020 database, stratified by food group and presence of mandatory items on food labels.

| Food group | Frequency | | Display of mandatory items on food labels* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Nutrition facts panel | | Ingredient list | | Allergen information | | Gluten information | | Lactose information | |
| | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % |
| Cereals, legumes, and tubers[§] | 1,092 | 14 | 1,086 | 99 | 1,008 | 92 | 897 | 82 | 1,073 | 98 | 80 | 7 |
| Vegetables[§] | 366 | 5 | 355 | 97 | 313 | 86 | 40 | 11 | 349 | 95 | 2 | 1 |
| Fruits[§] | 311 | 4 | 308 | 99 | 248 | 80 | 45 | 14 | 304 | 98 | 0 | 0 |
| Milk and dairy[§] | 804 | 10 | 790 | 98 | 799 | 99 | 765 | 95 | 798 | 99 | 418 | 52 |
| Meat, pork, poultry, and seafood[§] | 736 | 9 | 713 | 97 | 602 | 82 | 397 | 54 | 722 | 98 | 43 | 6 |
| Oils and fats[§] | 404 | 5 | 400 | 99 | 386 | 96 | 249 | 62 | 396 | 98 | 54 | 13 |
| Non-alcoholic beverages, sweets, and confectionery[§] | 1,966 | 25 | 1,935 | 98 | 1,945 | 99 | 1,483 | 75 | 1,941 | 99 | 510 | 26 |
| Sauces and ready-to-eat dishes[§] | 580 | 7 | 447 | 77 | 562 | 97 | 334 | 58 | 569 | 98 | 81 | 14 |
| Baby and infant foods | 72 | 1 | 72 | 100 | 72 | 100 | 53 | 74 | 72 | 100 | 12 | 17 |
| Alcoholic beverages | 941 | 12 | 21 | 2 | 887 | 94 | 275 | 29 | 885 | 94 | 1 | 0 |
| Mineral waters | 64 | 1 | 34 | 53 | 0 | 0 | 0 | 0 | 56 | 87 | 0 | 0 |
| Non-sugar sweeteners, colorings, flavorings, raising agents, and yeasts | 67 | 1 | 42 | 63 | 67 | 100 | 22 | 33 | 66 | 99 | 0 | 0 |
| Tea, herbs, and coffee | 307 | 4 | 66 | 21 | 190 | 62 | 18 | 6 | 300 | 98 | 0 | 0 |
| Vinegar and salt | 62 | 1 | 29 | 47 | 58 | 94 | 0 | 0 | 61 | 98 | 0 | 0 |
| Food supplements | 56 | 1 | 54 | 96 | 56 | 100 | 31 | 55 | 54 | 96 | 14 | 25 |
| Total | 7,828 | 100% | 6,352 | 81% | 7,193 | 92% | 4,609 | 59% | 7,646 | 98% | 1,215 | 16% |

[§]Food groups were classified according to the Brazilian legislation on nutrition labeling (30) in force during data collection (November 2020). *Mandatory information related to nutrition or health information, according to Brazilian legislation on food labeling (19–23) in force during data collection (November 2020). The absence of mandatory information does not necessarily mean a non-compliance with Brazilian legislation, considering that there are exceptions depending on the food category.

data collection are analyzed. An additional methodological measure to ensure the analysis of all food items sold in the supermarket was to have a coordinator present every day. This daily monitoring made it possible to help data collectors, minimizing possible errors and potential biases arising from failures. In addition, the coordinator managed and monitored the evolution of data collection, ensuring that all sectors of the supermarket and all food items were contemplated and photographed.

Studies adopting an in-store census method (32, 33) generally analyze a smaller number of food items than studies using data from online searches or existing databases based on crowdsourcing or information from food manufacturers (14, 34). These differences may be attributed to the characteristics of the method, in particular, the need for in-person visits to food sales locations. However, when using an in-store census method, it is possible to clearly determine the criteria for including foods in the database, that is, all those sold at the time of data collection. Furthermore, this strategy encompasses all foods available to consumers at the time of purchase, working with data collected in a real environment.

Transcription of the information on food labels is the costliest stage of the study in terms of time and human resources. The use of technology to extract data from photographs is still a challenge, representing an important future perspective for studies in the area. Optical Character Recognition (OCR) software can transcribe data from images; however, there are still barriers to its use for scientific research. Additionally, the legibility of labels is not always adequate, impairing automatic transcription of information. To date, one study used an artificial intelligence tool for the extraction of symbols from food labels (35), and one study used artificial intelligence to automatically extract written data (nutrition information panel and list of ingredients) from photographs (14). However, the photographs were captured from websites and the authors underscored the need for human validation to determine the accuracy of the extracted information (14).

We highlight two relevant methodological precautions regarding the treatment of data in food label databases. The first is the assessment and exclusion of duplicate foods in databases. Although the data collection app has measures to avoid duplicate entries, the arrangement of items in supermarkets, often in multiple locations, can lead to such occurrences. Another important factor for internal validation is quality control. Quality control can be carried out in several manners, depending on the purpose of the study, the time available to work on the database, and the technical capacity of the team. Procedures such as double-entry data tabulation, checking tabulated data in a subset or the entire sample, and performing concordance tests after checking are some of the possible strategies. Therefore, it is important to perform quality control procedures in the database, as well as in each study based on variables of interest. Furthermore, it is important to describe the quality control method adopted as proof of the methodological rigor of the study, as carried out by Aldhirgham et al. (15).

In the 2020 database, only duplicate entries were excluded. Absence of information such as the ingredients list and nutrient contents, among others, was not a reason for excluding foods from the database. Each study based on the database can establish criteria for including or excluding foods according to their variables of interest.

The analysis of quantitative data in food labeling, such as the nutrient declaration in the nutrition information panel, is generally well described in studies analyzing food labels, especially from a statistical point of view. However, the analysis of the ingredient list still seems to be a challenge. The ingredient list is an essential tool for assessing the nutritional quality of packaged foods. However, unlike the nutrient information, there is limited scientific literature focusing on the discussion of the list of ingredients (36, 37).

Although time-consuming, it is fundamental to systematically and individually analyze the ingredient list of all foods included in label studies, rather than solely conducting a search for predefined terms. A thorough analysis may identify potential nonconformities with current regulations in terms of food labels or food product composition. Additionally, it allows the detection of unexpected terms or ingredients. For example, regarding the presence of trans fat in packaged foods, if only terms related to hydrogenated vegetable fat are analyzed, the prevalence may be underestimated. Other ingredients containing trans fat may be listed in the ingredients list, as demonstrated by studies on the subject (26, 38). Two scoping reviews on food labeling studies underscored that the use of predefined terms to identify sweeteners and sugars in ingredient lists could underestimate the prevalence of these components in packaged foods (39, 40).

## 5 Limitations, strengths and perspectives for future research

It is important to note that this type of study may have some limitations, such as high costs, lengthy execution times, and challenges in updating the data (6). One of the challenging points of this approach is the periodic update of data collection, which must be conducted in-person and on-site. As previously mentioned, the cost and time required to collect and tabulate data through photographs is often a limitation. However, unlike other existing methods to collect information from food labels, in-store censuses allow analyzing how the information is available to food consumers at the time of purchase, in a real environment. There is a more complete picture of the reality, reducing the possibility of bias in the choice of samples for study purposes. Additionally, as an indirect result, in-store census methods may contribute to improving the quality of information provided to consumers, becoming relevant in the field of public health and nutrition.

Another limitation of the study is that data collection was carried out in one supermarket. However, in view of the continental dimensions of Brazil, our group chose a supermarket chain that is among the largest chains in the country. While due to this limitation it was not possible to capture regional variations in food availability considering the Brazilian territory, the chain sells a wide diversity of products and brands, 94% of which are sold nationwide, as previously discussed.

As perspectives for future research, it is noteworthy the relevance to building nationwide food labeling databases. This way, the information would be captured from real environments with the labeling information available to food consumers, both in-store and online platforms. It may also include other types of retail formats, such as cash-and-carry stores and—in countries where they are relevant (unlike in Brazil)—discounters. Additionally, it would enable the inclusion of regional variations of foods as well as local brands. Therefore, the packaged foods composition would be monitored from a public health perspective, through the development of cross-sectional and longitudinal studies using labeling information. However, it should be noted, as previously discussed, the challenges associated with building a nationwide food labeling database, especially regarding the costs and efforts needed to maintain and update the data in continental countries such as Brazil.

Additionally, technologies such as artificial intelligence and machine learning are becoming important tools for the construction of national food label databases through in-store census-type methodology, as they can automate time-consuming steps, such as data collection and tabulation. With the use of these tools, it would be possible to update the databases more frequently, covering a greater diversity of foods and brands marketed in the country. This approach could contribute to the monitoring of the composition of packaged foods through information available on labels. These tools may also prove valuable for analyzing qualitative label data, such as ingredients lists, by automating and accelerating the comprehensive assessment of this key labeling element.

## 6 Conclusion

This study underscores the essential role of research on food labeling in guiding the development and reformulation of public policies in food, nutrition, and health. The experience gathered with the process of building 4 food labeling databases in Brazil enabled methodological improvements to the research process, such as a more accurate data collection through food labels photographs, the inclusion of all packaged foods and beverages available for sale in the supermarket and the inclusion of more variables to the analysis. In view of the relevance of the topic, this study proposed a series of methodological recommendations related to data collection, data tabulation and data processing to be carefully considered, designed, and executed during planning, data collection, data tabulation, and database processing. Additionally, the experience permitted the identification of gaps and limitations related to the development of in-store census-type methods, such as the challenges to gather a representative sample of food labels, the difficulties on the transcription of the food labeling data, as well as the costs and efforts needed to maintain and update the data. Furthermore, greater rigor and detail are needed in the methods section of scientific articles on the subject, given that an important premise of the scientific method is replication. This methodological article underscores the importance of raising methodological discussions in the scientific literature to enhance the rigor of in-store census-type approaches.

## Data availability statement

The datasets presented in this article are not readily available because the authors do not have permission to share the dataset. Requests to access the datasets should be directed to MK, marianavskraemer@gmail.com.

## Author contributions

MK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. AF: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Validation, Writing – review & editing. MC: Investigation, Validation, Visualization, Writing – review & editing. TS: Investigation, Visualization, Writing – review & editing. BB: Investigation, Visualization, Writing – review & editing. EM: Investigation, Visualization, Writing – review & editing. MP: Investigation, Visualization, Writing – review & editing. PU: Investigation, Visualization, Writing – review & editing. GB: Investigation, Visualization, Writing – review & editing. NK: Visualization, Writing – review & editing. GA: Visualization, Writing – review & editing. RP: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) (grant number 88887.910128/2023-00) and by the Brazilian National Council for Scientific and Technological Development (CNPq) (grant number 303152/2015-0).

## Acknowledgments

The authors would like to thank all those involved in data collection and processing of NUPPRE Brazil 2010, 2011 and 2013 and NUPPRE/FoodSwitch Brazil 2020 food label databases.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2025.1568089/full#supplementary-material

## References

1. Popkin BM, Ng SW. The nutrition transition to a stage of high obesity and noncommunicable disease prevalence dominated by ultra-processed foods is not inevitable. *Obes Rev*. (2022) 23:e13366. doi: 10.1111/obr.13366

2. Lane M, Gamage E, Du S, Ashtree DN, McGuinness AJ, Gauci S, et al. Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses. *BMJ*. (2024) 384:e077310. doi: 10.1136/bmj-2023-077310

3. Kliemann N, Nahas AA, Vamos EP, Touvier M, Kesse-Guyot E, Gunter MJ, et al. Ultra-processed foods and cancer risk: from global food systems to individual exposures and mechanisms. *Br J Cancer*. (2022) 127:14–20. doi: 10.1038/s41416-022-01749-y

4. Juul F, Vaidean G, Parekh N. Ultra-processed foods and cardiovascular diseases: potential mechanisms of action. *Adv Nutr*. (2021) 12:1673–80. doi: 10.1093/advances/nmab049

5. Zhang YZ, Giovannucci EL. Ultra-processed foods and health: a comprehensive review. *Crit Rev Food Sci Nutr*. (2023) 63:10836–48. doi: 10.1080/10408398.2022.2084359

6. Pravst I, Hribar M, Žmitek K, Blažica B, Seljak BK, Kušar A. Branded foods databases as a tool to support nutrition research and monitoring of the food supply: insights from the Slovenian composition and labeling information system. *Front Nutr*. (2022) 8:576. doi: 10.3389/fnut.2021.798576

7. Food and Agriculture Organization of the United Nations. World Health Organization. Codex Alimentarius: General Standard for the Labelling of Prepackaged Foods. (2018). Available online at: https://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXS%2B1-1985%252FCXS_001e.pdf (Accessed August 20, 2024)

8. World Health Organization. Nutrition labelling: policy brief. (2022). Available online at: https://iris.who.int/bitstream/handle/10665/355295/9789240051324-eng.pdf?sequence=1 (Accessed August 20, 2024)

9. Open Food Facts. Open food facts: the free food products database. (2024). Available online at: https://world.openfoodfacts.org/ (Accessed August 20, 2024)

10. U.S. Department of Agriculture. FoodData Central. (2024). Available online at: https://fdc.nal.usda.gov/ (Accessed August 20, 2024)

11. Mintel. Food and drink. (2024). Available online at: https://www.mintel.com/insights/food-and-drink/. (Accessed August 20, 2024)

12. Westenbrink S, van der Vossen-Wijmenga W, Toxopeus I, Milder I, Ocké M. Leda, the branded food database in the Netherlands: data challenges and opportunities. *J Food Compos Anal*. (2021) 102:104044. doi: 10.1016/j.jfca.2021.104044

13. Maganja D, Miller M, Trieu K, Scapin T, Cameron A, Wu JHY. Evidence gaps in assessments of the healthiness of online supermarkets highlight the need for new monitoring tools: a systematic review. *Curr Atheroscler Rep*. (2022) 24:215–33. doi: 10.1007/s11883-022-01004-y

14. Ahmed M, Schermel A, Lee J, Weippert M, Franco-Arellano B, L'Abbé M. Development of the food label information program: a comprehensive Canadian branded food composition database. *Front Nutr*. (2021) 8:50. doi: 10.3389/fnut.2021.825050

15. Aldhirgham T, Alammari N, Aljameel GM, Alzuwaydi A, Almasoud SA, Alabbas NH, et al. The Saudi branded food database: first-phase development (branded beverage database). *J Food Compost Anal*. (2023) 120:105299. doi: 10.1016/j.jfca.2023.105299

16. Dunford E, Webster J, Metzler AB, Czernichow S, Mhurchu CN, Wolmaranset P, et al. International collaborative project to compare and monitor the nutritional composition of processed foods. *Eur J Prev Cardiol*. (2012) 19:1326–32. doi: 10.1177/1741826711425777

17. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and

crowdsourcing of national food composition data. *JMIR Mhealth Uhealth*. (2014) 2:e37. doi: 10.2196/mhealth.3230

18. Franco-Arellano B, Arcand J, Kim MA, Schermel A, L'Abbé MR. Progress towards eliminating industrially produced trans-fatty acids in the Canadian marketplace, 2013–2017. *Public Health Nutr*. (2020) 23:2257–67. doi: 10.1017/S1368980019004816

19. Brazilian Health Regulatory Agency. Resolution n° 259 of 2002. Approves the technical regulation of food labelling of packaged foods. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2002).

20. Brazilian Health Regulatory Agency. Resolution n° 360 of 2003. Approves the technical regulation of nutrition labelling of packaged foods. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2003).

21. Presidency of the Republic, Civil House. Law n° 10674 of 2003. Mandates that food products placed on the market disclose the presence of gluten. Brasília: Presidency of the Republic, Civil House (2003).

22. Brazilian Health Regulatory Agency. Resolution n° 26 of 2015. Provides requirements for mandatory labelling of the main foods that may cause food allergies. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2015).

23. Brazilian Health Regulatory Agency. Resolution n° 136 of 2017. Establishes the requirements for the mandatory declaration of lactose presence on food labels. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2017).

24. Brazilian Health Regulatory Agency. Resolution n° 727 of 2022. Establishes the requirements for labelling packaged foods. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2022).

25. Brazilian Health Regulatory Agency. Resolution n° 429 of 2020. Establishes requirements for the nutritional labeling of packaged food. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2020).

26. Silveira BM, Gonzalez-Chica DA, Proença RPC. Reporting of trans-fat on labels of Brazilian food products. *Public Health Nutr*. (2013) 6:2146–53. doi: 10.1017/S1368980013000050

27. Brazilian Association of Supermarkets. Ranking 2018. (2019). Available online at: https://www.abras.com.br/edicoes-anteriores/Main.php?MagNo=244#page/33. (Accessed August 20, 2024)

28. Presidency of the Republic, Civil House. Law n° 11265 of 2006. Regulates the marketing of food for infants and young children, as well as related childcare products. Brasília: Presidency of the Republic, Civil House (2006).

29. Brazilian Health Regulatory Agency. Resolution n° 54 of 2012. Provides requirements of the technical regulation governing supplementary nutritional information. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2012).

30. Brazilian Health Regulatory Agency. Resolution n° 359 of 2003. Approves the technical regulation of portions of packaged foods for nutritional labeling purposes. Brasília: Ministry of Health, Brazilian Health Surveillance Agency (2003).

31. Kraemer MVS, Fernandes AC, Ares G, Chaddad MCC, Pettigrew S, Scapin T, et al. Infant and children's exposure to food additives: an assessment of a comprehensive packaged food database. *J Food Compos Anal*. (2024) 134:106473. doi: 10.1016/j.jfca.2024.106473

32. Alcaire F, Giménez A, Ares G. Food additives associated with gut dysbiosis in processed and ultra-processed products commercialized in the Uruguayan market. *Food Res Int*. (2024) 191:114721. doi: 10.1016/j.foodres.2024.114721

33. Scapin T, Fernandes AC, Proenca RPC. Use of added sugars in packaged foods sold in Brazil. *Public Health Nutr*. (2018) 21:3328–34. doi: 10.1017/S1368980018002148

34. Bernstein JT, Christoforou AK, Weippert M, L'Abbé MR. Reformulation of sugar contents in Canadian prepackaged foods and beverages between 2013 and 2017 and resultant changes in nutritional composition of products with sugar reductions. *Public Health Nutr*. (2020) 23:2870–8. doi: 10.1017/S1368980020001159

35. Drole J, Pravst I, Eftimov T, Koroušić Seljak B. NutriGreen image dataset: a collection of annotated nutrition, organic, and vegan food products. *Front Nutr*. (2024) 11:1342823. doi: 10.3389/fnut.2024.1342823

36. Miller LMS, Cassady DL. The effects of nutrition knowledge on food label use. A review of the literature. *Appetite*. (2015) 92:207–16. doi: 10.1016/j.appet.2015.05.029

37. Anastasiou K, Miller M, Dickinson K. The relationship between food label use and dietary intake in adults: a systematic review. *Appetite*. (2019) 138:280–91. doi: 10.1016/j.appet.2019.03.025

38. Barros BYV, Proença RPC, Kliemann N, Hilleshein D, Souza AA, Cembranel F, et al. Trans-fat labeling in packaged foods sold in Brazil before and after changes in regulatory criteria for trans-fat-free claims on food labels. *Front Nutr*. (2022) 9:868341. doi: 10.3389/fnut.2022.868341

39. Milano E, Geraldo APG, Kraemer MVS, Ares G, Barros BV, Chaddad MCC, et al. Declaration of sweeteners on food labels: a scoping review of methodological issues and prevalence. *Nutr Rev*. (2024) 5:1–15. doi: 10.1093/nutrit/nuae175

40. Padovan M, Ares G, Scapin T, Kraemer MVS, Chaddad MCC, Fernandes AC, et al. Declaration of free sugars from fruits on food labels: a scoping review. *Br Food J*. (2024) 126:4342–56. doi: 10.1108/BFJ-05-2024-0547

# Frontiers in Nutrition

## Explores what and how we eat in the context of health, sustainability and 21st century food science

A multidisciplinary journal that integrates research on dietary behavior, agronomy and 21st century food science with a focus on human health.

---

## Discover the latest Research Topics

See more →


**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact