# Nutrition-health impact assessment of food items using nutrient profiles and dietary risk factors

Master thesis submitted to

Leiden University and Delft University of Technology

in partial fulfillment of the requirements for the degree of

**MASTER OF SCIENCE**

in Industrial Ecology

Leiden University: Faculty of Science, Institute of Environmental Sciences (CML)

TU Delft: Faculty of Technology, Policy and Management (TPM)

by

Noah Griffioen

Student number: s1794817 | 5431093

To be defended on 15 July 2022

**Graduation committee**

| | |
|---|---|
| First supervisor | Dr. Laura Scherer. Leiden University: Institute of Environmental Sciences (CML), Department of Industrial Ecology |
| Second supervisor | Prof. dr. Jessica Kiefte – de Jong. Leiden University. Leiden University Medical Center. Population health |

# Table of contents

## Abstract

The CONE-LCA framework is a promising framework that can be used to assess the impact of food items on human health and the environment. However, the nutrition-health assessment in the CONE-LCA framework quantifies the health impacts of food at the level of food *items* with the use of an indicator that is based on only few nutrients and otherwise broad food *groups*. Given that food items within food groups can have different impacts on health, disaggregating health impacts can be of added value and was explored here. An extended version of the Nutrients Rich Food index is used for both downscaling and cluster analysis.

The correlation coefficient at the aggregated levels with the NRF index (-0.93) was substantially higher than the original health impact indicator of food items with the NRF index (0.36). These results illustrate that the use of a downscaled indicator improves the health impact assessment of food items. The downscaling analysis showed that downscaled health impacts can differ substantially between food items within food groups. Furthermore, the cluster analysis showed that some food groups (e.g. fruits, vegetables and red meat) are highly variable in nutrient density and therefore prioritizing these food groups for clustering can be of added value in future epidemiological studies.

The improved nutrition health assessment model is not only relevant within the CONE-LCA framework but can be used on its own to assess health impacts of food items. The model may be very useful in policy applications as it enables comparison of food items on impact on human health and the environment, which can be the basis for sustainable dietary guidelines. Future research can focus on further development and validation of the model.

# 1 Introduction

Environmental sustainability and non-communicable diseases are inextricably linked through dietary choices (Reinhardt et al., 2020; Tilman & Clark, 2014). Non-communicable diseases are also known as chronic diseases which cannot be passed from person to person, such as cardiovascular diseases, cancers, and chronic respiratory diseases (World Health Organization, 2021). The average western dietary pattern is associated with negative health outcomes. Underconsumption of nourishing foods and overconsumption of harmful foods are the most important contributors to the health burden in the US (Afshin et al., 2019) as it results in non-communicable diseases (Gakidou et al., 2017). At the same time, current western dietary patterns drive detrimental environmental impacts including climate change, deforestation, land degradation, ocean acidification, and air pollution, all pressuring the earth's boundaries (Willet et al., 2019). Impacts on human health and the environment are further expected to increase due to the growth of the global population and we are already starting to see food supply security being threatened (Godfray et al., 2010). Therefore, there is a need for evidence-based policymaking to support a food production and consumption system that is sustainable and positively affects human health (Davis et al., 2014; Reinhardt, 2020). Consequently, the need for a joint environmental and health assessment of dietary choices has been acknowledged and literature is expanding (Doran-Browne, 2015; Tilman and Clark, 2014; Reinhardt et al., 2020; Springmann et al., 2018; Vega Mejia et al., 2018; Willet and Stampfer, 2019).

Although the impact of foods on human health and the environment has extensively been researched independently, research jointly assessing health and environmental impacts of dietary choices is scarce. The environmental impacts of diets have been researched extensively, in which health effects are often underexposed (Andrew et al., 2016; Behrens et al., 2017; Jones et al., 2016; Ritchie et al. 2018). On the other hand, health impacts of dietary choices have been researched thoroughly, but environmental impacts are often neglected (Afshin et al., 2019; Forouzanfar et al., 2015; Gakidou et al., 2017; Lim et al., 2012; Micha et al., 2011; Micha et al., 2017; Springmann et al., 2018; Willet and Stampfer, 2013). In the latter studies mostly wholesale diets are researched, including nationally recommended diets, national average diets, flexitarian, pescatarian and vegetarian diets. Although such comparisons are helpful to research the greatest potential beneficial impacts, their practical applications are limited because policies are bound to diverse food systems and resource availability and cultural preferences (Green et al., 2020; Reinhardt et al., 2020). Thus, literature suggests that more research should be conducted into health and environmental impacts of marginal dietary adjustments instead of wholesale comparisons between diets (Green et al., 2020; Reinhardt et al., 2020). Besides, Jones et al. (2016) recommended that future research should focus on streamlining the integration of nutrition into Life Cycle Analysis (LCA).

This research gap was filled when Stylianou et al. (2021) developed the Combined Nutritional and Environmental Life Cycle Assessment (CONE-LCA) framework to assess the impact of food items on human health and the environment. Resource use and environmental emissions, and dietary risks and benefits are quantified to estimate the impact of food on the environment and human health over its life cycle. The environmental impact assessment in the framework follows the traditional environmental LCA method. The nutritional assessment of foods builds on epidemiological studies (Stylianou et al., 2021).

Epidemiology is the study of the distribution of diseases and other health-related conditions in populations, and the application of this study to control health problems (National Research Council, 2012).

In the nutrition-health assessment of food items in this framework, dietary risk factors (DRFs) for only a few nutrients and otherwise broad food groups are considered. Although the Dietary Guidelines for Americans have shifted from nutrient requirements to food recommendations, food items within a food group differ in nutrient content and can therefore have different effects on health (U.S. Department of Agriculture, 2020; Lampe, 1999). For example, bananas contain more starch than oranges, and oranges contain more vitamin C than bananas (Khachatrian, 2021). Consequently, it can be of added value to differentiate between food items within food groups in health impact assessment. This may be of particular importance in developing countries where dietary diversity is low, and the populations often depend on staple foods (Sibhatu et al., 2015). Additionally, food items with different processing categories are combined within food groups (e.g., oat flakes and sweetened breakfast cereals or raw nuts and roasted and salted nuts) while the type of processing may have different effects on health (Fardet et al., 2015). Lastly, only few nutrients are included in the CONE-LCA framework and key nutrients, like vitamins, magnesium, iron, iodine, potassium, and zinc are missing (World Health Organization, 2017). Consequently, there is a need for an improved nutrition-health assessment method as part of the CONE-LCA framework.

The CONE-LCA framework builds on dietary risks reported in the Global Burden of Disease (GBD) 2016 study. Quantification of the health burden attributable to specific dietary risks has been researched extensively in GBD studies (Forouzanfar, 2015; Gakidou et al., 2017: Lim et al., 2012; Afshin et al., 2019). The GBD studies estimate these dietary risks based on data from epidemiological studies, which gathered their primary data based on the consumption of food groups as opposed to food items. Consequently, dietary risks in the GBD study, and therefore the indicators used in the CONE-LCA framework, are food group-based. If epidemiological studies would make use of clusters of food items with similar nutritional properties within these broad food groups, the quality of health impact estimates associated with food consumption could be improved. Another approach to refining dietary risks of food items is to downscale the health impacts of food groups to the level of food items, based on nutritional profiles of food items.

Ample literature on nutritional assessment of food items using nutrient profiles exists. Various nutrient profile (NP) models have been developed by the food industry, researchers and governments (Kourlaba et al., 2009) such as the Overall Nutritional Quality Index (Katz et al., 2010), Nutrient Rich Food Index (Fulgoni et al, 2009), Nutritional Quality Index (Sonesson, 2019) or Nutrient Balance Concept (Fern et al., 2015). NP models assess foods on nutrient quantity, nutrient quality and/or nutrient diversity. Nutrient quantity metrics are most commonly used in sustainability assessments (Green et al., 2020). They measure nutrient amounts, such as macronutrients, vitamins and minerals. These indices are informative, but they do not directly evaluate the impact of these foods on human health, as it only evaluates nutrient quantity. To assess the impact on human health, nutritional metrics should be translated to health metrics. Health metrics are used to quantify the impacts of nutrient consumption on human health. Disability-Adjusted Life Years (DALYs) are the most used health metric in sustainability studies, as nutrition and pollution-induced mortality and morbidity can be compared directly (Green et al., 2020). DALYs represent the burden of disease and are the sum of years of life lost due to premature mortality and years of healthy life lost due to disability (World Health Organization, 2013). The CONE-LCA framework could make use of such NP models to translate the nutrient content of food items to downscaled health impacts.

In summary, the nutrition-health assessment in the CONE-LCA framework quantifies the health impacts of food at the level of food items with the use of an indicator that is based on only few nutrients and otherwise broad food groups. Given that food items within a food group can have different impacts on health, disaggregation of food groups can be of added value and was explored in this thesis. The corresponding research question of this thesis is: How can the health impacts of food groups be disaggregated based on nutrient profiles to improve the nutritional health assessment of food items? The research objective of this thesis is to explore two approaches. Firstly, downscaling group-based health impact indicators to the level of food items based on the nutrient profiles of these food items. Secondly, prioritizing food groups with high variation in nutrient profiles for clustering. These clusters can then be used in subsequent epidemiological research.

The subsequent sub-questions are:
Sub-question 1: Which NP model is most suitable for this research?
Sub-question 2: Does the selected NP model improve the estimation of health impacts in the nutrition health assessment of food items?
Sub-question 3: What are the downscaled nutrition-related health impacts of food items compared to other food items within its food group?
Sub-question 4: Which food groups show high variability in nutritional content and could be prioritized for clustering?

The scope of this research is limited to the research boundaries of the study by Stylianou et al. (2021). The geographical scope of this research covers the consumption of food items in the US, since the impact indicators used in the CONE-LCA framework are US-based DRFs. Only impact on human health (expressed in μDALY or HENI) is covered by the indicator used in this thesis. The DRFs build on the dietary risks from the GBD 2016 study, which are constituted for 7 food groups and are based on a limited number of non-communicable diseases only. The number of food items analyzed in this thesis was limited to the availability of food items in the FoodData Central database. Only 144 of 167 food items emphasized on in the paper of Stylianou et al. (2021) were analyzed due to availability. Additionally, for the analysis of food items in the 7 food groups, a total of 908 food items were available in FoodData Central and were analyzed.

## 2 Methods

### 2.1 The CONE-LCA framework

In this section background information on the CONE-LCA framework will be discussed.

The CONE-LCA framework can be used to assess the impacts of food on human health and the environment over its lifecycle (Figure 1). The environmental impact assessment follows the traditional environmental LCA method. Starting from a functional unit, the related environmental emissions and resource extractions are analyzed, from which impact results are calculated, such as resources & ecosystem services, ecosystem quality and human health impacts & benefits. Health impacts related to nutritional effects of foods are assessed in parallel, with DALYs as a common endpoint metric. The preliminary CONE-LCA framework was tested in two case studies, on milk (Stylianou et al., 2016) and on fruit and vegetables (Stylianou et al., 2017). In 2021, the framework was further developed and tested in the assessment of 5,853 meals typically consumed in the US (Stylianou et al.,2021).

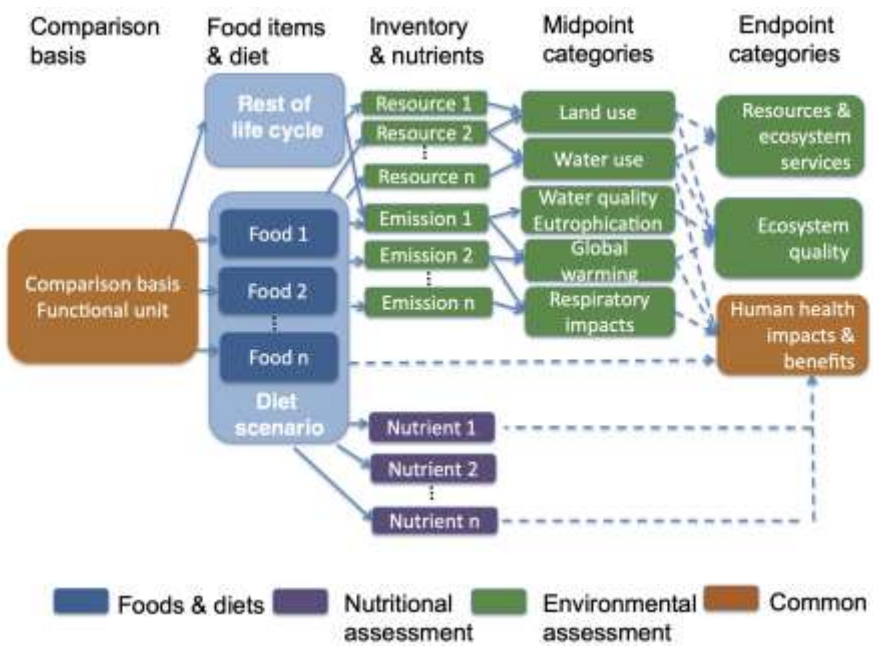


*Figure 1: CONE-LCA framework (Stylianou et al., 2016)*

In the GBD studies, DALYs associated with the consumption of 15 dietary risk components were identified across 195 countries. The primary data in the GBD study was retrieved from 94 systemic reviews, meta-analyses and pooled analyses dated from 1990 to 2016 (Gakidou, et al., 2017). The dietary risk components cover food groups, including nuts and seeds, processed meat, red meat, milk, whole grains, legumes, sugar-sweetened beverages (SSB), fruits and vegetables and nutrients including calcium, fibres, omega 3 fatty acids, polyunsaturated fatty acids (PUFAs), trans fatty acids (TFAs) and sodium. Building on these dietary risks from the GBD studies, Stylianou et al. (2021) developed DRFs to quantify health burden (mortality and morbidity) associated with a small intake shift from the baseline diet. These DRFs are quantified per consumed risk component for the average population and are expressed in µDALY per gram of intake. The health outcomes associated with these 15 dietary risk components are based on non-communicable diseases only. Communicable diseases and injuries are not considered in the calculations of these DALYs. Detailed information on the health endpoint methodology can be found in the supplementary materials of the study by Stylianou et al. (2021).

Integration of health impact indicators into LCA allows for comparison of environmental emissions and resource use, and dietary impacts of food on human health expressed in a common endpoint-unit, DALYs. Since DALYs can be abstract for non-experts to interpret, the Health Nutritional Index (HENI) score is used to translate the health impacts to a more tangible metric. “HENI is a continuous single score that quantifies the net minutes of healthy life gained (+) or lost (−) from all-cause mortality and morbidity per reference amount of food (for example, a standard serving size)” (Stylianou et al., 2021).

In the example of Figure 2, the environmental and health impacts of a serving of chicken wings are calculated. The right side represents the environmental LCA of this product. The left side represents the nutritional evaluation of the product. In a serving of chicken wings of 85 grams, an X amount of nutrients are present. These nutrients are, with the use of dietary risk factors (µDALY per gram of risk component), translated into health impacts (µDALY) and rescaled into the HENI score (minutes of life gained or lost) per serving.

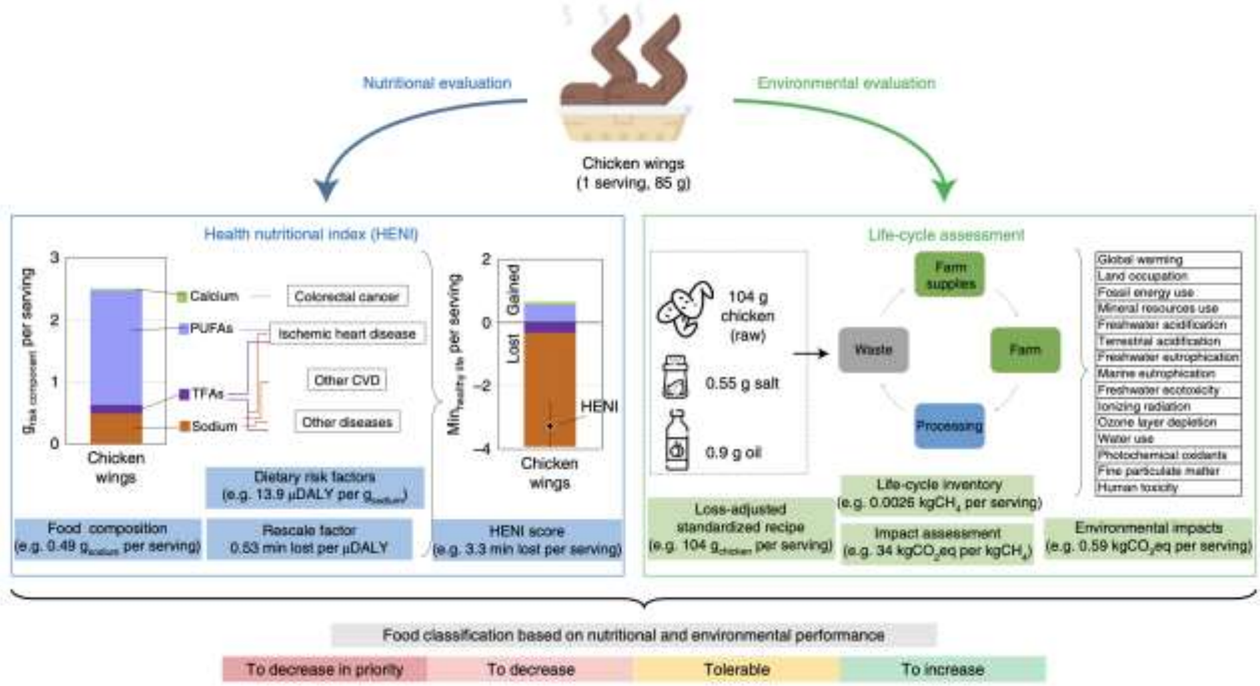


*Figure 2: Extended CONE-LCA framework (Stylianou et al., 2021)*

Important to note is that the food items are individually assessed in the environmental assessment. For example, for the environmental assessment of bananas and oranges, the life cycle stages of bananas and oranges are analyzed respectively. This individual assessment is missing in the nutritional evaluation. DRFs are based on only a few nutrients and otherwise broad food groups. Although food items in thousands of dishes are analyzed, in the health impact assessment these ingredients are assigned to the DRFs of the respective food *groups*. For example, both bananas and oranges are assigned to the DRF of “fruits” as a group.

## 2.2 Nutrient profile models

### 2.2.1 Background

Nutrient quantity of foods can be assessed using a NP model. NP models rank foods based on their nutrient content (Drewnowski and Fulgoni, 2008). The models usually consider nutrients to encourage which are known to be beneficial for health, or nutrients to limit which are known to be detrimental to health, or a combination of both (Drewnowski and Fulgoni, 2008). In this analysis, the use of two NP models, the Nutrient-Rich Foods (NRF) index and the Nutritional Quality Index (NQI), for assessing the health impacts of food items were explored.

### 2.2.2 Nutrient-Rich Foods Index

The NRF index was used in this study since it is a validated and commonly used nutritional quantity index and it is easily extendible (Fulgoni et al. 2009). The concept of this nutrient density score was first explored in 2005 and further developed to its current form (Drewnowski, 2005; Drewnowski and Fulgoni, 2008; Fulgoni et al., 2009; Maillot et al. 2007). A family of NRF indices were validated against the Healthy Eating Index (HEI), which is an extensively analyzed, accepted measure of diet quality (Fulgoni et al., 2009). Moreover, the NRF index has been validated against all-cause mortality (Streppel et al., 2014).

The NRF index measures nutrient quantity, calculating a joint score from nutrients to encourage and nutrients to avoid. The NRF index, calculated from both macro and micronutrients, attributes a score to food items based on their nutrient content. It sums nutrients that are beneficial for health and subtracts nutrients to limit into a joint score (NRFn.3), where n represents a variable number of beneficial nutrients and three nutrients to limit: saturated fat, added sugars and sodium (Drewnowski and Fulgoni, 2008). Scores are calculated based on recommended daily intake values (e.g. region-specific or global) and reference amounts (e.g. 100 g, 100 kcal or serving size) (Fulgoni et al., 2009). The advantage of this model is that it is easily extendible to other countries and various numbers of nutrients.

The NRF9.3 version scored best in the validation against the HEI (Fulgoni et al., 2009). In NRF9.3 the nine beneficial nutrients included are, protein, fibre, iron, calcium, potassium, magnesium and vitamins A, C and E and three nutrients to limit included are, sodium, added sugars and saturated fat. As an example, the NRF9.3 score is calculated for the US, by summing *the ratio of the nutrient in the food item (g) to the recommended daily value of the nutrient (g)* for beneficial nutrients and by subtracting this ratio for detrimental nutrients (Fulgoni et al., 2009) (Equation 1).

*Equation 1:* $$NRF9.3 = (protein\ g/50g + fibre\ g/25\ g + vitamin\ A\ IU/5000\ IU + vitamin\ C\ mg/60\ mg + vitamin\ E\ IU/30\ IU + calcium\ mg/1000\ mg + iron\ mg/18\ mg + magnesium\ mg/400\ mg + potassium\ mg/3500\ mg - saturated\ fat\ g/20\ g - added\ sugars\ g/50\ g - sodium\ mg/2400\ mg) \cdot 100$$

### 2.2.3 Nutritional Quality Index

Nutritional quality is dependent on dietary context. For example, in countries where calcium intake is typically low in the average diet, dairy is beneficial to health impacts. On the other hand, intake of products high in calcium will add little extra beneficial health impacts in countries with a high intake of dairy products (Sonesson et al., 2019), such as the Netherlands. Consequently, it is important to consider the dietary context in the nutrition-health assessment of foods.

The nutritional quality index (NQI) indicates the nutritional value of a product in a given dietary context. The NQI model uses a similar approach to the NRF9.3 model, including the same beneficial and detrimental nutrients (Sonesson et al., 2019). Yet, the NQI model assigns higher scores to nutrients that are deficient in the studied diet, by comparing the ratio of nutrients in the product to the ratio of nutrients in the diet and the total consumption ratio. Equations 2-4 are used to calculate the NQI index (Sonesson et al., 2019):

*Equation 2:* $$NQIproduct = \sum(NQIqualnutrient1\text{–}9) - \sum(NQIdisqualnutrient1\text{–}3)$$

*Equation 3:* $$NQIqualnutrient1 = (ratio\ in\ product/ratio\ in\ diet)/consumption\ ratio$$

*Equation 4:* $$NQIdisqualnutrient1 = (ratio\ in\ product/ratio\ in\ diet) \times consumption\ ratio$$

Where:

- Ratio in product = total content of nutrient in mass divided by total mass of product (g/g)
- Ratio in diet = total intake of nutrient in mass divided by total dietary intake in mass (g/g)
- Consumption ratio = ratio of dietary intake to dietary need for the nutrient

Similar to the NRF index, nutrients can be added and removed from the NQI index, which makes it easy to adjust the index to the aim of this study. This NQI model is the first presented version and is not verified yet (Sonesson et al., 2019). Therefore, in this study use of the NQI model to score a broad range of food items on nutritional quality will be explored and compared to the NRF model.

### 2.2.4 Approach

The original NRF and NQI models (Equation 5) were extended to reflect a more complete set of important nutrients. Literature highlighted a degree of arbitrariness in the exclusion of nutrients in other NP models (Ridoutt, 2021). To be as free as possible from this bias, all nutrients for which (1) US-specific recommended intake value was available and (2) the nutritional data was available in the FoodData Central database, were included in the model. This database provides detailed information on food items and their nutritional composition (USDA, 2019a). Based on these criteria, the original models were extended with 15 extra nutrients to encourage and one nutrient to limit (**Fout! Verwijzingsbron niet gevonden.**). Consequently, the extended models include 24 qualifying and four disqualifying nutrients and could be written as NRF24.4 and NQI24.4 (Equation 6). Moreover, nutrients did not meet the criteria and were excluded (appendix A). Additionally, data on added sugar was not available in the FoodData central database (USDA, 2019) and therefore added sugar was substituted by total sugar.

For the NQI model, ratio in product was retrieved from the food composition tables of FoodData Central (USDA, 2019a), which was also used for calculating the NRF Index. For the ratio in diet, total nutrient intake was retrieved from balance sheets from WWEIA (USDA, 2019b), where average consumption of adult males and females was used ((adult males + adult females)/2). To calculate total dietary intake in mass, the ratio between actual total protein intake and total protein supply was taken and applied to the total food supply food balance sheets from FAOstat (FAO, 2019). Total protein intake was retrieved from What We Eat In America (USDA, 2019b) and total protein supply from FAOstat food balance sheets (FAO, 2019). For the consumption ratio, the dietary intake is the same value as mentioned before. The dietary need for the nutrient is the daily value which was also used to calculate the NRF indices (**Fout! Verwijzingsbron niet gevonden.**).

In the equations of both models, the ratio between qualifying and disqualifying nutrients in the extended indices is different compared to the original ones. To resemble the ratio of qualifying and disqualifying nutrients in the original indices, a second version of the extended indices was tested, where more weight was attributed to disqualifying nutrients by multiplying the detrimental nutrients with a factor of 2 (Equation 7). Whether the extended formula with or without extra weight was used in this research, was based on the correlation analysis (chapter 2.3).

*Equation 5:* $NRF\ \&\ NQIproduct = \sum(qualnutrient1-9) - \sum(disqualnutrient1-3)$

*Equation 6:* $NRF\ \&\ NQIproduct = \sum(qualnutrient1-24) - \sum(disqualnutrient1-4)$

*Equation 7:* $NRF\ \&\ NQIproduct = \sum(qualnutrient1-24) - 2 \cdot \sum(disqualnutrient1-4)$

The NQI index assumes an exponential relation for qualifying nutrients and a linear relation for disqualifying nutrients, see Equations 3-4 (Sonesson et al., 2019). Therefore, the NQI can attribute a relatively high score to food items with both high contents of beneficial and detrimental nutrients. To test whether this assumption indeed improves the prediction of health impacts, two NQI versions that assume linear relations only or exponential relations only (Equations 8-9) were applied to all NQI models (Equations 5-7) and tested in the correlation analysis (chapter 2.3).

*Equation 8:* $NQIqualnutrient, NQIdisqualnutrient = \frac{ratio\ product}{ratio\ in\ diet} / consumption\ ratio$

*Equation 9:* $NQIqualnutrient, NQIdisqualnutrient = \frac{ratio\ product}{ratio\ in\ diet} \times consumption\ ratio$

*Table 1: Nutrients and recommended daily values.*

| Nutrient | Recommended Value | Source | Comment |
|---|---|---|---|
| *Nutrients to encourage* | | | |
| Fiber | 28g | (FDA, 2022) | |
| Protein | 50g | (FDA, 2022) | |
| Vitamin A (RAE) | 900mcg RAE | (FDA, 2022) | |
| Vitamin C (Ascorbic acid) | 90mg | (FDA, 2022) | |
| Vitamin E (alpha-tocopherol) | 15 mg alpha-tocopherol | (FDA, 2022) | |
| Magnesium | 420mg | (FDA, 2022) | |
| Calcium | 1300mg | (FDA, 2022) | |
| Potassium | 4700mg | (FDA, 2022) | |
| Iron | 18mg | (FDA, 2022) | |
| Zinc | 11mg | (FDA, 2022) | |
| Vitamin B1 (Thiamin) | 1.2 mg | (FDA, 2022) | |
| Vitamin B2 (Riboflavin) | 1.3mg | (FDA, 2022) | |
| Vitamin B9 (Folate) | 400mcg DFE | (FDA, 2022) | |
| Vitamin B12 (Cobalamin) | 2.4mcg | (FDA, 2022) | |
| Monounsaturated fat | 20g | (FDA, 2022) | |
| Polyunsaturated fatty acids | 18.9g | (WHO, 2010) | 6-11%E; $(6+11/2) \cdot 2000$ kcal $\cdot$ 1g fat / 9 kcal |
| Vitamin D | 20mcg | (FDA, 2022) | |
| Phosphorus | 1250mg | (FDA, 2022) | |
| Vitamin K | 120mcg | (FDA, 2022) | |
| Copper | 0.9mg | (FDA, 2022) | |
| Vitamin B3 (Niacin) | 16mg NE | (FDA, 2022) | |
| Selenium | 55mcg | (FDA, 2022) | |
| Vitamin B6 (Pyridoxine) | 1.7mg | (FDA, 2022) | |
| Choline | 550mg | (FDA, 2022) | |
| *Nutrients to limit* | | | |
| Total sugar | 125g | (IOM, 2005 in Fulgoni 2008) | |
| Sodium | 2300mg | (FDA, 2022) | |
| Saturated fat | 20g | (FDA, 2022) | |
| Cholesterol | 300mg | (FDA, 2022) | |

#### 2.2.4.1 NRF: Dietary context

As discussed previously, although dietary context is very important in the assessment of foods, the NRF model naturally does not consider this. Therefore, the NRF model was adjusted to attribute less weight to overconsumed nutrients in the US. Overconsumed nutrients are capped by the ratio of the recommended daily value to the actual consumption, under two conditions. Firstly, the nutrient must not be a nutrient to limit. There is a clear dose-response relationship in the consumption of detrimental nutrients (Qin et al., 2020). Where little extra beneficial health impacts are achieved after consuming more than 100% of the recommended value for beneficial nutrients, crossing 100% of the recommended daily value for detrimental nutrients is associated with additional negative health impacts. Secondly, the fat-soluble vitamins (e.g. vitamin A, D, E and K) are not capped. Fat-soluble vitamins are stored in the body and therefore consumption of more than 100% of the recommended DV is still associated with extra beneficial health impacts (Blomhoff et al, 1990; Card et al., 2013). For simplification in this research, several assumptions have been made in the capping method. It is assumed that consumption of more than 100% daily value is associated with no extra beneficial health effects. Secondly, for the nutrients to encourage, negative health impacts after crossing the daily maximum upper limit are ignored.

In the following example, the capping approach is illustrated: The recommended daily intake of copper is 0.9 mg and the actual US intake of copper is 1.3 mg. Thus, copper is overconsumed in the US. Copper is not detrimental and not fat-soluble. Therefore, overconsumption of this nutrient is associated with few additional health impacts. Consequently, copper quantity in food items is capped by a factor 0.69 (= 0.9 mg / 1.3 mg). This approach of capping is preferred over only capping at 100% DV because in the latter approach, capping or not is dependent on the functional unit. Therefore, downscaled health impacts could not be extrapolated to mixed dishes or daily food patterns since nutritional quantity is always capped for that fixed functional unit. This will be illustrated with an example: a FU of 100 g is used to calculate the health impacts and nutrients are capped at 100%. If one would want to use these downscaled health impacts in the CONE-LCA framework to calculate the total health impact of a mixed dish, the weight should be rescaled to the recipe. If one would simply rescale the health impacts with the weight used in the mixed dish, the health impacts of food items with high nutrient density could be underestimated. This is because the original health impact score, although rescaled, is used in which nutrients were capped at 100% for 100 g. But, when using only a small amount in a dish, the recommended DV would probably not have been crossed and therefore it should not have been capped. This also works the opposite way, when one eats more than 100 g per day and capping should be done but is not done so when capping 100% for 100 g. This example illustrates that capping at 100% is always bound to a specific functional unit.

## 2.3 Correlation analysis

### 2.3.1 Objective

The objectives of this analysis were to (1) select the most suitable NP model for further analyses and (2) to test whether the use of NP models improves the health impact assessment of food items.

### 2.3.2 Approach

First, all NP models described in section 2.2.4 were tested against the health impacts of the 144 food items emphasized in the paper of Stylianou et al. (2021). To select the most suitable versions of both the NRF and NQI models, all models were compared to each other based on Pearson correlation coefficients. Moreover, to justify whether the extended versions (24.4 and 24.4·2) improve the original versions (9.3), the original versions were also included in this analysis.

Afterwards, the two models with the highest Pearson correlation coefficients were tested in an additional correlation analysis at the level of 908 food items within seven food groups. The indicators used in the CONE-LCA framework are DRFs. These DRFs represent the average US-population-weighted risks (Table 2). A consumption weighted average nutrient index is calculated for each food group, which is considered equivalent to the DRF. The Pearson correlation coefficient between the nutrient index and dietary risk factors at the aggregated level was compared to the Pearson correlation coefficient between the nutrient index and the original health impact indicators of food items used by Stylianou et al. (2021). A higher correlation at the aggregated level would suggest that the use of a nutrient index to score food items improves the estimation of health impacts better compared to the use of the original aggregated indicator at the level of food items. The NP model with the highest correlation was selected for further analyses in this thesis.

*Table 2: Dietary risk factors in for all food groups (Stylianou et al., 2021) and number of items in food group.*

| Food group | Dietary risk factor (µDALY/g) | Number of items in food group |
|---|---|---|
| Fruits | -0.18 | 93 |
| Legumes | -0.23 | 51 |
| Milk | -0.01 | 30 |
| Nuts & seeds | -1.50 | 61 |
| Red meat | 0.10 | 295 |
| Vegetables | -0.08 | 346 |
| Whole grains | -0.34 | 32 |

### 2.3.3 Data on food items

Stylianou et al. (2021) highlighted 167 food items, from the Food and Nutrient Database for Dietary Studies (FNDDS) dataset within the FoodData Central database, in their paper. Food items in this database are marked by codes from the What We Eat In America (WWEIA) survey, which was conducted as a partnership between the U.S. Department of Agriculture (USDA) and the U.S. Department of Health and Human Services (DHHS) (USDA and DHHS, 2019). The 167 food codes were searched in this database to be used in this thesis. However, 21 food items were excluded because the corresponding WWEIA codes were not existing anymore in FoodData Central (Appendix B). Moreover, two food items were excluded because they were duplicated in the report of Stylianou et al. (2021) (Appendix B). The extended NRF and NQI indices were calculated for all remaining 144 food items.

For the additional correlation analysis, food items and their nutritional information were also retrieved from FoodData Central to be consistent with the research method of Stylianou et al. (2021). Moreover, the DRFs from the CONE-LCA framework were US-specific (Stylianou et al., 2021). Consequently, US consumption data was used in this research. The food groups, which include fruits, milk, nuts and seeds, red meat, vegetables, legumes, and whole grains, the corresponding DRFs and the associated health outcomes were retrieved from their study. The associated health outcomes of the dietary risk factors (in µDALYs) for these food groups were based on a limited number of non-communicable diseases only (Appendix C). Communicable diseases and injuries are not considered in the health endpoints. The inclusion and exclusion criteria for food items in a food group, defined by Stylianou et al. (2021) were used (Appendix C). Based on these criteria a total of 908 food items were selected from the FoodData Central database (Table 2) (USDA, 2019a). A detailed list of all items is included in Appendix D.

Stylianou et al. (2021) calculated DRFs based on consumption levels using National Health and Nutrition Examination Survey (NHANES) 2011–2016 (CNC, 2018). However, this raw data from the survey is not translated to average daily consumption values. Given the limited timeframe of this thesis, it was chosen to retrieve US consumption data elsewhere. Consequently, consumption data is retrieved from the GeNUS database (Smith and Matt, 2018). The GeNUS database was preferred over FAO food balance sheets because it predicts consumption more accurately. Firstly, because in this database broad food groups from the FAO food balance sheets are disaggregated into smaller sub-groups, which makes it possible to estimate consumption values for food items more accurately than when one can only equally divide the consumption data of the broad food group over food items. Secondly, consumption is estimated using the Global Dietary Database which is based on national household surveys (Smith et al., 2016). This gives a more realistic estimation of what is consumed than estimating consumption based only on food balance sheets.

The consumption data in GeNUS was sometimes very specific and sometimes aggregated to small food groups. In the latter case, the consumption value had to be distributed over the food items belonging to that group. For example, the category "fresh fruit, not elsewhere specified' includes 9 fruits. The total consumption value was equally divided over the 9 fruits. Of these 9 fruits only 2, pomegranate and tamarinds, were present in the FoodData Central database. In such a case, pomegranate, and tamarinds each received 1/9 of the consumption value from GeNUS since the other fruits are different sorts of fruit.

In this thesis, it is assumed that the consumption data from GeNUS applies to raw food items and is therefore assigned to the raw food items from FoodData Central. However, when a food item in FoodData Central was not present in raw form but only in cooked form, it was chosen to allocate the consumption values to the cooked item. In this way, the consumption weighted average can be calculated as accurately as possible. Excluding the consumption value at all would have given a distorted consumption weighted average.

## 2.4 Downscaling health impact indicators

### 2.4.1 Objective

The selected NP model from the correlation analysis is applied to score all food items in the food groups fruits, milk, nuts and seeds, red meat, vegetables, legumes and whole grains. These scores are used to downscale the health impacts of these food groups to food items.

### 2.4.2 Approach

The selected NP model was used applied to the food items (total of 908) in the seven food groups. For all food groups, a consumption weighted average is calculated with the following (Equation 10). The ratio of the selected indices to the consumption weighted average is used to calculate the downscaled HENI scores for all items (Equation 11).

*Equation 10:* $$Consumption\ weighted\ average\ nutrient\ index\ = \frac{\Sigma consumption \cdot \Sigma nutrient\ index}{\Sigma consumption}$$

*Equation 11:* $$HENI = -0.53 \cdot DRF\left(\frac{\mu DALYs}{g}\right) \cdot weight\ of\ item\ (g) \cdot \frac{nutrient\ index}{\substack{consumption\ weighted\ average \\ nutrient\ index}}$$

Red meat is the only food group that is associated with negative health impacts and therefore the corresponding DRF is negative. Therefore, an inverse ratio is used to calculate the HENI score for red meat. However, for the very unhealthy red meats with a negative NP score, this method yields a positive HENI score, which would suggest a positive health impact. Therefore, the HENI scores for food items with negative NP scores are calculated differently. The relation between the NP score and HENI score for positive items was calculated with a simple linear equation. To calculate the HENI score for the food items with negative NP scores, this equation was subtracted from an item with positive scores (Appendix D).

##### *2.4.2.1 Functional unit*

The NP scores are calculated per gram, and automatically, health impacts are also calculated per gram. In literature, the most commonly used functional units in the nutritional assessment are per mass unit (usually 100 g) or per energy unit (usually 100 kcal) (Grigoriadis et al., 2021). Therefore, health impacts are also rescaled to these functional units. It was chosen to calculate the initial NP scores and health impacts per gram so it can be built into the CONE-LCA framework and used for other research aims. For example, if one would calculate the health impact of a mixed dish, only multiplication by the weight in a recipe is needed.

## 2.5 Cluster analysis

### 2.5.1 Objective

In this analysis, it is explored whether food groups show high variability in NP scores. Clustering subsets food items with similar properties in such food groups can be of added value for improving nutrition health assessment. Thus, food groups with high variability were prioritized for the cluster analysis. The results of this analysis can be used in subsequent epidemiological studies.

### 2.5.2 Approach

For each food group, the minimum, maximum and coefficient of variation of the NP scores were calculated to explore whether clustering could be relevant for the food groups with high variation. A coefficient of variation measures relative variability in relation to the average and is calculated by the ratio between the standard deviation and the average (European Commission, n.d.). A coefficient of variation greater than one shows relatively high variability (Frost, 2020). Therefore, the food groups with a coefficient of variation greater than 1 were selected for cluster analysis. Moreover, if a coefficient of variation is close to 1 (e.g., >0.9), the range was considered as a second criterium for cluster analysis.

Jenks Natural Breaks classification method is used for this cluster analysis. A cluster analysis explores whether data can be grouped into subsets based on their similarities and differences (Johnstone et al., 2010). Jenks Natural Breaks is a one-dimensional data clustering method designed to find the optimal arrangement of values in a given number of clusters. The method splits the data into contiguous clusters in which the squared deviation within each cluster is minimized. The sum of the minimum squared deviation of each cluster is used to calculate the goodness of variance fit (GVF) (Equation 12).

*Equation 12:* $GVF = \Sigma minimum\ squared\ deviation\ /\ squared\ deviation\ of\ the\ dataset.$

GVF retains a value from 0 to 1, where 0 indicates no fit and 1 indicates a perfect fit (Zaiontz, n.d.). The analysis is performed with an excel Add-In (Zaiontz, n.d.). The model is used to explore a maximum of five clusters. No minimal cut-off GVF value to decide on the number of clusters was found in literature. Therefore, the cut-off was based on the fact that the GVF is a coefficient, just like a correlation coefficient which indicates no relation at 0 and a perfect relation at 1. For the Pearson correlation coefficient, literature indicates that a coefficient greater than 0.8 indicates a very good relation (Udovicic et al., 2007). Therefore, it was assumed that a GVF greater than 0.8 indicates a very good fit for this research. Consequently, a threshold GVF value of 0.8 in combination with a GVF difference of <0.1 to the GVF of a lower cluster, is chosen in determining the number of clusters. Moreover, the number of items in at least two clusters must be larger than 1. This is chosen to prevent positive and negative outliers from filling an entire cluster.

# 3 Results

## 3.1 Correlation analysis

Starting with the NRF model, the results show that the correlation between the adjusted NRF24.4·2 model and health impacts is 0.36 compared to 0.31 for the unweighted NRF24.4 version (Table 3). Furthermore, the NRF24.4·2 model correlates better with the health impacts than the original NRF9.3 model, although only with a slight difference of 0.02. Thus, the highest correlation was found in the model that was both extended and weighted. The correlation of the NRF24.4·2 model with health outcomes is significant (p-value <0.0001).

For the NQI model, the results show that the addition of weight to the extended models improves the correlation with health impacts in all formulas compared to the models that were only extended (Table 3). Nevertheless, versions that were both extended and weighted all correlate worse with health impacts than the NQI9.3 versions (unadjusted, exponential and linear). Moreover, the results show that the unadjusted NQI formulas correlate slightly better with health impacts than the adjusted linear and exponential formulas. Thus, of the extended NQI models, the highest correlation was found in the unadjusted NQI24.4·2 model. The correlation of the unadjusted NQI24.4·2 model with health outcomes is significant (p-value <0.0001).

Consequently, NRF24.4·2 and unadjusted NQI24.4·2 were selected for the additional correlation analysis. The unadjusted NQI24.4·2 will be referred to as NQI24.4·2 from now on.

The detailed results including nutrition sheets and NP indices of all 144 food items are added in Appendix E.

*Table 3: Pearson correlation coefficient of NP score with health impacts (HENI score). Highest scores of NRF model and NQI model marked in green.*

| | NP model | Pearson correlation coefficient |
|---|---|---|
| NRF | NRF24.4 | 0.31 |
| | NRF24.4·2 | 0.36 |
| | NRF9.3 | 0.34 |
| NQI | NQI24.4 | 0.29 |
| | NQI24.4·2 | 0.33 |
| | NQI9.3 | 0.35 |
| Exponen NQI | NQI24.4 | 0.27 |
| | NQI24.4·2 | 0.29 |
| | NQI9.3 | 0.31 |
| Linear NQI | NQI24.4 | 0.27 |
| | NQI24.4·2 | 0.31 |
| | NQI9.3 | 0.33 |

### 3.1.1 Correlation between aggregated health impacts and indicators

The DRFs retrieved from Stylianou et al. (2021) and consumption weighted average NP scores, calculated in the downscaling analysis, are displayed in table 4. The Pearson correlation coefficient between these DRFs (health impacts) and the consumption-weighted NRF scores is -0.93 (p-value 0.003). The Pearson correlation coefficient between these DRFs (health impacts) and the consumption-weighted NQI scores is -0.81 (p-value 0.03). Thus, the NRF model correlates better with health impacts than the NQI model and was selected for further analyses.

*Table 4: Dietary risk factors and consumption weighted average NP scores per food group.*

| Food group | Dietary risk factor (µDALY/g) | Consumption weighted average NRF (1/g) | Consumption weighted average NQI (1/g) |
|---|---|---|---|
| Fruits | -0.18 | 0.54 | 16.03 |
| Legumes | -0.23 | 1.63 | 42.89 |
| Milk | -0.01 | 0.31 | 17.95 |
| Nuts & seeds | -1.50 | 4.65 | 129.18 |
| Red meat | 0.10 | 1.48 | 71.28 |
| Vegetables | -0.08 | 1.17 | 26.27 |
| Whole grains | -0.34 | 1.65 | 38.82 |

## 3.2 Downscaling analysis

In the next sections, the results of the downscaled health impacts are summarized for red meat and vegetables. Detailed results of all food groups are added in appendix D.

### 3.2.1 Red meat

Downscaled health impacts for a subset of commonly consumed red meats from all animals in this food group are displayed in Table 5. The results show great variety in and between subgroups (e.g., salami scores more than 100 times worse than pork bacon, and sometimes beef scores better than pork and vice versa). Moreover, the relative ranking of food items is dependent on functional unit. For example, ham scores better per 100 grams while pork bacon scores better per 100 kcal.

*Table 5: Health impacts of a selected subset of commonly consumed red meats.*

| Main food description | HENI per g | HENI per 100 g | HENI per 100 kcal |
|---|---|---|---|
| **Pork** | | | |
| Pork, tenderloin, baked | -0.04 | -4.18 | -2.72 |
| Pork, spareribs, cooked | -0.07 | -6.81 | -1.73 |
| Chorizo | -0.10 | -9.67 | -2.79 |
| Ham, smoked or cured, cooked, lean and fat eaten | -0.13 | -13.40 | -9.00 |
| Pork bacon, smoked or cured, cooked | -0.14 | -13.53 | -2.89 |
| Salami | -17.41 | -1740.76 | -460.52 |
| **Beef** | | | |
| Beef steak, NS as to cooking method, NS as to fat eaten | -0.04 | -4.19 | -2.28 |
| Beef brisket, cooked, NS as to fat eaten | -0.05 | -4.87 | -2.26 |
| Ground beef, cooked | -0.06 | -6.15 | -2.37 |
| Frankfurter or hot dog, beef | -29.87 | -2987.48 | -902.56 |
| **Veal** | | | |
| Veal cutlet or steak, NS as to cooking method, lean only eaten | -0.05 | -4.98 | -2.75 |
| **Lamb** | | | |
| Lamb chop, NS as to cut, cooked, lean only eaten | -0.05 | -4.98 | -2.33 |
| Lamb, ribs, cooked, lean only eaten | -0.13 | -12.86 | -3.59 |
| **Goat** | | | |
| Goat ribs, cooked | -0.05 | -5.11 | -3.60 |

### 3.2.2 Vegetables

The downscaled health impacts for a selected subset of diverse vegetables are displayed in Table 6. Detailed results for all food items within vegetables can be found in appendix D. The results show that consuming spinach is associated with the highest beneficial health impacts and summer squash with the lowest beneficial health impacts compared to the other vegetables in this selection. Moreover, the canned vegetables in this selection are associated with lower beneficial health impacts than frozen or fresh vegetables of the same sort. In this food group again is visual that the relative performance of food items is dependent on the functional unit (e.g. raw spinach scores better per 100 kcal while fresh cooked spinach scores better per 100 g).

*Table 6: Health impacts of a selected subset of diverse vegetables prepared in different forms.*

| Main food description | HENI per g | HENI per 100 g | HENI per 100 kcal |
|---|---|---|---|
| **Asparagus** | | | |
| fresh, cooked, no added fat | 0.05 | 5.03 | 23.98 |
| frozen, cooked, no added fat | 0.07 | 6.52 | 36.20 |
| canned, cooked, no added fat | 0.04 | 4.34 | 22.85 |
| **Broccoli** | | | |
| raw | 0.10 | 9.98 | 29.37 |
| fresh, cooked, no added fat | 0.09 | 9.27 | 26.48 |
| frozen, cooked, no added fat | 0.07 | 6.80 | 24.27 |
| **Carrots** | | | |
| raw | 0.06 | 5.83 | 14.21 |
| fresh, cooked, no added fat | 0.05 | 5.36 | 12.47 |
| frozen, cooked, no added fat | 0.05 | 5.33 | 14.39 |
| canned, cooked, no added fat | 0.04 | 3.54 | 14.15 |
| **Spinach** | | | |
| raw | 0.24 | 23.94 | 104.11 |
| fresh, cooked, no added fat | 0.27 | 26.67 | 98.79 |
| frozen, cooked, no added fat | 0.25 | 24.79 | 72.90 |
| canned, cooked, no added fat | 0.21 | 20.79 | 90.38 |
| **Summer squash** | | | |
| fresh, cooked, no added fat | 0.02 | 2.33 | 11.08 |
| frozen, cooked, no added fat | 0.01 | 1.48 | 7.03 |
| canned, cooked, no added fat | 0.00 | 0.41 | 3.14 |

## 3.3 Cluster analysis

### 3.3.1 Variation in food groups

The coefficient of variation, minimum and maximum of NRF scores were calculated for all food groups per gram (Table 7). A coefficient of variation greater than 1 was found in red meats (20.60) and fruits (1.04). A coefficient of variation greater than 0.9 was found in vegetables (0.94). The range between maxima and minima is greater for vegetables than for fruits.

*Table 7: Coefficient of variation within food group, minimum and maximum NRF score and range. Coefficients of variation greater than 1 marked in green. Coefficient of variation greater than 0.9 marked in yellow.*

| Food group | Coefficient of variation | Min | Max | Range |
|---|---|---|---|---|
| Fruits | 1.04 | -1.65 | 3.35 | 5.00 |
| Legumes | 0.18 | 1.07 | 3.05 | 1.97 |
| Milk | 0.47 | -0.16 | 0.92 | 1.08 |
| Nuts & seeds | 0.52 | -2.16 | 20.66 | 22.83 |
| Red meat | 20.60 | -16.65 | 36.94 | 53.59 |
| Vegetables | 0.94 | -0.11 | 17.16 | 17.27 |
| Whole grains | 0.51 | 0.02 | 0.32 | 0.31 |

### 3.3.2 Cluster analysis

A cluster analysis was performed on fruits, red meat and vegetables. Results of the Jenks Natural Breaks analysis for these three food groups are summarized in Table 8. Detailed results of the analysis can be found in appendix F. The GVF threshold was crossed at three clusters for red meat and vegetables, and at four clusters for fruits. However, the second criterion, that in at least two clusters the number of items must be greater than 1, was not met at three clusters for red meat (Table 9). Consequently, the selected number of clusters is four for red meat.

*Table 8: GVF values from Jenks Natural Breaks analysis of red meat and vegetables. Scores at cut-off marked in green.*

| Food group | 2 clusters | 3 clusters | 4 clusters | 5 clusters |
|---|---|---|---|---|
| Fruit | 0.51 | 0.71 | 0.84 | 0.90 |
| Red meat | 0.68 | 0.84 | 0.91 | 0.95 |
| Vegetables | 0.72 | 0.83 | 0.91 | 0.94 |

*Table 9: Number of items in clusters based on the first criterion. If the second criterion is not met, the food group is marked in red.*

| Food group | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Fruits | 2 | 57 | 33 | 1 |
| Red meat | 1 | 293 | 1 | - |
| Vegetables | 244 | 54 | 48 | - |

Within the clusters, it was explored whether subgroups could be identified (Figure 3). Clusters are ascending order of nutrition score. In the food group fruits, the two unhealthiest processed fruits were found in cluster 1; canned fruits, other processed fruits, fresh citrus fruits and fresh prunus fruits in cluster 2; fresh berries and fresh tropical fruits in cluster 3, and only fresh guava in cluster 4. In the group of red meat, only brains filled cluster 1; ultra-processed beef and pork were found in cluster 2; processed beef and pork and all goat, lamb and veal meat in cluster 3; and only beef liver in cluster 4. In the group vegetables, beans, tomatoes, asparagus, carrots, lettuce, artichoke, and peppers were mainly found in cluster 1; broccoli and mustard greens in cluster 2; and spinach, kale, collards, and cress in cluster 3.

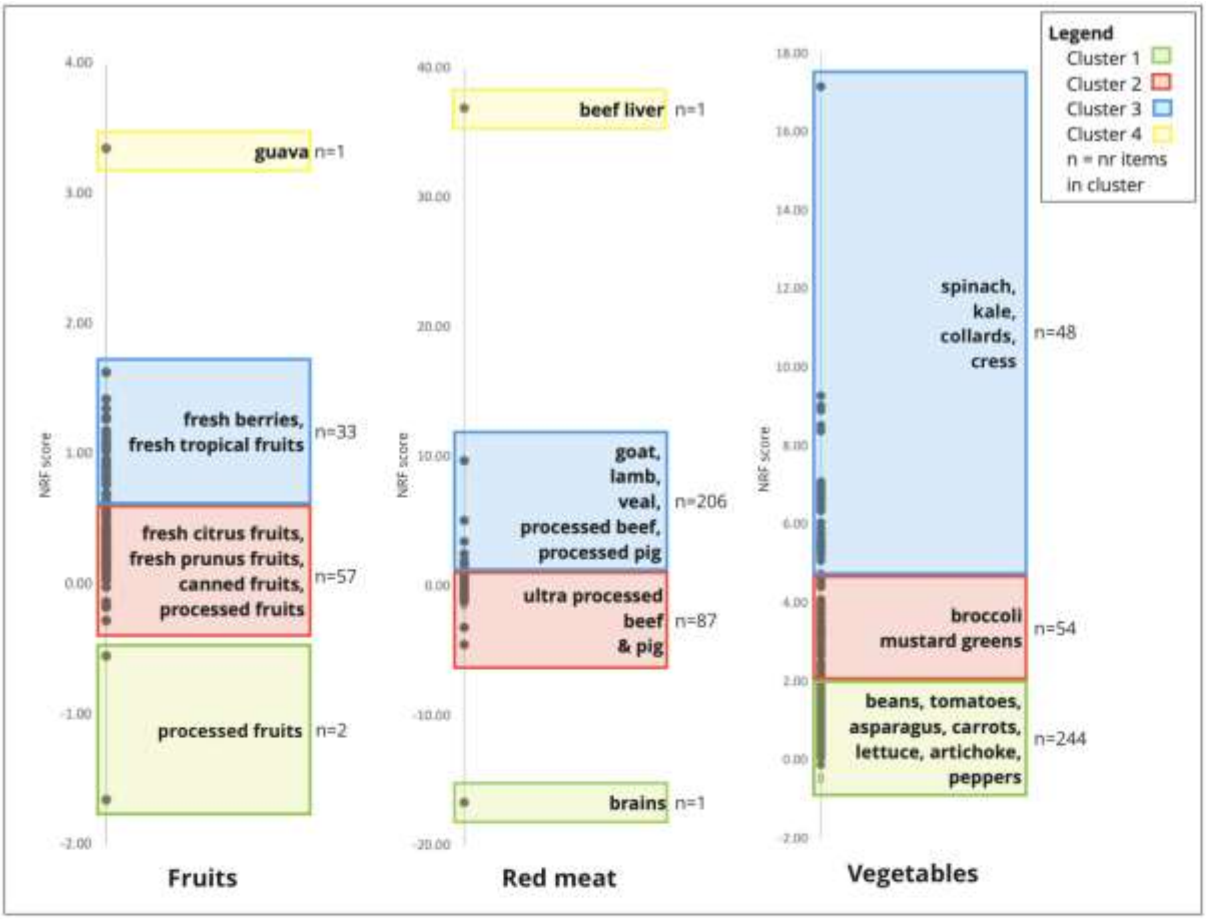


*Figure 3: Results of cluster analysis for fruits, red meat and vegetables. NRF scores of all items (dots) in food groups are displayed on Y-axis. In coloured boxes clusters are displayed including subgroups of items that were mainly identified in the cluster.*

# 4 Discussion

## 4.1 Key findings and interpretation

### 4.1.1 Correlation analysis

The results of the first analysis show that the addition of weight to detrimental nutrients, to resemble original ratios in the NP models, improves the correlation of the NP scores with health impacts. Regarding the NQI models, the linear NQI24.4·2 version correlates best with health impacts, with a correlation coefficient of 0.33 (P-value 0.0001). This confirms the use of the original NQI formula in literature. Regarding the NRF model versions, the NRF24.4·2 correlates best with health impacts with a correlation coefficient of 0.36 (P-value 0.0001).

Although the NRF24.4·2 model correlates better with health impacts than NQI24.4·2 in the first correlation analysis, the difference in correlation coefficients is very small (0.03). Yet, the results of the second correlation analysis also showed that the NRF24.4·2 model correlates better with health impacts than the NQI24.4·2 model (-0.93 vs. -0.81), with a larger difference in correlation coefficients (0.12). Besides these differences, other arguments were considered to determine which model was most suitable for this research. Namely, the NRF model is a verified method which is validated against the HEI. On the contrary, the NQI model is not yet verified at all. Moreover, the NRF model is a widely used method in many peer-reviewed articles (Drewnowski, 2005; Drewnowski and Fulgoni, Drewnowski et al., 2021; 2008; Fulgoni et al., 2009; Maillot et al. 2007; Ridoutt, 2021; Sugimoto et al., 2022). Also, its secondary use to assess other important parameters like affordability and sustainability is tested in literature (Drewnowski et al., 2021; Maillot et al. 2007; Ridoutt, 2021; Sugimoto et al., 2022). On the contrary, the NQI method is not widely used in peer-reviewed articles yet. Based on these considerations, the NRF model was considered most suitable for this research.

Finally, the correlation coefficient at the aggregated levels with the NRF index (-0.93) is substantially higher than the original health impact indicator of food items with the NRF index (0.36). This shows the added value of downscaling health impact indicators to the level of food items with the use of nutrient profiles of food items. It can be concluded that the NRF model captures the health impacts even though it assesses individual nutrients only.

### 4.1.2 Downscaling analysis

The food groups red meat and vegetables were selected to discuss in the report. The results show that health impacts differ substantially per item in the food group red meat. For example, in the subgroup pork, salami scores considerably worse than bacon. Moreover, across animal meats from the same body origin such as pork ribs, goat and lamb ribs health impacts differ. Besides this, the results indicate the importance of a functional unit.

The results in the vegetable food group show that health impacts differ substantially per vegetable sort. Moreover, it shows that not only the type of vegetable is important for its nutritious properties, but also its condition (raw/cooked, fresh/frozen/canned). Canned vegetables are often less beneficial for health than fresh or frozen items. Moreover, some vegetables are more nutritious when eaten raw, others are more nutritious when cooked. Concluding, health impacts can differ strongly between food items within food groups.

### 4.1.3 Cluster analysis

Coefficients of variation were greater than 1 in red meat and fruits, thus it can be concluded that these food groups show relatively high variability. Consequently, prioritizing these food groups for clustering can be relevant for epidemiological studies. Moreover, the coefficient of variation in vegetables was close to 1 and therefore the range was considered. This showed that, compared to fruits, the range is relatively large. Consequently, it was concluded vegetables should also be prioritized for clustering.

Following the Jenks Natural Breaks analysis, a GVF greater than 0.8 was found at four clusters for fruits and at three clusters for red meat and vegetables. However, based on a second criterium that prevents a single outlier from filling an entire cluster, the number of clusters for red meat was raised to four. Subgroups of food items with similar properties were found in these clusters. These results suggest that using these clusters in subsequent epidemiological research can be relevant.

## 4.2 Reflection on methodology

### 4.2.1 Data quality

All data used in this study originates from secondary sources. Data on food items and their nutritional composition were retrieved from FoodData Central (USDA, 2019a), which was in line with the research method of Stylianou et al. (2021). This database is widely used and is fit for the purpose of this research. Yet, this database is limited to certain nutrients and some important nutrients are missing (appendix A). Iodine is a particularly important missing nutrient in this database since iodine deficiencies are very common word wide (American Thyroid Association, 2021). Iodine is an essential mineral for thyroid gland function (National Institutes of Health, 2021). Moreover, in relation to total energy consumption, trans fats and added sugars are two nutrients, currently missing in the NP model, that are important to consider because of their detrimental health impacts and high consumption in the American diet (Ganguly and Pierce, 2015; Marriott et al., 2010; Remig et al., 2010; Vos et al., 2017). Nevertheless, most health impacts will be covered by all the other nutrients in the index, since food items that are high in trans fats or added sugars are often also low in beneficial nutrients and high in detrimental nutrients (Smith, 2020).

Despite gaps in consumption data of some specific food items, the GeNUS database is fit for the purpose of this research. This mainly results from high-quality, open-access, US-specific, detailed daily consumption data being scarce in the literature. Consumption data reliability can be improved by further development of such a database. Besides, data on total dietary mass in the US, which was needed for the NQI model, was not found in the literature and therefore assumptions have been made. The ratio between actual total protein intake and total protein supply was taken and applied to the total food supply food balance sheets from FAOstat (Chapter **Fout! Verwijzingsbron niet gevonden.**). Thus, it was assumed that the ratio between actual protein intake in the US and total protein supply in the US was the same as the ratio between actual dietary mass and total supply. The quality of the data could be improved with reliable literature on total dietary mass in the US.

The dietary risk factors developed by Stylianou et al. (2021) build on the dietary risks from the GBD studies. These dietary risks are based on rather few health outcomes (Appendix C). For example, the positive health impact of milk consumption builds on a decreased risk at colorectal cancer only, while consumption of milk is possibly associated with other health outcomes as well. Moreover, only non-communicable diseases are considered in the DRFs while infectious diseases are not considered at all. Nutrient deficiencies play an important role in immune system functioning and are therefore associated with communicable diseases (Bhaskaram, 2002). Also, overconsumption of certain nutrients is associated with certain health outcomes (Combet and Gray, 2019). Thus, it is important to balance out the impact of undernutrition and overnutrition on diseases. Consideration of more diseases in the GBD for the development of DRFs would improve the validity of the results.

### 4.2.2 Use of NP models

Dietary guidelines have shifted from nutrient requirements to food recommendations (U.S. Department of Agriculture, 2020). However, nutrient profiling methods naturally focus on isolated nutrients only. Consequently, scientists argue for a hybrid approach to nutrient profiling, where both the health impacts of food groups and isolated nutrients are considered (Drewnowski et al., 2019; Drewnowski and Fulgoni, 2020). In this research, such a hybrid approach is explored: Dietary risks of food groups, building on epidemiological studies, were combined with nutrient profiles of food items into a joint health metric, expressed in HENI. Still, for some scientists downscaling the health impacts to individual food items remains controversial because of these new recommendations (National Institutes of Health, 2022). Therefore, besides downscaling health impacts to individual food items, it is explored whether clusters of items with similar properties within food groups with high variation can be found. The result of this analysis can be used in subsequent epidemiological studies.

The NRF index is a validated and commonly used nutritional quantity index and it is easily extendible (Fulgoni et al. 2009), which makes it fit for the aim of this study. The NQI model uses a similar approach to the NRF9.3 model, including the same beneficial and detrimental nutrients (Sonesson et al., 2019). The advantage of the NQI over the NRF is that the original model already takes dietary context into account, although this can also be accomplished indirectly with the NRF model through capping. This is important considering that the health impacts of nutrients, among other things, depend on dietary context (Hess et al., 2017). However, the NQI model is not verified yet (Sonesson et al., 2019). Therefore, both models were tested in this research.

Using NP models like the NRF in the nutritional-health assessment of foods comes with limitations. These models only consider isolated nutrients in the assessment of foods. However, people do not consume isolated nutrients or even isolated food items, but combinations of foods that contain various nutrients and non-nutrients (Kourlaba and Panagiotakos, 2009). Non-nutrients are important no-calorie substances that play vital roles in our body, such as anti-inflammatory and antioxidant actions, increased insulin sensitivity and reduction of intestinal absorption of glucose and fat (Ribeiro et al., 2019). These vital non-nutrients are not captured in NP models. On the other hand, substances that are detrimental to health, like the carcinogenic features in red meat (Bouvard et al., 2015), are also not captured in these NP models. Furthermore, considering that total energy intake remains the same, high consumption of one food may be associated with low consumption of other foods (Kourlaba and Panagiotakos, 2009). Besides, nutrients interact with each other, influencing bioavailability and absorption (Combet and Gray, 2019; Kourlaba and

Panagiotakos, 2009). For example, fat promotes the absorption of vitamin D (Dawson-Hughes et al., 2015). Consequently, equal weighing of nutrients in these NP models may not be a valid method for assessing overall nutritional value. Different nutrients may not contribute equally to health impacts or to the nutritional value of a food item, which is also dependent on interactions and dietary context (Hess et al., 2017). This might be the biggest challenge from a methodological point of view. On top of this, accurately estimating the nutrient intake of food items is challenging. Firstly, because the nutrient content of food items varies due to seasonality (Kumar et al., 2015; Waswa et al., 2021). Secondly, because the recipes of mixed dishes can vary over time and across places (Afshin et al., 2019). All things considered, attribution of health impacts to food items or individual nutrients remains very complicated.

#### *4.2.2.1 Capping of nutrients in NRF*

Naturally, the NRF model does not take dietary context into account. Therefore, capping overconsumed nutrients in the US was integrated in the NRF model. In this way, more weight is assigned to under-consumed nutrients. Various aspects of the capping method can be improved to increase validity of the results. Firstly, once the recommended DV of a nutrient is crossed, the curve for beneficial health impacts flattens (Institute Of Medicine, 2011). For simplification in this research, it was assumed that consumption of a nutrient after the recommended DV was reached, was associated with no additional beneficial health impacts. Namely, nutrients are capped at 100% of the recommended DV. The method could be improved by incorporating a more realistic relationship between health impacts and consumption after reaching the recommended DV. Furthermore, fat-soluble vitamins are excluded from capping based on the assumption that excess of these vitamins are stored in fat tissue and can be used in periods with lower dietary intake or higher demand due to sickness etcetera (Blomhoff et al, 1990; Card et al., 2013). The fact that the relationship between vitamin intake and health impacts changes after reaching a certain intake of these vitamins is ignored in this research for simplification. The validity of the method would be improved when the flexibility in storage and absorption from fat tissue of these vitamins was considered. Also, MUFA and PUFA are capped following the basic capping method in this research. However, these nutrients compete for the same enzymes (Mariamenatu and Abdu, 2021). Therefore, in future research, a more comprehensive capping method could be explored to improve the validity of this method. Lastly, crossing the upper daily limit for some 'beneficial' nutrients (e.g. iron) is not only associated with little extra beneficial health impacts, but with detrimental health impacts from this point on (Qiao and Feng, 2013). This is especially important when one would study fortified foods. Since this research is based on unfortified foods, this fact is not considered. However, if this method would be applied to study fortified foods, it would be recommended to account for detrimental effects after reaching the upper daily limit. This applies especially to poorer countries where they have fortified programs and supplementation (World Health Organization, 2003).

Moreover, it was chosen to cap overconsumed nutrients according to the ratio between recommended DV and actual daily consumption over capping at 100% for a fixed FU. The latter was not preferred since this approach is highly dependent on the FU and rescaling impacts to dishes or daily diet would underestimate or overestimate the impacts.

#### 4.2.2.2 Assumptions energy intake

In this research, a normal energy intake of approximately 2000 kcal is assumed. However, in a country with high obesity prevalence, which is the case in the US, individuals often consume extreme energy intakes (Chooi et al., 2019). Although the nutrient intake of these individuals might be satisfactory, the extreme energy consumption will negatively affect their health impacts (Chooi et al., 2019). Currently, this is not considered in this research. The validity of the method could be improved by adding a correction factor on health impacts based on the obesity prevalence or average energy intake in a country. Similarly, physical activity affects the health impacts of dietary choices (Rhodes et al., 2017), but in this case positively. A similar correction factor could be applied to health impacts based on average physical activity in a country.

### 4.2.3 Method for cluster analysis

Jenks Natural Breaks works similar to K-means clustering. Yet, K-means is usually applied to multivariate data and Jenks Natural Breaks to univariate data. Therefore, the Jenks Natural Breaks method is fit for the aim of this research. For the cut-off GVF value, no literature was found and therefore assumptions have been made. The validity of the GVF cut-off value can therefore be improved when a minimal GVF value for a good fit is found in the literature.

### 4.2.4 Choice of functional unit

NP scores and health impacts are calculated per gram. In this way, the results can be incorporated into the CONE-LCA model or used for other purposes. Nevertheless, health impacts were also rescaled to a functional unit of 100 g and 100 kcal, which are common FUs in nutritional assessment of foods (Grigoriadis et al., 2021), to show what results will look like in a nutritional LCA. This showed the substantial impact of a FU on the results in comparative analysis. For example, pork bacon scored worse than lamb ribs with a FU of 100 g, while lamb ribs scored worse than pork bacon with a FU of 100 kcal. It was chosen to use these FUs to show the variance within a food group and make a case for downscaling and making food items easily comparable. Nevertheless, these results can give a distorted picture of the health impacts of foods that are consumed in very small or large amounts. Although the results are limited to these FUs in this research, in future research the health impacts can also be rescaled to serving sizes or portions to assess the role of a food item in an actual diet.

## 4.3 Societal relevance

The improved CONE-LCA model may be useful in policy applications, since healthy and sustainable diets are very relevant issues in politics. Healthy diets are also very relevant from an economical perspective, as lifestyle-induced health care costs are enormous and still rising (Edington, 2020). Consequently, the is a need for evidence-based policy making for a transition towards healthy diets that are also sustainable (Davis, D'Odorico, and Rulli, 2014; Reinhardt et al., 2020). The model makes it possible to compare food items on human health and environmental benefits, which can be the basis for sustainable dietary guidelines. Sustainability, however, encompasses not only environmental concepts but also social and economic issues. Considering food security and inequalities, a sustainable and healthy diet is from a policy point of view quite a challenge, since sustainable diets are not affordable for certain income groups (Green et al., 2020). Therefore, future research can build on this thesis by integrating the economic aspect into the model to assess trade-offs.

## 4.4 Academic relevance and future research

This work provides a comprehensive approach to study the impacts of marginal dietary shifts on human health. The research builds on the results of Stylianou et al. (2021) by refining the health impact assessment of food items. The research contributes to improvements in the nutritional-health assessment within the CONE-LCA framework since the results show that downscaling health impacts using nutrient profiles of food items, better predicts health impacts related to consumption of food items than the original indicators. LCA is a valued and often used tool within the field of Industrial Ecology (IE) (Guinee et al., 2002). Therefore, improvement of the CONE-LCA framework is highly relevant to the field of Industrial Ecology. This thesis served as a first exploration of how nutrition-health assessment within the CONE-LCA framework can be improved. Therefore, the results of this study are not intended to be prescriptive and should be interpreted with caution. Nevertheless, the added value of downscaling health impacts to the level of food items is demonstrated in this thesis.

The improved nutrition health assessment with downscaled health impact indicators is also highly relevant outside of the CONE-LCA framework and can be used on its own to assess the health impacts of food items. Although the scope of this research was limited to the US, the model can also be applied to other countries or to assess the health impacts of food items on a global level. The strength of this model lies in its wide applicability, as it can be adjusted to not only assess the health impacts of marginal dietary shifts, but also to compare broader dietary patterns within and across countries. Furthermore, the results of the cluster analysis showed that use of clusters of food items within food groups (e.g. fruits, vegetables and red meat) might be relevant for epidemiological research.

In future research, further development and validation of the nutrition health assessment method would be recommended. Future research should focus on: (i) further development of the NP model. Ideally, the NP model should account for interactions between nutrients. Moreover, the method for capping can be further improved. Also, incorporating important nutrients excluded in this research like iodine and trans fats would further improve the model; (ii) validating the NP model against a verified health assessment metric like the HEI; (iii) consideration of more diseases in the DRFs and specifically also communicable diseases; (iv) incorporating fish and seafood, dairy and processed foods into the model; (v) using clusters of food groups in epidemiological research; and (vi) affordability of healthy and sustainable diets.

# 5 Conclusion

In this thesis, the use of NP models to refine DRFs in the nutrition-health assessment of food items was explored. The results showed that the NRF24.4·2 correlates slightly better with health impacts than the original NRF9.3 model and the NQI models and was therefore selected for further analysis. The fact that the correlation coefficient at the aggregated levels with the nutrient index is substantially higher than the original health impact indicator of food items with the nutrient index, shows that NRF24.4·2 model captures health impacts even though it assesses individual nutrients only. Consequently, the added value of downscaling health impact indicators to the level of food items with the use of NP models was demonstrated. Concluding, in the nutritional health assessment of food items, the NRF24.4·2 model used in this research is a better predictor for health impacts than the method used by Stylianou et al. (2021) which is based on aggregated DRFs only.

Further analysis showed that downscaled health impacts can differ substantially between food items within food groups. Moreover, not only the type of food item is important for its nutritious properties, but also its condition (e.g. raw/cooked, fresh/frozen/canned/dried). Furthermore, since some food groups (e.g. fruits, vegetables and red meat) are highly variable in nutrient density, prioritizing these food groups for clustering can be of added value in future epidemiological studies.

The improved nutrition-health assessment model is not only relevant within the CONE-LCA framework but can be used on its own to assess the health impacts of food items. The model is widely applicable, as it can be adjusted to assess the health impacts of marginal dietary shifts as well as to compare broader dietary patterns within and across countries. The model may be very useful in policy applications, as it enables the comparison of beneficial impacts of food items on human health and the environment, which can be the basis for sustainable dietary guidelines. Future research can focus on further development and validation of the model as well as the incorporation of the economic dimension in the model for policy purposes.

Overall, this research provides a deepening step toward a further improved nutrition health assessment of food items, within the CONE-LCA framework and on itself.

**References**


Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., Mullany, E. C., Abate, K. H., Abbafati, C., Abebe, Z., Afarideh, M., Aggarwal, A., Agrawal, S., Akinyemiju, T., Alahdab, F., Bacha, U., Bachman, V. F., Badali, H., Badawi, A., . . . Murray, C. J. L. (2019). Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 393*(10184), 1958–1972.

American Thyroid Association. (2021). *Iodine Deficiency*. Retrieved June 13, 2022, from https://www.thyroid.org/wp-content/uploads/patients/brochures/IodineDeficiency_brochure.pdf

Bhaskaram, P. (2002). Micronutrient Malnutrition, Infection, and Immunity: an Overview. *Nutrition Reviews, 60*(suppl_5), S40–S45.

Behrens, P., Kiefte-de Jong, J. C., Bosker, T., Rodrigues, J. F. D., De Koning, A., & Tukker, A. (2017). Evaluating the environmental impacts of dietary recommendations. *Proceedings of the National Academy of Sciences, 114*(51), 13412–13417.

Blomhoff, R., Green, M. H., Berg, T., & Norum, K. R. (1990). Transport and Storage of Vitamin A. *Science, 250*(4979), 399–404.

Bouvard, V., Loomis, D., Guyton, K. Z., Grosse, Y., Ghissassi, F. E., Benbrahim-Tallaa, L., Guha, N., Mattock, H., & Straif, K. (2015). Carcinogenicity of consumption of red and processed meat. *The Lancet Oncology, 16*(16), 1599–1600.

Card, D. J., Gorska, R., Cutler, J., & Harrington, D. J. (2013). Vitamin K metabolism: Current knowledge and future research. *Molecular Nutrition & Food Research, 58*(8), 1590–1600.

Center for Disease Control and Prevention (CNC) and National Center for Health Statistics. (2019, July). *NHANES 2015–2016 Dietary Data*. Center for Disease Control and Prevention. Retrieved June 10, 2022, from https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Dietary&CycleBeginYear=2015

Chooi, Y. C., Ding, C., & Magkos, F. (2019). The epidemiology of obesity. *Metabolism, 92*, 6–10.

Davis, K. F., D'Odorico, P., & Rulli, M. C. (2014). Moderating diets to feed the future. *Earth's Future, 2*(10), 559–565.

Dawson-Hughes, B., Harris, S. S., Lichtenstein, A. H., Dolnikowski, G., Palermo, N. J., & Rasmussen, H. (2015). Dietary Fat Increases Vitamin D-3 Absorption*. Journal of the Academy of Nutrition and Dietetics, 115*(2), 225–230.

Doran-Browne, N.A., Eckard, R.J., Behrendt, R. (2015). Nutrient density as a metric for comparing greenhouse gas emissions from food production. *Climatic Change 129*, 73–87.

Drewnowski, A. (2005). Concept of a nutritious food: toward a nutrient density score. *The American Journal of Clinical Nutrition, 82*(4), 721–732.

Drewnowski, A., Dwyer, J., King, J. C., & Weaver, C. M. (2019). A proposed nutrient density score that includes food groups and nutrients to better align with dietary guidance. *Nutrition Reviews, 77*(6), 404–416.

Drewnowski, A., & Fulgoni, V.L. (2008). Nutrient profiling of foods: creating a nutrient-rich food index. *Nutrition Reviews, 66*(1), 23–39.

Drewnowski, A., & Fulgoni, V. L. (2020). New Nutrient Rich Food Nutrient Density Models That Include Nutrients and MyPlate Food Groups. *Frontiers in Nutrition, 7*.

Drewnowski, A., Smith, J., & Fulgoni, V. L. (2021). The New Hybrid Nutrient Density Score NRFh 4:3:3 Tested in Relation to Affordable Nutrient Density and Healthy Eating Index 2015: Analyses of NHANES Data 2013–16. *Nutrients, 13*(5), 1734.

Edington, D. W., Burton, W. N., & Schultz, A. B. (2020). Health and Economics of Lifestyle Medicine Strategies. *American Journal of Lifestyle Medicine, 14*(3), 274–277.

European Commission. (n.d.). *Coefficient of Variation*. Retrieved May 12, 2022 from https://datacollection.jrc.ec.europa.eu/wordef/coefficient-of-variation

Fardet, A., Rock, E., Bassama, J., Bohuon, P., Prabhasankar, P., Monteiro, C., Moubarac, J. C., & Achir, N. (2015). Current Food Classifications in Epidemiological Studies Do Not Enable Solid Nutritional Recommendations for Preventing Diet-Related Chronic Diseases: The Impact of Food Processing. *Advances in Nutrition, 6*(6), 629–638.

Farvid, M. S., Stern, M. C., Norat, T., Sasazuki, S., Vineis, P., Weijenberg, M. P., Wolk, A., Wu, K., Stewart, B. W., & Cho, E. (2018). Consumption of red and processed meat and breast cancer incidence: A systematic review and meta-analysis of prospective studies. *International Journal of Cancer, 143*(11), 2787–2799.

Food and Agriculture organization of the United Nations (FAO). (2019). *Food Balances (2010-).* FAOSTAT statistical database [Dataset]. https://www.fao.org/faostat/en/#data/FBS

Food and Drug Administration (FDA). (2022, February 25). *Daily Value on the New Nutrition and Supplement Facts Labels*. U.S. Food and Drug Administration. Retrieved February 14, 2022, from https://www.fda.gov/food/new-nutrition-facts-label/daily-value-new-nutrition-and-supplement-facts-labels

Fern, E. B., Watzke, H., Barclay, D. V., Roulin, A., & Drewnowski, A. (2015). The Nutrient Balance Concept: A New Quality Metric for Composite Meals and Diets. *PLOS ONE, 10*(7), e0130491.

Forkuoh, F., Boadi, N. O., Borquaye, L. S., & Afful, S. (2018). Risk Of Human Dietary Exposure To Organochlorine Pesticide Residues In Fruits From Ghana. *Scientific Reports, 8*(1).

Forouzanfar, M. H., Alexander, L., Anderson, H. R., Bachman, V. F., Biryukov, S., Brauer, M., Burnett, R., Casey, D., Coates, M. M., Cohen, A., Delwiche, K., Estep, K., Frostad, J. J., KC, A., Kyu, H. H., Moradi-Lakeh, M., Ng, M., Slepak, E. L., Thomas, B. A., . . . Murray, C. J. (2015). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet, 386*(10010), 2287–2323.

Frost, J. (2020, October 26). *Coefficient of Variation in Statistics*. Statistics By Jim. Retrieved June 10, 2022, from https://statisticsbyjim.com/basics/coefficient-variation/

Fulgoni, V. L., Keast, D. R., & Drewnowski, A. (2009). Development and Validation of the Nutrient-Rich Foods Index: A Tool to Measure Nutritional Quality of Foods. *The Journal of Nutrition, 139*(8), 1549–1554.

Gakidou, E., Afshin, A., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abdulle, A. M., Abera, S. F., Aboyans, V., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Abyu, G. Y., Adedeji, I. A., Adetokunboh, O., Afarideh, M., Agrawal, A., Agrawal, S., Ahmadieh, H., . . . Murray, C. J. L. (2017). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet, 390*(10100), 1345–1422.

Ganguly, R., & Pierce, G. N. (2015). The toxicity of dietary trans fats. *Food and Chemical Toxicology, 78,* 170–176.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science, 327*(5967), 812–818.

Grant, C. A., & Hicks, A. L. (2018). Comparative Life Cycle Assessment of Milk and Plant-Based Alternatives. *Environmental Engineering Science, 35*(11), 1235–1247.

Green, A., Nemecek, T., Chaudhary, A., & Mathys, A. (2020). Assessing nutritional, health, and environmental sustainability dimensions of agri-food production. *Global Food Security, 26*, 100406.

Grigoriadis, V., Nugent, A., & Brereton, P. (2021). Working towards a combined measure for describing environmental impact and nutritive value of foods: A review. *Trends in Food Science & Technology, 112,* 298–311.

Guinée, J.B. (Ed.), M. Gorrée, R. Heijungs, G. Huppes, R. Kleijn, A. de Koning, L. van Oers, A. Wegener Sleeswijk, S.Suh, H.A. Udo de Haes, J.A. de Bruijn, R. van Duin and M.A.J. Huijbregts. (2002). *Handbook on Life Cycle Assessment: Operational Guide to the ISO Standards.* Series: Eco-Efficiency in Industry and Science, Vol. 7. Springer. Dordrecht.

Hess, J., Rao, G., & Slavin, J. (2017). The Nutrient Density of Snacks. *Global Pediatric Health, 4*, 2333794X1769852.

Institute of Medicine, Board, Ross, C. A., Taylor, C. L., Yaktine, A. L., & Valle, H. D. B. (2011). *Dietary Reference Intakes for Calcium and Vitamin D*. National Academies Press.

Johnstone, E. C., Owens, D. C., & Lawrie, S. M. (2010). *Companion to Psychiatric Studies.* Churchill Livingstone.

Jones, A. D., Hoey, L., Blesh, J., Miller, L., Green, A., & Shapiro, L. F. (2016). A Systematic Review of the Measurement of Sustainable Diets. *Advances in Nutrition: An International Review Journal, 7*(4), 641–664.

Katz, D. L., Njike, V. Y., Rhee, L. Q., Reingold, A., & Ayoob, K. T. (2010). Performance characteristics of NuVal and the Overall Nutritional Quality Index (ONQI). *The American Journal of Clinical Nutrition, 91*(4), 1102S-1108S.

Khachatrian, E. (2021, March 1). *Banana vs Orange - Health impact and Nutrition Comparison.* Food Struct. Retrieved June 15, 2022, from https://foodstruct.com/compare/bananas-vs-oranges

Kourlaba, G., & Panagiotakos, D. B. (2009). Dietary quality indices and human health: A review. *Maturitas, 62*(1), 1–8.

Kumar, S., Sahoo, D., & Levine, I. (2015). Assessment of nutritional value in a brown seaweed Sargassum wightii and their seasonal variations. *Algal Research, 9*, 117–125.

Lampe, J. W. (1999). Health effects of vegetables and fruit: assessing mechanisms of action in human experimental studies. *The American Journal of Clinical Nutrition, 70*(3), 475s–490s.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., AlMazroa, M. A., Amann, M., Anderson, H. R., Andrews, K. G., Aryee, M., Atkinson, C., Bacchus, L. J., Bahalim, A. N., Balakrishnan, K., Balmes, J., Barker-Collo, S., Baxter, A., Bell, M. L., . . . Ezzati, M. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2224–2260.

Lukas, M., Palzkill, A., Rohn, H., & Liedtke, C. (2013). The nutritional footprint: an innovative management approach for the food sector. *WIT Transactions on Ecology and The Environment*.

Maillot, M., Darmon, N., Darmon, M., Lafay, L., & Drewnowski, A. (2007). Nutrient-Dense Food Groups Have High Energy Costs: An Econometric Approach to Nutrient Profiling. *The Journal of Nutrition, 137*(7), 1815–1820.

Mariamenatu, A. H., & Abdu, E. M. (2021). Overconsumption of Omega-6 Polyunsaturated Fatty Acids (PUFAs) versus Deficiency of Omega-3 PUFAs in Modern-Day Diets: The Disturbing Factor for Their “Balanced Antagonistic Metabolic Functions” in the Human Body. *Journal of Lipids, 2021*, 1–15.

Marriott, B. P., Olsho, L., Hadden, L., & Connor, P. (2010). Intake of Added Sugars and Selected Nutrients in the United States, National Health and Nutrition Examination Survey (NHANES) 2003–2006. *Critical Reviews in Food Science and Nutrition, 50*(3), 228–258.

Micha, R., Kalantarian, S., Wirojratana, P., Byers, T., Danaei, G., Elmadfa, I., Ding, E., Giovannucci, E., Powles, J., Smith-Warner, S., Ezzati, M., & Mozaffarian, D. (2011). Estimating the global and regional burden of suboptimal nutrition on chronic disease: methods and inputs to the analysis. *European Journal of Clinical Nutrition, 66*(1), 119–129.

Micha, R., Shulkin, M. L., Peñalvo, J. L., Khatibzadeh, S., Singh, G. M., Rao, M., Fahimi, S., Powles, J., & Mozaffarian, D. (2017). Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE). *PLOS ONE, 12*(4), e0175149.

National Institutes of Health. (2021, March 22*). Office of Dietary Supplements - Iodine*. Retrieved June 13, 2022, from https://ods.od.nih.gov/factsheets/Iodine-Consumer/

National Institutes of Health. (2022, June 2). *Office of Dietary Supplements - Vitamin D.* Retrieved June 15, 2022, from https://ods.od.nih.gov/factsheets/VitaminD-HealthProfessional/

National Research Council. (2012). *Analysis of Cancer Risks in Populations Near Nuclear Facilities: Phase 1.* Washington, DC: The National Academies Press.

Paulot, F., Jacob, D. J. (2014). Hidden cost of US agricultural exports: particulate matter from ammonia emissions. ammonia pollution from farming may exact hefty health costs. *Environ. Sci. Technol. 48*, 903–908.

Qiao, L., & Feng, Y. (2013). Intakes of heme iron and zinc and colorectal cancer incidence: a meta-analysis of prospective studies. *Cancer Causes & Control, 24*(6), 1175–1183.

Qin, P., Li, Q., Zhao, Y., Chen, Q., Sun, X., Liu, Y., Li, H., Wang, T., Chen, X., Zhou, Q., Guo, C., Zhang, D., Tian, G., Liu, D., Qie, R., Han, M., Huang, S., Wu, X., Li, Y., . . . Zhang, M. (2020). Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose–response meta-analysis of prospective cohort studies. *European Journal of Epidemiology, 35*(7), 655–671.

Reinhardt, S. L., Boehm, R., Blackstone, N. T., El-Abbadi, N. H., McNally Brandow, J. S., Taylor, S. F., & DeLonge, M. S. (2020). Systematic Review of Dietary Patterns and Sustainability in the United States. *Advances in Nutrition, 11*(4), 1016–1031.

Remig, V., Franklin, B., Margolis, S., Kostas, G., Nece, T., & Street, J. C. (2010). Trans Fats in America: A Review of Their Use, Consumption, Health Implications, and Regulation. *Journal of the American Dietetic Association, 110*(4), 585–592.

Ribeiro, P. V. D. M., Andrade, P. A., Hermsdorff, H. H. M., dos Santos, C. A., Cotta, R. M. M., Estanislau, J. D. A. S. G., Campos, A. A. D. O., & Rosa, C. D. O. B. (2019). Dietary non-nutrients in the prevention of non-communicable diseases: Potentially related mechanisms. *Nutrition, 66*, 22–28.

Ridoutt, B. (2021). An Alternative Nutrient Rich Food Index (NRF-ai) Incorporating Prevalence of Inadequate and Excessive Nutrient Intake. *Foods, 10*(12), 3156.

Ritchie, H., Reay, D. S., & Higgins, P. (2018). The impact of global dietary guidelines on climate change. *Global Environmental Change, 49*, 46–55.

Rhodes, R. E., Janssen, I., Bredin, S. S., Warburton, D. E., & Bauman, A. (2017). Physical activity: Health impact, prevalence, correlates and interventions. *Psychology & Health, 32*(8), 942–975.

Sibhatu, K. T., Krishna, V. V., & Qaim, M. (2015). Production diversity and dietary diversity in smallholder farm households. *Proceedings of the National Academy of Sciences, 112*(34), 10657–10662.

Smith, J. (2020, May 15). *How do processed foods affect your health?* Medical News Today. Retrieved July 8, 2022, from https://www.medicalnewstoday.com/articles/318630

Smith, M. R., & Matt. (2016, July 7). *Edible Food by Country and Year* [Dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/B1QO6O

Smith, M. R., Micha, R., Golden, C. D., Mozaffarian, D., & Myers, S. S. (2016). Global Expanded Nutrient Supply (GENuS) Model: A New Method for Estimating the Global Dietary Supply of Nutrients. *PLOS ONE, 11*(1), e0146976.

Sonesson, U., Davis, J., Hallström, E., & Woodhouse, A. (2019). Dietary-dependent nutrient quality indexes as a complementary functional unit in LCA: A feasible option? *Journal of Cleaner Production, 211*, 620–627.

Springmann, M., Wiebe, K., Mason-D'Croz, D., Sulser, T. B., Rayner, M., & Scarborough, P. (2018). Health and nutritional aspects of sustainable diet strategies and their association with environmental impacts: a global modelling analysis with country-level detail. *The Lancet Planetary Health, 2(*10), e451–e461.

Streppel, M. T., Sluik, D., van Yperen, J. F., Geelen, A., Hofman, A., Franco, O. H., Witteman, J. C. M., & Feskens, E. J. M. (2014). Nutrient-rich foods, cardiovascular diseases and all-cause mortality: the Rotterdam study. *European Journal of Clinical Nutrition, 68*(6), 741–747.

Stylianou, K. S., Fantke, P., Fulgoni, V. L., & Jolliet, O. (2017). Comprehensive Assessment of Human Health Impacts and Benefits of Fruits and Vegetables in a LCA Context. *The FASEB Journal, 31*, lb387-lb387.

Stylianou, K. S., Fulgoni, V. L., & Jolliet, O. (2021). Small targeted dietary changes can yield substantial gains for human health and the environment. *Nature Food, 2*(8), 616–627.

Stylianou, K. S., Heller, M. C., Fulgoni, V. L., Ernstoff, A. S., Keoleian, G. A., & Jolliet, O. (2016). A life cycle assessment framework combining nutritional and environmental health impacts of diet: a case study on milk. *The International Journal of Life Cycle Assessment, 21*(5), 734–746.

Sugimoto, M., Temme, E. H. M., Biesbroek, S., Kanellopoulos, A., Okubo, H., Fujiwara, A., Asakura, K., Masayasu, S., Sasaki, S., & Van't Veer, P. (2022). Exploring culturally acceptable, nutritious, affordable and low climatic impact diet for Japanese diets: proof of concept of applying a new modelling approach using data envelopment analysis. *British Journal of Nutrition,* 1–15.

Tilman, D., Clark, M.  (2014). Global diets link environmental sustainability and human health. *Nature 515*, 518–522.

Udovicic, M., Bazdaric, K., Bilic-Zulle, L., & Petrovecki, M. (2007). What we need to know when calculating the coefficient of correlation? *Biochemia Medica*, 10–15.

U.S. Department of Agriculture (USDA) and Agricultural Research Service. (2019a). *FoodData Central* [Dataset]. https://fdc.nal.usda.gov

U.S. Department of Agriculture (USDA) and Agricultural Research Service. (2019b). *Usual Nutrient Intake from Food and Beverages, by Gender and Age, What We Eat in America, NHANES 2013-2016*. Retrieved from https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/usual/Usual_Intake_gender_WWEIA_2013_2016.pdf

U.S. Department of Agriculture (USDA). (2019c, July 17). *Is corn a grain or a vegetable?* AskUSDA. Retrieved June 9, 2022, from https://ask.usda.gov/s/article/Is-corn-a-grain-or-a-vegetable

U.S. Department of Agriculture (USDA). (2019d, July 17). *What foods are in the Vegetable Group?* AskUSDA. Retrieved June 9, 2022, from https://ask.usda.gov/s/article/What-foods-are-in-the-Vegetable-Group

U.S. Department of Agriculture (USDA). (n.d*.). Beans, Peas, and Lentils*. MyPlate. Retrieved June 9, 2022, from https://www.myplate.gov/eat-healthy/protein-foods/beans-and-peas

U.S. Department of Agriculture (USDA) and the U.S. Department of Health and Human Services (DHHS). (2019). *What We Eat In America (WWEIA) Database*. Ag Data Commons. Retrieved June 10, 2022, from https://data.nal.usda.gov/dataset/what-we-eat-america-wweia-database

U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services (HHS). (2020, December). *Dietary Guidelines for Americans, 2020-2025.* Washington DC. 9th Edition. Available at https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf

Vega Mejía, N., Ponce Reyes, R., Martinez, Y., Carrasco, O., & Cerritos, R. (2018). Implications of the Western Diet for Agricultural Production, Health and Climate Change. *Frontiers in Sustainable Food Systems, 2*(88).

Vos, M. B., Kaar, J. L., Welsh, J. A., Van Horn, L. V., Feig, D. I., Anderson, C. A., Patel, M. J., Cruz Munos, J., Krebs, N. F., Xanthakos, S. A., & Johnson, R. K. (2017). Added Sugars and Cardiovascular Disease Risk in Children: A Scientific Statement From the American Heart Association. *Circulation, 135*(19).

Wang, B., Tsakiridis, E. E., Zhang, S., Llanos, A., Desjardins, E. M., Yabut, J. M., Green, A. E., Day, E. A., Smith, B. K., Lally, J. S. V., Wu, J., Raphenya, A. R., Srinivasan, K. A., McArthur, A. G., Kajimura, S., Patel, J. S., Wade, M. G., Morrison, K. M., Holloway, A. C., & Steinberg, G. R. (2021). The pesticide chlorpyrifos promotes obesity by inhibiting diet-induced thermogenesis in brown adipose tissue. *Nature Communications, 12*(1).

Waswa, L. M., Jordan, I., Krawinkel, M. B., & Keding, G. B. (2021). Seasonal Variations in Dietary Diversity and Nutrient Intakes of Women and Their Children (6–23 Months) in Western Kenya. *Frontiers in Nutrition, 8*.

World Health Organization. (2003, June 12). *New global alliance brings food fortification to world's poor*. Retrieved July 1, 2022, from https://www.who.int/news/item/12-06-2003-new-global-alliance-brings-food-fortification-to-world-s-poor

World Health Organization (WHO) and Food and Agriculture Organization of United Nations (FAO). (2010). *Interim Summary of Conclusions and Dietary Recommendations on Total Fat & Fatty*. https://www.who.int/nutrition/topics/FFA_summary_rec_conclusion.pdf?ua=1

World Health Organization. (2013*). WHO methods and data sources for global burden of disease estimates, 2000–2011. Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2013.4.* Department of Health Statistics and Information Systems WHO. https://www.who.int/healthinfo/statistics/GlobalDALYmethods_2000_2011.pdf

World Health Organization. (2017, December 19). *Nutrients*. Retrieved January 7, 2022, from https://www.who.int/elena/nutrient/en/

Willett, W. C., & Stampfer, M. J. (2013). Current Evidence on Healthy Eating. *Annual Review of Public Health, 34*(1), 77–95.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., World Cancer Research Fund/American Institute for Cancer Research. (2018). *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018.* World Cancer Research Fund. https://www.wcrf.org/wp-content/uploads/2021/02/Summary-of-Third-Expert-Report-2018.pdf

Zaiontz, C. (n.d.). *Jenks Natural Breaks.* Real Statistics. Retrieved May 12, 2022 from https://www.real-statistics.com/multivariate-statistics/cluster-analysis/jenks-natural-breaks/

# Appendices

A. Nutrients excluded from the nutrient profile model

All nutrients from which (1) US-specific recommended intake value was available and (2) the nutritional value was available in FoodData Central were included in the NP model. Based on these criteria, 12 nutrients were excluded and 1 nutrient was substituted (table A1).

Table A1: Nutrients substituted or excluded from NP model.

| Nutrient | Recommended Daily Value | | Available in Food Data central |
|---|---|---|---|
| *Excluded nutrients* | | | |
| Vitamin B7 (Biotin) | 30mcg | (FDA, 2022) | No |
| Chloride | 2300mg | (FDA, 2022) | No |
| Chromium | 35mcg | (FDA, 2022) | No |
| Iodine | 150mcg | (FDA, 2022) | No |
| Molybdenum | 45mcg | (FDA, 2022) | No |
| Vitamin B5 (Pantothenic acid) | 5mg | (FDA, 2022) | No |
| Manganese | 2.3mg | (FDA, 2022) | No |
| Trans fatty acids | <1%E | (FAO, 2010) | No |
| Lycopene | n/a | | Yes |
| Lutein + zeaxanthin | n/a | | Yes |
| Alcohol | n/a | | Yes |
| Theobromine | n/a | | Yes |
| Caffeine | n/a | | Yes |
| *Substituted nutrient* | | | |
| Added sugars | 50g | (FDA, 2022) | No |

B. Food items excluded from correlation analysis

The researched food items by Stylianou et al. (2021) were included in the correlation analysis. Stylianou et al. (2021) used WWEIA Food codes from the FoodData Central database for the 167 dishes. In this research 21 food items were excluded because they were not existing in the Food Data bank (table A1). Moreover, 2 food items were excluded because they were duplicated in the dataset of Stylianou (table A2).

Table A1: Food items excluded due to unavailability in FoodData Central.

| WWEIA code | Food Description |
|---|---|
| 24122120 | Chicken, breast, roasted, broiled, or baked, skin not eaten |
| 24144210 | Chicken, drumstick, fried, no coating, skin eaten, NS as to type of fat added in cooking |
| 25230310 | Chicken or turkey loaf, prepackaged or deli, luncheon meat |
| 27150100 | Shrimp curry |
| 27510225 | Cheeseburger, 1 medium patty, with condiments, on bun, from fast food / restaurant |

| | |
|---|---|
| 27510251 | Cheeseburger, 1 medium patty, with condiments, on white bun |
| 27510330 | Double cheeseburger (2 patties), with tomato and/or catsup, on bun |
| 27540150 | Chicken fillet, breaded, fried, sandwich with lettuce, tomato and spread |
| 27540190 | Chicken patty sandwich, with lettuce and spread |
| 41103020 | Lima beans, dry, cooked, fat not added in cooking |
| 54401080 | Salty snacks, corn or cornmeal base, tortilla chips |
| 56203010 | Oatmeal, cooked, regular, fat not added in cooking |
| 56203030 | Oatmeal, cooked, instant, fat not added in cooking |
| 58132310 | Spaghetti with tomato sauce and meatballs or spaghetti with meat sauce or spaghetti with meat sauce and meatballs |
| 58146110 | Pasta with meat sauce |
| 63135150 | Peach, cooked or canned, drained solids |
| 71201015 | White potato chips, regular cut |
| 71403000 | White potato, home fries |
| 72119224 | Kale, cooked, NS as to form, made with oil |
| 75207001 | Bean sprouts, cooked, from fresh, NS as to fat added in cooking |
| 81104560 | Vegetable oil-butter spread, reduced calorie, tub, salted |

Table A2: Food items excluded due duplication in dataset.

| **WWEIA code** | **Food Description** |
|---|---|
| 75111000 | Cucumber, raw |
| 74101000 | Tomatoes, raw |

C. In- and exclusion criteria of food groups, effective intake and associate health outcomes

Food items were selected from FoodData Central based on the in- and exclusion criteria defined by Stylianou et al. (2021) (Table A3). Moreover, food groups can compose of different items based on the country of reference (Drewnowski and Fulgoni, 2008). Therefore, here US guidelines for some items are clarified. Firstly, corn is considered a starchy vegetable when soft and a grain when fully mature (USDA, 2019c). Avocado is considered a vegetable (USDA, 2019d). Kidney beans, pinto beans, black beans, pink beans, black-eyed peas, garbanzo beans (chickpeas), split peas, pigeon peas, mung beans, and lentils are considered legumes (USDA, n.d.). Red meats include beef, veal, pork, lamb, mutton, horse, or goat meat (Farvid et al., 2018). However, the WWEIA database does not include horse meat (USDA, 2019a).
Moreover, effective intake and health outcomes associated with consumption of food groups are displayed in table A3.

Table A3: In and exclusion criteria, effective intake and associated health outcomes per food group (Stylianou et al., 2021).

| Food group | Criteria | Effective intake | Associated health outcomes |
|---|---|---|---|
| Fruits | Including fresh, frozen, cooked, canned, or dried. Excluding fruit juices and salted or pickled fruits. | <250 g/day | 10* |
| Legumes | Including fresh, frozen, cooked, canned, or dried. | <60 g/day | IHD |
| Milk | All milks including non-fat, low-fat, and full-fat milk. Excluding plant derivatives. | <435 g/day | CRC |
| Nuts and seeds | Including all nuts and seeds | <20.5 g/day | T2DM, IHD |
| Red meat | Beef, pork, lamb and goat. Excluding poultry, fish, eggs. | >22.5 g/day | T2DM, CRC |
| Vegetables | Including fresh, frozen, cooked, canned, or dried. Excluding legumes, salted or pickled vegetables, juices, and starchy vegetables. | <360 g/day | Hemorrhagic stroke, IHD, IS |
| Whole grains | Including whole grains from cereals, bread, rice, pasta, muffins, tortillas and other sources. | <125 g/day | T2DM, Hemorrhagic stroke, IHD, IS |

T2DM=Type 2 Diabetes mellitus; IHD=Ischemic heart disease; IS=Ischemic stroke; Other CVD=Other cardiovascular, CRC=Colorectal cancer and circulatory disease"
*The ten health outcomes associated with fruits are: T2DM; Esophageal cancer; Hemorrhagic stroke; IHD; IS; Larynx cancer; Lip and oral cavity cancer; Nasopharynx cancer; Other pharynx cancer; Tracheal, bronchus, and lung cancer

D. Results downscaling analysis

The results of the downscaling analysis can be found in the excel file below.

E. Results correlation analysis

The results of the correlation analysis can be found in the excel file below.

F. Results cluster analysis

The results of the cluster analysis can be found in the excel file below.