




Review

Data Integration for Diet Sustainability Analyses

Zach Conrad ^{1,2,*}, Alexandra Stern ³, David C. Love ^{4,5}, Meredith Salesses ⁶, Ashley Cyril ⁶, Acree McDowell ⁶ and Nicole Tichenor Blackstone ³

¹ Department of Health Sciences, William & Mary, Williamsburg, VA 23185, USA

² Global Research Institute, William & Mary, Williamsburg, VA 23185, USA

³ Division of Agriculture, Food, and Environment, Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA 02111, USA; alexandra.stern@tufts.edu (A.S.); nicole.blackstone@tufts.edu (N.T.B.)

⁴ Department of Environmental Sciences and Engineering, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21202, USA; dlove8@jhu.edu

⁵ Center for a Livable Future, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21202, USA

⁶ College of Arts & Sciences, William & Mary, Williamsburg, VA 23185, USA; mksalesses@email.wm.edu (M.S.); acyril@email.wm.edu (A.C.); eamcdowell@email.wm.edu (A.M.)

* Correspondence: zconrad@wm.edu; Tel.: +1-757-231-2037

Abstract: Diet sustainability analyses are stronger when they incorporate multiple food systems domains, disciplines, scales, and time/space dimensions into a common modeling framework. Few analyses do this well: there are large gaps in food systems data in many regions, accessing private and some public data can be difficult, and there are analytical challenges, such as creating linkages across datasets and using complex analytical methods. This article summarizes key data sources across multiple domains of food system sustainability (nutrition, economic, environment) and describes methods and tools for integrating them into a common analytic framework. Our focus is the United States because of the large number of publicly available and highly disaggregated datasets. Thematically, we focus on linkages that exist between environmental and economic datasets to nutrition, which can be used to estimate the cost and agricultural resource use of food waste, interrelationships between healthy eating and climate impacts, diets optimized for cost, nutrition, and environmental impacts, and others. The limitations of these approaches and data sources are described next. By enhancing data integration across these fields, researchers can be better equipped to promote policy for sustainable diets.

Keywords: sustainability; food system; integration; NHANES; environment; economic; climate change; diets; nutrition; natural resource use



Citation: Conrad, Z.; Stern, A.; Love, D.C.; Salesses, M.; Cyril, A.; McDowell, A.; Blackstone, N.T. Data Integration for Diet Sustainability Analyses. *Sustainability* **2021**, *13*, 8082. <https://doi.org/10.3390/su13148082>

Academic Editors:
Adam Drewnowski and
Timothy Griffin

Received: 14 June 2021
Accepted: 15 July 2021
Published: 20 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Food systems should promote food security and nutrition for current and future generations without compromising the economic, social, and environmental bases that support them [1]. Food systems transformation is integral to achieving the United Nations (UN) Sustainable Development Goals for 2030, including but not limited to greater access to safe and nutritious food (target 2.1), more sustainable agricultural systems (target 2.4), a reduction in premature mortality from non-communicable diseases (target 3.4), the promotion of safe and secure working environments (target 8.8), and the sustainable management of natural resources (target 12.2) [2]. To achieve these ambitious targets, the UN established the Decade of Action on Nutrition, which calls for increased investments across six critical action areas, including sustainable food systems [3]. To inform UN initiatives, identify synergies, and avoid unintended consequences, it will be critical to build research capacities that address multiple domains of sustainability simultaneously [4–6].

More integrated research approaches are needed to understand how diet patterns influence sustainability outcomes, known as diet sustainability analyses. Yet, persistent

barriers prevent advancements in this area of research [7]. Reviews of food systems studies found that most focus on just one or two domains of sustainability and overlook other domains [8,9]. For example, diet sustainability studies often overlook economic and social domains [8], while food security studies often overlook the environment domain [9]. Integrating analytic methods from multiple disciplines into a common modeling framework can be technically and conceptually demanding, and the requisite data are dispersed across federal agencies and other repositories. To fill this gap and facilitate advancements in diet sustainability research and policy, the objectives of this article are to (1) summarize key data sources from multiple domains of sustainability, and (2) describe methods and tools for integrating these data into a common analytic framework. We seek to share these data, methods, and tools with other researchers who are invited to use them in their own research, improve upon them, and apply them to other contexts.

2. Data Sources Overview

The following sections describe publicly available data sources across three sustainability domains—nutrition, micro-economic, and environment—that can be linked de novo or have pre-established linkages. We focus on the United States (US) because of the large number of publicly available and highly disaggregated food system datasets. While the majority of sustainability studies have focused on international or global contexts [8], the approaches used in the US can inform future efforts for food systems analyses in other countries.

Within the nutrition domain, we describe the primary sources of nationally representative data on food intake, as well as supplementary data sources that can be used to estimate the intake of food groups, ingredients, and nutrients. Throughout this article we refer to foods and beverages collectively as foods. Within the economic domain, we summarize data on consumer food prices. Within the environment domain, we summarize data sources on agricultural resource use (land, pesticides, fertilizer nutrients, and irrigation water) and environmental impacts (greenhouse gas emissions and cumulative energy demand). The subsequent sections describe how food intake data can be used to estimate diet quality using available tools (nutrition domain); how data on food intake and consumer food prices can be combined to estimate consumer food expenditures for food-at-home (FAH), food-away-from-home (FAFH), consumer food waste, and inedible portions (economic domain); and how data on food intake and environmental indicators can be integrated and analyzed using computational simulation models.

3. Food and Nutrition Data Sources

3.1. National Health and Nutrition Examination Survey (NHANES)

The National Health and Nutrition Examination Survey (NHANES) is the nation's most detailed source of nationally representative, individual-level dietary data. Staff at federal agencies have linked the NHANES data with supplementary databases [10,11] to provide additional detail for food groups (Food Patterns Equivalents Database) [12], individual foods (Food and Nutrient Database for Dietary Studies) [13], ingredients (Food Patterns Equivalents Ingredients Database and Food Commodity Intake Database) [14,15], and nutrients (Food Data Central) [16]. The NHANES collects cross-sectional data on health status, health behaviors, and demographic characteristics from approximately 5000 non-institutionalized individuals per year using in-person surveys, physical examinations, and laboratory tests conducted in mobile examination centers (MECs) with trained staff. Data have been collected continuously since 1999 and are released in two-year cycles (Table 1) [17]. The NHANES uses a stratified, clustered, four-stage sampling design. Some demographic groups are oversampled to increase the reliability and precision of subgroup analyses, and individuals are assigned survey weights that reduce the potential for bias from differential probabilities of selection and nonresponse [18].

Table 1. Data sources for diet sustainability analyses.

Data Source	Data Availability	Content	Coverage	File Formats	Hosting Institution
Food and nutrition					
National Health and Nutrition Examination Survey (NHANES)	1971–1975, 1976–1980, 1982–1984, 1988–1994, 1999–2018 (continuous)	Dietary data, health behaviors, anthropometric measurements, laboratory tests, and demographics	1971–1994: 16,000–40,000 respondents per period; 1999–2018: ~5000 respondents per year, >4600 foods	SAS export files (.xpt)	U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics
Food Data Central (FDC)	NA	Nutrient contents for database foods	>8000 foods and ~150 nutrients and food components	Comma separated values (.csv) and Microsoft Access (.accdb)	U.S. Department of Agriculture, Agricultural Research Service
Loss-adjusted Food Availability (LAFA)	1970–2019	Food availability data, and loss and waste rates.	~200 commodities	Microsoft Excel (.xlsx)	U.S. Department of Agriculture, Economic Research Service
Food composition					
Food and Nutrient Database for Dietary Studies (FNDDS)	2001–2018	Coding structure and recipes for NHANES foods	Nine primary food categories and 65 subcategories	Complete data provided as executable files for SAS (sas.exe) and Microsoft Access (access.exe). Some data are provided in Microsoft Excel files (.xlsx)	U.S. Department of Agriculture, Agricultural Research Service, Food Surveys Research Group
Food Patterns Equivalents Database (FPED) and Food Patterns Equivalents Ingredients Database (FPID) ¹	1999–2018	Conversion factors to estimate foods groups from NHANES foods	Nine primary food categories and 37 subcategories	Microsoft Excel (.xls), new for 2017–2018; Microsoft Access executable files (access.exe); SAS executable files (SAS.exe)	U.S. Department of Agriculture, Agricultural Research Service, Food Surveys Research Group
Food Commodity Intake Database (FCID)	2001–2010	Recipes for NHANES foods; conversion from NHANES foods to agricultural commodities	~500 commodities	Comma separated values (.csv)	US Environmental Protection Agency, Office of Pesticide Programs
Food Intake Converted to Retail Commodities Database (FICRCD)	1994–2008	Conversion from NHANES foods to agricultural commodities	65 commodities	SAS export (.xpt), Microsoft Access executable files (access.exe)	U.S. Department of Agriculture, Food Surveys Research Group

Table 1. Cont.

Data Source	Data Availability	Content	Coverage	File Formats	Hosting Institution
Economic					
Center for Nutrition Policy and Promotion Prices Database (CNPP Prices Database)	2001–2004	Food prices for NHANES foods	>4600 foods	Microsoft Excel (.xlsx)	U.S. Department of Agriculture, Food and Nutrition Service
Purchase-to-Plate Price Tool	Not publicly available as of this writing	Food prices for NHANES foods	~4600 foods	Microsoft Excel (.xlsx), possibly others	U.S. Department of Agriculture, Economic Research Service
Consumer Price Index	1913–2021	Average change in consumer prices	~15 food groups	Microsoft Excel (.xlsx)	U.S. Department of Labor, Bureau of Labor Statistics, Division of Consumer Prices and Price Indexes
National Household Food Acquisition and Purchase Survey (FoodAPS)	2013	Dietary data, health behaviors, and demographics	>4800 respondents	SAS (.sas), Stata (.dta), comma-separated values (.csv)	U.S. Department of Agriculture, Economic Research Service
Agriculture and environment					
Census of Agriculture	1840–2017	Characteristics of farms, ranches, and their operators	~2 million farms and ranches	Comma-separated values (.csv)	U.S. Department of Agriculture, National Agricultural Statistics Service
Agricultural Surveys	1850–2021	Characteristics of farms, ranches, and their operators	Up to 81,000 farms and ranches	Comma-separated values (.csv)	U.S. Department of Agriculture, National Agricultural Statistics Service
Irrigation and Water Management Surveys	2013 and 2018	Irrigation water application rates	~35,000 farmers and ranchers per five-year period	Comma-separated values (.csv)	U.S. Department of Agriculture, National Agricultural Statistics Service
Database of Food Impacts on the Environment for Linking to Diets (dataFIELD)	2001–2010	Greenhouse gas emissions and cumulative energy demand	~500 commodities	Microsoft Excel (.xlsx)	University of Michigan, Center for Sustainable Systems

¹ Includes MyPyramid Equivalents Database (MPED) that applies to NHANES 1999–2002. Does not include MPED that applies to the Continuing Survey of Food Intake by Individuals (CSFII) 1994–1996 and 1998.

The dietary component of the NHANES is What We Eat In America (WWEIA), which uses an in-person 24-h dietary recall administered by a trained interviewer [19]. The computer-assisted Automated Multiple Pass Method (AMPM) has been used since 2002 to minimize respondent burden and increase reliability and validity [20,21], and since that time approximately 80% of the sample completes a subsequent 24-h dietary recall administered by telephone 3–10 days after the first interview [19]. The AMPM includes five steps [22]. In the first step, respondents are asked to list the name of each food consumed from “midnight to midnight” on the preceding day, without indicating the amounts consumed. In many cases, respondents report consuming mixed dishes that include multiple ingredients, such as lasagna (subsequent sections describe supplementary databases that can be used to disaggregate these mixed dishes into their component ingredients). Next, seven probes about specific food groups are used to help respondents remember any omitted foods. Third, respondents are asked to indicate the time and eating occasion of each consumed food, which helps to identify any other forgotten foods. Fourth, respondents are asked to indicate the amounts of each food consumed, which are usually reported in their as-consumed amounts such as one banana, one slice of bread, or one sandwich, and visual prompts are used to help improve the accuracy of reporting. As part of this step, the Food and Nutrient Database for Dietary Studies (FNDDS), described below, is used to convert these foods from their as-consumed amounts to gram weights. The respondents are also asked to report whether each food was consumed at home or away from home. The last step provides respondents with specific cues to help them remember any foods not reported thus far, such as foods consumed in the car, in meetings, while shopping, or in other easily forgotten locations or situations. Approximately 4500 unique foods (and mixed dishes) are captured using WWEIA, and a portion of these foods are updated for each NHANES period to reflect new products and reformulations.

3.2. Food Data Central (FDC)

Food Data Central (FDC) provides nutrient content data for >8000 foods, which are derived from U.S. Department of Agriculture (USDA) contracted analyses, the scientific literature, calculations, and the food industry (Table 1) [23]. Approximately 150 nutrients and other components are represented, although not all of these are indicated for each food. A subset of approximately 3000 of these foods along with 65 nutrients and other components are aggregated in various combinations to represent each NHANES food, thereby providing nutrient content information for all 4500 foods included in the NHANES [24]. This linkage between the NHANES and FDC is made possible by the Food and Nutrient Database for Dietary Studies (FNDDS), which serves as a crosswalk between these two datasets, and is discussed below. FDC was launched in April 2019 and includes two extant data types (the USDA National Nutrient Database for Standard Reference and the USDA Global Branded Food Products Database) and two novel data types (Foundation Foods and Experimental Foods) [23].

The USDA National Nutrient Database for Standard Reference, more recently known as the Standard Reference Legacy Release, provides aggregated food composition data from USDA contracted analyses, the scientific literature, and calculations acquired or completed through 2018; this was the predominant data type for food composition until FDC was launched. The USDA Global Branded Food Products Database provides food composition data from branded and private label foods acquired through a public–private partnership between the USDA and the food industry, with data updated monthly. Foundation Foods provides food composition data and extensive metadata on the number of samples, sampling location, date of collection, analytical approaches, and agricultural information such as genotype and production practices [25]. Approximately 100 foods are currently included in Foundation Foods, but this data type represents the primary focus of USDA’s efforts to expand FDC in the future. Experimental Foods provides food composition data for the foods produced, acquired, or evaluated using alternative agricultural management systems, experimental genotypes, analytic protocols, or other innovative

conditions, and includes foods that may not be commercially available [26]. These foods will be linked with information on genetics, environmental inputs and outputs, supply chains, and economics [26].

3.3. Loss-Adjusted Food Availability (LAFA) Data Series

The Loss-adjusted Food Availability (LAFA) data series is based on the Food Availability data series, which provides information on approximately 200 minimally processed foods (i.e., commodities) available for human consumption in the US and Armed Forces overseas, not adjusted for loss and waste [27]. The per capita availability for each food is computed by estimating the difference between supply (production, imports, and beginning stocks) and disappearance (feed and seed, exports, ending stocks, and industrial uses), and the residual is divided by the US population. Data on supply and disappearance are directly measured or estimated using sampling and statistical methods [27]. To develop the LAFA data series, the USDA acquired food loss and waste rates from published reports and discussions with commodity experts, and applied these rates to each commodity in the Food Availability data series [28]. The USDA has ongoing efforts to continually improve and update these loss and waste estimates through partnerships and contracted analyses with industry groups and academic institutions [27]. The data for each food are presented in balance sheets that indicate the rates of loss from farm to retail, at the retail level, and several types of loss at the consumer level including inedible portions, cooking loss, and uneaten food [28]. After adjustment for losses, these estimates can be best understood as a proxy for food consumption and date back to 1970. The loss and waste rates can be manually linked with the Food Commodity Intake Database (FCID), which can then be used as a crosswalk to the NHANES to estimate individual-level food loss and waste [29]. These estimation procedures are described in subsequent sections.

4. Food Composition Data Sources

4.1. Food and Nutrient Database for Dietary Studies (FNDDS)

The FNDDS is a database that serves several critical functions that facilitate the NHANES data collection and analysis. First, it provides an eight-digit code for each food reported as consumed by NHANES respondents (here we refer to these as NHANES codes) to ensure that each food has a unique numerical identity in addition to its text description (Table 1). As such, it is the underlying database for the AMPM, to ensure that when NHANES respondents report eating a given food, that food is associated with a unique eight-digit code in the FNDDS. Second, since the majority of NHANES foods represent mixed dishes with multiple components, the FNDDS provides recipes that allow each NHANES food to be disaggregated into its component ingredients [24]. Each of these component ingredients in each NHANES food is linked with a unique five-digit code that represents a food in FDC. For example, if an NHANES respondent reported consuming pizza, the FNDDS would assign that an eight-digit code and link that to several five-digit codes in FDC that represent dough, mozzarella cheese, and tomato sauce. As FDC provides nutrient values for each of these component foods (e.g., dough, mozzarella cheese, and tomato sauce), this linkage thereby provides a way to estimate the nutrient content of each NHANES food by summing the nutrient content of their component ingredients from FDC. These linkages are established by USDA staff, and the final nutrient intake files for NHANES foods are made publicly available as downloadable files on the NHANES website [30]. A standard ontology has not been established; therefore, researchers interchangeably refer to these NHANES foods as “WWEIA foods” or “FNDDS foods”, which can be confusing to those not familiar with the intricacies of these data sources; here, we refer to them as NHANES foods. Third, the FNDDS converts NHANES foods from the forms in which they were reported as consumed by respondents (volume or conventional serving sizes, such as one slice of bread or one orange) into gram amounts, and these are the units provided in the publicly accessible NHANES data files [30]. The FNDDS categorizes these foods into nine primary food categories and 65 secondary food categories,

and the coding structure is updated for each NHANES period to reflect new products and reformulations.

4.2. Food Patterns Equivalents Database (FPED)

The Food Patterns Equivalents Database (FPED) converts each NHANES food (in mass) into one or more food groups (in serving sizes), based on the food groups included in the Dietary Guidelines for Americans: cup equivalents of fruit, vegetables, and dairy; ounce equivalents of grains and protein foods; teaspoon equivalents of added sugars; gram equivalents of oils and solid fats; and number of alcoholic drinks (Table 1) [12]. For example, according to the FPED, 100 g of cheese pizza (NHANES code 58106210) contains the equivalent of 0.11 cups of vegetables, 0.66 cups of dairy, 1.87 ounces of grains, 0.58 teaspoons of added sugars, 1.84 g of oils, and 8.04 g of solid fats. These food groups are further divided into 37 subgroups (for example, the dairy group includes milk, yogurt, and cheese, and the grain group includes refined grains and whole grains). The FPED is constructed by USDA staff in several steps, which are summarized here but provided in greater detail elsewhere [31]. First, internal recipe files derived from food labels, cookbook information, and standardized USDA handbooks are used to construct the Food Patterns Equivalents Ingredients Database (FPID), which converts each FDC food into the equivalent serving size for each of the FPED subgroups. Second, the FNDDS is used to group FDC foods, along with their FPID conversions, into various combinations to represent each NHANES food. The FPED has been updated for each NHANES period since 2005–2006 to reflect new products and reformulations [31]. The previous version of the FPED was the MyPyramid Equivalents Database (MPED), which links with data from the NHANES 1999–2004 (as well as with the Continuing Survey of Food Intake for Individuals 1994–1996 and 1998, not discussed here). Some food groups and conversions differ between the MPED and the FPED; therefore, these are not directly comparable [31].

4.3. Food Commodity Intake Database (FCID)

The Food Commodity Intake Database (FCID) provides information on the amount of approximately 500 commodity-level ingredients in each NHANES food (Table 1) [15]. While the FNDDS can be used to disaggregate the bun from a hamburger, the FCID can estimate the amount of wheat in that bun. This commodity-level resolution allows researchers to manually link FCID ingredients with food loss and waste rates provided in the LAFA data series [28], environmental impacts, and agricultural resource use [32], which will be discussed in subsequent sections. The FCID was developed by the US Environmental Protection Agency (US EPA) in conjunction with the Dietary Exposure Evaluation Model (DEEM) to estimate dietary exposure to pesticides, but researchers can use the FCID without the DEEM to disaggregate NHANES foods into their primary ingredients when commodity-level resolution is needed. The FCID links with NHANES data from 1999–2010, and the EPA does not have imminent plans for further updates.

4.4. Food Intakes Converted to Retail Commodities Database (FICRCD)

The Food Intakes Converted to Retail Commodities Database (FICRCD) [33] is similar to the FCID in that it can be used to convert dietary intakes from the NHANES into food commodities. Developed jointly by the USDA Economic Research Service (ERS) and Agricultural Research Service (ARS), the FICRCD links production and consumption by providing conversions of NHANES foods to 65 retail-level commodities such as fluid milk, fruits, vegetables, and meats (Table 1). The conversions are based on food preparation and cooking or processing losses. The FICRCD differs from the FCID in multiple ways including defined purpose, methods for disaggregating foods, and accessibility. The FICRCD was created for the broad purpose of converting NHANES foods into 65 retail-level commodities (i.e., food forms that appear in the grocery store), whereas the FCID was developed specifically to assess dietary pesticide exposure from agricultural commodities (i.e., food forms that appear on the farm). Commodities in the FCID are differentiated

by the cooking or processing method, and fat or water content because this can impact pesticide residues. For example, the FCID uses three separate agricultural commodities to represent cow's milk: milk fat, milk water, and nonfat-milk solids. In the FICRCD, NHANES foods are disaggregated into the retail-level commodities for cow's milk: whole, 2% fat, 1% fat, and skim milk.

Other differences in the databases appear in temporal coverage, the level of documentation, and file formats. The current version of the FICRCD connects with the NHANES 2007–2008, whereas the current version of the FCID connects with the NHANES 2009–2010. The FICRCD is accompanied by a 58-page user guide that provides detailed documentation of conversion factors and guiding principles for categorizing commodities [34], and data are provided in comma-separated values files [33]. The FCID does not provide documentation for use and development but has information on the main page of the website and in the frequently asked questions section, and data are provided in SAS and Microsoft Access formats [15].

5. Economic Data Sources

5.1. Center for Nutrition Policy and Promotion Prices Database (CNPP Prices Database)

The USDA Center for Nutrition Policy and Promotion (CNPP) Prices Database provides national average retail prices for each food reported as consumed in the NHANES 2001–2004 except alcohol (Table 1) [35]. All the foods were priced as if they were purchased at retail outlets for at-home consumption, such as at supermarkets, grocery stores, convenience stores, supercenters, farmers' markets, and other food stores, rather than at restaurants, cafeterias, or vending machines. These prices were derived from the 2001–2004 National Consumer Panel Homescan data [36], which provides information on food prices and other food attributes collected from participating households throughout the country; these are known as panel data or home-based scanner data [37]. Data on all the food purchased for at-home consumption is collected via handheld scanner devices or cell phone applications from a nationally representative sample of American households [37].

The CNPP Prices Database uses price data from approximately 700,000 food products collected from approximately 8500 households in the Homescan panel each year from 2001–2004 to derive prices for each NHANES food [37]. Each price was manually matched with an NHANES food by CNPP staff, and in some cases online cookbooks and other materials were used for disaggregation purposes. This resulted in approximately 75 price observations per food for about 90% of NHANES foods, and the remaining 10% represented foods that were consumed infrequently and in small quantities. Foods were converted from their purchased forms to their as-consumed forms by subtracting inedible portions, as well as moisture and fat loss and gains from cooking, using adjustment factors from FDC, USDA handbooks, and proxy matches. Finally, the multiple prices within each NHANES food were averaged [37].

5.2. Purchase-to-Plate Price Tool (PPPT)

The Purchase-to-Plate Price Tool (PPPT) provides national average retail prices for each food reported as consumed in the NHANES 2011–2012, based on the data collected in 2013 (Table 1) [38]. Similar to the CNPP Prices Database, the PPPT prices only represent the consumed portion of food [38]. Unlike Homescan, which is owned by Nielsen and based on the National Consumer Panel (i.e., panel data), the PPPT prices were derived from InfoScan, which is owned by Information Resources, Inc., and includes prices for approximately 350,000 products recorded by checkout scanners (i.e., store data). These data represent approximately 50% of all the retail food sales in the US [39]. Additionally, unlike the CNPP Prices Database, the PPPT uses the Purchase-to-Plate Crosswalk (PPC) to match price data with the FNDDS and other USDA-derived recipes using machine learning [40]. Similar to the CNPP Prices Database, the PPPT applies national average retail prices to all the foods reported as consumed by NHANES participants, regardless of whether it was consumed at home or away from home. The PPPT will also be extended to foods reported

as consumed in the NHANES 2013–2014 by matching to price data from 2015 [38]. The PPPT is not available to the public at the time of writing.

5.3. Consumer Price Index (CPI)

The Consumer Price Index (CPI) is a monthly measure of the average change in price of a market basket of goods and services. Approximately 75 foods that are most commonly purchased by consumers are included in the CPI and are represented in approximately 15 food categories (Table 1) [41]. Data are acquired from a random sample of retail outlets through a monthly survey, and these data are verified with store managers. The CPI can be used to inflate or deflate food prices to align with the year of dietary data collection [42,43].

5.4. National Household Food Acquisition and Purchase Survey (FoodAPS)

The National Household Food Acquisition and Purchase Survey (FoodAPS) is a cross-sectional, multi-stage survey that collected data on the foods acquired and factors that affect food acquisition decisions from April 2012 through January 2013 (Table 1) [44]. It is the only data source that includes nationally representative household-level expenditures for FAH and FAFH. The final sample consisted of 4826 households. Household food acquisition data were collected from a primary respondent using two in-person interviews, three telephone interviews, scanned food barcodes, and food receipts [45]. Eighty-five percent of FoodAPS foods were matched with an NHANES food from 2011–2012, 11% were matched with an SR food, <1% were assigned a new food code, and 3% were not assigned any food code [45].

6. Agricultural and Environmental Data Sources

6.1. Farm, Ranch, and Operator Characteristics

The Census of Agriculture collects data on the characteristics of farms, ranches, and their operators, and provides these data at the national, state, and county levels (Table 1) [46]. Data are collected every five years and include crop and livestock yields, land use, and total production. The USDA's goal is to account for any operation in which \geq USD 1000 of agricultural products are normally produced and sold, and survey participation is mandatory by federal statute [47]. Data are collected by mail, internet, telephone, and personal enumeration. To reduce the nonresponse bias, special efforts are made to collect data from all large or unique operations and Native American operators. Data gaps are filled by statistical imputation and the data are further calibrated to reduce bias from nonresponse, under coverage, and misclassification. Approximately 1.5 million operations provide data, and imputation and calibration methods account for an additional 500,000 operations [47].

USDA Agricultural Surveys collect annual data on the characteristics of farms, ranches, and their operators at the national, state, and county levels [48]. Data are collected throughout the year depending on the production structure of each crop, and include crop and livestock yields, land use, total production, and chemical applications [49]. Approximately 65,000–81,000 producers are surveyed every year, and producers with larger operations have a greater likelihood of being selected. Over 75% of interviews are conducted by phone and the remainder are conducted by mail and in person [49].

6.2. Agricultural Irrigation Water

USDA Irrigation and Water Management Surveys (formerly called Farm and Ranch Irrigation Surveys) collect data on national annual application rates of irrigation water (Table 1) [50]. All the producers who indicated irrigation activity on their Census of Agriculture reporting form are contacted by mail, and surveys can be submitted by mail, online, telephone, or in person [51]. The completion of surveys is mandatory. Surveys are conducted every five years and approximately 35,000 producers are surveyed for each data release. Data gaps are filled by statistical imputation and the data are further calibrated to reduce bias from nonresponse, under coverage, and misclassification [51]. Researchers should be aware that the application rates for irrigation water are reported for irrigated

land rather than total land; therefore, these rates should be adjusted for the amount of land that does not receive applications in order to estimate the average application rates for specific crops.

6.3. Environmental Impacts of Food Production

The database of Food Impacts on the Environment for Linking to Diets (dataFIELD) contains information on the greenhouse gas emissions (GHGE) and cumulative energy demand (CED) associated with the production of agricultural commodities and minimally processed ingredients (Table 1) [52]. Data were collected through a systematic review of food environmental life cycle assessments (LCAs) published between 2005 and 2016. Most data are from peer-reviewed journal articles (64%) and based on European production systems (63%). Emissions and energy use estimates do not include transportation or activities beyond the farm unless otherwise specified. The estimates were averaged across studies and connected to FCID commodities to estimate the environmental impacts of diets [53].

7. Other Data Sources

This article discusses a selection of data sources and methods that researchers can draw upon to conduct diet sustainability assessments, but it is not intended to be exhaustive. Researchers interested in state-level estimates of fruit and vegetable intake and other behavioral risk factors can utilize the Behavioral Risk Factor Surveillance System (BRFSS), the largest telephone-based health survey in the world [54]. Data on behavioral and social environmental risk factors among parent–teen dyads, such as food group intake and food environments, is available in the Family Life, Activity, Sun, Health, and Eating (FLASHE) survey [55]. For information on the school meals program in the US, including nutrient content and costs of meals, and student participation, dietary intake, and plate waste, researchers can use the School Nutrition and Meal Cost Study. In addition to the FNDDS and the FPED, NHANES users can utilize the What We Eat In America (WWEIA) Food Categories to categorize foods [56]. Consumer food spending can be estimated using the Consumer Expenditure Survey (CEX) [57].

8. Data Integration: Diet Quality

Diet quality is a multidimensional, quantitative construct that represents the healthfulness of diet patterns as an overall score and is typically measured using an index that captures the daily intake (or availability) of food groups, foods, and nutrients. Scoring algorithms are used to compute scores for each of these dietary components based on consumption amounts relative to a predefined standard, and total scores are computed by summing the scores for each component. Many different diet quality indices are available for use in research applications, each has its own strengths and limitations, and there is no single gold standard [58].

Here, we discuss several of the most commonly used indices in US studies that assess diet quality in divergent ways, but nonetheless similarly predict health outcomes [59,60]. The Healthy Eating Index (HEI-2015) [61,62] measures adherence with the 2015–2020 Dietary Guidelines for Americans [63], the Alternative Healthy Eating Index (AHEI-2010) evaluates the intake of food groups and nutrients that are associated with chronic disease risk [64], and the Nutrient-Rich Foods Index (NRF9.3) assesses the nutrient density of dietary patterns [60,65]. Researchers may consider using multiple indexes to comprehensively evaluate diet quality, which has been described [66] and demonstrated [64,67,68] by others.

8.1. Healthy Eating Index (HEI)

The Healthy Eating Index (HEI) evaluates the degree of adherence to the Dietary Guidelines for Americans (DGA), and is updated to reflect the recommendations in each version of the DGA [61]. The HEI was originally developed in 1995 and was updated in

2005, 2010, and 2015. As of this writing the HEI-2015 [69] is the most current version but the HEI-2020 is expected to be released soon. It is recommended that researchers use a single version of the HEI when evaluating diet quality across different years [69]. The HEI-2015 includes nine components to encourage (total fruit, whole fruit, total vegetables, greens and beans, whole grains, dairy, total protein foods, seafood and plant proteins, and the ratio of unsaturated to saturated fats) and four components to limit (refined grains, sodium, added sugars, and saturated fats) [69]. The density method is used to standardize the consumption amounts for each component to a 1000 kcal basis [70]. Not all the components are scored similarly, with some being scored from 0–5 and some being scored 0–10, and intermediate intakes are scored proportionally. Components are scored differently from one another for a variety of reasons, such as to ensure face validity, to follow precedent, and to represent a range of observed intakes that vary between the components [69]. Reverse scoring is applied to components to limit to ensure that higher scores are more favorable [69]. The scores for each component are summed to compute a total score for each respondent, with a maximum of 100. Researchers can choose from five distinct analytic methods to compute HEI scores based on the structure of their data and the purpose of their research [71].

8.2. Alternative Healthy Eating Index (AHEI)

The Alternative Healthy Eating Index (AHEI) measures the intake of dietary components associated with chronic disease risk. It was originally developed in 2002 [72] and was updated in 2005 and 2010 [64]. The AHEI-2010 includes six components to encourage (vegetables, fruit, whole grains, nuts and legumes, long-chain ω -3 fats, total polyunsaturated fats) and five components to limit (sugar-sweetened beverages and fruit juice, red and processed meat, sodium, alcohol, and trans fats). The trans fat content of foods in the NHANES is incomplete; therefore, researchers have omitted this component when computing AHEI scores [32,73,74]. This omission is unlikely to affect overall scores because the intake of *trans* fats has decreased dramatically since 1999 [73]. Each component is scored from 0–10 and the components to limit are reverse scored to ensure that higher scores are more favorable. Higher scores are awarded for a moderate consumption of alcohol. The component scores are summed for each individual to compute an overall score with a maximum of 110 if the trans fat component is included or 100 if the trans fat component is excluded. The AHEI is not energy adjusted but researchers can perform this adjustment [70], which will be necessary for studies that include groups with different energy needs, such as children and adults, by standardizing to the mean or median energy intake of the source population (mean = 1849 kcal/day, median = 1800 kcal/day) on which the AHEI was initially constructed [64] (for example, see Bernstein et al. [75] and Conrad et al. [32]). Researchers using the AHEI to evaluate diet quality in non-adult populations should consider modifying the alcohol scoring standards to ensure that the maximum number of points (10) are awarded for zero consumption and zero points are awarded for any consumption [32].

8.3. Nutrient-Rich Foods Index (NRF)

The Nutrient-Rich Foods (NRF) index assesses the nutrient density of dietary patterns by comparing the intake of nutrients to encourage to the intake of nutrients (and one food component) to limit [60,65]. The NRF index can also be used to evaluate the nutrient density of foods [60], which is not discussed here. There are multiple versions of NRF that are differentiated by the types of nutrients to encourage, and all versions include the same three nutrients (and one food component) to limit, which are as follows: saturated fat, added sugar, and sodium. Validation analyses demonstrated that NRF9.3 performed the best against the HEI-2005 and includes the following nine nutrients to encourage: protein; fiber; vitamins A, C, and E; calcium; iron; magnesium; and potassium [60]. The intake of each nutrient is measured per 100 kcal of each food and is evaluated against its Daily Reference Value (based on 2000 kcal/day) established by the US Food and Drug Administration and capped at 100%. The scores for each nutrient to limit are summed and

then subtracted from the sum of nutrients to encourage [60,65]. The minimum NRF9.3 score for each food is -300 and the maximum score is 900 . The total scores for each respondent are computed by averaging their food scores weighted by the consumption amount of each food. Recently, a new scoring system to measure nutrient density was developed, known as the Nutrient Rich Food hybrid (NRFh) score, that measures the intake of food groups as well as nutrients, and was validated against the HEI-2015 [76].

9. Data Integration: Food Loss and Waste

Food loss and waste (retail loss, inedible portions, and consumer waste) can be estimated through a three-step procedure that links the LAFA data series, the FCID, and the NHANES [28,32]. The first step requires applying simple algebra to the food loss/waste rates and the available data in the LAFA data series to disaggregate inedible portions from cooking loss and uneaten food, which are otherwise aggregated under the heading “loss at the consumer level” (these calculations are described in detail elsewhere [77]). Data are not available in the LAFA data series to disaggregate cooking loss from uneaten food; therefore, it can be assumed that these collectively represent consumer food waste. In the second step, hand-coding is used to link each food in the LAFA data series with a distinct food in the FCID based on the similarity of their descriptions. Others have provided a framework for this procedure that involves two investigators performing these matches independently with infrequent differences resolved through discussion and consensus [28]. Successful matches can be achieved for $>90\%$ of FCID foods and the remainder can be reasonably excluded from analyses due to infrequent and minute intake by the general population [28].

The third step requires linking the FCID with the NHANES, and the EPA has already established this linkage for 2001–2010. The linkage to subsequent NHANES waves is incomplete because new food codes have been added during that time and some food codes have been discontinued, with progressively fewer links as further NHANES waves are released. Still, investigators aiming to estimate loss/waste for the NHANES waves from 2011–2012 onward have several options for doing so. The first option is to proceed with incomplete linkages for these later waves, but this may only be defensible if a sufficient number of waves for which there are complete linkages are included in the analyses and the investigators are careful to mention that this method will underestimate loss/waste. Others have demonstrated that this method underestimated Total Food Demand (sum of retail loss, inedible portions, consumer waste, and consumed food) by 11% for the NHANES 2005–2016, although this did not appear to vary by quintiles of diet quality [32]. Yet, the defensibility of this approach diminishes for each subsequent NHANES wave due to progressively fewer matches. The second option is to impute the missing loss/waste values. A simple method is to use the average of all the foods within each food category weighted by the consumption amount of each food within each food category. The validity of this approach increases as more food categories are established, owing to the greater likelihood that highly differentiated food categories will represent the individual foods within those categories. A useful approach for creating these food categories is to adopt the FNDDS coding scheme, which uses the first few digits in each NHANES food code to categorize each food into progressively more differentiated food categories, which can result in >40 different food categories [43]. Others have demonstrated that this approach underestimated daily consumer food costs by only 4.5% for the NHANES 2001–2016 [43]. The third option is for researchers to establish new FCID–NHANES linkages for 2011 onward, which the authors of the present paper are pursuing.

10. Data Integration: Food Prices

Researchers aiming to acquire food prices for NHANES foods face several barriers to doing so, yet all can be overcome to some degree. These barriers relate to under coverage, inflation, accounting for the cost of loss/waste, and accounting for the price difference between food-at-home (FAH) and food-away-from-home (FAFH).

10.1. Undercoverage and Inflation

The most recent version of the CNPP Prices Database only aligns with the NHANES 2003–2004, which presents problems of under coverage and inflation (these issues also pertain to the PPPT but to a lesser degree). As discussed above, NHANES food codes are modified over time; therefore, successful linkages with other databases erode with each new NHANES wave unless efforts are made to iteratively establish new linkages. Researchers will not be able to ignore the severe under coverage that results from linking the CNPP Prices Database with more recent NHANESs, which will prohibit valid analyses. Instead, researchers can impute these missing values by taking the average of all the foods within each food category weighted by the consumption amount of each food within each food category, as discussed above. Again, the validity of this approach increases as more food categories are established, and researchers are advised to use the FNDDS coding scheme for this purpose. Researchers may also want to establish new linkages for 2004 onward. This approach will not address the issue of inflation, but researchers can use the CPI to inflate food prices to align with the relevant year of dietary data collection in the NHANES [42,43,78]. A limitation of this approach is that food price inflation data are available for only 15 major food categories; therefore, this may result in over-generalized estimates for certain foods.

10.2. Food Loss and Waste

The price that consumers pay for food includes the cost of the consumed portion, inedible portion, and wasted portion, but the prices supplied by the CNPP Prices Database and the PPPT only represent the consumed portion [37,38]. Food waste and inedible portions account for approximately 26 and 16% of the weight of purchased food, respectively [32]; therefore, the failure to account for these portions will underestimate total food expenditures. Others have demonstrated that food waste and inedible portions account for 27 and 14% of total food expenditures, respectively [43]. Researchers can use the data sources and approaches discussed above to estimate the cost of loss and waste for each NHANES food by multiplying the unit price (e.g., price per gram) of the consumed portion by the amount lost and wasted.

10.3. Food-away-from-Home

Food prices vary markedly depending on whether they were purchased for at-home consumption (FAH) or away-from-home consumption (FAFH) because substantial value is added for consumer experience and convenience at FAFH outlets. However, the CNPP Prices Database and the PPPT do not include FAFH prices; therefore, they assign FAH prices to all the foods reported as consumed by NHANES participants. Recent data from the USDA ERS demonstrate a dramatic increase in consumer spending on FAFH over the last few decades, which now represents approximately 50% of total food spending [79]; therefore, researchers using the CNPP Prices Database or the PPPT to estimate food prices may want to adjust the price of FAFH to avoid underestimating total expenditures. An expert panel to the USDA ERS has suggested theoretical options for this adjustment using the FoodAPS [80], which is the only source of data that differentiates spending on FAH from FAFH at the individual level. A simple implementation of this concept has been demonstrated by others [43] and is depicted in Figure 1. First, the NHANES provides information about whether a food was consumed at home vs. away from home, and these data can be linked to the CNPP Prices Database or the PPPT to estimate the FAH price of each FAFH as well as the amount consumed. Second, data from the FoodAPS can be used to derive a coefficient that represents the ratio between the average price paid for each FAH to the average price paid for each FAFH for each major food category. Finally, this coefficient can be multiplied by the price of each FAFH in the linked NHANES–CNPP Prices Database file (or the NHANES–PPPT file) to derive its adjusted price. To increase the data resolution, researchers may want to derive FAH-to-FAFH price ratios for each food rather than each food group, which will require hand-matching the 15% of FoodAPS codes

that are not already matched with NHANES foods. Researchers should also be mindful that these derived FAH-to-FAFH price ratios only represent data from 2012–2013 owing to the last year that FoodAPS data were collected.

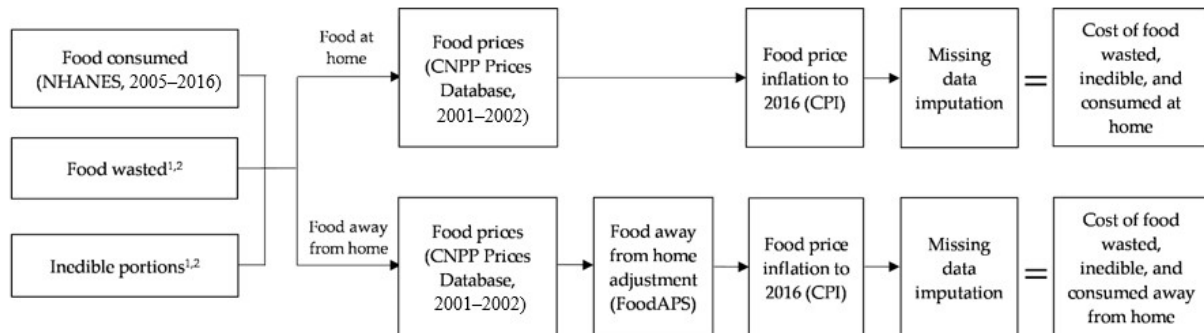


Figure 1. Methodology for estimating the cost of food purchased, wasted, inedible, and consumed at home and away from home. Adapted with permission from: Conrad, Zach. (2020). Daily cost of consumer food wasted, inedible, and consumed in the United States, 2001–2016. *Nutrition Journal*, 19:35. CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>, accessed on 31 May 2021). Minor changes were made to the original version to more clearly represent food loss and waste, and missing data imputation. ¹ Conrad et al. (2018). Relationship between diet quality, food waste, and environmental sustainability. *PLoS ONE*, 13:e0195405. ² Conrad, Zach. (2019). Daily cost of consumer food wasted, inedible, and consumed in the United States, 2001–2016. *Nutrition Journal*, 19:35. NHANES: National Health and Nutrition Examination Survey. CNPP: Center for Nutrition Policy and Promotion, U.S. Department of Agriculture. CPI: Consumer Price Index. FoodAPS: National Household Food Acquisition and Purchase Survey.

11. Data Integration: Biophysical Modeling

Data on food intake, loss/waste, agricultural chemical application rates, and water irrigation rates can be integrated by inputting them into computer models such as Foodprint [81], which can be used to estimate the amount of agricultural land, fertilizer nutrients, pesticides, and irrigation water needed to meet specific dietary patterns (Figure 2) [32]. Foodprint can also be used to estimate the number of people that can be fed a nutritionally adequate diet on a given area of land (i.e., population carrying capacity) [81,82]. Foodprint is a generalized biophysical simulation model that represents a given geographic locale as a closed food system and can be modified to represent food systems at any spatial scale [81,83–85]. Users must parameterize the model to reflect the agricultural conditions of the desired locale, which include crop and pasture yields, livestock output (e.g., milk produced per cow), the availability of agricultural land for specific purposes (e.g., cropland, pasture, and non-productive land), and whether local climatic and soil conditions can support specific crops (e.g., tropical fruits).

Users input data on the daily per capita consumption of 22 food groups in their as-consumed forms (grains; dark green vegetables; red and orange vegetables; dry beans, lentils, and peas; starchy vegetables; other vegetables; fluid milk and yogurt; cheese and other dairy; soy milk; nuts; tofu; beef; pork; chicken; turkey; eggs; seafood; plant oils; dairy fats; lard and tallow; and sweeteners) [81]. The embedded computations transform these foods back to raw agricultural crops (grains, fruits, vegetables, legumes, nuts, sweeteners, feed grains and oilseeds, hay, cropland pasture, and permanent pasture) and the associated amount of agricultural land needed to produce them by modeling their stepwise transformation as they progress through the various stages of a given food system. The transformation parameters include population size, food processing conversions, loss/waste, livestock feed requirements, crop and livestock yields, the availability of agricultural land, and the suitability of agricultural land for food production. The embedded calculations also account for multi-use crops (i.e., crops that are used to produce multiple products from equivalent mass) and multi-use cropland (cropland used to produce multiple crops during different parts of the year) [81].

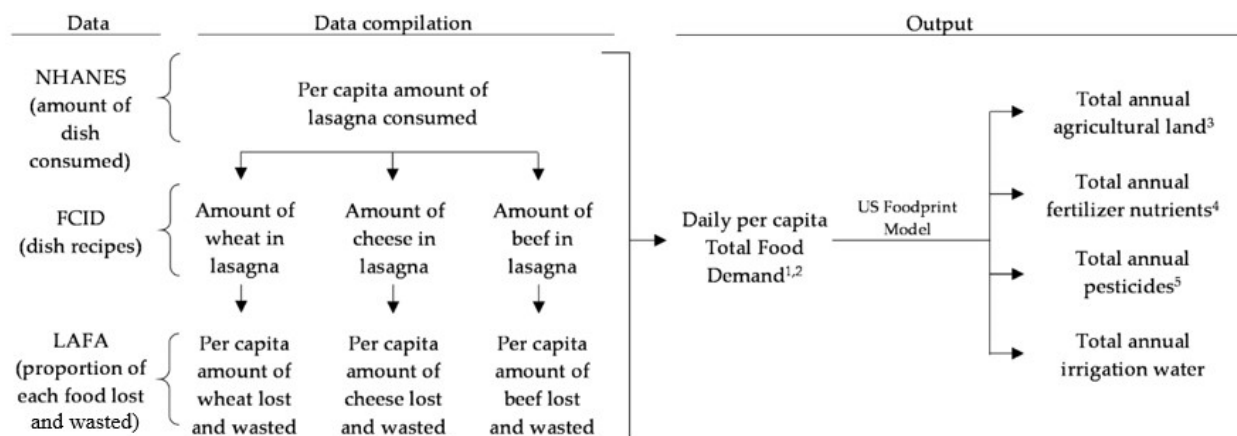


Figure 2. Methodology for estimating agricultural resource use of self-selected diets through biophysical simulation modeling. Adapted with permission from: Conrad, Zach; Blackstone, Nicole Tichenor; Roy, Eric. (2020). Healthy diets can create environmental trade-offs, depending on how diet quality is measured. *Nutrition Journal*, 19:117. CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>, accessed on 31 May 2021). No changes were made to the original version. LAFA, Loss-adjusted Food Availability data series. FCID, Food Commodity Intake Database. NHANES, National Health and Nutrition Examination Survey. ¹ Includes retail loss, inedible portions, consumer waste, and consumed food. ² Meat and mixed meat dishes (beef and beef mixed dishes; pork and pork mixed dishes; poultry and poultry mixed dishes; seafood and seafood mixed dishes; meat sandwiches, burgers, sausages, and hotdogs; bacon; and other meat dishes); eggs and egg dishes; dairy (milk and cream, cheese); soup; grains and mixed grain dishes (bread; breakfast cereal; pancakes, waffles, and French toast; pastas and grain mixtures; pizza and calzones; and grain-based desserts); nuts and seeds; fruits and vegetables in mixed dishes (whole fruit and mixed fruit dishes; fruit/vegetable juice; dark green vegetables; yellow and orange vegetables; tomatoes and tomato mixtures; legumes; other vegetables); potatoes and potato mixed dishes; margarine, table oils, and salad dressings; salty snacks; Mexican dishes; other foods and dishes. ³ Grains, fruits, vegetables, legumes, nuts, sweeteners, feed grains and oilseeds, hay, permanent pasture, and cropland pasture. ⁴ Sum of nitrogen, phosphorus (P₂O₅), and potash (K₂O). ⁵ Sum of insecticides, herbicides, and fungicides.

Foodprint is a publicly available spreadsheet model that does not require highly specialized computer software; therefore, users can modify it to suit their research needs [81]. As discussed above, others have modified the model for different spatial scales that include national [83], state [85], and sub-state levels [84]. Others have used regression models and time-series data on food intake, crop yields, and population size to project agricultural land use and population carrying capacity to 2030 [82]. Although Foodprint was not originally designed to produce variance estimates, users can utilize Microsoft Excel's Visual Basic for Applications (VBA) programming language to write built-in macros that draw from the variance estimates produced from individual-level dietary data analyses, such as from the NHANES [32].

12. Data Integration: Life Cycle Assessment Modeling

The GHGE and CED of individual self-selected diets in the US can be estimated using the dataFIELD, which connects environmental impact estimates from LCA studies to agricultural commodities in the FCID, thereby providing a linkage to NHANES consumption data [53,86,87]. As discussed above, researchers created the dataFIELD by conducting a systematic review of food LCAs to identify the environmental impacts associated with FCID commodities, and the impacts were averaged across studies for each commodity [53]. On average, each commodity is represented by 11 data points. Data gaps were filled by averaging the impacts of similar commodities. Conversion factors from the FICRCD [33] and USDA handbooks [88] were applied to align the impacts with the weight basis description of FCID commodities. To estimate the impacts from food waste, researchers linked FCID commodities with LAFA loss and waste rates, as described above [53]. The dataFIELD is publicly available in a Microsoft Excel format, and can be readily modified for a variety of research purposes [52]. Modifications might include excluding data points not relevant

to the study purpose, weighting data points based on consumption or import data, and altering the system boundaries to align with the scope of the research.

13. Limitations

13.1. Dietary Data

Measurement error is inherent in all scientific endeavors but poses unique challenges when studying dietary patterns. Unlike energy intake, which can be objectively measured using the doubly labeled water technique, there is no objective way to measure dietary patterns. Ultimately, researchers must rely on self-reported food intake to understand what, when, where, how, and why people eat, which is subjective. Although a priori sampling methods and post hoc statistical techniques will mitigate bias, they will not eliminate it. All surveys, including the NHANES, suffer from a social desirability bias that occurs when respondents conform their responses to be seemingly more favorable, such as over-reporting the intake of foods perceived to be healthy and under-reporting the intake of foods perceived to be unhealthy. Reactivity can also introduce bias, which occurs when respondents anticipate the data collection and alter their food intake accordingly. Dietary collection instruments such as 24-h recalls and food frequency questionnaires rely on memory, which is not infallible. Despite all these challenges, self-reported food intake continues to provide a rich source of information on dietary patterns from large populations [89].

13.2. Food Loss and Waste

Food loss and waste data from the LAFA data series are useful to food systems researchers but they are not without limitations. The LAFA data series provides a single estimate of the proportion of each food lost and wasted that does not vary temporally, geographically, or by type of food outlet (although different rates are provided for each processing type of each food, such as dried, canned, frozen, fresh, and juice) [27]. This limits analyses across time and place and requires that the food-specific loss/waste rates be applied consistently to FAH and FAFH. When linking these data to the NHANES, researchers should be cognizant that any variance in the final estimates will be due to inter-individual differences in food intake rather than different food loss/waste rates across individuals. It is also possible that some NHANES respondents consumed a portion of their meal away from home but later consumed the leftovers at home, which would misclassify loss/waste estimates of FAH and FAFH. Finally, the lack of uncertainty values provided by the LAFA data series could result in overly narrow variance estimates when merged with the NHANES or other survey-based data.

13.3. Food Prices

To the best of our knowledge, there is currently no publicly available data source that provides contemporary, individual-level information on dietary patterns and scanner-based food prices differentiated by purchase location, which severely limits comprehensive sustainability analyses. To address this need, we have presented a method that links the following four data sources: the NHANES (dietary patterns), the CNPP Prices Database or the PPPT (FAH prices), the FoodAPS (FAFH prices), and the CPI (food price inflation) [43]. Despite its utility, researchers should be aware of the limitations of this approach. Data from the NHANES can be linked with the CNPP Prices Database or the PPPT at the food level, which provides a high degree of resolution, but data from the CPI are only available for approximately 15 food groups, which may produce overly generalized inflation estimates. Users should use the most recent price data available and minimize the time period that is being inflated/deflated to limit uncertainty. A similar issue of generalizability arises when applying FAFH prices from the FoodAPS to dietary intake data from the NHANES, although researchers may be able to overcome this barrier by establishing novel food-level linkages for the remaining 15% of FoodAPS foods that have not already been linked to the NHANES. Finally, measurement error is embedded in all of these data sources but

computing combined variance estimates resulting from these data linkages remains a persistent challenge that requires further scientific study.

13.4. Agricultural Resource Use

Foodprint is a generalized tool that integrates data from diverse sources to estimate agricultural resource use and population carrying capacity and does not produce stratified estimates for multiple scenarios (e.g., subpopulations or agricultural subsystems) within a single simulation. However, Foodprint is highly modifiable and researchers can re-parameterize the model for each scenario, run separate simulations for each scenario, and then compare the outputs using z-tests [32]. Scenarios may represent diets differentiated by diet quality or other characteristics, and outcomes may represent the use of agricultural resources such as land, irrigation water, fertilizer nutrients, and pesticides [32]. Another method to conduct stratified analyses is to modify the VBA code to produce individual-level estimates that can then be stratified and tested using *t*-tests or Wald tests.

Foodprint also represents a closed system, meaning that the food demand of a given population is met by the agricultural system of that locale rather than imports. Therefore, the demand for foods that cannot be produced within that locale is proportionally apportioned to the other foods within that same food group according to their availability in the LAFA data series [81]. With finite planetary resources, there is an increasing need to evaluate the degree to which individual geopolitical locales can provide enough food for their own populations within the limitations of their biophysical systems. The carrying capacity can be increased with international imports, but not all locales can similarly and simultaneously increase their imports due to the finite resources of the planet. Although food imports are critical for stabilizing the seasonal fluctuations in availability and price, and their share of domestic consumption has increased over time, the majority of American food demand is still met with domestic production (89% by volume and 85% by value) [90]. Researchers interested in modifying Foodprint to account for international trade patterns may be able to do so by incorporating data on the import share of consumption for each food [90–92].

13.5. Environmental Impacts of Agriculture

The dataFIELD provides an essential resource for estimating the GHGE and CED of US foods and diets because it is highly transparent, comprehensive, and connected to the FCID. However, there are limitations to this database due to its breadth and the limited availability of LCA data [52]. The dataFIELD represents compiled estimates from a wide range of LCAs with varying geographies, system boundaries, units of analysis, and assumptions [53]. The database manages some of this variability by standardizing the system boundary (i.e., cradle-to-farmgate or processor gate), but this does not account for differences in what is included in the boundary, how the systems are modeled, or the methods used for calculating emissions. For example, an LCA of apples might calculate the emissions of GHGs from managed soils differently than an LCA of beef. It is unclear how these differences might impact the overall results, partially because of the large number of studies included. The dataFIELD is unique, however, in that it provides uncertainty estimates for each commodity. This can help researchers more appropriately account for the limitations described here.

Other limitations stem from the lack of LCAs performed in the US. Most of the data in the dataFIELD are from European systems, where agricultural production methods might differ compared to US production systems, depending on the product [52]. The database does not include estimates for other important environmental impacts such as effects on biodiversity, land use, or water and air quality. Researchers have developed parallel methods to evaluate the water scarcity footprints of individual diets, and although these data are not integrated into the dataFIELD, they are publicly available [93]. Ideally, information on variation in food consumption would expand to include details on supply

chains and sourcing, and this could be linked to regionally specific impact estimates to more accurately reflect the environmental implications of diets [53].

13.6. Interpreting Uncertainty

Science is a systematized process of understanding the natural world, which requires explicit recognition that it is never possible to measure every observation and outcome with absolute certainty. Tightly controlled clinical studies with small sample sizes can often achieve greater internal validity than population-based modeling studies, but they often suffer from lower external validity. Data integration science offers tools to link clinical and population data, perform extrapolation procedures, and draw implications for larger scales and future conditions, and is particularly useful for evaluating sustainability outcomes as they relate to food systems. These tools are routinely used in various other fields such as climate and weather science, medicine and health care, transportation and public infrastructure planning, commercial development, business, and many others. The aphorism “All models are wrong but some are useful”, often attributed to the statistician George E. P. Box [94], is applicable. Food systems scientists performing data integration procedures should be explicit about limitations, get creative with solutions, and rigorously test assumptions. To do this, it is critical that we cultivate collaborative and scientifically rigorous environments where meaningful advancements in data integration can proceed.

14. Conclusions

Building research capacities that simultaneously address multiple domains of diet sustainability are critically needed to inform global health, environmental, and development initiatives. These advancements can only occur by addressing persistent barriers that slow data integration and knowledge transfer. To fill these gaps, this article discusses key data sources from multiple sustainability domains and disciplines and describes methods and tools for integrating and analyzing these data. Other researchers are invited to use these data, tools, and methods in their own research, to improve upon them, and to apply them to other contexts. Researchers should pursue this agenda with the recognition that it is never possible to measure every observation and outcome with complete certitude, while at the same time bringing all of their resources to bear to reduce bias and uncertainty.

Author Contributions: Conceptualization, Z.C.; methodology, Z.C., A.S., D.C.L. and N.T.B.; data curation, Z.C., A.S., D.C.L., M.S., A.C., A.M. and N.T.B.; writing—original draft preparation, Z.C. and A.S.; writing—review and editing, Z.C., A.S., D.C.L., M.S., A.C., A.M. and N.T.B.; visualization, Z.C.; supervision, Z.C.; project administration, Z.C.; funding acquisition, Z.C. and D.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: Z.C. was supported by the Commonwealth Center for Energy and the Environment at William & Mary. D.C.L. was supported by the U.S. Department of Agriculture under an INFEWS grant [#2018-67003-27408].

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. High Level Panel of Experts on Food Security and Nutrition. Nutrition and Food Systems; Rome. 2017. Available online: <http://www.fao.org/cfs/cfs-hlpe/hlpe-reports/report-12-elaboration-process/en/> (accessed on 31 May 2021).
2. United Nations. Sustainable Development Goals. 2021. Available online: <https://sustainabledevelopment.un.org/?menu=1300> (accessed on 31 May 2021).
3. United Nations. *The United Nations Decade of Action on Nutrition*; Food and Agriculture Organization: Rome, Italy, 2016. Available online: <https://www.un.org/nutrition/> (accessed on 5 January 2021).
4. United Nations. *Sustainable Healthy Diets: Guiding Principles*; Food and Agriculture Organization; World Health Organization: Rome, Italy, 2019. Available online: <http://www.fao.org/3/ca6640en/ca6640en.pdf> (accessed on 31 May 2021).
5. Kennedy, E.; Raiten, D.; Finley, J. A view to the future: Opportunities and challenges for food and nutrition sustainability. *Curr. Dev. Nutr.* **2020**, *4*, nzaa035. [CrossRef] [PubMed]

6. Finley, J.W.; Dimick, D.; Marshall, E.; Nelson, G.C.; Mein, J.R.; Gustafson, D.I. Nutritional sustainability: Aligning priorities in nutrition and public health with agricultural production. *Adv. Nutr.* **2017**, *8*, 780–788. [[CrossRef](#)] [[PubMed](#)]
7. Finley, J.; Jaacks, L.M.; Peters, C.J.; Ort, D.R.; Aimone, A.M.; Conrad, Z.; Raiten, D.J. Perspective: Understanding the intersection of climate/environmental change, health, agriculture, and improved nutrition—A case study: Type 2 diabetes. *Adv. Nutr.* **2019**, *10*, 731–738. [[CrossRef](#)] [[PubMed](#)]
8. Reinhardt, S.L.; Boehm, R.; Blackstone, N.T.; El-Abbadi, N.H.; McNally Brandow, J.S.; Taylor, S.F.; DeLonge, M.S. Systematic review of dietary patterns and sustainability in the United States. *Adv. Nutr.* **2020**, *11*, 1016–1031. [[CrossRef](#)]
9. Farmery, A.K.; Brewer, T.D.; Farrell, P.; Kottage, H.; Reeve, E.; Thow, A.M.; Andrew, N.L. Conceptualising value chain research to integrate multiple food system elements. *Glob. Food Secur.* **2021**, *28*, 100500. [[CrossRef](#)]
10. Ahluwalia, N.; Dwyer, J.; Terry, A.; Moshfegh, A.; Johnson, C. Update on NHANES dietary data: Focus on collection, release, analytical considerations, and uses to inform public policy. *Adv. Nutr.* **2016**, *7*, 121–134. [[CrossRef](#)] [[PubMed](#)]
11. Ahuja, J.K.; Moshfegh, A.J.; Holden, J.M.; Harris, E. USDA food and nutrient databases provide the infrastructure for food and nutrition research, policy, and practice. *J. Nutr.* **2013**, *143*, 241S–249S. [[CrossRef](#)]
12. U.S. Department of Agriculture; Agricultural Research Service. Food Patterns Equivalents Database (FPED). 2021. Available online: www.ars.usda.gov/Services/docs.htm?docid=23871 (accessed on 31 May 2021).
13. U.S. Department of Agriculture; Agricultural Research Service. Food and Nutrient Database for Dietary Studies. 2021. Available online: www.ars.usda.gov/News/docs.htm?docid=12089 (accessed on 31 May 2021).
14. U.S. Department of Agriculture; Agricultural Research Service. Food Patterns Ingredients Database (FPID). 2021. Available online: www.ars.usda.gov/Services/docs.htm?docid=23871 (accessed on 31 May 2021).
15. US Environmental Protection Agency. Food Commodity Intake Database (FCID). 2001–2010. Available online: <http://fcid.foodrisk.org/#> (accessed on 31 May 2021).
16. U.S. Department of Agriculture; Agricultural Research Service. FoodData Central. 2021. Available online: <https://fdc.nal.usda.gov/> (accessed on 31 May 2021).
17. U.S. Department of Health and Human Services; Centers for Disease Control and Prevention (CDC). About the National Health and Nutrition Examination Survey. 2017. Available online: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm (accessed on 31 May 2021).
18. Chen, T.; Clark, J.; Riddles, M.; Mohadjer, L.; Fakhouri, T. The National Health and Nutrition Examination Survey, 2015–2018: Sample Design and Estimation Procedures; National Center for Health Statistics. 2020. Available online: <https://www.cdc.gov/nchs/nhanes/AnalyticGuidelines.aspx#sample-design> (accessed on 31 May 2021).
19. U.S. Department of Agriculture; Agricultural Research Service; Food Surveys Research Group. What We Eat in America: Overview. 2021. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/wweianhanes-overview/> (accessed on 31 May 2021).
20. U.S. Department of Agriculture; Agricultural Research Service; Food Surveys Research Group. Automated Multiple Pass Method. 2021. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/ampm-usda-automated-multiple-pass-method/> (accessed on 31 May 2021).
21. Moshfegh, A.J.; Rhodes, D.G.; Baer, D.J.; Murayi, T.; Clemens, J.C.; Rimpler, W.V.; Paul, D.R.; Sebastian, R.S.; Kuczynski, K.J.; Ingwersen, L.A.; et al. The US department of agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am. J. Clin. Nutr.* **2008**, *88*, 324–332. [[CrossRef](#)]
22. Steinfeldt, L.; Anand, J.; Murayi, T. Food reporting patterns in the USDA Automated Multiple-Pass Method. *Procedia Food Sci.* **2013**, *2*, 145–156. [[CrossRef](#)]
23. U.S. Department of Agriculture; Agricultural Research Service. FoodData Central: About Us. 2021. Available online: <https://fdc.nal.usda.gov/about-us.html> (accessed on 31 May 2021).
24. U.S. Department of Agriculture; Agricultural Research Service; Food Surveys Research Group. Food and Nutrient Database for Dietary Studies: Documentation. 2017–2018. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds/> (accessed on 31 May 2021).
25. U.S. Department of Agriculture; Agricultural Research Service. Food Data Central, Foundation Foods: Documentation and User Guide. 2020. Available online: https://fdc.nal.usda.gov/docs/Foundation_Foods_Documentation_Oct2020.pdf (accessed on 31 May 2021).
26. U.S. Department of Agriculture; Agricultural Research Service. FoodData Central, Experimental Foods: Documentation and User Guide. 2020. Available online: https://fdc.nal.usda.gov/docs/Experimental_Foods_Documentation_Oct2020.pdf (accessed on 31 May 2021).
27. U.S. Department of Agriculture; Economic Research Service. Loss-Adjusted Food Availability (LAFA) Data Series Documentation. 2020. Available online: <https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/loss-adjusted-food-availability-documentation/> (accessed on 31 May 2021).
28. Conrad, Z.; Niles, M.T.; Neher, D.A.; Roy, E.D.; Tichenor, N.E.; Jahns, L. Relationship between food waste, diet quality, and environmental sustainability. *PLoS ONE* **2018**, *13*, e0195405. [[CrossRef](#)]
29. U.S. Department of Agriculture; Economic Research Service. Food Availability Data Series Documentation. 2021. Available online: <https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/food-availability-documentation/> (accessed on 31 May 2021).

30. U.S. Department of Health and Human Services; Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Data Files. 2021. Available online: <https://www.cdc.gov/nchs/nhanes/Default.aspx> (accessed on 3 May 2021).
31. U.S. Department of Agriculture; Agricultural Research Service. *Food Patterns Equivalents Database (FPED) 2017–2018: Methodology and User Guide*; U.S. Department of Agriculture: Washington, DC, USA, 2020.
32. Conrad, Z.; Blackstone, N.T.; Roy, E.D. Healthy diets can create environmental trade-offs, depending on how diet quality is measured. *Nutr. J.* **2020**, *19*, 117. [CrossRef]
33. U.S. Department of Agriculture; Agricultural Research Service. Food Intakes Converted to Retail Commodities Database. 2021. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/ficrcd-overview/> (accessed on 30 May 2021).
34. Bowman, S.A.; Martin, C.L.; Carlson, J.L.; Clemens, J.C.; Lin, B.-H.; Moshfegh, A.J. Food Intakes Converted to Retail Commodities Databases 2003–2008: Methodology and User Guide; U.S. Department of Agriculture, Agricultural Research Service. 2013. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/ficrcd-methodology/> (accessed on 30 May 2021).
35. U.S. Department of Agriculture; Center for Nutrition Policy and Promotion. CNPP Prices Database. 2001–2004. Available online: <https://www.fns.usda.gov/resource/cnpp-data> (accessed on 31 May 2021).
36. Nielsen and IRI; National Consumer Panel. 2001–2004 National Consumer Panel Homescan Data. 2021. Available online: <https://www.ncponline.com/panel/US/EN/Login.htm> (accessed on 22 January 2021).
37. Carlson, A.; Lino, M.; Juan, W.; Marcoe, K.; Bente, L.; Hiza, H.A.B.; Guenther, P.M.; Leibtag, E. *Development of the CNPP Prices Database*; U.S. Department of Agriculture; Center for Nutrition Policy and Promotion: Washington, DC, USA, 2008. Available online: <https://fns-prod.azureedge.net/sites/default/files/resource-files/PricesDatabaseReport.pdf> (accessed on 31 May 2021).
38. Carlson, A.; Kuczynski, K.; Pannucci, T.; Koegel, K.; Page, E.T.; Tornow, C.E.; Zimmerman, T.P. *Estimating Prices for Foods in the National Health and Nutrition Examination Survey: The Purchase to Plate Price Tool*; U.S. Department of Agriculture; Economic Research Service: Washington, DC, USA, 2020. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=99294> (accessed on 31 May 2021).
39. Levin, D.; Noriega, D.; Dicken, C.; Okrent, A.M.; Harding, M.; Lovenheim, M. *Examining Food Store Scanner Data: A Comparison of the IRI InfoScan Data with Other Data Sets, 2008–2012*; U.S. Department of Agriculture; Economic Research Service: Washington, DC, USA, 2018. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=90354> (accessed on 31 May 2021).
40. Carlson, A.; Page, E.T.; Zimmerman, T.P.; Tornow, C.E.; Hermansen, S. *Linking USDA Nutrition Databases to IRI Household-Based and Store-Based Scanner Data*; U.S. Department of Agriculture; Economic Research Service: Washington, DC, USA, 2019. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=92570> (accessed on 31 May 2021).
41. U.S. Department of Agriculture; Economic Research Service. Consumer Price Index: Overview. 2020. Available online: <https://www.bls.gov/opub/hom/cpi/home.htm> (accessed on 31 May 2021).
42. Rehm, C.D.; Monsivais, P.; Drewnowski, A. Relation between diet cost and Healthy Eating Index 2010 scores among adults in the United States 2007–2010. *Prev. Med.* **2015**, *73*, 70–75. [CrossRef]
43. Conrad, Z. Daily cost of consumer food wasted, inedible, and consumed in the United States, 2001–2016. *Nutr. J.* **2020**, *19*, 35. [CrossRef]
44. U.S. Department of Agriculture; Economic Research Service. National Household Food Acquisition and Purchase Survey. 2012–2013. Available online: <https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey/> (accessed on 31 May 2021).
45. U.S. Department of Agriculture; Economic Research Service. National Household Food Acquisition and Purchase Survey: User’s Guide to Survey Design, Data Collection, and Overview of Datasets. 2019. Available online: <https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey/> (accessed on 31 May 2021).
46. U.S. Department of Agriculture; National Agricultural Statistics Service. Census of Agriculture. Washington, DC, USA, 2021. Available online: <https://www.nass.usda.gov/AgCensus/> (accessed on 31 May 2021).
47. U.S. Department of Agriculture; National Agricultural Statistics Service. Census of Agriculture Methodology. Washington, DC, USA, 2021. Available online: <https://www.nass.usda.gov/Publications/AgCensus/2017/index.php> (accessed on 31 May 2021).
48. U.S. Department of Agriculture; National Agricultural Statistics Service. Agricultural Surveys. 2019. Available online: http://www.nass.usda.gov/Quick_Stats/ (accessed on 31 May 2021).
49. U.S. Department of Agriculture; National Agricultural Statistics Service. About NASS Agricultural Surveys. 2020. Available online: https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/index.php (accessed on 31 May 2021).
50. U.S. Department of Agriculture; National Agricultural Statistics Service. Farm and Ranch Irrigation Survey. 2003–2013. Available online: https://www.agcensus.usda.gov/Publications/Irrigation_Survey/ (accessed on 31 May 2021).
51. U.S. Department of Agriculture; National Agricultural Statistics Service. Irrigation and Water Management Survey: Statistical Methodology. 2021. Available online: https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Farm_and_Ranch_Irrigation_Survey/index.php (accessed on 31 May 2021).
52. Center for Sustainable Systems; University of Michigan. Database of Food Impacts on the Environment for Linking to Diets. 2017. Available online: <http://css.umich.edu/page/datafield> (accessed on 31 May 2021).

53. Heller, M.C.; Willits-Smith, A.; Meyer, R.; Keoleian, G.A.; Rose, D. Greenhouse gas emissions and energy use associated with production of individual self-selected US diets. *Environ. Res. Lett.* **2018**, *13*, 044004. [CrossRef]
54. U.S. Department of Health and Human Services; Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System. 2020. Available online: <https://www.cdc.gov/brfss/index.html> (accessed on 31 May 2021).
55. National Cancer Institute; National Institutes of Health. Family Life, Activity, Sun, Health, and Eating (FLASHE) Study. 2017. Available online: <https://cancercontrol.cancer.gov/brp/hbrb/flashe.html> (accessed on 31 May 2021).
56. U.S. Department of Agriculture; Agricultural Research Service; Food Surveys Research Group. What We Eat in America Food Categories. 2021. Available online: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/dmr-food-categories/> (accessed on 4 March 2021).
57. U.S. Department of Labor; Bureau of Labor Statistics. Consumer Expenditure Survey. 2021. Available online: <http://www.bls.gov/cex/> (accessed on 31 May 2021).
58. Miller, V.; Webb, P.; Micha, R.; Mozaffarian, D. Defining diet quality: A synthesis of dietary quality metrics and their validity for the double burden of malnutrition. *Lancet Planet. Health* **2020**, *4*, e352–e370. [CrossRef]
59. Morze, J.; Danielewicz, A.; Hoffmann, G.; Schwingshackl, L. Diet quality as assessed by the Healthy Eating Index, alternate Healthy Eating Index, dietary approaches to stop hypertension score, and health outcomes: A second update of a systematic review and meta-analysis of cohort studies. *J. Acad. Nutr. Diet.* **2020**, *120*, 1998–2031.e1915. [CrossRef] [PubMed]
60. Fulgoni, V.L.; Keast, D.R.; Drewnowski, A. Development and validation of the nutrient-rich foods index: A tool to measure nutritional quality of foods. *J. Nutr.* **2009**, *139*, 1549–1554. [CrossRef] [PubMed]
61. National Cancer Institute; National Institutes for Health. Overview and Background of the Healthy Eating Index. 2020. Available online: <https://epi.grants.cancer.gov/hei/> (accessed on 31 May 2021).
62. Reedy, J.; Lerman, J.L.; Krebs-Smith, S.M.; Kirkpatrick, S.I.; Pannucci, T.E.; Wilson, M.M.; Subar, A.F.; Kahle, L.L.; Toozee, J.A. Evaluation of the Healthy Eating Index-2015. *J. Acad. Nutr. Diet.* **2018**, *118*, 1622–1633. [CrossRef]
63. U.S. Department of Health and Human Services; U.S. Department of Agriculture. Dietary Guidelines for Americans 2015–2020. Washington, DC, USA, 2015. Available online: <http://health.gov/dietaryguidelines/> (accessed on 20 April 2020).
64. Chiuve, S.E.; Fung, T.T.; Rimm, E.B.; Hu, F.B.; McCullough, M.L.; Wang, M.; Stampfer, M.J.; Willett, W.C. Alternative dietary indices both strongly predict risk of chronic disease. *J. Nutr.* **2012**, *142*, 1009–1018. [CrossRef]
65. Drewnowski, A. The nutrient rich foods index helps to identify healthy, affordable foods. *Am. J. Clin. Nutr.* **2010**, *91*, 1095S–1101S. [CrossRef] [PubMed]
66. Reedy, J.; Subar, A.F. 90th anniversary commentary: Diet quality indexes in nutritional epidemiology inform dietary guidance and public health. *J. Nutr.* **2018**, *148*, 1695–1697. [CrossRef]
67. Harmon, B.E.; Boushey, C.J.; Shvetsov, Y.B.; Ettienne, R.; Reedy, J.; Wilkens, L.R.; Le Marchand, L.; Henderson, B.E.; Kolonel, L.N. Associations of key diet-quality indexes with mortality in the multiethnic cohort: The dietary patterns methods project. *Am. J. Clin. Nutr.* **2015**, *101*, 587–597. [CrossRef]
68. Liese, A.D.; Krebs-Smith, S.M.; Subar, A.F.; George, S.M.; Harmon, B.E.; Neuhauser, M.L.; Boushey, C.J.; Schap, T.R.E.; Reedy, J. The dietary patterns methods project: Synthesis of findings across cohorts and relevance to dietary guidance. *J. Nutr.* **2015**, *145*, 393–402. [CrossRef] [PubMed]
69. Krebs-Smith, S.M.; Pannucci, T.E.; Subar, A.F.; Kirkpatrick, S.I.; Lerman, J.L.; Toozee, J.A.; Wilson, M.M.; Reedy, J. Update of the Healthy Eating Index: HEI-2015. *J. Acad. Nutr. Diet.* **2018**, *118*, 1591–1602. [CrossRef] [PubMed]
70. Willett, W.C.; Howe, G.R.; Kushi, L.H. Adjustment for total energy intake in epidemiologic studies. *Am. J. Clin. Nutr.* **1997**, *65*, 1220S–1228S. [CrossRef]
71. National Cancer Institute; National Institutes for Health. Healthy Eating Index: Overview of the Methods and Calculations. 2020. Available online: <https://epi.grants.cancer.gov/hei/hei-methods-and-calculations.html> (accessed on 28 January 2021).
72. McCullough, M.L.; Feskanich, D.; Stampfer, M.J.; Giovannucci, E.L.; Rimm, E.B.; Hu, F.B.; Spiegelman, D.; Hunter, D.J.; Colditz, G.A.; Willett, W.C. Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance. *Am. J. Clin. Nutr.* **2002**, *76*, 1261–1271. [CrossRef]
73. Wang, D.D.; Leung, C.W.; Li, Y.; Ding, E.L.; Chiuve, S.E.; Hu, F.B.; Willett, W.C. Trends in dietary quality among adults in the United States, 1999 through 2010. *JAMA Intern. Med.* **2014**, *174*, 1587–1595. [CrossRef]
74. Conrad, Z.; Karlsen, M.; Chui, K.; Jahns, L. Diet quality on meatless days: National Health and Nutrition Examination Survey (NHANES), 2007–2012. *Public Health Nutr.* **2017**, *20*, 1564–1573. [CrossRef] [PubMed]
75. Bernstein, A.M.; Bloom, D.E.; Rosner, B.A.; Franz, M.; Willett, W.C. Relation of food cost to healthfulness of diet among US women. *Am. J. Clin. Nutr.* **2010**, *92*, 1197–1203. [CrossRef] [PubMed]
76. Drewnowski, A.; Fulgoni, V.L., 3rd. New nutrient rich food nutrient density models that include nutrients and myplate food groups. *Front. Nutr.* **2020**, *7*, 107. [CrossRef] [PubMed]
77. Conrad, Z.; Reinhardt, S.; Boehm, R.; McDowell, A. Higher diet quality is associated with higher diet costs when eating at home and away from home: National Health and Nutrition Examination Survey, 2005–2016. *Public Health Nutr.* **2021**, 1–29. [CrossRef] [PubMed]
78. Rehm, C.D.; Monsivais, P.; Drewnowski, A. The quality and monetary value of diets consumed by adults in the United States. *Am. J. Clin. Nutr.* **2011**, *94*, 1333–1339. [CrossRef] [PubMed]

79. U.S. Department of Agriculture; Economic Research Service. Food Expenditure Series: Normalized Food Expenditures by All Purchasers and Household Final Users. 2021. Available online: <https://www.ers.usda.gov/data-products/food-expenditure-series/> (accessed on 31 May 2021).
80. Muth, M.K.; Giombi, K.C.; Bellemare, M.; Ellison, B.; Roe, B.; Smith, T. Expert Panel on Technical Questions and Data Gaps for the ERS Loss-Adjusted Food Availability (LAFA) Data Series. 2018. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=92408> (accessed on 31 May 2021).
81. Peters, C.; Picardy, J.; Darrouzet-Nardi, A.; Wilkins, J.; Griffin, T.; Fick, G. Carrying capacity of U.S. Agricultural land: Ten diet scenarios. *Elementa* **2016**, *4*. [[CrossRef](#)]
82. Conrad, Z.; Johnson, L.K.; Peters, C.J.; Jahns, L. Capacity of the US food system to accommodate improved diet quality: A biophysical model projecting to 2030. *Curr. Dev. Nutr.* **2018**, *2*, nzy007. [[CrossRef](#)] [[PubMed](#)]
83. Canning, P.; Rehkamp, S.; Hitaj, C.; Peters, C. *Resource Requirements of Food Demand in the United States*; U.S. Department of Agriculture; Economic Research Service: Washington, DC, USA, 2020. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=98400> (accessed on 31 May 2021).
84. Galzki, J.C.; Mulla, D.J.; Peters, C.J. Mapping the potential of local food capacity in southeastern Minnesota. *Renew. Agric. Food Syst.* **2015**, *30*, 364–372. [[CrossRef](#)]
85. Peters, C.J.; Wilkins, J.L.; Fick, G.W. Testing a complete-diet model for estimating the land resource requirements of food consumption and agricultural carrying capacity: The New York state example. *Renew. Agric. Food Syst.* **2007**, *22*, 145–153. [[CrossRef](#)]
86. Willits-Smith, A.; Aranda, R.; Heller, M.C.; Rose, D. Addressing the carbon footprint, healthfulness, and costs of self-selected diets in the USA: A population-based cross-sectional study. *Lancet Planet. Health* **2020**, *4*, e98–e106. [[CrossRef](#)]
87. Rose, D.; Heller, M.C.; Willits-Smith, A.M.; Meyer, R.J. Carbon footprint of self-selected US diets: Nutritional, demographic, and behavioral correlates. *Am. J. Clin. Nutr.* **2019**, *109*, 526–534. [[CrossRef](#)] [[PubMed](#)]
88. USDA Economic Research Service. Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products. 1992. Available online: <https://www.ers.usda.gov/publications/pub-details/?pubid=41881> (accessed on 31 May 2021).
89. Subar, A.F.; Freedman, L.S.; Toozé, J.A.; Kirkpatrick, S.I.; Boushey, C.; Neuhauser, M.L.; Thompson, F.E.; Potischman, N.; Guenther, P.M.; Tarasuk, V.; et al. Addressing current criticism regarding the value of self-report dietary data. *J. Nutr.* **2015**, *145*, 2639–2645. [[CrossRef](#)] [[PubMed](#)]
90. Congressional Research Service. US Food and Agricultural Imports: Safeguards and Selected Issues. 2020. Available online: <https://crsreports.congress.gov/product/pdf/R/R46440/2> (accessed on 31 May 2021).
91. U.S. Department of Agriculture; Foreign Agricultural Service. Global Agricultural Trade System. 2021. Available online: <https://apps.fas.usda.gov/gats/ExpressQuery1.aspx> (accessed on 31 May 2021).
92. U.S. Department of Agriculture; Economic Research Service. Loss-Adjusted Food Availability (LAFA) Data Series. 2020. Available online: <https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/food-availability-per-capita-data-system/#Loss-Adjusted%20Food%20Availability> (accessed on 31 May 2021).
93. Heller, M.C.; Willits-Smith, A.; Mahon, T.; Keoleian, G.A.; Rose, D. Individual US diets show wide variation in water scarcity footprints. *Nat. Food* **2021**, *2*, 255–263. [[CrossRef](#)]
94. Box, G.E. Science and statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [[CrossRef](#)]