# Iron Absorption and Food Substitution in Diet Modeling

Approaches to Designing Healthy, Sustainable, and Acceptable Diets



Dominique van Wonderen

## Propositions

1. Effective diet plans are best modelled at food-item and individual level, not at food-group and population level. (this thesis)

2. Diet modeling and food recommender systems are inseparable for practical dietary guidance. (this thesis)

3. In PhD research, finding a gap in literature is harder than conducting the study.

4. ChatGPT makes writing abstracts an obsolete skill.

5. The need to sell research undermines its critical rigor.

6. Consuming extra protein to build a six-pack belongs on the list of environmental guilts.

7. Friendship diversity enhances food diversity.

Propositions belonging to the thesis, entitled

Iron Absorption and Food Substitution in Diet Modeling

Dominique van Wonderen
Wageningen, 17 September 2026

# Iron Absorption and Food Substitution in Diet Modeling

Approaches to Designing Healthy, Sustainable, and Acceptable Diets

Dominique van Wonderen

**Thesis committee**

**Promotors**
Prof. Dr S.L.J.M. de Leeuw
Professor of Operations Research and Logistics
Wageningen University & Research

Dr A. Melse-Boonstra
Associate Professor in Global Nutrition
Wageningen University & Research

**Co-promotor**
Dr J.C. van Lemmen-Gerdessen
Assistant Professor in Operations Research and Logistics
Wageningen University & Research

**Other members**
Prof. Dr J.M. Geleijnse, Wageningen University & Research
Dr A. Eilander, Unilever, Wageningen
Dr J. Wagenaar, Tilburg University
Prof. B. Tekinerdogan, Wageningen University & Research

This research was conducted under the auspices of
Wageningen School of Social Sciences (WASS)

# Iron Absorption and Food Substitution in Diet Modeling

Approaches to Designing Healthy, Sustainable, and Acceptable Diets

Dominique van Wonderen

**Thesis**
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 17 September 2026
at 10.30 a.m. in the Omnia Auditorium.

Dominique van Wonderen
Iron Absorption and Food Substitution in Diet Modeling: Approaches to Designing Healthy, Sustainable, and Acceptable Diets
209 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2026)
With references, with summary in English

DOI 10.18174/681853

# Contents

**Chapter 1 General introduction** **1**
1.1 Nutrition security . . . 2
1.2 Diet modeling . . . 2
1.3 Research challenges . . . 3
1.3.1 Iron absorption . . . 3
1.3.2 Food substitution . . . 4
1.4 Aim and objectives . . . 5

**IRON ABSORPTION** **8**

**Chapter 2 Modeling iron absorption** **11**
2.1 Background . . . 13
2.2 Method . . . 14
2.2.1 Meal data . . . 14
2.2.2 Iron absorption . . . 15
2.2.3 Food group and nutrient estimations . . . 16
2.2.4 Diet model . . . 17
2.2.5 Evaluation of results . . . 18
2.3 Results . . . 19
2.4 Discussion . . . 23
2.A Appendix - Meals . . . 26
2.B Appendix - Iron absorption . . . 29
2.C Appendix - Diet model . . . 31
2.D Appendix - Statistical analysis . . . 35
2.E Appendix - Sensitivity analyses . . . 37
2.E.1 Top N menu plans . . . 37
2.E.2 Parameter input iron absorption equations . . . 38
2.E.3 Meals used for diet modeling . . . 41
2.F Appendix - Nutrient content menu plans . . . 43

**Chapter 3 Nonlinear absorption equations** **45**
3.1 Background . . . 47
3.2 Methods . . . 48
3.2.1 Input data . . . 48
3.2.2 Diet modeling . . . 48
3.2.3 Evaluation of results . . . 50

3.3 Results . . . 51
3.4 Discussion . . . 54
3.A Appendix - Nutrient absorption equations . . . 57
3.A.1 Iron . . . 57
3.A.2 Zinc . . . 58
3.B Appendix - Nonlinear programming . . . 59
3.B.1 Theory . . . 59
3.B.2 Evaluation of performance . . . 60
3.C Appendix - Piecewise linear approximation . . . 61
3.C.1 Theory . . . 61
3.C.2 Parameters . . . 67
3.C.3 Evaluation of accuracy . . . 75
3.D Appendix - Diet modeling . . . 76
3.D.1 Notation . . . 76
3.D.2 Base models and parameters . . . 77
3.D.3 Nonlinear programming . . . 83
3.D.4 Piecewise linear aproximation . . . 87
3.D.5 Modeling scenarios . . . 92
3.E Appendix - Additional results . . . 94
3.E.1 Accuracy piecewise linear approximation . . . 94
3.E.2 Performance mixed-integer diet model with 30 min time limit . . . 95
3.E.3 Micronutrient content mixed-integer menu plans . . . 96

**FOOD SUBSTITUTION** **97**

**Chapter 4 Within-food-group substitution** **101**
4.1 Introduction . . . 103
4.2 Materials and Methods . . . 106
4.2.1 Data . . . 107
4.2.2 Diet modeling . . . 108
4.3 Results . . . 110
4.3.1 Within-food-group optimization . . . 110
4.3.2 Lowered nutrient goals . . . 115
4.3.3 Between-food-group vs. Between-and-within-food-group optimization . . . 117
4.4 Discussion . . . 118
4.5 Conclusion . . . 121
4.A Appendix - Calculation of dietary change . . . 122
4.B Appendix - Custom food group classification . . . 123
4.C Appendix - Diet model . . . 124
4.C.1 Notation . . . 124

4.C.2 Base model . . . 125
4.C.3 Modeling experiments . . . 128
4.C.4 Parameter input . . . 130

**Chapter 5 Food substitution by machine learning 133**
5.1 Introduction . . . 135
5.2 Background . . . 136
5.2.1 Optimizing consumer acceptance in diet modelling . . . 136
5.2.2 Recipe completion algorithms . . . 137
5.3 Methodology . . . 138
5.3.1 Recipe completion model . . . 138
5.3.2 Diet modelling . . . 140
5.3.3 Data . . . 146
5.4 Results and Discussion . . . 150
5.4.1 Performance recipe completion model . . . 150
5.4.2 Comparison of the RC and FGF diet models . . . 153
5.5 Conclusion . . . 160
5.A Appendix - Qualitative evaluation . . . 161

**Chapter 6 General discussion 165**
6.1 Positioning within existing knowledge . . . 166
6.2 Societal relevance . . . 167
6.3 Remaining challenges and future research . . . 169
6.3.1 Iron absorption . . . 169
6.3.2 Food substitution . . . 171
6.4 Conclusion . . . 177

**References 179**

**Summary 197**

**Acknowledgements 199**

**Authorship statement 203**

**Completed training and supervision plan 207**

# Chapter 1

General introduction



## 1.1 Nutrition security

Over the past two decades, the number of undernourished individuals worldwide has decreased by nearly 50%, driven by rapid economic development and advancements in agricultural productivity (UNDP, 2025). However, since 2015, this trend has stagnated, with a sharp increase in undernourishment rates between 2019 and 2021. In 2023, an estimated 29% of the global population experienced moderate or severe food insecurity, with 11% facing severe food insecurity (FAO, IFAD, UNICEF, WFP and WHO, 2024). Meanwhile, micronutrient malnutrition, also referred to as hidden hunger, is estimated to affect more than 30% of the global population (FAO & WHO, 2011), with children and women of reproductive age being the most affected (Stevens et al., 2022), and surpassing chronic hunger in prevalence (Lenaerts & Demont, 2021). Furthermore, the rates of overweight and obesity are steadily climbing, with 38% of the population classified as overweight and 14% as obese in 2020 (World Obesity Federation, 2023). This triple burden of malnutrition – comprising undernutrition, micronutrient deficiencies, and overnutrition – represents a significant challenge to public health (Gómez et al., 2013). At the same time, environmental health is undermined by unsustainable food production and consumption, particularly the overconsumption of animal-based foods (Aiking & de Boer, 2020; Rust et al., 2020). Consequently, achieving the United Nations Sustainable Development Goal for 2030 to "end hunger, achieve food security, improve nutrition, and promote sustainable agriculture" (UN, 2016) will require substantial and concerted efforts.

## 1.2 Diet modeling

While improving food accessibility is a critical step toward enhancing nutrition security (Ericksen, 2008), it is equally important to understand what constitutes healthy and sustainable diets, as well as how such diets can be adopted in everyday life (Spronk et al., 2014). Diet modeling serves as a valuable tool to support the design of healthy, sustainable, and acceptable diets by optimizing the selection and quantities of food items, food groups, or meals (van Dooren, 2018; Gazan et al., 2018b). In this context, health refers to meeting dietary requirements for energy, macronutrients, and micronutrients; sustainability involves reducing environmental impact in line with planetary boundaries; and acceptability relates to alignment with consumers' tastes, habits, and cultural norms. As an interdisciplinary field, diet modeling applies operations research techniques to address diet-related challenges. Over time, it has been applied in various contexts, including the development of food-based dietary guidelines, the creation of menu plans for hospitals and schools, the formulation of food baskets delivered by aid organizations, and the provision of personal nutrition advice (van Dooren, 2018; Gazan et al., 2018b). The rise in computational power has enabled researchers to process larger datasets and apply more complicated modeling techniques for solving diet problems. Nevertheless, challenges re-

main in improving the quality of model outputs (van Dooren, 2018; Gazan et al., 2018b). In this thesis, we focus on modeling iron absorption and food substitution. The challenges involved, along with their relevance, are explained in the next section.

## 1.3 Research challenges

### 1.3.1 Iron absorption

Iron deficiency is a major contributor to the global disease burden, particularly affecting children, premenopausal women, and individuals in low- and middle-income countries (Pasricha et al., 2021). Those following a vegetarian or vegan diet also risk deficiencies (Haider et al., 2018), as nonheme iron from plant sources is less readily absorbed than heme iron from animal products. Heme iron absorption ranges from 15% to 40%, whereas nonheme iron absorption is considerably lower, between 1% and 15% (Hunt, 2002). The absorption of nonheme iron is influenced by several factors, including one's current iron status and the interactions between nutrients and other food components during consumption. Researchers have quantified these relationships by developing nonheme iron absorption equations (Armah et al., 2013; Conway et al., 2007; Hallberg & Hulthén, 2000; De Carli et al., 2018; Collings et al., 2013; Reddy et al., 2000; Fairweather-Tait et al., 2017). These diet-dependent absorption equations can be categorized into two types: single-meal equations, which estimate iron absorption for each meal separately, and complete diet equations, which estimate absorption for the entire diet (De Carli et al., 2018). However, many diet modeling studies do not use these equations and instead apply a constant absorption factor, such as 18% for omnivorous diets (IOM, 2006) and 10% for vegetarian diets (Eustachio Colombo et al., 2021). The impact of this simplification on the evaluation of iron adequacy, and consequently on the healthiness of modeled diets, has not yet been investigated. Accurate estimation of iron absorption is essential, as iron is often a limiting nutrient, particularly in vegetarian and vegan diets, which play an important role in the transition toward more sustainable, plant-based food consumption.

However, modeling iron absorption with equations is technically more complex than applying a constant factor. As the absorption equations are log-transformed (Armah et al., 2013; Conway et al., 2007; Hallberg & Hulthén, 2000; Reddy et al., 2000), incorporating them makes the decision problem nonlinear. To address such nonlinearities, methods like nonlinear programming and piecewise linear approximation are commonly used. Both methods have limitations, as they may yield inaccurate results under certain conditions (Kronqvist et al., 2019; Geißler et al., 2011). In the case of nonlinear programming, a solver may struggle to find the optimal solution or may require extensive computational time, particularly when dealing with integer problems as opposed to continuous ones[1]

[1]Integer diet problems include integer variables, such as the decision to include or exclude specific meals, whereas continuous diet problems involve only continuous variables, such as food quantities.

(Kronqvist et al., 2019). Piecewise linear approximation, by contrast, may return inaccurate results due to poor approximation of the nonlinear function. Despite this limitation, studies have shown that piecewise linear approximation can achieve faster and more reliable results than nonlinear programming (Geißler et al., 2011). In the context of modeling nonlinear absorption equations, piecewise linear approximation has not yet been applied, nor has the relative performance of these methods been investigated.

### 1.3.2 Food substitution

Food substitution involves replacing specific foods with acceptable alternatives to create, for instance, a healthier and more environmentally sustainable diet. One of the main difficulties in modeling food substitution is ensuring the acceptability of these changes, as the success of any dietary change depends on how willing people are to adopt and sustain these new habits (Bos et al., 2013). In diet modeling, common approaches to maintaining acceptability include minimizing the differences between observed and optimized diets, imposing constraints on substitutions between- and within food groups, and regulating food item quantities (van Dooren, 2018).

Substitution within food groups, also referred to as food group filtering, is not that often applied. Many studies have modeled diets at food group level, maintaining a fixed distribution of food items within each food group (Gazan et al., 2018a; Tompa et al., 2022; Horgan et al., 2016; Vieux et al., 2018; Perignon et al., 2016; Dussiot et al., 2022a; Mariotti et al., 2021; Green et al., 2015; Reynolds et al., 2019; Rocabois et al., 2022). However, given the variability in nutrient and sustainability profiles among items within food groups (Rippin et al., 2021), we expected that optimizing foods both between and within food groups, could enhance the nutritional adequacy and sustainability of diets. Moreover, between-and-within-food-group substitution could offer greater acceptability, as this may require less dietary change compared to between-food-group optimization alone. These potential benefits of within-food-group substitution on the nutritional adequacy, sustainability and acceptability of diets have not been investigated before.

In addition to food-group-based substitution, machine learning has emerged as an alternative method for substituting food items. Given the difficulty in quantifying and translating individual preferences into specific constraints, recent studies have focused on learning these constraints directly from data, aiming to provide more personalized food suggestions and, consequently, greater acceptability of diet plans (Ahmadi et al., 2020; Ghobadi & Mahmoudzadeh, 2021; Fajemisin et al., 2024).

The application of machine learning to identify acceptable food alternatives draws on techniques from various research areas, including food recommender systems, food substitution algorithms, and recipe completion algorithms. Food recommender systems generate personalized food suggestions based on individual preferences and historical consumption patterns, including those of users with similar profiles (Tran et al., 2021). Food substi-

tution algorithms aim to replace food items with healthier alternatives that are similar in characteristics. The degree of similarity can be determined based on factors such as flavor, texture, nutrient profile, food classification, and recipe context (Shirai et al., 2021). Meanwhile, recipe completion algorithms seek to identify complementary food items that are frequently consumed together in order to complete a recipe (Cueto et al., 2019). The complexity of these algorithms can range from straightforward linear regression models (De Clercq et al., 2016) to more sophisticated large language models (Devlin, 2018) and knowledge graph embeddings (Shirai et al., 2021).

Currently, data-driven methods for modeling consumer acceptability are rarely applied in diet modeling and have only been tested in scenarios with a limited number of food groups or food items (Ahmadi et al., 2020; Ghobadi & Mahmoudzadeh, 2021; Fajemisin et al., 2024). Their effectiveness in improving the acceptability of modeled diets in real-world applications has not yet been investigated, particularly in the complex context of menu planning where food item compatibility must be considered.



## 1.4 Aim and objectives

By evaluating different approaches for modeling iron absorption and food substitution, this thesis aimed to improve the nutritional adequacy, sustainability, and acceptability of modeled diets. The first objective was to determine how best to model iron absorption. To this end, we examined differences in model outcomes when using constant factors versus diet-dependent absorption equations for omnivorous, vegetarian, and vegan menu plans (Chapter 2). Next, we evaluated whether nonlinear absorption equations are most effectively implemented using nonlinear programming or piecewise linear approximation (Chapter 3). We conducted this evaluation for both mixed-integer and continuous diet problems. In addition to iron, we demonstrated this for zinc, a mineral for which deficiency also poses a global challenge and whose absorption is likewise affected by meal composition (Singh & Prasad, 2023).

The second objective was to determine how best to model food substitution. Specifically, we assessed the extent to which within-food-group substitution can improve the nutritional adequacy, environmental sustainability (as indicated by greenhouse gas emissions), and acceptability of modeled diets compared to considering only between-food-group substitution (Chapter 4). Furthermore, we investigated whether data-driven rather than food-group-based substitution can improve the acceptability and nutritional adequacy of modeled diets. Chapter 5 introduces a new method for data-driven substitution that combines diet modeling with recipe completion. We evaluated its performance in a menu-planning context, comparing the outcomes with those produced by traditional food-group-based substitution methods.

An overview of all research objectives, methods, and data used, along with the corresponding publications, is given in Table 1.1. Finally, Chapter 6 concludes this thesis by summarizing the key findings and contributions, and discussing limitations, directions for future research, and research implications.

**Table 1.1:** Overview of research objectives, methods, data, and publications by chapter.

| Chapter | Objective | Method[1] | Data[2] | Publication |
|---|---|---|---|---|
| | **Iron absorption** | | | |
| 2 | Compare the use of constant absorption factors with diet-dependent absorption equations and assess the differences in absorbable iron estimates for omnivorous, vegetarian, and vegan menu plans. | MILP, MOP, GP | NHANES 17-18, FNDDS 17-18, FPED 17-18, WWEIA 17-18, Phytate[3], DRV | van Wonderen, D., Melse-Boonstra, A., & Gerdessen, J. C. (2023). Iron bioavailability should be considered when modeling omnivorous, vegetarian, and vegan diets. *The Journal of Nutrition*, 153(7), 2125-2132. |
| 3 | Demonstrate how absorbable iron and zinc can be optimized using piecewise linear approximation and compare its performance with that of nonlinear programming for both a mixed-integer and continuous diet problem. | MILP, PLA, MINLP, MOP, GP | NHANES 17-18, FNDDS 17-18, FPED 17-18, WWEIA 17-18, Phytate[3], DRV | van Wonderen, D., Gerdessen, J. C., Kirst, P., & Melse-Boonstra, A. (2025). Diet optimization: modeling iron and zinc absorption by nonlinear programming and piecewise linear approximation using NHANES. *The American Journal of Clinical Nutrition*, 122(3), 853-858. |
| | **Food substitution** | | | |
| 4 | Examine how within-food-group optimization can improve the nutritional adequacy, environmental sustainability, and acceptability of modeled diets compared to between-food-group optimization alone. | LP, MOP, GP | NHANES 17-18, FNDDS 17-18, WWEIA 17-18, DRV, LAFA, dataFIELD, FCID | van Wonderen, D., Gerdessen, J. C., Biesbroek, S. & Melse-Boonstra, A. (2025). Within-food-group optimization improves nutritional adequacy, sustainability, and acceptability of modeled diets. *Frontiers in Nutrition*, 12, 1648055. |
| 5 | Implement a data-driven method for substituting food items using a recipe completion algorithm, with performance evaluated in a menu planning context and compared to traditional food group-based substitution methods. | MILP, MOP, GP | NHANES 13-18, FNDDS 13-18, WWEIA 13-18, DRV | van Wonderen, D., Gerdessen, J. C., Melse-Boonstra, A., & Onwezen, M. C. (2026). Using diet optimization and machine learning for the design of healthy and acceptable menu plans. *European Journal of Operational Research*, 328(2), 668-679. |

[1] Abbreviations: GP, Goal Programming; LP, Linear Programming; MILP, Mixed Integer Linear Programming; MINLP, Mixed Integer Non Linear Programming; MOP, Multi Objective Programming; PLA, Piecewise Linear Approximation.

[2] Abbreviations: dataFIELD, US database of Food Impacts on the Environment for Linking to Diets (Heller et al., 2018); DRV, US Dietary Reference Values (USDA & US HHS, 2020); FCID, US Food Commodity Intake Database (US EPA, 2001-2010); FNDDS, US Food and Nutrient Database for Dietary Studies (USDA & ARS, 2020a); FPED, US Food Patterns Equivalents Database (USDA & ARS, 2020b); LAFA, Loss-Adjusted Food Availability (USDA ARS, 2022); NHANES, US National Health and Nutrition Examination Survey (CDC & NCHS, 2018); WWEIA, What we Eat In America (USDA & ARS, 2020c).

[3] Larvie & Armah (2021)

# Part 1

Iron Absorption

# Chapter 2

## Iron bioavailability should be considered when modeling omnivorous, vegetarian, and vegan diets



This chapter is published as:

van Wonderen, D., Melse-Boonstra, A., & Gerdessen, J. C. (2023). Iron bioavailability should be considered when modeling omnivorous, vegetarian, and vegan diets. *The Journal of Nutrition*, 153(7), 2125-2132.

# Abstract

**Background:** To lower environmental impact of human food consumption, replacement of animal proteins with plant-based proteins is encouraged. However, the lower iron bioavailability of plant-based foods is rarely considered when designing healthy and sustainable diets by using diet modeling. The estimated absorbable iron content of vegetarian and vegan menu plans might therefore be too optimistic.

**Objective:** Our aim was to investigate and compare the impact of various methods to estimate absorbable iron intake on the nutritional adequacy of omnivorous, vegetarian and vegan menu plans designed for women of reproductive age.

**Methods:** A diet model was developed to design menu plans consisting of a selection of meals that best complied with nutritional requirements. Meals used for modeling were created based on food intake data from the National Health and Nutrition Examination Survey. For each meal, absorbable iron concentrations were estimated by using two constant absorption factors (18% and 10%) and two diet-dependent absorption equations (Conway and Hallberg). For each absorption method and diet type, we used the diet model to design the optimal menu plan. Retrospectively, menu plans were evaluated by estimating the absorbable iron content by using the other absorption methods.

**Results:** Retrospective diet-dependent absorbable iron estimates were consistently lower than estimates based on constant absorption factors. Using diet-dependent estimates increased absorbable iron by optimizing enhancer and inhibitor concentrations.

**Conclusion:** Iron bioavailability should be considered when modeling diets.

## 2.1 Background

To reduce the environmental impact of human food consumption, multiple stakeholders have advocated for transitioning from animal-based to plant-based diets (Aiking & de Boer, 2020; Graça et al., 2019; Hemler & Hu, 2019; Sabate & Soret, 2014; Vinnari & Vinnari, 2014; Willett et al., 2019). However, it is still a subject of debate what trade-offs of partially or fully plant-based diets are to be expected with regard to human health. Some research showed that vegetarian diets do contribute to public health as such diets are associated with a lower risk of chronic diseases (Graça et al., 2019; Hemler & Hu, 2019; Vinnari & Vinnari, 2014). In contrast, other researchers expressed concern that vegetarian diets are more susceptible to nutritional inadequacies due to lower nutrient bioavailability (Andrewski et al., 2022; Broekema et al., 2020; Elorinne et al., 2016).



Iron is one of the nutrients of greatest concern in vegetarian diets (Hunt, 2003; Martınez-Navarrete et al., 2002; Zimmermann & Hurrell, 2007), especially for women of reproductive age (Lopez et al., 2016). Non-heme iron found in plant-based foods has a relatively low absorption (1-15%) in humans compared to heme iron found in animal-sourced foods (15-40%) (Hunt, 2002). This is due to the higher content of phytate and polyphenols in vegetarian diets, which inhibit the absorption of non-heme iron. Additionally, meat protein, which enhances the absorption of non-heme iron, is absent in vegetarian diets. To some extent, these effects are mitigated by the higher amounts of vitamin C in vegetarian diets, which enhances the absorption of non-heme iron. Nonetheless, the bioavailability of iron in vegetarian diets is substantially lower than in omnivorous diets (Collings et al., 2013; Conway et al., 2007; De Carli et al., 2018; Hallberg & Hulthén, 2000; Hunt, 2003; Hurrell & Egli, 2010; Hurrell et al., 2006; Reddy et al., 2000).

To optimize the nutritional adequacy and sustainability of diets, diet modeling is a quantitative technique that is commonly applied (van Dooren, 2018). For the design of vegetarian and vegan diets, however, the lower bioavailability of iron is rarely taken into account as most diet models assume the same constant absorption factor as for omnivorous diets. Only a few examples exist in literature where the applied iron absorption factor was adjusted based on the type of diet. For instance, Eustachio Colombo et al. (2021) used a lower iron absorption factor for the design of vegetarian and vegan diets. Barré et al. (2018) as well as Dussiot et al. (2022b) developed a diet model to reduce meat consumption using diet-dependent absorption equations for iron, zinc, vitamin A and protein quality.

In contrast to constant iron absorption factors, diet-dependent absorption equations estimate iron absorption based on enhancer and inhibitor concentrations within a specific diet. These equations are therefore considered to be more accurate. A question still left unanswered though, is to what extent menu plans based on either constant iron absorption factors or diet-dependent iron absorption equations differ from each other. Accordingly,

the objective of the present study was to compare the estimated absorbable iron content of omnivorous, vegetarian and vegan diets based on different methods to estimate absorbable iron intake for women of reproductive age.

## 2.2 Method

A diet model was developed to design various two-week menu plans: omnivorous, lacto-ovo-vegetarian (in this paper referred to as vegetarian), and vegan. Each menu plan consisted of a selection of meals that complied best with nutritional requirements. Meals were created based on dietary intake data of foods frequently consumed in the United States of America. For each meal, the bioavailability of iron was estimated by applying both constant absorption factors and diet-dependent absorption equations as explained below in more detail.

### 2.2.1 Meal data

Meals were created from dietary intake data provided by the National Health and Nutrition Examination Survey (NHANES) 2017-2018 (CDC & NCHS, 2018). NHANES reports on demographics, food items and quantities consumed, meal type, and corresponding nutrient values that are based on the Food and Nutrient Database for Dietary Studies (FNDDS) 2017-2018 (USDA & ARS, 2020b). To create a database of meals, we aggregated consumption data by respondent, interview day and time of eating occasion. An example of such a meal is given in Table 2A.1 (Appendix 2A).

To obtain an example dataset of meals, the following filters were applied (Figure 2A.1): 1) Consumption for respondents below the age of 18 years was filtered out (5,856 respondents remained); 2) Only breakfast, lunch, dinner, and snack meals were retained; 3) Meals that contained food items that were consumed by less than 10 respondents were removed. Food items were counted by their first word before a comma or parenthesis of the USDA food name description. For example, “Rice, white, with corn, no added fat” and “Rice, white, with corn, fat added” were both counted as “Rice”; 4) Meals that contained only drinks were excluded. In addition, meals that contained 400 g of drinks or more were filtered out; 5) Breakfast and lunch meals were retained when they contained food items belonging to at least two of the following food groups: Fruit, Vegetables, Legumes, Grains, Dairy, Meat and seafood, Eggs, and Nuts; 6) Dinner meals were kept either when they contained food items belonging to at least two of the aforementioned food groups and consisted of four food items or more, or when they contained food items belonging to at least two of the following food groups: Vegetables, Legumes, Grains, Meat and seafood; 7) Meals containing food items with artificially increased nutrient density (commercial breakfast cereals and oatmeal-based porridges, nutritional powder mix and drinks, nutrition bars, and energy drinks) were removed; 8) Meals with very large portion sizes were removed. Portion sizes were regarded large when food group (Dairy, Fruits, Grains, Protein foods,

Vegetables) amounts of one meal exceeded the daily recommended amount (USDA & US HHS, 2020); 9) Meals were discarded when their caloric intake was lower than 400 kcal for lunch and dinner meals or larger than 1500 kcal; 10) Duplicate meals were aggregated by taking the median value of food item quantities for meals that consisted of similar food items; 11) Meals with nonheme iron absorption estimates above 50% (based on either of the two diet-dependent absorption equations, see next section) were excluded.

After the applied filtering steps, 8,887 omnivorous meals remained of which 4,850 meals were classified as vegetarian and 1,541 meals as vegan (Table 2A.2). Meals were considered vegetarian when they did not contain meat, fish, or poultry. Meals that did not contain meat, fish, poultry, eggs, or dairy were labeled as vegan. This classification was established by using the food groups of the Food Patterns Equivalents Database (FPED) (Bowman et al., 2022).



### 2.2.2 Iron absorption

Absorbable iron was estimated for each meal by applying different iron absorption estimation methods. Two constant absorption factors were utilized: 18% as used by the Institute of Medicine (IOM, 2006) (**IOM**) to set the nutritional requirement on iron intake for premenopausal women, and 10% as applied by Eustachio Colombo et al. (2021) (**Colombo**) to account for the lower bioavailability of vegetarian and vegan diets (IOM, 2001).

Furthermore, two diet-dependent absorption equations were deployed. The equation of Hallberg & Hulthén (2000) (**Hallberg**) was selected as it was evaluated as one of the most accurate iron absorption models by multiple reviewing articles (Armah et al., 2013; Conway et al., 2007; De Carli et al., 2018). Secondly, Conway et al.'s 2007 (**Conway**) equation was selected because it has the advantage that it only requires information on food groups. Estimation of nutrients that are not present in food composition databases can be avoided this way. The detailed formulation of both equations is presented in Appendix 2B.

Hallberg's equation estimates the absorption of heme and nonheme iron for a meal based on the amount of phytate; vitamin C; meat, fish, and poultry (MFP); tannic acid, calcium, soy, eggs; and alcohol. For Conway's equation, the following food groups were used: high vitamin C fruit or fruit juice; MFP; soy; beans and lentils excluding soy; wholegrain cereal; black tea infusion; dairy excluding cream and cottage cheese; cheese; eggs; and nuts. Conway's equation only estimates nonheme iron absorption though. Therefore, heme iron absorption was assumed to be 25% (Armah et al., 2015a). In addition, iron absorption was adjusted for the serum ferritin concentration within an individual. In this study, serum ferritin was set at 15 $\mu$g/L, which is the cut-off value used by IOM (2001) to calculate Dietary Reference Intakes for iron deficient individuals.

After estimating heme and nonheme iron absorption, total absorbable iron (2.1) and total iron absorption (2.2) were calculated as follows:

$$\begin{aligned}&\text{Total absorbable iron (mg)} = \\ &\quad \text{Nonheme iron absorption (\%)} \times \text{Nonheme iron content (mg)} + \\ &\quad \text{Heme iron absorption (\%)} \times \text{Heme iron content (mg)}\end{aligned} \tag{2.1}$$

$$\text{Total iron absorption (\%)} = \frac{\text{Total absorbable iron (mg)}}{\text{Total iron (mg)}} \tag{2.2}$$

### 2.2.3 Food group and nutrient estimations

To estimate absorbable iron for all meals, we needed to estimate the food groups and nutrients used by the equations. The food groups used in Conway's equation were estimated based on the food categories available from the What we Eat In America (WWEIA) database (USDA & ARS, 2020c) and Food Patterns Equivalents Database (FPED) (Bowman et al., 2022). For example, MFP was calculated by taking the amount of WWEIA category "MFP". For WWEIA category "mixed dishes", MFP was estimated by multiplying FPED's equivalent "Total meat, poultry, seafood, organ meats, and cured meat" by a conversion factor 28.35 g/oz eq. (Armah et al., 2015a).

For Hallberg's equation, nutrients that were not available in FNDDS and therefore had to be estimated included (non)heme iron, tannic acid, and phytate. Heme iron was estimated to be half of the total iron content of MFP (Monsen, 1988). Tannic acid was assumed to be zero as no appropriate databases were available to make reliable estimations. Phytate was estimated by applying the method reported by Larvie & Armah (2021). The phytate content of a specific meal was estimated by multiplying the average phytate content of corresponding food groups by their FPED's unit equivalents. Subsequently, unit equivalents were converted to grams. For example, for a 155g pizza that contains 2.91 oz eq. refined grains, the phytate content was estimated to be 2.91 oz eq. $\cdot$ 28.35 g/oz eq. [1] $\cdot$ 1.97 mg/g [2] = 162.5 mg.

Finally, added sugar was estimated as it is constrained in the diet model (Table 2.1) and is not included in FNDDS. Added sugar was estimated by multiplying FPED equivalent "added sugar" by conversion factor 4.2 g/tsp. eq. (Cioffi et al., 2018).

[1] weight / unit equivalent refined grains (Larvie & Armah, 2021)
[2] phytate content refined grains (Larvie & Armah, 2021)

**Table 2.1:** Dietary reference intakes (women 19–50y) for nutrients considered by the diet model (IOM, 2006).

| Macronutrient | Unit | LB | UB |
|---|---|---|---|
| Energy | kcal/day | 2000 | 2400 |
| Protein | E% | 10 | 35 |
| Total carbohydrates | E% | 45 | 65 |
| Dietary fibre | g | 25 | — |
| Added sugars | E% | — | 10 |
| Total fat | E% | 20 | 35 |
| Total saturated fatty acids | E% | — | 10 |
| Sodium | mg | — | 2300 |

| Micronutrient | Unit | EAR | RDA | UL |
|---|---|---|---|---|
| Calcium[3] | mg | 800 | 1000 | 2500 |
| Iron (Absorbable iron)[4] | mg | 8.1 (1.5) | 18 (3.2) | 45 (8.1) |
| Zinc | mg | 6.8 | 8 | 40 |
| Vitamin A | $\mu$g | 500 | 700 | 3000 |
| Vitamin B1 | mg | 0.9 | 1.1 | — |
| Vitamin B2 | mg | 0.9 | 1.1 | — |
| Vitamin B3 | mg | 11 | 14 | — |
| Vitamin B6 | mg | 1.1 | 1.3 | 100 |
| Vitamin B9 | $\mu$g | 320 | 400 | 1000 |
| Vitamin C | mg | 60 | 75 | 2000 |

[1] Abbreviations: LB, Lower Bound; UB, Upper Bound; EAR, Estimated Average Requirement; RDA, Recommended Daily Allowance; UL, tolerable Upper Level.
[2] (USDA & US HHS, 2020)
[3] (IOM, 2011)
[4] Absorbable iron was calculated by multiplying the dietary reference intakes with IOM's (2006) absorption factor of 18%.

### 2.2.4 Diet model

A mixed-integer linear goal programming model was developed to construct a two-week menu plan consisting of a selection of meals that best complied with nutritional requirements. For the selection of meals, a binary decision variable was defined that states whether or not a meal is used for a meal moment on a certain day of the two-week menu plan. One breakfast, lunch, and dinner meal and two snacks should be selected each day. To secure menu diversity, dinner and lunch meals could only be selected once every two weeks. Breakfast meals and snacks could be selected twice every two weeks. Fur-

thermore, food item use for dinner and lunch meals was restricted. For example, each vegetable could only be selected twice per week.

To guarantee the nutritional adequacy of the menu plan, we placed lower and upper bounds on macronutrient intakes based on the dietary reference intakes (DRI) provided by IOM (2006) (Table 2.1). The DRIs were adopted for women (19-50y) as iron deficiency is most commonly observed in women of reproductive age, and dietary iron bioavailability is, therefore, most critical for this age group (Lopez et al., 2016). Micronutrient intakes were maximized for those nutrients for which recommended dietary allowances (RDA) were available, with the exception of vitamin B12. Vitamin B12 is mainly derived from meat and fish (Sanders, 1999), and would therefore constitute a bottleneck nutrient in the design of vegetarian and vegan diets. To allow the model to optimize absorbable iron (serum ferritin 15 $\mu$g/L), it was therefore decided to leave vitamin B12 out of the objective function. The objective function of the model was to minimize the largest deviation below the recommended dietary allowance (RDA). For a fair comparison between nutrients, deviations were normalized by scaling them between 0 and 1. Finally, tolerable upper intake levels were applied as upper bounds to prevent unhealthily high intake of micronutrients. Full model definitions are given in Appendix 2C (Tables 2C.1 and 2C.2).

The diet model was applied to create three types of two-week menu plans: omnivorous, vegetarian, and vegan. Additionally, four different methods to estimate iron absorption (IOM, Colombo, Conway, Hallberg) were used by the diet model. This finally resulted in $3 \times 4 = 12$ menu plans in total.

### 2.2.5 Evaluation of results

The nutritional adequacy of the twelve menu plans was evaluated based on the EARs and RDAs for women (19-50y) presented in Table 2.1. Furthermore, the absorbable iron content of the menu plans was compared with the content estimated by each of the other iron absorption methods. Thus for the omnivorous menu plan optimized based on IOM's estimate, referred to as IOM's omnivorous menu plan, the absorbable iron content was also calculated by applying Colombo's, Hallberg's and Conway's absorption methods. These absorbable iron estimates that were calculated after the menu plans had been optimized will be referred to as retrospective estimates (e.g., Hallberg's retrospective estimate for IOM's menu plan). The statistical test used to compare the mean absorbable iron content of the meals selected for each menu plan is described in Appendix 2D.

Additionally, the equation of Armah et al. (2013) (**Armah**) was selected for retrospective evaluation to be able to compare estimates based on single-meal equations and complete-diet equations. Single-meal equations (Conway and Hallberg) are based on consumption data of foods that are digested together during a single-meal moment. As such, the iron absorption factor of each meal can be calculated. In contrast, complete-diet equations (Armah) are fitted on the average daily dietary intake from one day or more and, therefore

cannot be used to estimate the iron absorption of single-meals. The detailed formulation of Armah's equation is presented in Appendix 2B.

Furthermore, it has been examined to which extent the optimized menu plans are unique (Appendix 2E.1), and a sensitivity analysis was conducted to investigate the impact of a different serum ferritin concentration, and enhancer and inhibitor concentrations in the diet on the estimation of absorbable iron (Appendix 2E.2). Lastly, we explored to what extent the meals used as data input for the diet model affect the nutritional adequacy of optimized menu plans (Appendix 2E.3).

## 2.3 Results



To investigate the impact of the applied iron absorption method on the nutritional adequacy of menu plans, we used diet modelling to create menu plans based on four different absorption methods (IOM, Colombo, Conway, and Hallberg) for three types of diets (omnivorous, vegetarian, and vegan). The mean absorbable iron content of the meals used as input data for the diet model is displayed in Table 2A.3. An overview of the average nutrient content of these menu plans is given in Figure 2F.1 (Appendix 2F). The menu plans complied with all RDAs except for the vegan menu plans of Colombo and Hallberg. Colombo's vegan menu plan could not satisfy the RDA for iron, calcium and vitamin A. Hallberg's vegan menu plan did not satisfy the RDA for iron, calcium and zinc. In the current section, only the absorbable iron content of the menu plans will be further analysed.

When evaluating the menu plans based on IOM's constant absorption factor, the most notable observation is that IOM's absorbable iron estimates were significantly higher than the retrospective estimates based on the diet-dependent absorption equations (Conway, Hallberg and Armah) (Figure 2.1A, Table 2D.1). All IOM's menu plans complied with the RDA for iron when the absorbable iron content was estimated by applying IOM's constant absorption factor. There were even 13 other omnivorous, 9 other vegetarian and 1 other vegan menu plans that satisfied the RDA for iron (Table 2E.1, Figure 2E.1). However, when the absorbable iron content of IOM's menu plans was estimated retrospectively by applying the diet-dependent absorption equations, menu plans did not even satisfy the EAR except for Armah's retrospective estimate for IOM's omnivorous menu plan.

A similar pattern was observed for Colombo's menu plans (Figure 2.1B). Colombo's absorbable iron estimates were consistently higher than the retrospective estimates based on the diet-dependent absorption equations. All differences were statistically significant, except for Hallberg's and Armah's retrospective estimates for Colombo's omnivorous menu plan (Table 2D.1). Nonetheless, modelling with a lower constant absorption did increase the retrospective diet-dependent estimates with the result that the EAR could now be satisfied but not the RDA.

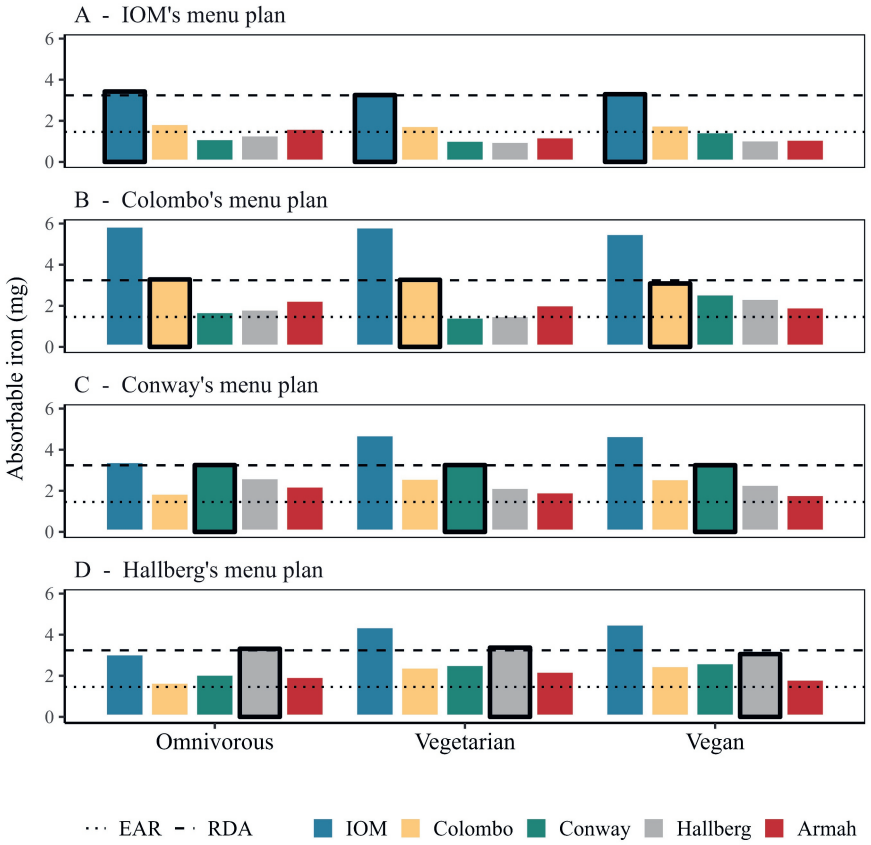



**Figure 2.1:** Average absorbable iron content (mg/day) compared to the EAR and RDA (women 19-50y). The absorption of iron was estimated using a serum ferritin concentration of 15 $\mu$g/L. Figures A-D represent the different absorption methods that have been applied by the diet model to optimize absorbable iron quantities (e.g., Figure A shows the menu plans that are based on IOM's absorption factor). The colored bars display which absorption method was used to estimate absorbable iron after the menu plans had been generated (e.g., Hallberg's retrospective estimate for IOM's menu plan). The outlined bars indicate the situation where the absorption method used by the diet model is the same as the method used to estimate absorbable iron retrospectively.

The absorbable iron estimates by diet-dependent equations were higher when Conway's and Hallberg's equations were used by the diet model. All menu plans complied with the RDA for iron except for Hallberg's vegan menu plan, which only satisfied the EAR (Figure 2.1C-D). The absorbable iron content of Conway's vegan menu plan was factor 1.2 higher than Conway's retrospective estimate for Colombo's vegan menu plan. Similarly, the absorbable iron content of Hallberg's vegan menu plan was a factor 1.3 higher than Hallberg's estimate for Colombo's vegan menu plan. When making the same comparison for IOM's vegan menu plan, Conway's and Hallberg's estimates were factor 2.2 and 2.8 higher, respectively. Higher diet-dependent absorption estimates were also observed for the omnivorous and vegetarian menu plans.

When comparing retrospective estimates based on Armah's complete-diet equation and Conway's and Hallberg's single-meal equations, it can be observed that Armah's estimates were slightly higher for IOM's and Colombo's omnivorous and vegetarian menu plans (Figure 2.1A-B). In contrast, Armah's estimates were slightly lower for the vegan menu plans. As regards the menu plans based on the diet-dependent equations (Figure 2.1C-D), Armah's estimates were also slightly lower than the other retrospective estimates (e.g., Armah's retrospective estimates for Conway's omnivorous menu plan is slightly lower than Hallberg's retrospective estimate). However, none of these differences were significant, except for Armah's and Conway's retrospective estimates for Hallberg's vegan menu plan (Table 2D.1).

Differences in absorbable iron estimates between the various menu plans become visible through differences in enhancer and inhibitor quantities (Figure 2.2). For example, Conway's absorbable iron estimate was higher for Conway's vegan menu plan (3.2 mg) than for IOM's vegan menu plan (1.5 mg) (Figure 2.1) because the enhancer concentration was higher (Figure 2.2A), and inhibitor concentrations were lower with the exception of soy (Figure 2.2C). Similarly, Hallberg's estimate for Hallberg's vegan menu plan (3.1 mg) was higher than for IOM's vegan menu plan (1.1 mg) (Figure 2.1) as the concentrations of Hallberg's enhancers were higher (Figure 2.2B) and the concentrations of inhibitors phytate and calcium were lower (Figure 2.2D).

The sensitivity analysis (where the content of enhancers and inhibitors in the meals were raised and lowered by 50%) showed that vitamin C and calcium had the largest impact on the estimation of iron absorption for Hallberg's equation (Table 2E.2). For Conway's equation, the most influential foods were high vitamin C fruit or juice, beans and lentils, wholegrain cereals and nuts.

Yet, the estimation of iron absorption was mostly affected by the applied serum ferritin concentration (Table 2E.2). Modelling with a serum ferritin level of 30 instead of 15 $\mu$g/L lowered absorbable iron estimates substantially (Figure 2E.2). Conway's and Hallberg's vegetarian and vegan menu plans complied no longer with the RDA. They only satisfied the EAR for iron. When the diet model was allowed to select meals containing

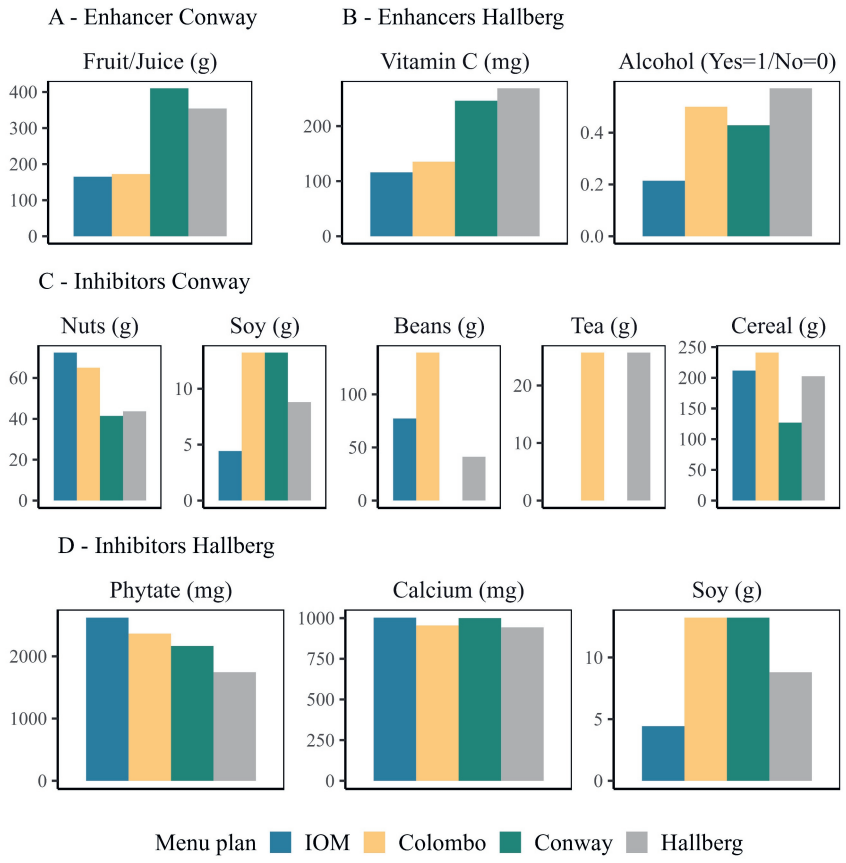



**Figure 2.2:** Average enhancer and inhibitor content (unit/day) of the vegan menu plans. Enhancers and inhibitors used by Conway's and Hallberg's equations are presented in figures A-D respectively.

food items with artificially increased nutrient density and meals with nonheme iron absorption estimates above 50% (Figure 2E.3), Conway's and Hallberg's vegetarian menu plans did satisfy the RDA (Figure 2E.4C-D). The vegan menu plans still only satisfied the EAR.

In summary, the absolute values of absorbable iron estimates based on diet-dependent equations greatly depended on the applied parameter input, but this was not the case for the relative outcomes. Regardless of the input data, retrospective diet-dependent absorbable iron estimates were consistently lower than the estimates based on the constant absorption factors.

## 2.4 Discussion



Our study demonstrated that the applied method to estimate iron absorption has a substantial impact on the estimated absorbable iron content of menu plans. All menu plans based on IOM's and Colombo's constant iron absorption factors failed to meet the RDA or even the EAR for iron when absorbable iron concentrations were evaluated retrospectively by the diet-dependent absorption equations of Conway, Hallberg and Armah. This applied to all types of menu plans ranging from omnivorous to vegan diets. In contrast to the constant absorption factors, the diet-dependent absorption equations ensured the optimization of enhancer and inhibitor concentrations with the result that absorbable iron estimates were higher. As such, all menu plans based on the diet-dependent absorption equations complied with the RDA except for Hallberg's vegan plan, which only satisfied the EAR.

Nonetheless, our sensitivity analysis (Appendix 2E) showed that these results can best be interpreted relatively. Although retrospective diet-dependent absorbable iron estimates were consistently lower than the estimates based on constant absorption factors, the absolute value of absorbable iron estimates heavily depended on the applied parameter input and the absorption equation used.

Of all parameter input, iron absorption estimates by diet-dependent equations are mostly affected by the serum ferritin concentration (Armah et al., 2013; EFSA, 2015). Increasing the serum ferritin level from 15 to 30 $\mu$g/L lowered the estimation of absorbable iron with the result that the RDA for vegetarian and vegan diets could no longer be satisfied (Figure 2E.2). The requirement for absorbable iron is similar for those with low or high iron status, whereas absorptive capacity reduces at higher serum ferritin concentration (Fairweather-Tait et al., 2017), hence it becomes more difficult to satisfy the same requirement. This raises the question which serum ferritin concentration is most appropriate for diet modelling. IOM (2001) applies a serum ferritin concentration of 15 $\mu$g/L to set their dietary reference intake for iron, while the European Food Safety Authority (EFSA, 2015) uses a concentration of 30 $\mu$g/L. More recently, Galetti et al. (2021) advice an even higher

threshold value of 50 $\mu$g/L to define the onset of functional iron inadequacy, especially in young women. Although there are no universally adopted guidelines which serum ferritin concentration to use for setting dietary requirements, it could be argued that the appropriate concentration for diet modelling depends on whether the goal is to establish functional adequacy (15 $\mu$g/L), maintain some iron reserves (30 $\mu$g/L), or build iron stores (50 $\mu$g/L). The latter two goals imply the use of lower estimates for iron bioavailability, and hence translate to higher dietary reference intake recommendations that are harder to achieve.

Furthermore, the level of iron absorption is affected by enhancer and inhibitor concentrations. For this study, several of these parameters had to be estimated. For example, Hallberg's equation required estimates of tannic acid. As regards tannic acid, there was no database available to make these estimations. Therefore tannic acid concentrations were assumed to be zero. However, the sensitivity analysis showed that raising the tannic acid content to 29 mg, which was the average content in meals of Hallberg and Hulthén's second study (2000), lowered nonheme iron absorption by 31% (Table 2E.2). In that case, current absorption estimates would be too high with the result that it would have been more difficult to satisfy the RDA for iron.

Lastly, the outcome of iron absorption estimates depends on the applied absorption equation. For instance, the average nonheme iron absorption of the vegan meals was 29% higher for Hallberg's equation than for Conway's equation (Table 2A.3). A relevant question to answer is which absorption equation is most accurate. In literature, there has been an ongoing debate whether single-meal or complete-diet equations are more accurate. Multiple studies have reported that single-meal equations exaggerate the effect of enhancers and inhibitors on nutrient absorption compared to complete-diet equations (Armah et al., 2015a, 2013; Collings et al., 2013; Dainty et al., 2014; EFSA, 2015; Fairweather-Tait et al., 2017; Reddy et al., 2000). On the other hand, De Carli et al. (2018) reviewed multiple absorption equations and observed that some of the single-meal models outperformed the complete-diet equations that had claimed to be more precise. Our results, however, did only show minor differences between retrospective estimates based on Conway's and Hallberg's single-meal equations and Armah's complete-diet equation (Figure 2.1). All differences (except one) were not statistically significant (Table 2D.1). In contrast, Conway's, Hallberg's and Armah's absorbable iron estimates were all (except two) significantly lower than the estimates based on the constant absorption factors. We suggest further research is necessary to settle the debate about which equation is most appropriate for diet modeling.

Another reason why our results can best be interpreted relatively is because the nutrient content of the optimized menu plans heavily depends on the meals available to the diet model. In this study, we filtered consumption data to obtain the meals for diet modelling. However, these filtering steps can determine whether dietary guidelines can be satisfied or

not. Our analysis showed that the RDA for all vegetarian menu plans could be satisfied when meals containing food items with artificially increased nutrient density and meals with nonheme iron absorption estimates above 50% were included, even when the modelled serum ferritin concentration amounted 30 $\mu$g/L (Figure 2E.4). Without these nutrient-dense meals, only the EAR could be satisfied (Figure 2E.2).

Apart from the fact that the meals used for diet modelling impact the nutritional adequacy of menu plans, these meals also affect the acceptability of designed menu plans. Many diet models design menu plans by adding food items to a so-called 'food basket' (Gazan et al., 2018b). However, the acceptability of the combination of these food items when preparing a meal is not considered this way. Yet, meal acceptability is key to encourage consumers to eat more healthily and sustainably (Forde & Decker, 2022). We aimed to improve the acceptability of menu plans by modelling with consumer-reported meals instead of separate food items. Despite the applied filtering steps to obtain acceptable meals from consumption data, the designed menu plans still contained meals that may not be acceptable to all consumers. For example, one dinner meal in IOM's omnivorous menu plan consisted of raw avocado, tortilla chips and water. As such, the use of recipe data instead of consumption data could be investigated for designing menu plans, but even then menu plans may not be acceptable to all consumers. So far, no study has presented the ultimate solution to secure consumer acceptability (van Dooren, 2018).

In summary, we have investigated the impact of several common methods to estimate absorbable iron intake on the nutritional adequacy of omnivorous, vegetarian and vegan menu plans. Absorbable iron estimates based on constant absorption factors were consistently higher than the retrospective absorbable iron estimates based on diet-dependent equations. To reduce the risk of overestimating the absorbable iron content of menu plans, we advise applying diet-dependent absorption equations for the design of omnivorous, vegetarian and vegan menu plans.

# 2.A Appendix - Meals

**Table 2A.1:** Example of formulating meals from dietary intake data; consumption data is aggregated by respondent, interview day and time of eating occasion.

| Respondent | Interview day | Time of eating occasion | Food item | Quantity (g) | Iron (mg) |
|---|---|---|---|---|---|
| 1 | 1 | 9 AM | Cheese, American | 21 | 0.20 |
| 1 | 1 | 9 AM | Egg omelet or scrambled egg, from fast food / restaurant | 110 | 1.81 |
| 1 | 1 | 9 AM | Muffin, English | 58 | 2.32 |
| 1 | 1 | 9 AM | Grits, regular or quick, made with water, no added fat | 180 | 0.82 |
| 1 | 1 | 9 AM | Jelly, all flavors | 14 | 0.03 |
| 1 | 1 | 9 AM | Coffee, brewed | 360 | 0.04 |
| | | | Total | 743 | 5.22 |

| Meal | Food items | Iron (mg) |
|---|---|---|
| 1 | Cheese, American;<br>Egg omelet or scrambled egg, from fast food / restaurant;<br>Muffin, English;<br>Grits, regular or quick, made with water, no added fat;<br>Jelly, all flavors;<br>Coffee, brewed | 5.22 |

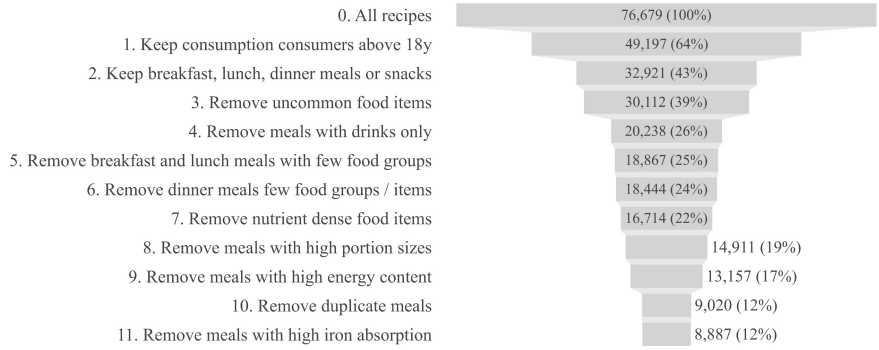



**Figure 2A.1:** Flowchart of the filtering steps applied to NHANES consumption data to obtain the meals used for diet modelling. The bars reflect the count of meals that remained after the applied filtering steps.



**Table 2A.2:** Count and percentage of breakfast, lunch, dinner meals and snacks used for diet modeling.

| Diet | Breakfast | Lunch | Dinner | Snack | Total |
|---|---|---|---|---|---|
| Omnivorous | 2,047 (23%) | 1,688 (19%) | 1,669 (19%) | 3,483 (39%) | 8,887 (100%) |
| Vegetarian | 1,242 (14%) | 349 (4%) | 215 (2%) | 3,044 (34%) | 4,850 (55%) |
| Vegan | 129 (1%) | 86 (1%) | 49 (1%) | 1,277 (14%) | 1,541 (17%) |

**Table 2A.3:** Iron content, estimated absorbable iron content, and enhancer and inhibitor quantities of the meals used for diet modelling.[1][2][3]

| | Omnivorous | Vegetarian | Vegan |
|---|---|---|---|
| Count | 8,865 | 4,828 | 1,532 |
| Total iron (mg) | 2.9 (2.2) | 2.0 (1.9) | 1.6 (1.8) |
| Heme iron (mg) | 0.5 (0.8) | - | - |
| Nonheme iron (mg) | 2.4 (1.9) | 2.0 (1.9) | 1.6 (1.8) |
| **Conway** | | | |
| Total iron absorption (%) | 11.9 (5.1) | 9.6 (3.8) | 10.7 (3.5) |
| Heme iron absorption (%) | 25 | - | - |
| Nonheme iron absorption (%) | 10.3 (4.3) | 9.6 (3.8) | 10.7 (3.5) |
| Total absorbable iron (mg) | 0.34 (0.33) | 0.16 (0.17) | 0.15 (0.18) |
| Enhancers | | | |
| High vitamin C fruit or juice (g) | 10.9 (48.4) | 11.1 (46.8) | 9.7 (41.9) |
| MFP (g) | 31.1 (46.2) | - | - |
| Inhibitors | | | |
| Nuts (g) | 1.9 (9.0) | 3.2 (11.7) | 6.6 (16.4) |
| Soy (g) | 0.3 (5.9) | 0.4 (6.9) | 0.3 (5.9) |
| Beans and lentils (g) | 3.4 (26.8) | 2.3 (21.8) | 3.3 (25.8) |
| Cheese (g) | 2.9 (10.5) | 2.2 (9.7) | - |
| Eggs (g) | 10.4 (33.3) | 10.1 (32.4) | - |
| Tea (g) | 13.8 (60.5) | 13.1 (57.5) | 8.6 (47.6) |
| Cereal (g) | 27.8 (57.6) | 21.6 (49.7) | 12.8 (44.0) |
| Dairy (g) | 18.7 (63.5) | 23.9 (68.8) | - |
| **Hallberg** | | | |
| Total iron absorption (%) | 13.2 (9.5) | 10.5 (9.5) | 13.8 (11.7) |
| Heme iron absorption (%) | 28.2 (9.6) | - | - |
| Nonheme iron absorption (%) | 11.8 (9.4) | 10.5 (9.5) | 13.8 (11.7) |
| Total absorbable iron (mg) | 0.35 (0.38) | 0.13 (0.13) | 0.14 (0.15) |
| Enhancers | | | |
| Vitamin C (mg) | 13.5 (25.4) | 10.8 (23.5) | 11.8 (22.9) |
| Raw MFP (g) | 40.4 (60.0) | - | - |
| Alcohol (Yes=1/No=0) | 0.02 (0.15) | 0.02 (0.13) | 0.02 (0.15) |
| Inhibitors | | | |
| Phytate (mg) | 193.8 (218.2) | 197.3 (262.7) | 248.0 (332.3) |
| Eggs (g) | 10.4 (33.3) | 10.1 (32.4) | - |
| Calcium (mg) | 193.4 (186.4) | 149.4 (168.4) | 51.6 (68.9) |
| Soy (g) | 0.3 (5.9) | 0.4 (6.9) | 0.3 (5.9) |

[1] Results are only displayed for the meals with an iron content greater than zero.
[2] Iron absorption was estimated for a serum ferritin concentration of 15 $\mu$g/L.
[3] The table displays averages with standard deviations in parentheses.

## 2.B Appendix - Iron absorption

Hallberg & Hulthén (2000; 2000b; 2006) presented their equation to estimate nonheme iron absorption separated into multiple factors. Each factor expresses the effect of various dietary components. In Equation 2B.1, all these factors are combined into one single formula. The equation to estimate heme iron absorption is presented in Equation 2B.2. Conway et al.'s (2007) and Armah et al.'s (2013) equations to estimate nonheme iron absorption are given by Equations 2B.3 and 2B.4 respectively.

$$
\begin{aligned}
\text{Nonheme iron absorption (\%)} = \Bigg[ & 22.1 \\
& \cdot 10^{-0.3 \cdot \log(1+\text{Phytate})} \\
& \cdot \left(1 + 0.01 \cdot \text{VitC} + \log(1+\text{Phytate}) \cdot 10^{0.8875 \cdot \log(1+\text{VitC})}\right) \\
& \cdot \min\left\{1,\ (1 + 0.01 \cdot \text{MFP}) \cdot 10^{\,0.4515 - (0.715 - 0.1825 \cdot \log(1+\text{VitC})) \cdot \log(1+\text{TannicAcid})}\right\} \\
& \cdot \left(0.4081 + \frac{0.5919}{1 + 10^{-2.919 \cdot (2.022 - \log(1+\text{Calcium}))}}\right) \\
& \cdot (1 + 0.00628 \cdot \text{MFP} \cdot (1 + 0.006 \cdot \text{Phytate})) \\
& \cdot (1 - 0.022 \cdot \min\{20, \text{Soy}\}) \cdot (1 - 0.27 \cdot \min\{3, \text{Eggs}\}) \\
& \cdot \begin{cases} 1.25, & \text{if Alcohol} > 1 \\ 1, & \text{otherwise} \end{cases} \\
\Bigg] & \cdot \left(\frac{23}{\text{SF}}\right)^{0.94049}
\end{aligned}
\tag{2B.1}
$$

$$
\begin{aligned}
\text{Heme iron absorption (\%)} = & \\
& 10^{\,1.9897 - 0.3092 \cdot \log(\text{SF})} \\
& \cdot \left(0.4081 + \frac{0.6059}{1 + 10^{-2.919 \cdot (2.022 - \log(1+\text{Calcium}))}}\right)
\end{aligned}
\tag{2B.2}
$$

Where Phytate (mg), VitC (vitamin C, mg), MFP (uncooked meat, fish, and poultry, g), Tannic acid (mg), Calcium (mg), and Soy (g) denote their respective amounts; Eggs is the number of eggs (1 egg = 60 g); Alcohol is the number of glasses (1 glass = 10 g alcohol); and SF is serum ferritin ($\mu$g/L).

Nonheme iron absorption (%) =

$$10^{\begin{pmatrix} 0.9385 + 0.001 \cdot \text{Fruit} + 0.0009 \cdot \text{MFP} - 0.0052 \cdot \text{Beans} - 0.0016 \cdot \text{Cereal} - 0.002 \cdot \text{Tea} \\ -0.0012 \cdot \text{Dairy} - 0.0041 \cdot \text{Cheese} - 0.0032 \cdot \text{Eggs} - 0.0062 \cdot \text{Soy} - 0.0087 \cdot \text{Nuts} \end{pmatrix}} \cdot \left(\frac{21.7}{\text{SF}}\right)^{0.94049} \tag{2B.3}$$

Where Fruit (g) refers to high vitamin C (>30 mg/100g) fruit or juice; MFP (g) is meat, fish, and poultry; Beans (g) are beans and lentils, excluding soybeans; Cereal (g) is wholegrain cereals; Tea (g) is black tea infusion; Dairy (g) includes all dairy products; Cheese (g) excludes cream and cottage cheese; Soy (g) is soybeans and flour porridge; Eggs (g) and Nuts (g) refer to their respective food items.

Ln nonheme iron absorption (%) =

$$\begin{aligned} &6.294 - 0.709 \cdot \ln(\text{SF}) + 0.119 \cdot \ln(\text{VitC}) \\ &+ 0.006 \cdot \ln(\text{MFP} + 0.1) - 0.055 \cdot \ln(\text{Tea} + 0.1) \\ &- 0.247 \cdot \ln(\text{Phytate}) - 0.137 \cdot \ln(\text{Calcium}) - 0.083 \cdot \ln(\text{nhIron}) \end{aligned} \tag{2B.4}$$

Where SF is serum ferritin ($\mu$/L); VitC is vitamin C (mg); MFP is meat, fish, and poultry (g); Tea is number of cups; Phytate and Calcium are in mg; and nhIron is nonheme iron (mg).

## 2.C Appendix - Diet model

For each day of the two-week menu plan, one meal is selected for breakfast, one for lunch and one for dinner (2C.1) (Table 2C.1). Two snacks are selected per day (2C.2). Each dinner and lunch meal can only be selected once per two weeks (2C.3). Breakfast meals and snacks can be selected twice every two weeks (2C.4). In addition, menu diversity is secured by restricting the selection of food items for dinner and lunch meals (2C.5) (Table 2C.2). For the selected meals, the nutrient content (2C.6) and the average nutrient content of the two-week menu plan (2C.7) are calculated. Lower (2C.8) and upper (2C.9) bounds are placed to restrict the consumption of macronutrients. Intake of micronutrients is restricted by the tolerable upper intake levels (2C.10). Equation 2C.11 sets the nutrient goals based on the recommended daily allowances (RDA). For a fair comparison between nutrients, deviations are normalized by scaling them between 0 and 1. In addition, both the upper intake levels and RDA goals are adjusted to reflect the absorbable nutrient content instead of the total content. With Equation 2C.12, the maximum deviation below the RDA is determined. Equations 2C.13 and 2C.14 set the domain of the variables. The objective function of the model is to minimize the largest deviation below the RDA (2C.15).

The diet model was solved by applying mixed-integer linear programming using General Algebraic Modeling System (GAMS) software (version 40.1.1; GAMS Development Corp.).

**Table 2C.1:** Model indices, parameters, variables and constraints.

| Indices | |
|---|---|
| n | Nutrient n |
| nmacro | Subset of macronutrients |
| nmicro | Subset of micronutrients |
| r | Meal r |
| i | Food item i |
| m | Meal moment m |

| Parameters | |
|---|---|
| NDays | Number of days to cover by the menu plan |
| $EAR_n$ | Estimated average requirement per day for each micronutrient n |
| $RDA_n$ | Recommended daily allowance for each micronutrient n |
| $UL_n$ | Tolerable upper intake level for each micronutrient n |
| $LB_n$ | Lower bound on intake level for each macronutrient n |
| $UB_n$ | Upper bound on intake level for each macronutrient n |
| $fAbs_n$ | Assumed absorption (%) factor for each micronutrient n |
| $fDiversity_i$ | Number of times a food item i may be added to the menu plan |
| $mealType_{rm}$ | Boolean to indicate whether meal r belongs to meal moment m (yes:1; no:0) |
| $mealFooditems_{rmi}$ | Boolean to indicate whether food item i belongs to meal r with meal moment m (yes:1; no:0) |
| $nContentAbs_{rmn}$ | Nutrient n content for each meal r with meal moment m. For iron, the nutrient content is expressed as absorbable iron in mg. |

| Variables | |
|---|---|
| $X_{rm}$ | Number of times that meal r with meal moment m is added to the menu plan |
| $belowRDA_n$ | Normalised deviation below the recommended daily allowance for micronutrient n. |
| maxBelowRDA | Maximal deviation below RDA |
| $nIntake_{rmn}$ | Nutrient n intake for selected meals r with meal moment m |
| $avgIntake_n$ | Average nutrient n intake per day |

Constraints

$$\sum_r (X_{rm} \cdot mealType_{rm}) = NDays \qquad \forall m \neq \text{'snack'} \quad (2C.1)$$

$$\sum_r (X_{rm} \cdot mealType_{rm}) = 2 \cdot NDays \qquad m = \text{'snack'} \quad (2C.2)$$

$$X_{rm} \leq \frac{NDays}{14} \qquad \forall r,\ m = \text{'dinner', 'lunch'} \quad (2C.3)$$

$$X_{rm} \leq \frac{NDays}{7} \qquad \forall r,\ m = \text{'breakfast', 'snack'} \quad (2C.4)$$

$$\sum_r (X_{rm} \cdot mealFoodItems_{rmi}) \leq fDiversity_i \qquad \forall i,\ m = \text{'dinner', 'lunch'} \quad (2C.5)$$

$$nIntake_{rmn} = nContentAbs_{rmn} \cdot X_{rm} \qquad \forall r, m, n \quad (2C.6)$$

$$avgIntake_n = \frac{\sum_{r,m} nIntake_{rmn}}{NDays} \qquad \forall n \quad (2C.7)$$

$$avgIntake_n \geq LB_n \qquad \forall nmacro \quad (2C.8)$$

$$avgIntake_n \leq UB_n \qquad \forall nmacro \quad (2C.9)$$

$$avgIntake_n \leq UL_n \cdot \frac{fAbs_n}{100} \qquad \forall nmicro \quad (2C.10)$$

$$avgIntake_n + RDA_n \cdot \frac{fAbs_n}{100} \cdot belowRDA_n \geq RDA_n \cdot \frac{fAbs_n}{100} \qquad \forall nmicro \quad (2C.11)$$

$$maxBelowRDA \geq belowRDA_n \qquad \forall nmicro \quad (2C.12)$$

$$belowRDA_n,\ maxBelowRDA \geq 0 \qquad \forall n \quad (2C.13)$$

$$X_{rm} \in \mathbb{Z} \qquad \forall r, m \quad (2C.14)$$

$$\min\{maxBelowRDA\} \quad (2C.15)$$

**Table 2C.2:** Number of times food items from a certain category may be added to the menu plan for lunch and dinner.

| Category | Times per week[1] |
|---|---|
| Alcohol | - |
| Beans, peas, legumes | 3 |
| Cheese | - |
| Coffee | - |
| Condiments and sauces | - |
| Dairy | - |
| Drinks | - |
| Eggs | - |
| Fats and oils | - |
| Fruit | - |
| Grains | 4 |
| MFP | 3 |
| Milk substitutes | - |
| Mixed dishes | 2 |
| Nuts and seeds | - |
| Other | - |
| Potatoes | 4 |
| Processed soy products | 2 |
| Snacks and sweets | - |
| Sugars | - |
| Tea | - |
| Vegetables | 2 |

[1] '-' indicates no constraint is added for this type of food category in the diet model.

## 2.D Appendix - Statistical analysis

We compared the mean absorbable iron estimates of the menu plans based on the constant absorption factors and the retrospective estimates of diet-dependent equations (Table 2D.1). As the meals selected by the diet model were not assigned to specific days, we did not compare the mean daily absorbable iron intake but the mean absorbable iron content of the selected meals. The alternative hypothesis was that the constant absorption estimates were significantly greater than the retrospective diet-dependent estimates. Additionally, we compared Armah's estimates with Conway's and Hallberg's retrospective estimates with a two-sided alternative hypothesis. For both comparisons, estimates were considered significantly different when $P < 0.05$.

The Shapiro-Wilk test and Levene's test, respectively, showed that estimates were not normally distributed, and variances were unequal. Therefore, we applied the Brunner Munzel test (Neubert & Brunner, 2007), also called the generalized Wilcoxon test, which allows ties and adjusts for unequal variances. The statistical analyses were performed using R software (version 4.2.1; R Core Team).

**Table 2D.1:** Pairwise comparisons of retrospective absorbable iron estimates.[1]

| Method[2] | Diet | Est 1[3] | Mean 1[4] | SD 1 | Est 2 | Mean 2 | SD 2 | AH | P[5] | |
|---|---|---|---|---|---|---|---|---|---|---|
| Col | Omn | Col | 0.66 | 0.73 | Arm | 0.46 | 0.49 | > | 6.66E-02 | |
| Col | Omn | Col | 0.66 | 0.73 | Con | 0.35 | 0.45 | > | 1.57E-02 | * |
| Col | Omn | Col | 0.66 | 0.73 | Hal | 0.38 | 0.38 | > | 6.63E-02 | |
| Col | Vegan | Col | 0.62 | 0.34 | Arm | 0.4 | 0.22 | > | 3.51E-08 | * |
| Col | Vegan | Col | 0.62 | 0.34 | Con | 0.52 | 0.49 | > | 1.80E-02 | * |
| Col | Vegan | Col | 0.62 | 0.34 | Hal | 0.48 | 0.41 | > | 2.76E-04 | * |
| Col | Vt | Col | 0.65 | 0.69 | Arm | 0.42 | 0.44 | > | 1.73E-02 | * |
| Col | Vt | Col | 0.65 | 0.69 | Con | 0.3 | 0.43 | > | 5.36E-03 | * |
| Col | Vt | Col | 0.65 | 0.69 | Hal | 0.31 | 0.32 | > | 2.34E-02 | * |
| IOM | Omn | IOM | 0.69 | 0.79 | Arm | 0.33 | 0.34 | > | 3.74E-04 | * |
| IOM | Omn | IOM | 0.69 | 0.79 | Con | 0.23 | 0.2 | > | 7.80E-07 | * |
| IOM | Omn | IOM | 0.69 | 0.79 | Hal | 0.27 | 0.21 | > | 5.24E-05 | * |
| IOM | Vegan | IOM | 0.66 | 0.58 | Arm | 0.23 | 0.2 | > | 2.57E-10 | * |
| IOM | Vegan | IOM | 0.66 | 0.58 | Con | 0.3 | 0.39 | > | 2.27E-07 | * |
| IOM | Vegan | IOM | 0.66 | 0.58 | Hal | 0.22 | 0.25 | > | 2.63E-10 | * |
| IOM | Vt | IOM | 0.65 | 0.95 | Arm | 0.25 | 0.37 | > | 3.23E-05 | * |
| IOM | Vt | IOM | 0.65 | 0.95 | Con | 0.22 | 0.42 | > | 4.24E-07 | * |
| IOM | Vt | IOM | 0.65 | 0.95 | Hal | 0.21 | 0.31 | > | 1.71E-06 | * |
| Con | Omn | Hal | 0.53 | 0.6 | Arm | 0.45 | 0.43 | < > | 6.76E-01 | |
| Con | Vegan | Hal | 0.47 | 0.41 | Arm | 0.37 | 0.23 | < > | 6.64E-01 | |
| Con | Vt | Hal | 0.44 | 0.39 | Arm | 0.4 | 0.31 | < > | 7.05E-01 | |
| Col | Omn | Con | 0.35 | 0.45 | Arm | 0.46 | 0.49 | < > | 2.57E-01 | |
| Col | Omn | Hal | 0.38 | 0.38 | Arm | 0.46 | 0.49 | < > | 8.67E-01 | |
| Col | Vegan | Con | 0.52 | 0.49 | Arm | 0.4 | 0.22 | < > | 4.70E-01 | |
| Col | Vegan | Hal | 0.48 | 0.41 | Arm | 0.4 | 0.22 | < > | 6.12E-01 | |
| Col | Vt | Con | 0.3 | 0.43 | Arm | 0.42 | 0.44 | < > | 2.88E-01 | |
| Col | Vt | Hal | 0.31 | 0.32 | Arm | 0.42 | 0.44 | < > | 7.44E-01 | |
| Hal | Omn | Con | 0.42 | 0.53 | Arm | 0.4 | 0.44 | < > | 9.98E-01 | |
| Hal | Vegan | Con | 0.54 | 0.44 | Arm | 0.38 | 0.26 | < > | 8.06E-03 | * |
| Hal | Vt | Con | 0.52 | 0.45 | Arm | 0.45 | 0.38 | < > | 2.03E-01 | |
| IOM | Omn | Con | 0.23 | 0.2 | Arm | 0.33 | 0.34 | < > | 2.77E-01 | |
| IOM | Omn | Hal | 0.27 | 0.21 | Arm | 0.33 | 0.34 | < > | 9.97E-01 | |
| IOM | Vegan | Con | 0.3 | 0.39 | Arm | 0.23 | 0.2 | < > | 9.68E-01 | |
| IOM | Vegan | Hal | 0.22 | 0.25 | Arm | 0.23 | 0.2 | < > | 5.42E-01 | |
| IOM | Vt | Con | 0.22 | 0.42 | Arm | 0.25 | 0.37 | < > | 6.80E-01 | |
| IOM | Vt | Hal | 0.21 | 0.31 | Arm | 0.25 | 0.37 | < > | 9.48E-01 | |

[1] The following abbreviations are used for the table values and headers: >, greater; < >, two-sided; AH, Alternative hypothesis; Arm, Armah; Con, Conway; Est, Estimate; Com, Colombo; Hal, Hallberg; Omn, Omnivorous; Vt, Vegetarian.

[2] Absorption method used for diet modelling.

[3] Absorption method used to estimate iron absorption retrospectively.

[4] Mean retrospective absorbable iron estimates of the meals selected by the diet model.

[5] '*' indicates differences are statistically significant.

# 2.E Appendix - Sensitivity analyses

### 2.E.1 Top N menu plans

When all nutrient goals (RDA's) can be satisfied, alternative optimal solutions may exist. To obtain the top N of best menu plans, we reran the diet model several times. Meals that had been selected in the first run, were excluded from the second model run. For the third model run, meals selected in both the first and second model runs were excluded and so forth. For all menu plans, this procedure was repeated 20 times as long as there were enough meals left to obtain a feasible solution (Table 2E.1, Figure 2E.1). For execution speed, the maximum model run time was set at 3 minutes. This implies that the found solutions may not be globally optimal but they are close to the optimal solution.



**Table 2E.1:** Number of menu plans that satisfied all RDAs.[1]

| Diet type | Absorption method | Nr of menu plans | Bottleneck nutrient[2] |
|---|---|---|---|
| Omnivorous | IOM | 14 | Ca, Fe[3], Vit. A, Vit. B9 |
| Omnivorous | Colombo | 3 | Ca, Fe, Vit. A, Vit. B9 |
| Omnivorous | Conway | 3 | Ca, Fe, Vit. B9 |
| Omnivorous | Hallberg | 8 | Ca, Fe, Vit. A, Vit. B9 |
| Vegetarian | IOM | 10 | Ca, Fe, Vit. A, Vit. B9 |
| Vegetarian | Colombo | 2 | Ca, Fe, Vit. A |
| Vegetarian | Conway | 1 | Ca, Fe, Zn, Vit. A, Vit. B9 |
| Vegetarian | Hallberg | 1 | Ca, Fe, Zn |
| Vegan | IOM | 2 | Ca, Fe, Vit. A, Vit. B9 |
| Vegan | Colombo | 0 | Ca, Fe, Vit. A |
| Vegan | Conway | 1 | Ca, Fe, Zn, Vit. A |
| Vegan | Hallberg | 0 | Ca, Fe, Zn |

[1] Excluding vitamin B12.

[2] This column indicates which micronutrients were the bottleneck for the first menu plan that could not satisfy all RDAs.

[3] The absorbable iron content was estimated by using a serum ferritin concentration of 15 $\mu$g/L.

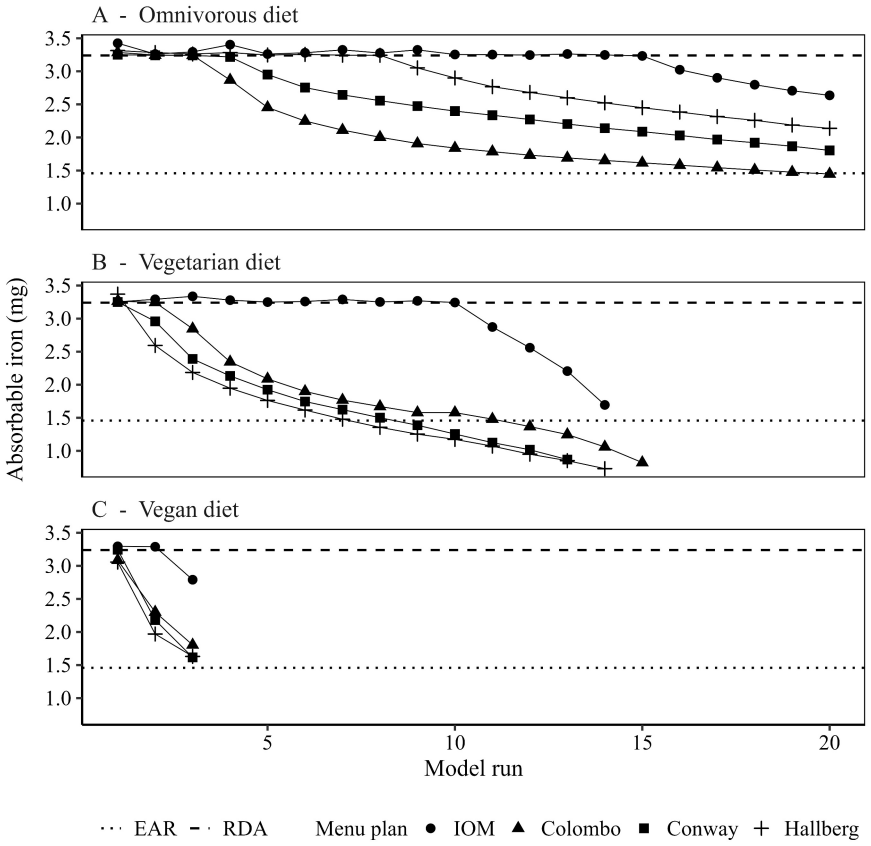



**Figure 2E.1:** Average absorbable iron content (mg/day) for the top N best menu plans compared to the EAR and RDA (women 19-50y). The absorption of iron was estimated using a serum ferritin concentration of 15 $\mu$g/L.

### 2.E.2 Parameter input iron absorption equations

To investigate the sensitivity of parameter input used for the estimation of nonheme iron absorption by Conway's and Hallberg's equations, the average enhancer and inhibitor content of the meals selected for the optimal menu plans was calculated. This constitutes the base case. The serum ferritin concentration for the base case amounted 15 $\mu$g/L. For each enhancer and inhibitor, the content was raised and lowered by 50% while leaving other parameter input unchanged. The tannic acid content, which was assumed to be 0, was raised to 14 and 29 mg. This is equivalent to the average tannic acid content of the meals used in respectively the first and second study of Hallberg and Hulthén (Hallberg & Hulthén, 2000). The serum ferritin concentration was raised to 30 $\mu$g/L, which is the value

used by the European Food Safety Authority (EFSA) to set dietary guidelines (EFSA, 2015). Finally, nonheme iron absorption was estimated for each scenario and compared to the base case (Table 2E.2).

**Table 2E.2:** Sensitivity analysis on the impact of parameter input on nonheme iron absorption estimates by Conway's and Hallberg's equations.

| | | Change in nonheme iron absorption (%) | |
|---|---|---|---|
| | Mean content | Lower content (-50%) | Higher content (+50%) |
| Conway | | | |
| Serum ferritin ($\mu$g/L) | 15 | - | -47.9[1] |
| Enhancer | | | |
| Fruit/Juice (g) | 49.0 | -6.0 | 5.8 |
| MFP (g) | 5.0 | -0.5 | 0.5 |
| Inhibitor | | | |
| Beans (g) | 16.0 | 10.0 | -9.1 |
| Cereal (g) | 43.8 | 8.4 | -7.8 |
| Cheese (g) | 0.9 | 0.4 | -0.4 |
| Dairy (g) | 12.2 | 1.7 | -1.7 |
| Eggs (g) | 3.8 | 1.4 | -1.4 |
| Nuts (g) | 6.5 | 6.7 | -6.3 |
| Soy (g) | 0.7 | 0.5 | -0.5 |
| Tea (g) | 3.0 | 0.7 | -0.7 |
| Hallberg | | | |
| Serum ferritin ($\mu$g/L) | 15 | - | -47.9 |
| Enhancer | | | |
| Alcohol (Yes=1/No=0) | 0.1 | 0.0 | 0.0 |
| MFP (g) | 6.5 | -5.3 | 5.3 |
| Vitamin C (mg) | 33.2 | -22.1 | 21.4 |
| Inhibitor | | | |
| Calcium (mg) | 203.6 | 49 | -10.4 |
| Eggs (g) | 3.8 | 0.9 | -0.9 |
| Phytate (mg) | 311.4 | 14.5 | -6.4 |
| Soy (g) | 0.7 | 0.8 | -0.8 |
| Tannic acid (mg) | 0.0 | - | -7.2 / -31.4[2] |

[1] The serum ferritin concentration was raised to 30 $\mu$g/L.

[2] The tannic acid content was raised to 14 and 29 mg, which is equivalent to the average tannic acid content of the meals used in respectively the first and second study of (Hallberg & Hulthén, 2000).

In addition, we repeated our main analysis (Section 2.2) with a serum ferritin concentration of 30 $\mu$g/L (Figure 2E.2) to investigate how menu plans would change when modelling with a higher serum ferritin concentration. We only repeated our main analysis for input parameter serum ferritin as diet-dependent iron absorption estimates are mostly affected by serum ferritin (Armah et al., 2013; EFSA, 2015).

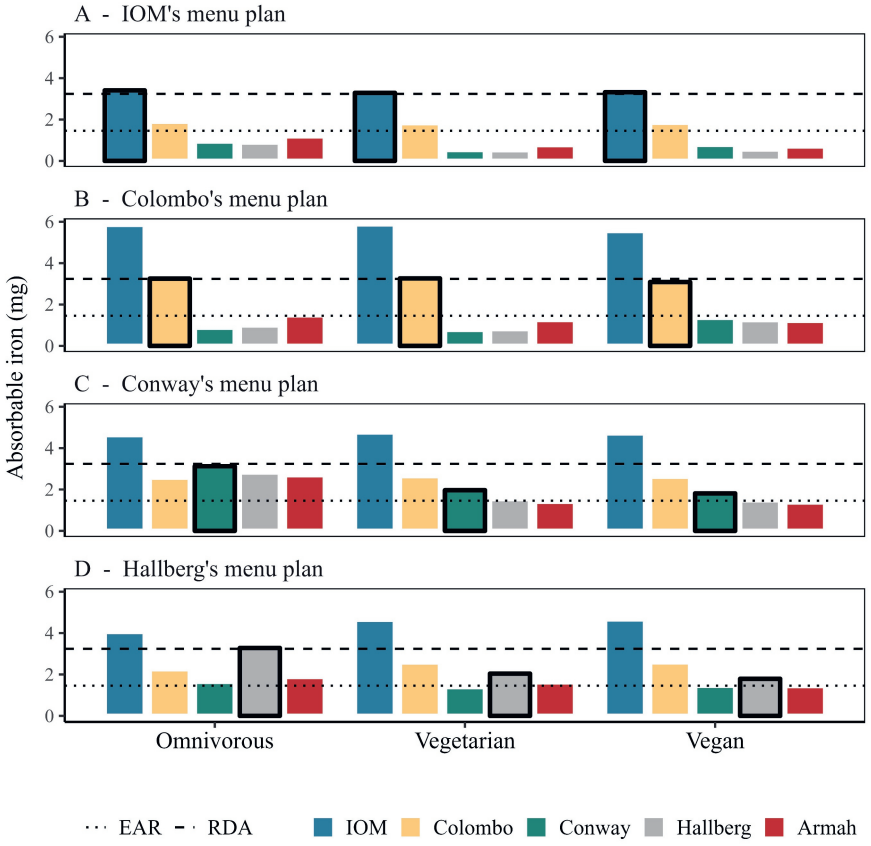



**Figure 2E.2:** Average absorbable iron content (mg/day) compared to the EAR and RDA (women 19-50y). The absorption of iron was estimated using a serum ferritin concentration of 30 $\mu$g/L.

### 2.E.3 Meals used for diet modeling

In our main analysis (Section 2.2.1), we applied several filtering steps to obtain meals based on consumption data (Figure 2A.1). These filtering steps were applied for two reasons: 1) to avoid menu plans with unrealistic meals (e.g. dinner meals that consist of fruit only), and 2) to avoid menu plans with unrealistically high nutrient content. To demonstrate the impact of these filtering decisions on the absorbable iron content of menu plans, we repeated our main analysis, with the only difference that we did not filter meals containing food items with artificially increased nutrient density and meals with nonheme iron absorption estimates above 50% (Figures 2E.3 and 2E.4).




**Figure 2E.3:** Flowchart of the filtering steps applied to NHANES consumption data to obtain the meals used for diet modelling. The bars reflect the count of meals that remained after the applied filtering steps. Meals containing food items with artificially increased nutrient density and meals with nonheme iron absorption estimates above 50% were not filtered.

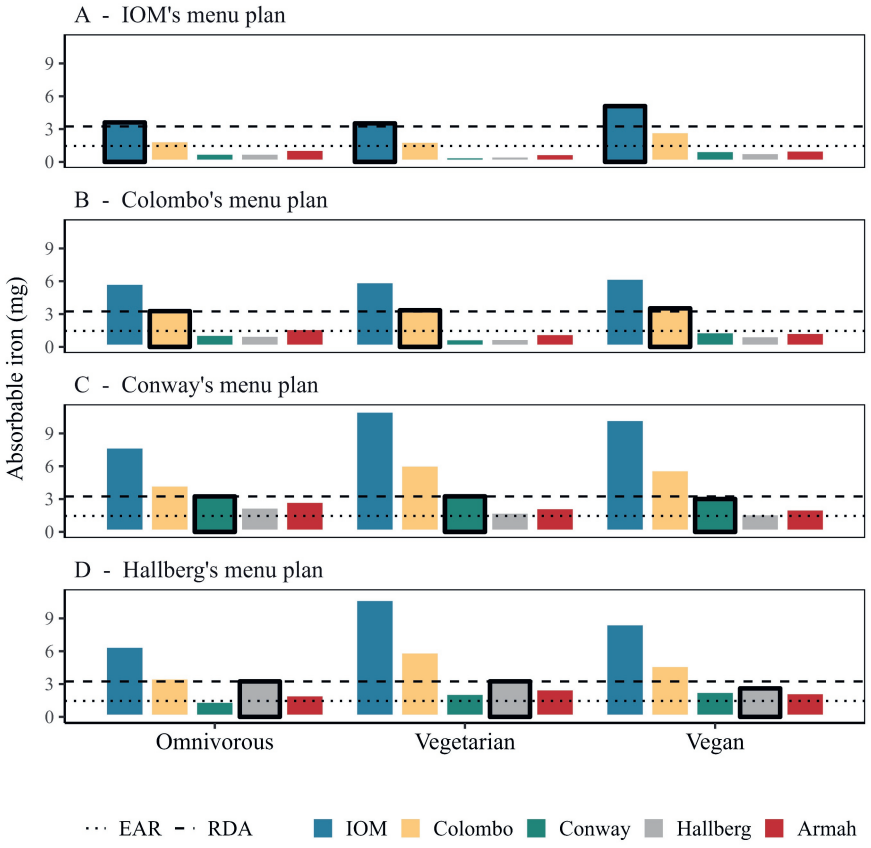



**Figure 2E.4:** Average absorbable iron content (mg/day) compared to the EAR and RDA (women 19-50y). The absorption of iron was estimated using a serum ferritin concentration of 30 $\mu$g/L. Meals containing food items with artificially increased nutrient density and meals with nonheme iron absorption estimates above 50% were not removed from the dataset used for diet modelling.

## 2.F Appendix - Nutrient content menu plans

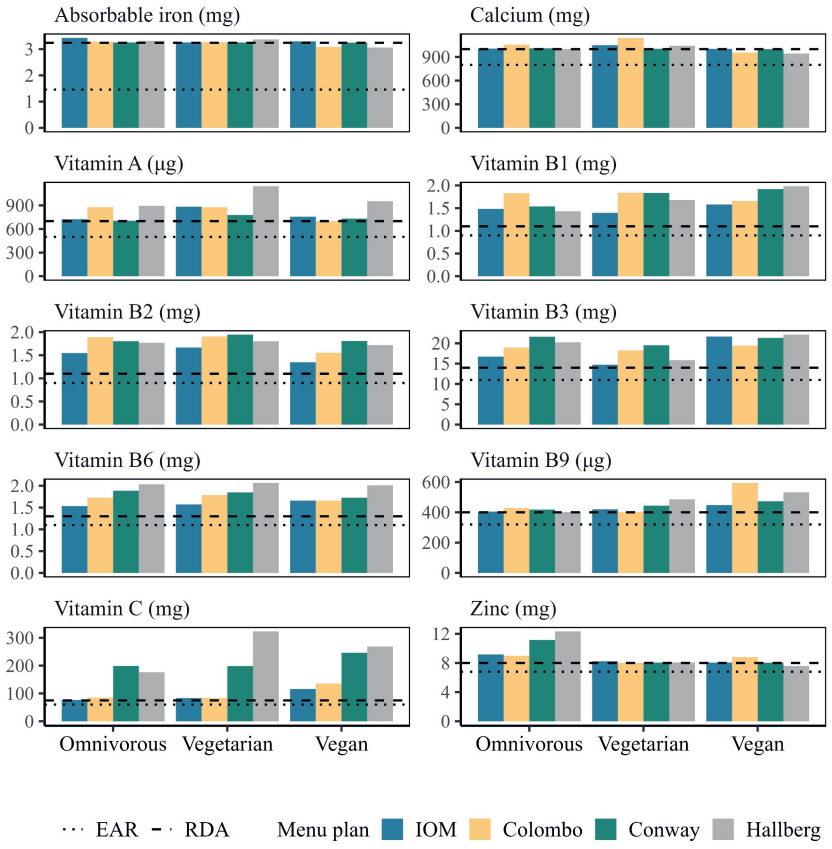



**Figure 2F.1:** Average nutrient content (unit/day) compared to the EAR and RDA (women 19-50y). The absorbable iron content was calculated with a serum ferritin concentration of 15 $\mu$g/L.

# Chapter 3

## Diet optimization: modeling iron and zinc absorption by nonlinear programming and piecewise linear approximation using NHANES



This chapter is published as:

van Wonderen, D., Gerdessen, J. C., Kirst, P., & Melse-Boonstra, A. (2025). Diet optimization: modeling iron and zinc absorption by nonlinear programming and piecewise linear approximation using NHANES. *The American Journal of Clinical Nutrition*, 122(3), 853-858.

# Abstract

**Background:** The accuracy of the calculation of absorbable nonheme iron and zinc content in diet-model-generated menu plans can be improved by using nonlinear absorption equations. The resulting diet models cannot be solved with standard linear programming software.

**Objective:** The aim of this study is to evaluate the effectiveness of nonlinear programming (NLP) and piecewise linear approximation (PLA) for solving diet models with nonlinear equations for nonheme iron and zinc absorption.

**Methods:** A mixed-integer and a continuous diet model were developed to optimize absorbable iron and zinc intake, using different absorption equations available from literature. Model input data were obtained from NHANES. Various diet plans were then generated applying both NLP and PLA techniques. Evaluation criteria included solution quality and computational efficiency.

**Results:** For the mixed-integer diet model, PLA found accurate solutions within minutes, outperforming NLP in consistency and solution quality. NLP frequently hit the 1-hour time limit and did not always find the best observed solution. In the worst cases, NLP either found no solution or the deviation was as large as 2.1 mg for absorbable iron. For absorbable zinc, the maximum deviation was only 0.2 mg. For the continuous diet model, NLP and PLA performed equally well in most cases.

**Conclusion:** This study provides practical examples for researchers who seek to improve the accuracy of their diet models through the implementation of nonheme iron and zinc absorption equations using either NLP or PLA.

## 3.1 Background

Absorption of iron and zinc varies among individuals and diets, influenced by factors such as nutrient status, the concentration of these minerals in consumed meals, and their interactions with other dietary components. Iron is present in two forms: heme and nonheme, with heme iron being more easily absorbed than nonheme iron. The absorption of nonheme iron can be enhanced by dietary components such as vitamin C, while phytates inhibit its uptake. Similarly, zinc absorption is affected by dietary components like phytates (Singh & Prasad, 2023). Several equations have been proposed in the literature to estimate the absorption of nonheme iron (De Carli et al., 2018) and zinc (Miller et al., 2007, 2013) (Appendix 3A).

In diet modeling—a quantitative technique for designing healthy and sustainable diets—nutrient absorption is often approximated using a constant factor (van Wonderen et al., 2023). This approach, however, does not account for the influence of dietary enhancers and inhibitors. Incorporating absorption equations can improve accuracy but introduces nonlinearity, which cannot be handled by standard linear programming software. Consequently, alternative techniques such as nonlinear programming (**NLP**) or piecewise linear approximation (**PLA**) are required.



When applying NLP, it is important to validate model outcomes, as nonlinear solvers may produce inaccurate results under certain circumstances. Validation strategies include using multiple solvers and experimenting with different solver settings (McCullough & Vinod, 2003). For example, solution quality and computational efficiency can be improved by providing the algorithm with a good initial solution or by initiating it from multiple starting points (György & Kocsis, 2011). Compared to NLP, PLA can offer advantages such as shorter model runtimes and greater robustness in finding an optimal solution. The main challenge with PLA lies in selecting appropriate segments for approximating the nonlinear function. Insufficient segmentation may result in the loss of crucial nonlinear characteristics, thereby reducing the accuracy of model outcomes. Conversely, excessive segmentation can substantially prolong model runtime (Geißler et al., 2011). A more detailed background on NLP and PLA is provided in Appendices 3B.1 and 3C.1, respectively.

Previous studies (Barré et al., 2018; Dussiot et al., 2022b) that incorporated nonlinear absorption equations into diet modeling have used NLP with a local solver. However, the effectiveness of NLP in this application has not yet been assessed, and no comparative analysis with PLA has been performed. In this study, we demonstrate how NLP and PLA can be used to model nonheme iron and zinc absorption, and analyze the performance of both approaches. We do this for both a mixed-integer diet model, which combines meals into daily menus, and a continuous diet model, which optimizes food quantities from observed diets, with the former being more challenging to solve (Lee & Leyffer,

2011) (Appendix 3B.1.1). It is important to note that this paper does not aim to cover all available methods for NLP and PLA, explore every performance improvement strategy, or provide diet plans intended for real-world application. Instead, our aim is to provide practical examples for other researchers seeking to enhance the accuracy of their diet models through the implementation of nonheme iron and zinc absorption equations using either NLP or PLA.

## 3.2 Methods

### 3.2.1 Input data

The mixed-integer and continuous diet models used meals and observed consumption data, respectively, as model inputs. Both meal data and observed consumption data were derived from dietary intake data provided by the National Health and Nutrition Examination Survey (NHANES) 2017-2018 (CDC & NCHS, 2018). The NHANES dataset provides data on demographics, the types and amounts of food consumed, and their corresponding nutrient values. For the mixed-integer diet model, meals were obtained by aggregating consumption data by respondent, interview day, and eating occasion. Filtering steps were applied to enhance the realism of the meals, such as excluding those with unconventionally high or low caloric intake (van Wonderen et al., 2023). A total of 9,016 omnivorous, 4,901 vegetarian, and 1,590 vegan meals were available for selection during modeling. For the continuous diet model, individual daily consumption data were used. Only women of reproductive age (19-50y) were included. Respondents with a daily energy intake below 1,200 kcal or above 3,000 kcal were excluded, resulting in a final sample of 882 women. The average daily intake of 2,732 food items was then calculated based on their reported consumption.

Nutrients and food groups missing in NHANES but necessary for calculating iron and zinc absorption were estimated as follows (van Wonderen et al., 2023): heme iron was assumed to constitute half of the total iron content in meat, fish, and poultry; phytate was estimated using the method reported by Larvie & Armah (2021); and food groups were derived from the food categories available in the What we Eat In America (WWEIA) database (USDA & ARS, 2020c) and the Food Patterns Equivalents Database (FPED) (Bowman et al., 2022).

Given that no human participants or animals were involved in this study and the data used for modeling were publicly available, ethical approval was not necessary.

### 3.2.2 Diet modeling

We tested the performance of nonlinear programming (NLP) and piecewise linear approximation (PLA) for both a mixed-integer and a continuous diet model. The mixed-integer

diet model generated a two-week menu plan consisting of a diverse set of meals. The model selected one breakfast, lunch, and dinner, along with two snacks for each day (decision variable $X_i = 1$ if meal $i$ is selected, 0 otherwise). To ensure menu diversity, restrictions were placed on how often a meal could be selected each week and how frequently food items from a specific category could be included in the menu plan. The continuous diet model adapted observed consumption patterns through the optimization of food item quantities (decision variable $X_i =$ amount of food $i$). Constraints on modifications to food items and WWEIA food subgroups were applied to limit deviation from the observed diet. For both models, macronutrient intake was constrained by lower and upper bounds, while for micronutrient intake deviations below the Recommended Dietary Allowances (RDA) were minimized (objective function). Once all RDAs were met, absorbable iron or zinc was further maximized to eliminate alternative optimal solutions and guarantee a unique solution.

To calculate absorbable iron, we implemented the nonheme iron absorption equations developed by Armah et al. (2013) (**Armah**) and Conway et al. (2007) (**Conway**), modeling with a serum ferritin concentration of 30 $\mu$g/L (EFSA, 2015). Armah's equation estimates absorption based on the complete diet, whereas Conway's equation estimates absorption for each single meal. Heme iron absorption was assumed to be 25% (Armah et al., 2015b). Zinc absorption was estimated using the complete diet equation by Miller et al. (2007) (**Miller**), and later updated by Hambidge et al. (2010) (Appendix 3A). These equations were selected because they can be approximated linearly relatively easily compared to other absorption equations found in the literature. We evaluated each absorption equation independently. For instance, when modeling nonheme iron absorption using Armah's equation, we applied a constant absorption factor for zinc.

To implement these nonlinear absorption equations, both NLP and PLA were applied. For NLP, two different nonlinear solvers were applied: local solver Linear Interactive and Discrete Optimizer (**LINDO**) (GAMS, 2024b) and global solver Solving Constraint Integer Programs (**SCIP**) (Vigerske & Gleixner, 2018). In our evaluation of LINDO, we explored the efficiency of both the multistart and initialization functionalities. For SCIP, we focused solely on the impact of providing the models with an initial solution. Background information on these techniques is provided in Appendix 3B. For PLA, the linear CPLEX solver (GAMS, 2024a) was applied. Appendix 3C provides an explanation of how the parameters were obtained to linearize the nonlinear absorption equations. The mathematical notation and parameter input for the NLP and PLA diet models are provided in Appendix 3D. All diet models were executed within the General Algebraic Modeling System (GAMS) software framework (version 40.1.1; GAMS Development Corp.), with each model tested under two time constraints: 30 minutes and 1 hour.

With the NLP and PLA mixed-integer diet models, three types of two-week menu plans were created: omnivorous, vegetarian, and vegan. With the continuous diet models, diet

plans were generated with a varying degree of allowed deviations from the observed diet. In the most restrictive case, food item changes were limited to 50%, while food group changes were limited to 40%, to which we refer as the I50-G40 scenario. Other scenarios included I75-G50 and I100-G60. We tested these scenarios since the size of the problem and the search space can also affect how well NLP and PLA perform (Lee & Leyffer, 2011).

The notation and all model formulations are provided in Appendices 3D.1–3D.4. A summary of the experimental setup is given in Table 3.1, while a full overview is available in Appendix 3D.5.

**Table 3.1:** Modeling aspects and options.[1][2]

| Modeling aspects | Options |
|---|---|
| Absorption equation | Iron:<br>– Armah<br>– Conway<br>Zinc:<br>– Miller |
| Diet model | – Mixed-integer<br>– Continuous |
| Diet plan | Mixed-integer:<br>– Vegan<br>– Vegetarian<br>– Omnivorous<br><br>Continuous:<br>– I50-G40[3]<br>– I75-G50<br>– I100-G60 |
| Method | – NLP<br>– PLA |
| Solver | NLP:<br>– LINDO<br>– SCIP<br><br>PLA:<br>– CPLEX |
| Solver settings | – Default<br>– Initial solution<br>– Multistart |
| Time limit | – 30 min<br>– 1 h |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); NLP, nonlinear programming; LINDO, Linear Interactive and Discrete Optimizer; Miller, Hambidge et al. (2010); PLA, piecewise linear approximation; SCIP, Solving Constraint Integer Programs.

[2] A full overview of modeling experiments is given in Appendix 3D.5.

[3] I indicates the percentage of allowed food item changes, while G represents the allowed food group changes.

### 3.2.3 Evaluation of results

The accuracy of PLA was evaluated by comparing the absorbable nonheme iron and zinc estimates obtained from the linearized functions with those from the original nonlinear functions. For the single meal equation of Conway, the absorbable nonheme iron content was computed and compared for each meal in the meal dataset. For the complete

diet equations of Armah and Miller, absorption was calculated and compared for 1,000 simulated two-week menu plans (Appendix 3C.3).

The performance of the NLP and PLA diet models was evaluated based on the computational time required for each solution and how closely their outcomes matched the optimal solution. In this study, the best solution refers to the solution with the highest absorbable iron or zinc content.

## 3.3 Results

We begin by briefly evaluating the accuracy of linearizing the nonlinear iron and zinc absorption equations. Next, we compare the performance of the nonlinear programming (NLP) and piecewise linear approximation (PLA) approaches for the mixed-integer diet model, starting with iron and followed by zinc. The results for the continuous diet model are presented afterward.



The comparison between absorbable nonheme iron and zinc estimates obtained from the nonlinear and linearized functions showed that, on average, the linearized functions yielded similar results to the nonlinear functions. The maximum deviations between NLP and PLA were 0.1 mg of absorbable iron for Conway, 0.2 mg of absorbable iron for Armah, and 0.1 mg of absorbable zinc for Miller (Appendix 3E.1).

Analyzing the optimized menu plans generated by the mixed-integer diet model for absorbable iron, we first observed that PLA consistently retrieved the optimal solution for Conway and the highest solution observed for Armah (Table 3.2). In contrast, the NLP approaches showed substantial variation in performance. For example, LINDO was unable to find a solution even after one hour for the vegetarian and omnivorous menu plans based on Conway's equation. For the vegetarian menu plan based on Armah's equation, the solutions provided by LINDO and SCIP were lower by 0.9 mg and 0.5 mg, respectively, than the highest observed solution. The second main observation is that PLA had shorter model runtimes compared to the NLP approaches. The longest observed runtime for PLA was 4.5 minutes (Conway Omnivorous), whereas NLP frequently reached the time limit of 1 hour.

Comparison between local solver LINDO and global solver SCIP reveals that SCIP outperformed LINDO in estimating higher absorbable nonheme iron, with differences ranging from 0.4 mg to 2.1 mg for menu plans based on Conway's equation, while model runtimes were comparable. When analyzing menu plans based on Armah's equation, SCIP's estimates of absorbable nonheme iron were similar to LINDO's, yet the model runtimes were often much longer. For example, with the vegan menu plan, SCIP reached the 1-hour time limit, while LINDO obtained the solution within a few minutes.

**Table 3.2:** Comparison of model runtime and average absorbable iron and zinc content of the optimized menu plans (mg/day) based on Armah's, Conway's, and Miller's equations for the nonlinear programming (NLP) and piecewise linear approximation (PLA) mixed-integer diet models.[1]

| Equation | Model | Vegan | | Vegetarian | | Omnivorous | |
|---|---|---|---|---|---|---|---|
| | | Runtime[2] | Absorbed[3] | Runtime | Absorbed | Runtime | Absorbed |
| **Iron** | | | | | | | |
| Armah | PLA[4] | 0:00:29 | **1.6** | 0:00:19 | **2.2** | 0:01:36 | **3.2** |
| | NLP | | | | | | |
| | LINDO | 0:07:27 | **1.6** | 0:00:39 | 1.3 | 0:07:28 | **3.2** |
| | LINDO+initial | 0:10:36 | **1.6** | 0:15:41 | **2.2** | 1:00:13 | 2.8 |
| | LINDO+initial+multistart | 0:03:13 | **1.6** | 0:03:52 | **2.2** | 0:31:60 | **3.2** |
| | SCIP | 1:00:28 | **1.6** | 1:00:07 | 1.7 | 1:00:12 | 3.1 |
| | SCIP+initial | 1:00:32 | **1.6** | 1:00:07 | **2.2** | 1:00:12 | **3.2** |
| Conway | PLA | 0:00:54 | **1.8** | 0:01:50 | **2.0** | 0:04:34 | **3.1** |
| | NLP | | | | | | |
| | LINDO | 0:00:29 | 1.4 | 1:00:13 | - | 1:00:12 | - |
| | LINDO+initial | 1:02:02 | 0.8 | 1:04:05 | 0.5 | 1:20:34 | 0.9 |
| | LINDO+initial+multistart | 1:01:49 | 0.8 | 1:04:01 | 0.5 | 1:20:45 | 0.9 |
| | SCIP | 0:01:38 | **1.8** | 1:00:26 | 1.9 | 1:10:39 | 3.0 |
| | SCIP+initial | 0:02:47 | **1.8** | 1:00:22 | 1.9 | 1:00:46 | 0.9 |
| **Zinc** | | | | | | | |
| Miller | PLA | 0:00:31 | **2.9** | 0:00:16 | **4.2** | 0:00:53 | **5.1** |
| | NLP | | | | | | |
| | LINDO | 0:01:15 | **2.9** | 0:02:01 | 4.1 | 0:03:04 | 5.0 |
| | SCIP | 1:00:08 | 2.7 | 1:00:07 | **4.2** | 1:00:10 | **5.1** |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); NLP, nonlinear programming; LINDO, Linear Interactive and Discrete Optimizer; Miller, Hambidge et al. (2010); PLA, piecewise linear approximation; SCIP, Solving Constraint Integer Programs.

[2] Runtime is displayed in h:mm:ss. The limit for the model's runtime was set to 1 hour.

[3] Per type of menu plan, highest absorbable iron and zinc intake (mg/day) estimates are highlighted in bold.

[4] For the menu plans obtained by the PLA model, absorbable iron and zinc intake estimates based on the original absorption equations were calculated retrospectively.

The attempt to enhance performance by providing the NLP approaches with an initial solution did not always improve results. In certain instances, absorbable nonheme iron estimates were even lower compared to the solutions where no initial solution was provided (Armah Omnivorous LINDO, Conway Vegan LINDO, Conway Omnivorous SCIP). LINDO's multistart procedure did not notably improve the estimates of absorbable nonheme iron either. However, for menu plans based on Armah's equation, it effectively

replaced the provided initial solution with better alternatives generated by the algorithm, thereby reducing the model's runtime.

Furthermore, increasing the model runtime limit only marginally improved absorbable nonheme iron estimates. When compared to a 30-minute time limit (Appendix 3E.2), extending the runtime to 1 hour enhanced results only for Conway Vegetarian SCIP (+ startpoint) with 1.4 mg.

Analyzing the modeling results for absorbable zinc, we observed that absorbable zinc intake estimates were comparable between the PLA and NLP models. SCIP's vegan menu plan showed the greatest difference, with absorbable zinc being 0.2 mg lower than in PLA. When analyzing nutrients other than iron and zinc, we found that menu plans with similar absorbable iron or zinc content generally had similar nutrient profiles (Appendix 3E.3).

For the continuous diet model optimizing absorbable iron, PLA and LINDO performed similarly, producing similar model runtimes and iron estimates (Table 3.3). However, SCIP failed to find a solution within one hour. For zinc, all three methods—PLA, LINDO, and SCIP—delivered similar results. Varying the allowed changes in food item and food group quantities did not affect these results.

To summarize, for the mixed-integer diet model, PLA produced accurate modeling solutions with greater consistency and shorter model runtimes than NLP. NLP solution quality varied, and attempts to improve them with an initial solution or a multistart approach were mostly unsuccessful. For the continuous diet model, PLA and NLP showed comparable performance in all cases except for one.

**Table 3.3:** Comparison of model runtime and average absorbable iron and zinc content of the optimized diet plans (mg/day) based on Armah's and Miller's equations for the nonlinear programming (NLP) and piecewise linear approximation (PLA) continuous diet models.[1]

| Equation | Model | I50%–G40%[2] | | I75%–G50% | | I100%–G60% | |
|---|---|---|---|---|---|---|---|
| | | Runtime[3] | Absorbed | Runtime | Absorbed | Runtime | Absorbed |
| **Iron**[4] | | | | | | | |
| Armah | PLA[5] | 0:00:04 | 1.9 | 0:00:04 | 2.3 | 0:00:06 | 2.6 |
| | NLP | | | | | | |
| | LINDO | 0:00:20 | 1.9 | 0:00:17 | 2.3 | 0:00:24 | 2.6 |
| | SCIP | 1:00:01 | – | 1:00:02 | – | 1:00:02 | – |
| **Zinc** | | | | | | | |
| Miller | PLA | 0:00:20 | 4.1 | 0:00:23 | 4.5 | 0:00:21 | 4.7 |
| | NLP | | | | | | |
| | LINDO | 0:00:20 | 4.1 | 0:00:23 | 4.5 | 0:00:23 | 4.7 |
| | SCIP | 0:00:39 | 4.1 | 0:00:38 | 4.5 | 0:00:40 | 4.7 |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); NLP, nonlinear programming; LINDO, Linear Interactive and Discrete Optimizer; Miller, Hambidge et al. (2010); PLA, piecewise linear approximation; SCIP, Solving Constraint Integer Programs.

[2] Diet plans were generated with a varying degree of allowed deviations from the observed diet. I represents the limit for food item changes, while G indicates the limit for food group changes.

[3] Runtime is displayed in h:mm:ss. The limit for the model's runtime was set to 1 hour.

[4] Note that Conway's equation could not be applied for the continuous diet model, as it requires meals as inputs instead of complete diets.

[5] For the diet plans obtained by the PLA model, absorbable iron and zinc intake estimates based on the original absorption equations were calculated retrospectively.

## 3.4 Discussion

In this study, we demonstrated the implementation of NLP and PLA for both a mixed-integer and a continuous diet model, where nonheme iron and zinc absorption were modeled using Armah's, Conway's, and Miller's equations. PLA always provided accurate solutions, outperforming NLP in consistency and model runtime, in particular for the mixed-integer diet model. These results are in line with existing literature indicating that PLA can offer advantages over NLP in terms of speed and solution quality (Geißler et al., 2011).

However, the advantage of PLA over NLP may prove less pronounced in other diet modeling contexts. Solver performance can be greatly affected by problem size, such as the number of variables or how much constraints narrow the search space (Lee & Leyffer, 2011). Although we examined several scenarios with different problem sizes, results may differ in other contexts. As such, the current results cannot be generalized to all use cases, but they serve as an example that NLP does not always provide the best observed

solution. The same holds true for PLA (Geißler et al., 2011), although this was not found in the present study.

The gap between PLA and NLP might also have appeared narrower if alternative nonlinear solvers had been employed. Our evaluation was restricted to solvers LINDO and SCIP, which were freely available with our GAMS license. Nonetheless, experiments have demonstrated significant variations in performance among solvers (Kronqvist et al., 2019). Therefore, considering alternative solvers might prove beneficial, though challenges may arise regarding costs and ease of implementation. Not all solvers are freely available or compatible with every interface, which may require learning a new modeling language. In any case, testing multiple solvers and settings is advisable when applying NLP to evaluate model outcomes (McCullough & Vinod, 2003).

Furthermore, the difference in performance between PLA and NLP may not be practically meaningful. For instance for the mixed-integer diet model, the absorbable zinc content for NLP was at most 0.2 mg lower than for PLA, which may be negligible for some applications. Similarly, while computational efficiency is valuable when producing thousands of menu plans, a one-hour runtime may be acceptable for creating just one.



Another aspect to consider with PLA is the function that needs to be linearized. Armah's and Conway's nonheme iron absorption equations could be approximated with univariate PLA, and Miller's equation with bivariate PLA (Appendix 3C.1). However, more complicated functions, such as the Hallberg & Hulthén (2000) equation for nonheme iron absorption or the more recent equation by Miller et al. (2013) for zinc absorption, require a multivariate PLA approach, which exceeds the complexity of univariate and bivariate cases (Geißler et al., 2011). Future research could explore whether the advantages observed for PLA in this study extend to approximating more complicated absorption functions.

As a sidenote, it is important to realize that when modeling with single meal absorption equations and pre-existing meals, there is no need to solve the model using NLP or PLA. Nutrient absorption can be calculated for each meal beforehand, thus enabling mixed-integer linear programming, as demonstrated in a previous publication (van Wonderen et al., 2023) using the iron absorption equations of Conway et al. (2007) and Hallberg & Hulthén (2000). However, for complete diet equations based on average or total dietary intake, such as Armah et al. (2013) and Miller et al. (2013) equations, or when not using pre-existing meals, the model cannot be solved linearly, requiring the use of NLP or PLA.

Lastly, on a more general note, the diet models in this study were designed to compare the performance of NLP and PLA in a theoretical exercise. Therefore, we kept the models relatively simple. Before practical application, the models should be extended to consider a wider range of nutrients and additional factors such as sustainability, consumer acceptability, and affordability. Nevertheless, the findings of this study can inform

improvements to existing diet models. For instance, one of the food pattern models developed by the Dietary Guidelines Advisory Committee estimates the impact of reducing the intake of animal-based foods (USDA DGA, 2025). This methodology could be enhanced by incorporating diet-dependent absorption equations to more accurately estimate the intake of absorbable iron and zinc, which tend to be lower in diets with higher proportions of plant-based foods (Hunt, 2002). In this way, absorbable iron and zinc intake could be either maximized or constrained to meet a specified minimum level. For the performance of PLA, it does not matter whether iron and zinc are maximized or constrained. Given the absence of integer variables in this model, NLP is likely to perform well. However, for mixed-integer diet models, we recommend using either PLA or multiple NLP solvers to validate results.

In summary, we demonstrated the implementation of NLP and PLA in the context of modeling nonheme iron and zinc absorption for a mixed-integer and a continuous diet model. In this study, the main advantage of NLP was its ease in implementation, specifically when modeling multivariate nonlinear functions. When applying NLP, it is advisable to validate solution quality, for example, by testing different solvers. The main advantages of PLA were the quality and consistency of modeling outcomes, while requiring relatively low computational time. To conclude, this study provides practical examples for researchers aiming to improve diet model accuracy by implementation of nonheme iron and zinc absorption equations using either NLP or PLA.

# 3.A Appendix - Nutrient absorption equations

### 3.A.1 Iron

The formula to calculate the total absorbable iron content is given by Equation 3A.1. Heme iron absorption was assumed to be 25% (Armah et al., 2015b).

$$\begin{aligned}&\text{Total absorbable iron (mg)} = \\ &\quad \text{Nonheme iron absorption (\%)} \times \text{Nonheme iron content (mg)} + \\ &\quad \text{Heme iron absorption (\%)} \times \text{Heme iron content (mg)}\end{aligned} \tag{3A.1}$$

Armah et al.'s (2013) (**Armah**) and Conway, Powell & Geissler's (2007) (**Conway**) equations to estimate nonheme iron absorption are given by Equations 3A.2-3A.3 and 3A.4-3A.5 respectively. For Armah's complete diet equation, iron absorption is calculated based on average intake per day. For Conway's single meal equation, iron absorption is calculated for each single meal. We assume the serum ferritin concentration to be 30 $\mu$g/L (EFSA, 2015) in this research.



*Armah's equation*

$$\begin{aligned}&\text{Ln nonheme iron absorption (\%)} = \\ &\quad 6.294 - 0.709 \cdot \ln(\text{SF}) + 0.119 \cdot \ln(\text{VitC}) \\ &\quad + 0.006 \cdot \ln(\text{MFP} + 0.1) - 0.055 \cdot \ln(\text{Tea} + 0.1) \\ &\quad - 0.247 \cdot \ln(\text{Phytate}) - 0.137 \cdot \ln(\text{Calcium}) - 0.083 \cdot \ln(\text{nhFe})\end{aligned} \tag{3A.2}$$

$$\text{Nonheme iron absorption (\%)} = e^{\ln(\text{Nonheme iron absorption})} \tag{3A.3}$$

Where SF is serum ferritin ($\mu$/L); VitC is vitamin C (mg); MFP is meat, fish, and poultry (g); Tea is number of cups; Phytate and Calcium are in mg; and nhFe is nonheme iron (mg). The additional terms of +0.1 for MFP and Tea are included to prevent taking the logarithm of zero ($\ln(0) = -\infty$). However, in our research, we model all dietary factors using $\ln(\max\{\text{dietary factor}, 1\})$. This choice reflects our belief that the presence of small amounts of MFP should not negatively affect absorption. Similarly, we argue that minor quantities of Tea should not enhance absorption. For example, $\ln(0.1) = -2.3$, which in terms of magnitude is comparable to $\ln(10) = 2.3$. Using $\ln(\max\{\text{dietary factor}, 1\})$ ensures that the impact of a nutrient on absorption is minimal or nonexistent when it is scarce or absent.

*Conway's equation*

$$\begin{aligned}\text{Log nonheme iron absorption} = \\ 0.9385 + 0.001 \cdot \text{Fruit} + 0.0009 \cdot \text{MFP} - 0.0052 \cdot \text{Beans} - 0.0016 \cdot \text{Cereal} \\ - 0.002 \cdot \text{Tea} - 0.0012 \cdot \text{Dairy} - 0.0041 \cdot \text{Cheese} - 0.0032 \cdot \text{Eggs} \\ - 0.0062 \cdot \text{Soy} - 0.0087 \cdot \text{Nuts}\end{aligned} \tag{3A.4}$$

$$\text{Nonheme iron absorption (\%)} = 10^{\log(\text{Nonheme iron absorption})} \cdot \left(\frac{21.7}{\text{SF}}\right)^{0.94049} \tag{3A.5}$$

Where Fruit (g) refers to high vitamin C (>30 mg/100g) fruit or juice; MFP (g) is meat, fish, and poultry; Beans (g) are beans and lentils, excluding soybeans; Cereal (g) is wholegrain cereals; Tea (g) is black tea infusion; Dairy (g) includes all dairy products; Cheese (g) excludes cream and cottage cheese; Soy (g) is soybeans and flour porridge; Eggs (g) and Nuts (g) refer to their respective food items.

### 3.A.2 Zinc

Miller, Krebs & Hambidge's (2007) (**Miller**) equation to estimate absorbable zinc intake is given by Equation 3A.6. We used the updated coefficients reported by Hambidge et al. (2010).

*Miller's equation*

$$\begin{aligned}\text{Zn}_A = 0.5 \cdot \Bigg\{ & 0.091 + \text{Zn} + 0.033 \left(1 + \frac{\text{Phy}}{0.68}\right) \\ & - \sqrt{\left(0.091 + \text{Zn} + 0.033 \left(1 + \frac{\text{Phy}}{0.68}\right)\right)^2 - 4 \cdot 0.091 \cdot \text{Zn}} \Bigg\}\end{aligned} \tag{3A.6}$$

Where $\text{Zn}_A$ is absorbable zinc (mmol), Zn is total dietary zinc (mmol), and Phy is total dietary phytate (mmol). Molar masses of 65.4 and 660 g/mol were used for zinc and phytate, respectively.

## 3.B Appendix - Nonlinear programming

### 3.B.1 Theory

Nonlinear programming (NLP) is a branch of mathematical optimization dealing with optimization problems that involve nonlinear objective functions, nonlinear constraints, or both (Claassen et al., 2007). One way to classify the algorithms used for solving nonlinear problems is to divide them into two categories: local solvers and global solvers.

Local solvers aim at finding a solution that is optimal within a local neighborhood of the search space. For a maximization problem, local solvers employ an algorithmic strategy that leads them uphill. To prevent the algorithm from halting at the first peak encountered, heuristics are applied to explore neighboring regions in search of other maxima, thereby lowering the risk of returning a local maximum that is not globally optimal (Hoos & Tsang, 2006). The risk of obtaining a local optimum can further be reduced by employing a multistart approach. This technique lowers the likelihood of being trapped in a local optimum by initiating the algorithm from multiple starting points (György & Kocsis, 2011). Additionally, providing the algorithm with an initial solution that is close to a global optimum can improve its effectiveness in identifying a global optimum. Conversely, an initial solution far from a global optimum may reduce model performance (Lourenço et al., 2003). It is also advisable to experiment with default solver settings and use multiple algorithms for solving the same problem to verify model outcomes (McCullough & Vinod, 2003). Nevertheless, one needs to remain careful when interpreting results from local solvers, as their solutions are only optimal under certain conditions.

Global solvers, on the other hand, aim to find a global optimum across the entire search space within termination tolerances. By partitioning the search space, they carry out a systematic exploration, resulting in the establishment of upper and lower bounds for the globally optimal value (Burer & Letchford, 2012). Unlike local solvers, global solvers do not require multistart techniques since they encompass the entire search space. Nonetheless, offering the algorithm a strong starting point can significantly reduce runtime (Kazimipour et al., 2013).

The advantage of local solvers lies in their speed, as they focus on specific areas within the solution space, though at the expense of the risk of finding solutions that are only locally optimal. Conversely, global solvers have the capability to provide a solution while also indicating how close that solution is to the true optimal objective value, but may require longer computational time, especially for large search spaces (Ploskas & Sahinidis, 2022).

#### *3B.1.1 Mixed-integer nonlinear programming*

In addition to nonlinear functions, the presence of integer variables adds another layer of difficulty. In diet modeling, integer variables can be used to model the selection of food items or meals. This allows for constraints, such as limiting the number of times a meal appears in a menu plan or ensuring that food item quantities are either zero or above a specified threshold. Both nonlinear and mixed-integer linear problems are NP-hard, meaning their computational complexity grows exponentially with problem size, potentially rendering them unsolvable in practice. The combination of nonlinearity and integrality, known as mixed-integer nonlinear programming (MINLP), makes these problems even more challenging to solve. To address this, a distinct class of solvers has been developed (Lee & Leyffer, 2011). In this study, both NLP and MINLP are referred to as NLP for simplicity.

### 3.B.2 Evaluation of performance

To validate NLP model outcomes, we tested two different nonlinear solvers: local solver Linear Interactive and Discrete Optimizer (LINDO) (GAMS, 2024b) and global solver Solving Constraint Integer Programs (SCIP) (Vigerske & Gleixner, 2018). In our evaluation of LINDO, we explored the efficiency of both the multistart and initialization functionalities. For SCIP, we focused solely on the impact of providing the models with an initial solution. For conducting the multistart procedure in parallel, we adhered to LINDO's default settings. This procedure initiates the local solver from multiple intelligently generated starting points, for which the default number of multistarts is established at 5, as experiments have shown this number to be sufficient for most models (GAMS, 2024b). As for providing the model with an initial solution, we used a solution where total iron intake was optimized instead of absorbable iron intake.

# 3.C Appendix - Piecewise linear approximation

## 3.C.1 Theory

PLA involves approximating nonlinear functions with line segments for univariate functions, triangles for bivariate functions, and n-simplices for multivariate functions. A function is classified as univariate, bivariate, or multivariate based on whether it has one, two, or more dependent variables, respectively. Note that a function with multiple inputs can sometimes be treated as a univariate function if its nonlinear components can be separated linearly (Geißler et al., 2011). For example, the bivariate function $\log(a \cdot b)$ can be rewritten as the sum of two univariate terms: $\log(a) + \log(b)$.

Compared to NLP, PLA may offer advantages such as speed, robustness, scalability, and the ability to find a global optimum, under the appropriate conditions. PLA is typically effective where the number of nonlinear functions and their dependent variables is limited, making accuracy in the approximation the main challenge. Insufficient segmentation may result in the loss of crucial nonlinear characteristics, thereby reducing the accuracy of model outcomes. Conversely, an excessive number of segments can substantially prolong model runtime (Geißler et al., 2011).



Two commonly employed approaches for approximating nonlinear functions are the $\delta$-formulation and $\lambda$-formulation (Claassen et al., 2007), also known as the incremental method and convex-combination method (Babayev, 1997; Geißler et al., 2011), respectively. In the subsequent sections, we present two concise numerical examples to illustrate the functionality of the $\delta$-formulation for univariate cases and the $\lambda$-formulation for bivariate cases.

### *3C.1.1 Univariate approach*

The $\delta$-formulation is outlined in Table 3C.1 and will be clarified through the example depicted in Figure 3C.1. The red line in Figure 3C.1 represents the nonlinear function $f(X)$, which is linearly approximated by three line segments $\varphi(X)$. The lengths $d_s$ and slopes $a_s$ are provided for each line segment $s$. Furthermore, the coordinates of the left-most point are $(c^x, c^y)$.

$X = 3$ can be reached from $c^x$ by traversing the first line segment completely: $X = c^x + d_1 = 1 + 2 = 3$ (Figure 3C.1A). To compute $\varphi(X)$ for $X = 3$, we traverse the first line segment as well: $\varphi(3) = c^y + a_1 \cdot d_1 = 3 + 2 \cdot 2 = 7$. To calculate $\varphi(X)$ for a point positioned between two breakpoints, we introduce $\delta_s$. This $\delta_s$ denotes whether $X$ has traversed ($\delta_s = 1$), lies on ($0 < \delta_s < 1$), or has not encountered ($\delta_s = 0$) a given line segment. For instance, at $X = 4.5$, we traverse the first line segment ($\delta_1 = 1$) and halt midway through the second line segment ($\delta_2 = 0.5$) (Figure 3C.1B). These $\delta$ values are confirmed by Constraint 3C.1 (Table 3C.1): $X = c^x + \sum_s d_s\delta_s = c^x + d_1\delta_1 + d_2\delta_2 + d_3\delta_3 =$

$1 + 2 \cdot 1 + 3 \cdot 0.5 + 4 \cdot 0 = 4.5$. Next, applying Constraint 3C.2, we compute $\varphi(X)$ as: $\varphi(4.5) = c^y + \sum_s a_s d_s \delta_s = c^y + a_1 d_1 \delta_1 + a_2 d_2$.

However, an alternative $\delta$ vector may also result in $X = 4.5$, yet produce a different outcome than $\varphi(4.5) = 8$. Consider the $\delta$ vector [0 0.5 0.5], for instance. Although it satisfies Constraint 3C.1 and leads to $X = 4.5$, the corresponding outcome from Constraint 3C.2, $\varphi(4.5) = 2.5$, is incorrect. This occurs because, in that scenario, the path does not represent a direct journey from $c^x$ to $X = 4.5$, but rather follows a fragmented route (Figure 3C.1C). To ensure a continuous path beginning at the start of the function, we introduce the binary variable $Y_s$, which indicates whether a line segment $s$ has been fully traversed ($Y_s = 1$) or not ($Y_s = 0$). For example, when $\delta_2 = 0.5$, then $Y_1 = 1$ (Constraint 3C.4) and $\delta_1 = 1$ (Constraint 3C.3), indicating that the first line segment has been fully covered. Additionally, when $\delta_2 = 0.5$, then $Y_2 = 0$ (Constraint 3C.3) and $\delta_3 = 0$ (Constraint 3C.4), showing that the second segment is only partially traversed and the third is not encountered at all. Note that these constraints are only required when solving nonconvex or nonconcave optimization problems, as is the case in this study. For more information on convexity and concavity, we refer to Dahl (2010).

### *3C.1.2 Bivariate approach*

The model formulation of the bivariate $\lambda$-formulation is given in Table 3C.2. The nonlinear function is represented by a grid (Figure 3C.2A), which is subdivided into triangles (Figure 3C.2B). Each grid point $(i, j)$ has a corresponding top-left triangle $T_{ij}$ and bottom-right triangle $B_{ij}$. The point $(X, Y)$, which is to be approximated (red dot), lies within one of these triangles (Constraint 3C.8) and is estimated using the weighted sum of the coordinates of the triangle's corner points (black dots). Only the weights $\lambda_{ij}$ of these corner points can be greater than or equal to 0, with all other $\lambda$-values set to 0 (Constraint 3C.9). This is demonstrated in Figure 3C.2C, in which point $(X, Y)$ lies in triangle $B_{22}$ ($B_{22} = 1$, all other $B_{ij}$ and $T_{ij}$ are 0). Substituting $B_{22}$ into Constraint 3C.9 permits only $\lambda_{22}$, $\lambda_{32}$, and $\lambda_{21}$ to be greater than 0, which are the corner points of triangle $B_{22}$. Constraint 3C.10 ensures that the sum of the $\lambda$-values equals 1. The $\lambda$-values are then calculated by expressing $X$ and $Y$ as weighted sums of the $x$ and $y$ coordinates of the triangle's corner points (Constraints 4–5, Figure 3C.2D). Finally, $\varphi(X, Y)$ can be determined (Constraint 3C.13).

**Table 3C.1:** $\delta$-formulation for univariate piecewise linear approximation.

| Notation | |
|---|---|
| Sets | |
| $s$ | Line segments for piecewise linear approximation |
| Parameters | |
| $c^x$ | $X$ value of the left-most point of the first line segment |
| $c^y$ | $\varphi(X)$ value of the left-most point of the first line segment |
| $a_s$ | Slope of line segment $s$ |
| $d_s$ | Length of line segment $s$ |
| Variables | |
| $X$ | Variable $X$ |
| $\varphi(X)$ | Piecewise linear approximation of nonlinear function $f(X)$ |
| $\delta_s$ | Delta for line segment $s$ |
| $Y_s$ | Binary variable to indicate whether $\delta_s$ equals 1 or not (yes: 1; no: 0) |

Constraints

$$X = c^x + \sum_s d_s \delta_s \tag{3C.1}$$

$$\varphi(X) = c^y + \sum_s a_s d_s \delta_s \tag{3C.2}$$

$$Y_s \leq \delta_s \qquad \forall\ s \tag{3C.3}$$

$$Y_s \geq \delta_{s+1} \qquad \forall\ s < S \tag{3C.4}$$

$$\delta_s \in [0, 1] \qquad \forall\ s \tag{3C.5}$$

$$Y_s \in \{0, 1\} \qquad \forall\ s \tag{3C.6}$$

$$X,\ \varphi(X) \in \mathbb{R} \tag{3C.7}$$

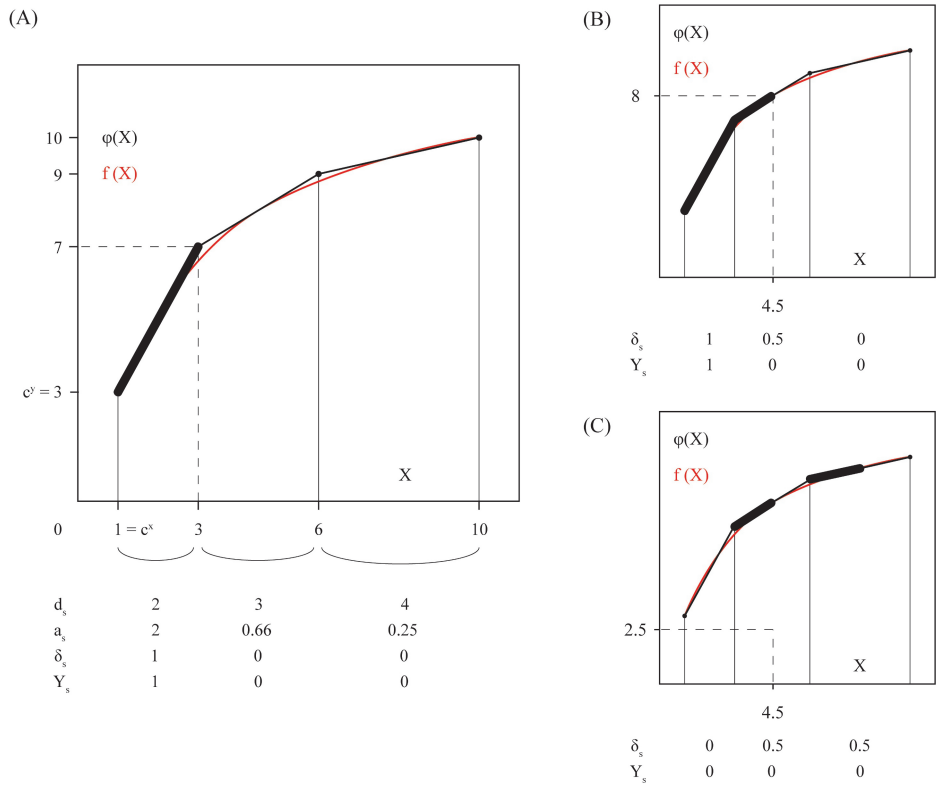



**Figure 3C.1:** Illustrative example of $\delta$-formulation for univariate piecewise linear approximation. The nonlinear function $f(X)$ is approximated by the linear line segments of the function $\varphi(X)$, where $d_s$ and $a_s$ represent the length and slope of line segment $s$, respectively. Furthermore, the coordinates of the left-most point are $(c^x, c^y)$. To accurately determine $\varphi(X)$ for a given $X$, the line segments must be traversed sequentially from the start of the first line segment up to $X$, as illustrated in Figures A and B. This requirement is enforced by the variables $\delta_s$ and $Y_s$, and the constraints provided in Table 3C.1. Omitting these constraints can lead to incorrect results, as shown in Figure C.

**Table 3C.2:** $\lambda$-formulation for bivariate piecewise linear approximation.

| Notation | |
|---|---|
| Sets | |
| $i$ | Vertical gridlines |
| $j$ | Horizontal gridlines |
| Parameters | |
| $x_i$ | X coordinate of gridline $i$ |
| $y_j$ | Y coordinate of gridline $j$ |
| $f_{ij}$ | $f(X,Y)$ value of grid point with coordinates $(i,j)$ |
| Variables | |
| $\varphi(X,Y)$ | Piecewise linear approximation of nonlinear function $f(X,Y)$ |
| $T_{ij}$ | 1 if point $(X,Y)$ is in the top left triangle of grid point $(i,j)$, 0 otherwise |
| $B_{ij}$ | 1 if point $(X,Y)$ is in the bottom right triangle of grid point $(i,j)$, 0 otherwise |
| $\lambda_{ij}$ | Weight of grid point $(i,j)$ in computing $\varphi(X,Y)$ |

Constraints

$$\sum_i \sum_j (T_{ij} + B_{ij}) = 1 \tag{3C.8}$$

$$\lambda_{ij} \leq B_{ij} + B_{i-1,j} + B_{i,j+1} + T_{ij} + T_{i+1,j} + T_{i,j-1} \qquad \forall i,j \tag{3C.9}$$

$$\sum_i \sum_j \lambda_{ij} = 1 \tag{3C.10}$$

$$X = \sum_i \sum_j x_i \lambda_{ij} \tag{3C.11}$$

$$Y = \sum_i \sum_j y_j \lambda_{ij} \tag{3C.12}$$

$$\varphi(X,Y) = \sum_i \sum_j f_{ij} \lambda_{ij} \tag{3C.13}$$

$$\lambda_{ij} \in [0,1] \qquad \forall i,j \tag{3C.14}$$

$$T_{ij} \in \{0,1\} \qquad \forall i = 2 \dots I, \quad j = 1 \dots J-1 \tag{3C.15}$$

$$B_{ij} \in \{0,1\} \qquad \forall i = 1 \dots I-1, \quad j = 2 \dots J \tag{3C.16}$$

$$X,\ Y,\ \varphi(X,Y) \in \mathbb{R} \tag{3C.17}$$

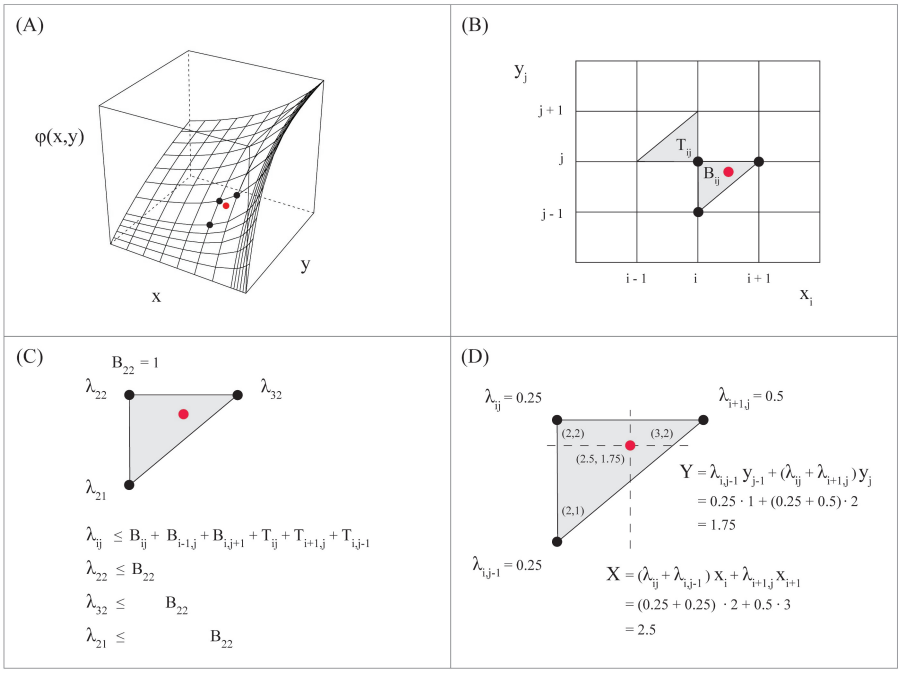



**Figure 3C.2:** Illustrative example of λ-formulation for bivariate piecewise linear approximation. The nonlinear function $f(X, Y)$ is approximated by a grid $\varphi(X, Y)$, where a specific point on the grid (red dot) is calculated as the weighted sum of the three nearest grid points (black dots), as shown in Figure A. To determine weights $\lambda_{ij}$, the three closest grid points must first be identified by locating the triangle (top-left $T_{ij}$ or bottom-right $B_{ij}$) in which the point resides (Figure B). The constraints in Table 3C.2 ensure that these grid points can have weights greater than 0 (Figure C). The exact computation of these weights is detailed in Figure D.

### 3.C.2 Parameters

How parameter values were obtained to linearize the nonlinear absorption equations is explained in the next two subsections.

#### *3C.2.1 Iron*

For Armah's and Conway's nonheme iron absorption equations, univariate PLA can be applied since the input variables are linearly separable (Constraints 3C.2, 3C.4). In Armah's equation, this required linearization of the univariate function $ln(X)$ for various dietary factors, whereas Conway's equation required linearizing the function $\log_{10}(X)$.

First, we determined the maximum range of all dietary factors $X$. Then, points were generated across this range using the nonlinear functions. We used the R package Segmented (Muggeo, 2003) to fit line segments to the generated data, with breakpoints manually determined based on visual observation. This method provided all necessary parameters for the PLA diet models (Tables 3C.3 and3C.4). The visualizations of the linear approximations are included in Figures 3C.3 and 3C.4. Note that these parameters are valid only when $X \geq 1$. To account for values between 0 and 1, additional line segments must be introduced.



#### *3C.2.2 Zinc*

Miller's equation is a bivariate function that utilizes dietary zinc and phytate intake as inputs. The grid size, employed to approximate this nonlinear function, was determined based on the maximum range of zinc and phytate intake. Different numbers and sizes of grid cells were evaluated for their accuracy in estimating absorbable zinc (Table 3C.5, Figure 3C.5). After testing several options with a maximum deviation below 0.1 mg, we chose the 15 by 15 for its relatively fast runtime. Half of the grid cells were allocated to the region covering the $80^{\text{th}}$ percentile of zinc and phytate intake (Figure 3C.6). This higher density in this area ensures a more accurate approximation of absorbable zinc intake. Table 3C.6 contains the parameter values used in the zinc PLA diet model, while Figure 3C.7 provides the visual representation.

**Table 3C.3:** Parameters for univariate piecewise linear approximation of the natural logarithm to linearize Armah's iron absorption equation.[1]

| $c^x$ | $c^y$ | $s$ | $d_s$ | $a_s$ |
|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 0.015244 |
| | | 2 | 1 | 0.719470 |
| | | 3 | 2 | 0.337060 |
| | | 4 | 2 | 0.199350 |
| | | 5 | 4 | 0.130400 |
| | | 6 | 10 | 0.073503 |
| | | 7 | 30 | 0.029894 |
| | | 8 | 50 | 0.013755 |
| | | 9 | 100 | 0.006759 |
| | | 10 | 98 | 0.003863 |
| | | 11 | 102 | 0.002914 |
| | | 12 | 200 | 0.002087 |
| | | 13 | 400 | 0.001269 |
| | | 14 | 500 | 0.000793 |
| | | 15 | 500 | 0.000568 |

**Table 3C.4:** Parameters for univariate piecewise linear approximation of the base 10 logarithm to linearize Conway's iron absorption equation.[1]

| $c^x$ | $c^y$ | $s$ | $d_s$ | $a_s$ |
|---|---|---|---|---|
| 0 | -2.301 | 1 | 0.005 | 0 |
| | | 2 | 0.005 | 75.353 |
| | | 3 | 0.04 | 16.847 |
| | | 4 | 0.05 | 5.6668 |
| | | 5 | 0.15 | 2.6361 |
| | | 6 | 0.25 | 1.1462 |
| | | 7 | 0.25 | 0.6759 |
| | | 8 | 0.25 | 0.5067 |
| | | 9 | 1 | 0.3130 |
| | | 10 | 1 | 0.1610 |
| | | 11 | 1 | 0.1274 |
| | | 12 | 2 | 0.0905 |
| | | 13 | 4 | 0.0552 |
| | | 14 | 5 | 0.0343 |
| | | 15 | 5 | 0.0250 |
| | | 16 | 8 | 0.0183 |
| | | 17 | 12 | 0.0129 |
| | | 18 | 20 | 0.0090 |
| | | 19 | 40 | 0.0056 |

[1] Parameters: $c^x$ and $c^y$, x and y values of the left-most point of the first line segment; s, index for line segments for piecewise linear approximation; $d_s$, length of line segment s; $a_s$, slope of line segment s.

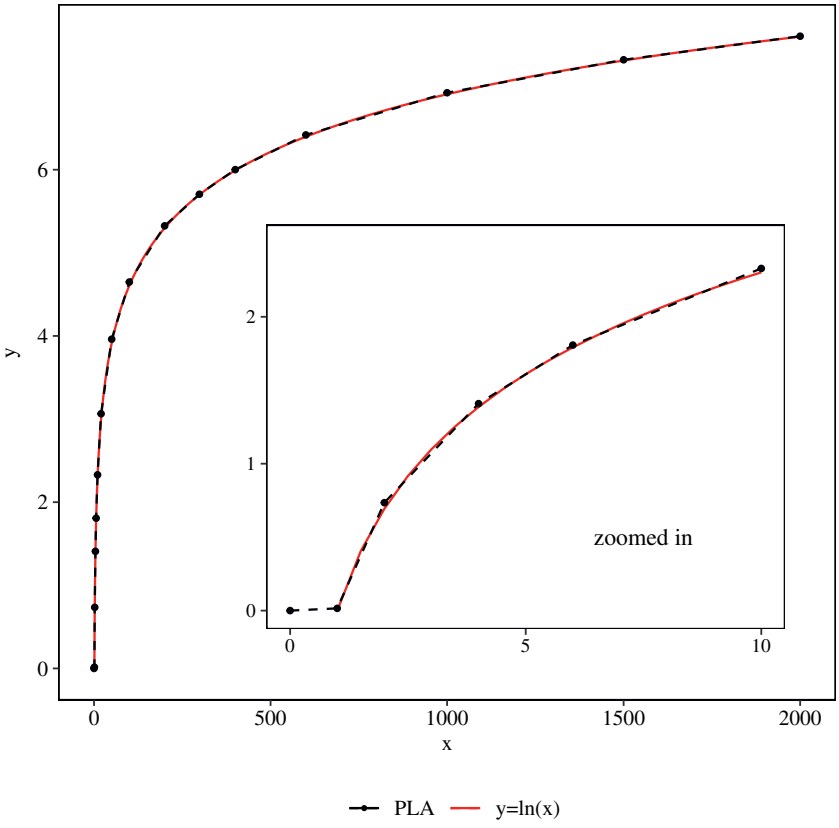



**Figure 3C.3:** Piecewise linear approximation of the natural logarithm to linearize Armah's iron absorption equation. $X$ represents the dietary factors in Armah's equation (Equations 3A.2 and 3A.3), such as vitamin C (mg), calcium (mg), nonheme iron absorption (%), and absorbable nonheme iron (mg), where $Y = ln(X)$.

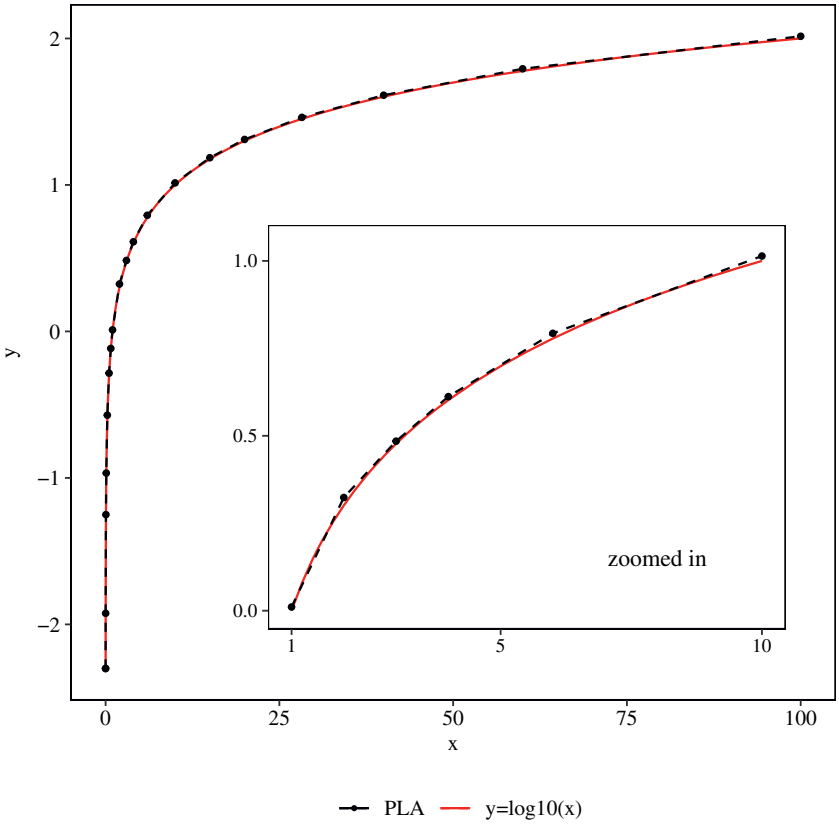



**Figure 3C.4:** Piecewise linear approximation of the base 10 logarithm to linearize Conway's iron absorption equation. $X$ represents nonheme iron absorption (%) (Equations 3A.4 and 3A.5) and absorbable nonheme iron (mg), where $Y = log_{10}(X)$.

**Table 3C.5:** Accuracy of bivariate piecewise linear approximation to linearize Miller's absorbable zinc intake equation for different parameter settings.

| Grid | Equal grid cells | Percentile[1] | Coverage (%)[2] | MAE (mg)[3] | Max absolute error (mg) |
|---|---|---|---|---|---|
| 6x6 | yes | 100 | 100 | 0.08 | 0.88 |
| | no[4] | 70 | 40 | 0.07 | 0.24 |
| 8x8 | yes | 100 | 100 | 0.05 | 0.70 |
| | no | 70 | 40 | 0.04 | 0.16 |
| 10x10 | yes | 100 | 100 | 0.03 | 0.57 |
| | no | 80 | 60 | 0.03 | 0.11 |
| 12x12 | yes | 100 | 100 | 0.02 | 0.46 |
| | no | 65 | 40 | 0.02 | 0.09 |
| 15x15 | yes | 100 | 100 | 0.01 | 0.35 |
| | no | 80 | 50 | 0.01 | 0.06 |
| 20x20 | yes | 100 | 100 | 0.01 | 0.23 |
| | no | 80 | 50 | 0.01 | 0.04 |
| 25x25 | yes | 100 | 100 | 0.01 | 0.16 |
| | no | 75 | 40 | 0.00 | 0.03 |
| 30x30 | yes | 100 | 100 | 0.00 | 0.12 |
| | no | 70 | 40 | 0.00 | 0.02 |

[1] Percentile of zinc and phytate intake.
[2] Percentage of grid cells used to cover the percentile of zinc and phytate intake.
[3] Mean absolute error (MAE) of absorbable zinc intake (mg) between nonlinear and approximated functions.
[4] Results are only displayed for the best-performing combination of parameters percentile and coverage.

**Table 3C.6:** Parameters for bivariate piecewise linear approximation to linearize Miller's absorbable zinc intake equation.[1]

| $z$ | $p$ | $x_z$ | $y_p$ | $f_{zp}$ | $z$ | $p$ | $x_z$ | $y_p$ | $f_{zp}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.0000 | 0.0000 | 0.0000 | 11 | 1 | 0.5198 | 0.0000 | 5.5322 |
| 2 | 1 | 0.0313 | 0.0000 | 1.3878 | 12 | 1 | 0.6547 | 0.0000 | 5.6251 |
| 3 | 1 | 0.0625 | 0.0000 | 2.5114 | 13 | 1 | 0.7895 | 0.0000 | 5.6846 |
| 4 | 1 | 0.0938 | 0.0000 | 3.3504 | 14 | 1 | 0.9244 | 0.0000 | 5.7258 |
| 5 | 1 | 0.1250 | 0.0000 | 3.9417 | 15 | 1 | 1.0593 | 0.0000 | 5.7559 |
| 6 | 1 | 0.1563 | 0.0000 | 4.3511 | 1 | 2 | 1.1942 | 0.0000 | 5.7787 |
| 7 | 1 | 0.1875 | 0.0000 | 4.6382 | 2 | 2 | 0.0000 | 0.2570 | 0.0000 |
| 8 | 1 | 0.2188 | 0.0000 | 4.8461 | ... | ... | ... | ... | ... |
| 9 | 1 | 0.2500 | 0.0000 | 5.0005 | 14 | 15 | 1.0593 | 17.5007 | 3.1778 |
| 10 | 1 | 0.3849 | 0.0000 | 5.3667 | 15 | 15 | 1.1942 | 17.5007 | 3.3583 |

[1] Parameters: $z$ and $p$ are the indices for the vertical and horizontal gridlines; $x_z$ and $y_p$ are the coordinates for total dietary zinc and phytate intake (mmol); $f_{zp}$ is the absorbable zinc intake (mg). For this 15 by 15 grid, 50% of the grid cells encompassed the $80^{th}$ percentile of zinc and phytate intake.

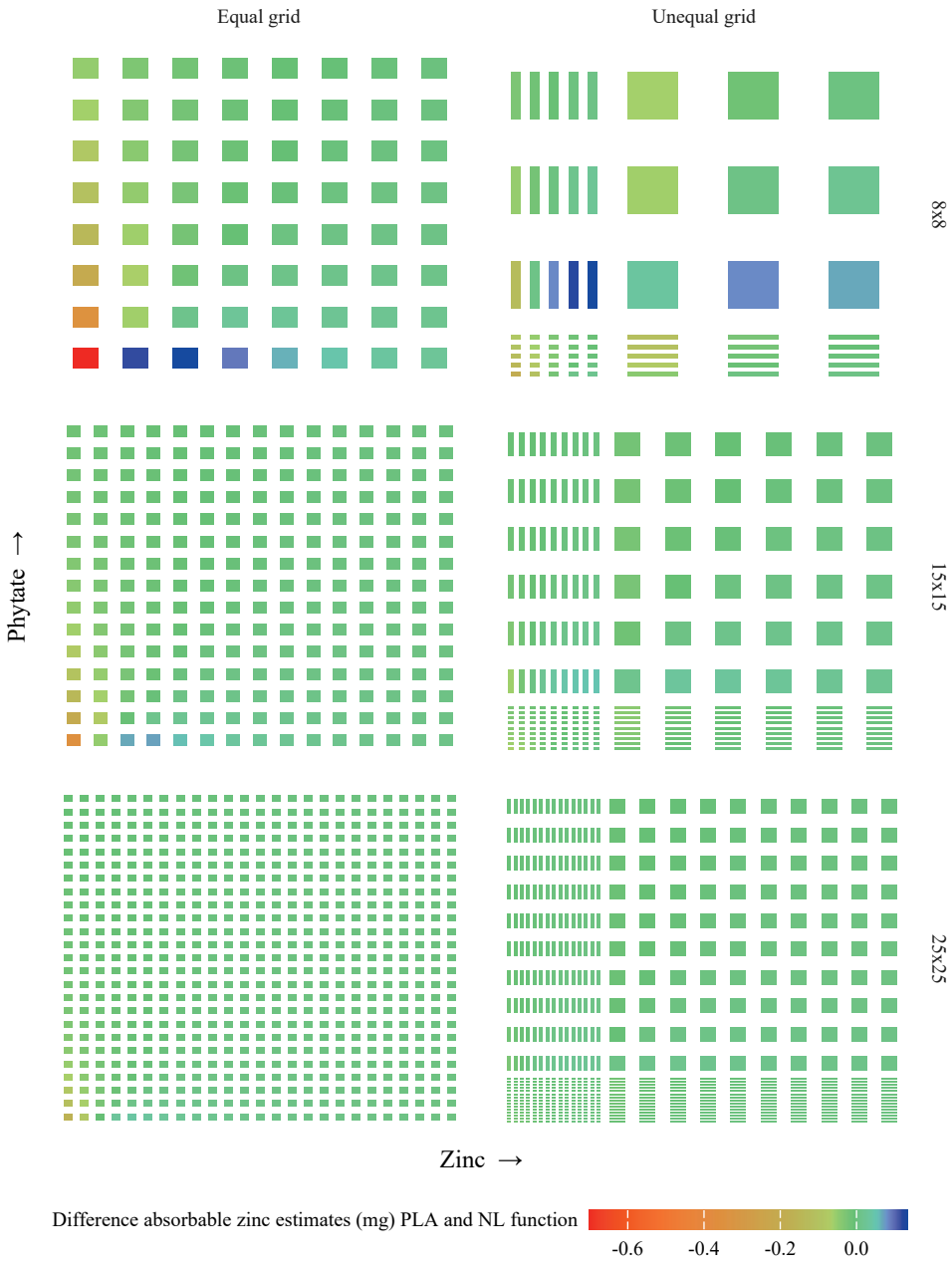



**Figure 3C.5:** Accuracy of the bivariate piecewise linear approximation to linearize Miller's absorbable zinc intake equation for different grids. For grids with unequal cell sizes, 50% of the grid cells encompass the $80^{\text{th}}$ percentile of zinc and phytate intake.

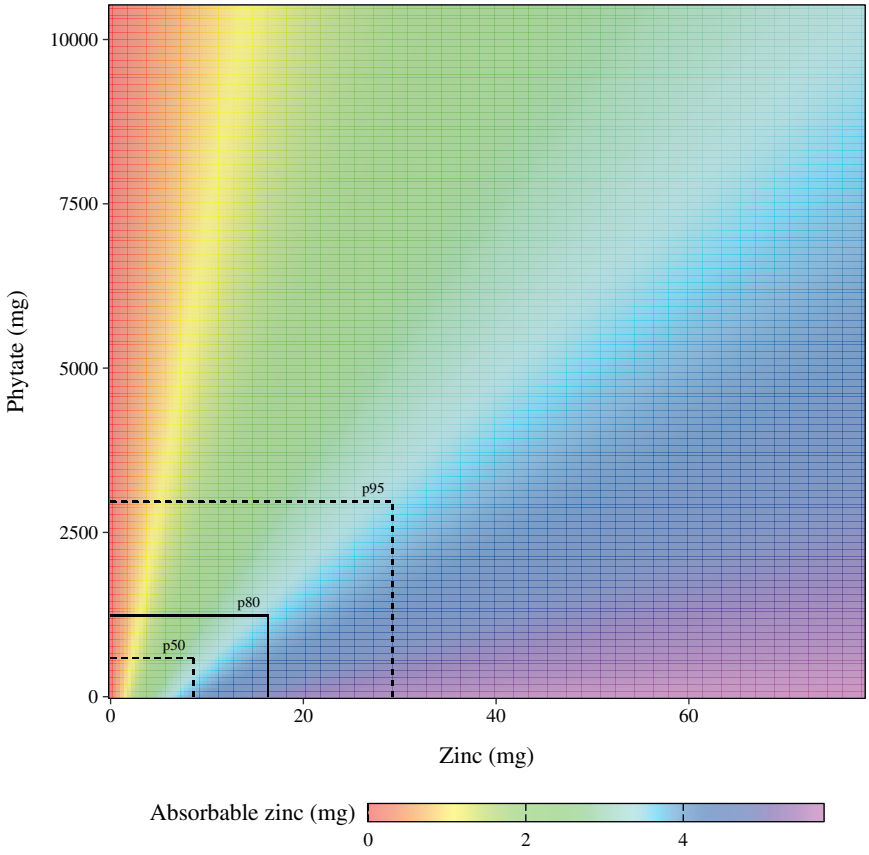



**Figure 3C.6:** 2D plot of absorbable zinc intake as a function of total dietary zinc and phytate intake according to Miller's equation, displaying the estimated $50^{\text{th}}$, $80^{\text{th}}$, and $95^{\text{th}}$ percentiles of dietary intake.

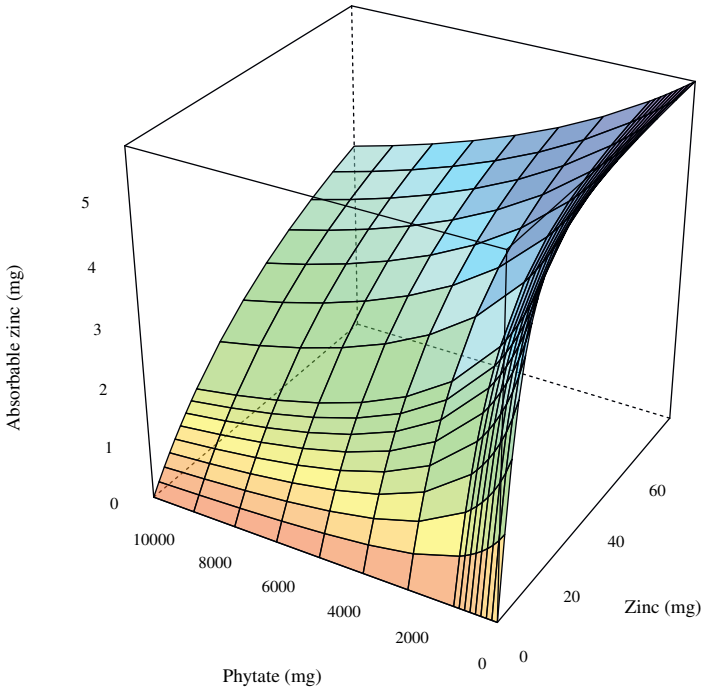



**Figure 3C.7:** 3D plot of absorbable zinc intake as a function of total dietary zinc and phytate intake according to Miller's equation.

### 3.C.3 Evaluation of accuracy

We assessed the accuracy of PLA for both Conway, Armah, and Miller by comparing the absorbable nonheme iron and zinc estimates obtained from the nonlinear and linearized functions. For the single meal equation of Conway, the absorbable nonheme iron content was computed for each meal or snack with a nonheme iron content greater than 0. For Armah's and Miller's equations, a different approach was required, as absorption can only be calculated for complete diets, not individual meals. Therefore, we simulated 1,000 two-week menu plans by randomly selecting breakfast, lunch, dinner, and snacks from the meal dataset. From this dataset, only meals with a ranking among the top 1000 in absorbable nonheme iron or zinc content were selected. The nonheme iron and zinc estimates of the meals and diets, calculated using both nonlinear and linearized functions, were rounded to one decimal place to align with the precision of the dietary intake data (CDC & NCHS, 2018). Results are displayed in Section 3E.1.

# 3.D Appendix - Diet modeling

## 3.D.1 Notation

| Indices | |
|---|---|
| i | Food item i |
| g | Food group g |
| m | Meal moment m |
| n | Nutrient n |
| nmacro | Subset of macronutrients |
| nmicro | Subset of micronutrients |
| nabs | Nonheme iron, absorbable nonheme iron, and all nutrients that affect nonheme iron absorption. |
| p | Phytate coordinate of grid point |
| r | Meal r |
| s | Line segment s for piecewise linear approximation |
| z | Zinc coordinate of grid point |

| Parameters | |
|---|---|
| $\varepsilon$ | Arbitrary small value close to 0 and positive |
| $a_s$ | Slope of line segment s |
| $c^x$ | X value of the left-most point of the first line segment |
| $c^y$ | Y value of the left-most point of the first line segment |
| $d_s$ | Length of line segment s |
| $EAR_n$ | Estimated average requirement per day for each micronutrient n |
| eqSF | Logarithm (base 10) of factor to adjust nonheme iron absorption for serum ferritin concentration: $\log_{10}\left((21.7/SF)^{0.94049}\right)$ |
| $f_{zp}$ | Absorbable zinc intake corresponding to total zinc intake z and phytate intake p |
| $fDiversity_i$ | Number of times a food item i may be added to the menu plan |
| fChangeItem | Allowed food item change within food groups |
| fChangeGroup | Allowed food group change |
| $groupItems_{gi}$ | Boolean to indicate whether food item i belongs to food group g (yes:1; no:0) |
| $LB_n$ | Lower bound on intake level for each macronutrient n |
| ln100 | Natural logarithm of 100 |
| lnSF | Natural logarithm of serum ferritin concentration ($\mu$g/L) |
| log100 | Logarithm (base 10) of 100 |
| $mealFoodItems_{rmi}$ | Boolean to indicate whether food item i belongs to meal r with meal moment m (yes:1; no:0) |
| $mealType_{rm}$ | Boolean to indicate whether meal r belongs to meal moment m (yes:1; no:0) |
| $nContent_{rmn}$ | Nutrient n content for each meal r with meal moment m |
| $nContent_{in}$ | Nutrient n content for food item i |
| NDays | Number of days to cover by the menu plan |
| $obsIntake_i$ | Observed daily intake of food item i |
| $RDA_n$ | Recommended daily allowance for each micronutrient n |
| SF | Serum ferritin concentration ($\mu$g/L) |
| $UB_n$ | Upper bound on intake level for each macronutrient n |
| $UL_n$ | Tolerable upper intake level for each micronutrient n |

Parameters

| | |
|---|---|
| $x_z$ | X coordinate of gridline z (zinc) |
| $y_p$ | Y coordinate of gridline p (phytate) |

Variables

| | |
|---|---|
| $\delta_{(rm)ns}$ | Delta for line segment s for nutrient n (and meal r with meal moment m) |
| $\lambda_{zp}$ | Lambda of grid point with coordinates z and p |
| $absNhFe_{rm}$ | Estimated nonheme iron absorption for meal r with meal moment m |
| $absNhFeIntake_{rm}$ | Estimated absorbable nonheme iron intake for meal r with meal moment m |
| $avgIntake_n$ | Average nutrient n intake per day |
| $avgIntakeAdj_n$ | Average nutrient n intake per day adjusted for absorbable iron intake |
| $avgAbsNhFe$ | Average estimated nonheme iron absorption |
| $avgAbsZnIntake$ | Average estimated absorbable zinc intake |
| $B_{zp}$ | 1 if point (average zinc and phytate intake) is in the top left triangle of grid point (p,z), 0 otherwise |
| $belowRDA_n$ | Normalised deviation below the recommended daily allowance for micronutrient n |
| $lnAbsNhFe$ | Natural logarithm of average estimated nonheme iron absorption |
| $lnIntake_n$ | Natural logarithm of average nutrient n intake per day |
| $logAbsNhFe_{rm}$ | Logarithm (base 10) of nonheme iron absorption for meal r with meal moment m |
| $logIntake_{rmn}$ | Logarithm (base 10) of nutrient n intake for meals r with meal moment m |
| $maxBelowRDA$ | Maximal deviation below RDA |
| $nIntake_{rmn}$ | Nutrient n intake for total number of selected meals r with meal moment m |
| $nIntake_{in}$ | Nutrient n intake for food item i |
| $nIntakeAdj_{rmn}$ | Nutrient n intake for total number of selected meals r with meal moment m adjusted for absorbable iron intake |
| $nIntakeMeal_{rmn}$ | Nutrient n intake for single meal r with meal moment m |
| $Qplus_i$ | Increase of food item i compared to observed intake |
| $Qmin_i$ | Decrease of food item i compared to observed intake |
| $Qdiff_i$ | Absolute difference between observed and optimized intake for food item i |
| $T_{zp}$ | 1 if point (average zinc and phytate intake) is in the bottom right triangle of grid point (p,z), 0 otherwise |
| $X_{rm}$ | Number of times that meal r with meal moment m is added to the menu plan |
| $X_i$ | Quantity of food item i |
| $Y_{(rm)ns}$ | Binary variable to indicate whether $\delta_s$ equals 1 or not (yes:1; no:0) for nutrient n (and meal r with meal moment m) |
| $Z_{rm}$ | Binary variable to indicate whether meal r with meal moment m is added to the menu plan (yes:1; no:0) |

### 3.D.2 Base models and parameters

#### *3D.2.1 Mixed-integer diet model*

For each day of the two-week menu plan, one meal is selected for breakfast, one for lunch, one for dinner (B1) (Table 3D.1), and two for snacks (B2). To secure menu diversity, dinner and lunch meals can only be selected once per two weeks (B3), while breakfast meals and snacks can be selected twice every two weeks (B4). Furthermore, food item

use for dinner and lunch meals was restricted (B5). For example, each vegetable can only be selected twice per week (Table 3D.3). For the selected meals, the nutrient intake (B6) and the average nutrient intake of the two-week menu plan (B7) are calculated. To guarantee the nutritional adequacy of the menu plan, lower (B8) and upper (B9) bounds are placed to restrict the consumption of macronutrients, which are based on the dietary reference intakes (DRI) provided by IOM (2006) (Table 3D.4). The DRIs were adopted for women (19-50y) as iron deficiency is most commonly observed in women of reproductive age, and dietary iron bioavailability is, therefore, most critical for this age group (Lopez et al., 2016). Intake of micronutrients is restricted by the tolerable upper intake levels (B10). Constraint B11 sets the nutrient goals based on the recommended daily allowances (RDA). For a fair comparison between nutrients, deviations are normalized by scaling them between 0 and 1. With Constraint B12, the maximum deviation below the RDA is determined. Constraints B13 and B14 set the domain of the variables. The objective function of the model is to minimize the largest deviation below the RDA (B15).

### *3D.2.2 Continuous diet model*

The continuous diet model optimizes food item quantities while enforcing restrictions to maintain similarity to the observed diet (Table 3D.2). First, the absolute change in food items is calculated (C1-2). Then, an upper limit is imposed on food item changes within each food group (C3). Additionally, lower and upper bounds are set for the quantities of each food group (C4-5). Constraints C6-7 calculate the nutrient content of the optimized diet. The remaining constraints align with those in the mixed-integer diet model.

**Table 3D.1:** Constraints mixed-integer diet model.

| Constraints | |
|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall\, m \neq \text{snack}$ (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m = \text{snack}$ (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall\, r;\ m = \text{dinner}, \text{lunch}$ (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall\, r;\ m = \text{breakfast}, \text{snack}$ (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall\, i;\ m = \text{dinner}, \text{lunch}$ (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall\, r, m, n$ (B6) |
| $avgIntake_n = \frac{\sum_{r,m} nIntake_{rmn}}{NDays}$ | $\forall\, n$ (B7) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall\, nmacro$ (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall\, nmacro$ (B9) |
| $avgIntake_n \leq \text{UL}_n$ | $\forall\, nmicro$ (B10) |
| $avgIntake_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall\, nmicro$ (B11) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall\, nmicro$ (B12) |
| $avgIntake_n,\ belowRDA_n,\ maxBelowRDA,\ nIntake_{rmn} \geq 0$ | $\forall\, n$ (B13) |
| $X_{rm} \in \mathbb{N}$ | $\forall\, r, m$ (B14) |
| $\min \left\{ maxBelowRDA + \varepsilon \sum_n belowRDA_n \right\}$ | (B15) |

**Table 3D.2:** Constraints continuous diet model.

| Constraints | |
|---|---|
| $X_i - obsIntake_i = Qplus_i - Qmin_i$ | $\forall i$ **(C1)** |
| $Qdiff_i = Qplus_i + Qmin_i$ | $\forall i$ **(C2)** |
| $\sum_i (groupItems_{gi} \cdot Qdiff_i) \leq fChangeItem \cdot \sum_i (groupItems_{gi} \cdot obsIntake_i)$ | $\forall g$ **(C3)** |
| $\sum_i (groupItems_{gi} \cdot X_i) \geq (1 - fChangeGroup) \cdot \sum_i (groupItems_{gi} \cdot obsIntake_i)$ | $\forall g$ **(C4)** |
| $\sum_i (groupItems_{gi} \cdot X_i) \leq (1 + fChangeGroup) \cdot \sum_i (groupItems_{gi} \cdot obsIntake_i)$ | $\forall g$ **(C5)** |
| $nIntake_{in} = \text{nContent}_{in} \cdot X_i$ | $\forall i, n$ **(C6)** |
| $avgIntake_n = \sum_i nIntake_{in}$ | $\forall n$ **(C7)** |
| $avgIntake_n \geq \text{LB}_n$ | $\forall nmacro$ (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall nmacro$ (B9) |
| $avgIntake_n \leq \text{UL}_n$ | $\forall nmicro$ (B10) |
| $avgIntake_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall nmicro$ (B11) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall nmicro$ (B12) |
| $avgIntake_n$, $belowRDA_n$, $maxBelowRDA$, $nIntake_{in}$, $X_i \geq 0$ | $\forall n$ **(C8)** |
| $\min \left\{ maxBelowRDA + \varepsilon \sum_n belowRDA_n \right\}$ | (B15) |

*3D.2.3 Parameters*

**Table 3D.3:** The frequency at which a meal may be chosen per week, along with the number of times food items from a given category can be included in the menu plan.

| Category | Times per week[1] |
|---|---|
| Meals | |
| Breakfast | 2 |
| Lunch | 1 |
| Dinner | 1 |
| Snack | 2 |
| | |
| Foods | |
| Alcohol | - |
| Beans, peas, legumes | 3 |
| Cheese | - |
| Coffee | - |
| Condiments and sauces | - |
| Dairy | - |
| Drinks | - |
| Eggs | - |
| Fats and oils | - |
| Fruit | - |
| Grains | 4 |
| Meat, fish, poultry | 3 |
| Milk substitutes | - |
| Mixed dishes | 2 |
| Nuts and seeds | - |
| Other | - |
| Potatoes | 4 |
| Processed soy products | 2 |
| Snacks and sweets | - |
| Sugars | - |
| Tea | - |
| Vegetables | 2 |

[1] '-' indicates no constraint is added for this type of food category in the diet model.

**Table 3D.4:** Dietary reference intakes (women 19–50y) for nutrients considered by the diet models (IOM, 2006).

| Macronutrient | Unit | LB | UB |
|---|---|---|---|
| Energy[2] | kcal/day | 2000 | 2400 |
| Protein | E% | 10 | 35 |
| Total carbohydrates | E% | 45 | 65 |
| Dietary fibre | g | 25 | — |
| Added sugars[3] | E% | — | 10 (—) |
| Total fat | E% | 20 | 35 |
| Total saturated fatty acids | E% | — | 10 |
| Sodium[3] | mg | — | 2300 (—) |

| Micronutrient[4] | Unit | EAR | RDA | UL |
|---|---|---|---|---|
| Calcium[5] | mg | 800 | 1000 | 2500 |
| Iron (Absorbable iron)[6] | mg | 8.1 (1.5) | 18 (3.2) | 45 (8.1) |
| Zinc (Absorbable zinc)[6] | mg | 6.8 (3.3) | 8 (3.8) | 40 (19.2) |
| Vitamin A | $\mu$g | 500 | 700 | 3000 |
| Vitamin B1 | mg | 0.9 | 1.1 | — |
| Vitamin B2 | mg | 0.9 | 1.1 | — |
| Vitamin B3 | mg | 11 | 14 | — |
| Vitamin B6 | mg | 1.1 | 1.3 | 100 |
| Vitamin B9 | $\mu$g | 320 | 400 | 1000 |
| Vitamin C | mg | 60 | 75 | 2000 |

[1] Abbreviations: LB, Lower Bound; UB, Upper Bound; EAR, Estimated Average Requirement; RDA, Recommended Daily Allowance; UL, tolerable Upper Level.

[2] (USDA & US HHS, 2020)

[3] For the continuous diet model, no upper bounds were set for added sugar and sodium intake to prevent them becoming the bottleneck nutrients instead of absorbable iron or zinc.

[4] Vitamin B12 was not considered, as it could constitute a bottleneck nutrient in the design of vegetarian and vegan diets, while optimizing absorbable iron intake is of interest to the current research.

[5] (IOM, 2011)

[6] Absorbable iron and zinc were calculated by multiplying the dietary reference intakes with IOM's (2006) absorption factors of 18% and 48%, respectively.

### 3.D.3 Nonlinear programming

#### *3D.3.1 Armah*

To model nonheme iron absorption with Armah's equation, adjustments and extensions are made to the base diet models. For conciseness, only the formulations for the mixed-integer case are provided. Constraint NA1 (Table 3D.5, Equations 3A.2-3A.3) computes nonheme iron absorption based on average daily intake, and Constraint NA2 (Equation 3A.1) calculates total absorbable iron intake. Now, optimization can be directed towards absorbable iron intake instead of total iron intake (NA3-6). As model performance is based on absorbable iron content, we continue maximizing it when all dietary guidelines are met (NA8).

#### *3D.3.2 Conway*

Adjustments and extensions made to the base models to estimate nonheme iron absorption with Conway's equation are highlighted in Table 3D.6. In the calculation of iron absorption for individual meals (NC4, Equations 3A.4-3A.5), we utilize the nutritional composition specific to that meal (NC3). In order to accomplish this, we introduce binary variable Z to signify meal selection, as opposed to variable X, which denotes the frequency of meal selection (NC1-2). Subsequently, we proceed to compute both the total absorbable iron intake (NC5, Equation 3A.1) and the average daily intake (NC8). Constraints NC6-7 are implemented to optimize absorbable iron intake rather than total iron intake.

#### *3D.3.3 Miller*

To model absorbable zinc intake with Miller's equation, we adjusted the base models by adding Constraint NM1 (Table 3D.7, Equation 3A.6), which computes absorbable zinc intake based on average daily intake of total zinc and phytate.

**Table 3D.5:** Constraints mixed integer nonlinear programming model using Armah's equation to estimate nonheme iron absorption.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall\, m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall\, r;\ m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall\, r;\ m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall\, i;\ m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall\, r, m, n$ | (B6) |
| $avgIntake_n = \frac{\sum_{rm} nIntake_{rmn}}{NDays}$ | $\forall\, n$ | (B7) |
| $avgAbsNhFe = \exp\Big(6.294 - 0.709 \ln(SF) + 0.119 \ln\left(\max\left\{\text{avgIntake}_{\text{VitC}}, 1\right\}\right) + 0.006 \ln\left(\max\left\{\text{avgIntake}_{\text{MFP}}, 1\right\}\right) - 0.055 \ln\left(\max\left\{\text{avgIntake}_{\text{Tea}}, 1\right\}\right) - 0.247 \ln\left(\max\left\{\text{avgIntake}_{\text{Phy}}, 1\right\}\right) - 0.137 \ln\left(\max\left\{\text{avgIntake}_{\text{Ca}}, 1\right\}\right) - 0.083 \ln\left(\max\left\{\text{avgIntake}_{\text{nhFe}}, 1\right\}\right)\Big)$ | | (**NA1**) |
| $avgIntake_{absFe} = 0.25 \cdot avgIntake_{hFe} + 0.01 \cdot avgAbsNhFe \cdot avgIntake_{nhFe}$ | | (**NA2**) |
| $avgIntakeAdj_{Fe} = avgIntake_{absFe}$ | | (**NA3**) |
| $avgIntakeAdj_n = avgIntake_n$ | $\forall\, n$ not `Fe` | (**NA4**) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall\, nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall\, nmacro$ | (B9) |
| $avgIntakeAdj_n \leq \text{UL}_n$ | $\forall\, nmicro$ | (**NA5**) |
| $avgIntakeAdj_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall\, nmicro$ | (**NA6**) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall\, nmicro$ | (B12) |
| $absNhFe$, $avgAbsNhFe$, $avgIntake_n$, $avgIntakeAdj_n$, $belowRDA_n$, $maxBelowRDA$, $nIntake_{rmn} \geq 0$ | $\forall\, r, m, n$ | (**NA7**) |
| $X_{rm} \in \mathbb{N}$ | $\forall\, r, m$ | (B14) |
| $\min\left\{maxBelowRDA + \varepsilon_1 \sum_n belowRDA_n - \varepsilon_2 \cdot avgIntake_{absFe}\right\}$ | | (**NA8**) |

**Table 3D.6:** Constraints mixed integer nonlinear programming model using Conway's equation to estimate nonheme iron absorption.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall\, m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall\, r; m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall\, r; m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall\, i; m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall\, r, m, n$ | (B6) |
| $Z_{rm} \cdot \frac{NDays}{14} \geq X_{rm}$ | $\forall\, r, m$ | (**NC1**) |
| $Z_{rm} \leq X_{rm}$ | $\forall\, r, m$ | (**NC2**) |
| $nIntakeMeal_{rmn} = \text{nContent}_{rmn} \cdot Z_{rm}$ | $\forall\, r, m, n$ | (**NC3**) |
| $absNhFe_{rm} = 10\hat{}(0.9385 + 0.001\, nIntakeMeal_{r,m,Fruit} + 0.0009\, nIntakeMeal_{r,m,MFP} - 0.0052\, nIntakeMeal_{r,m,Beans} - 0.0016\, nIntakeMeal_{r,m,Cereal} - 0.002\, nIntakeMeal_{r,m,Tea} - 0.0012\, nIntakeMeal_{r,m,Dairy} - 0.0041\, nIntakeMeal_{r,m,Cheese} - 0.0032\, nIntakeMeal_{r,m,Eggs} - 0.0062\, nIntakeMeal_{r,m,Soy} - 0.0087\, nIntakeMeal_{r,m,Nuts}) \cdot (21.7/SF)^{0.94049}$ | $\forall\, r, m$ | (**NC4**) |
| $nIntake_{r,m,absFe} = 0.25\, nIntake_{r,m,hFe} + 0.01\, absNhFe_{rm} \cdot nIntake_{r,m,nhFe}$ | $\forall\, r, m$ | (**NC5**) |
| $nIntakeAdj_{r,m,Fe} = absFeIntake_{rm}$ | $\forall\, r, m$ | (**NC6**) |
| $nIntakeAdj_{rmn} = nIntake_{rmn}$ | $\forall\, r; m; n$ not `Fe` | (**NC7**) |
| $avgIntake_n = \frac{\sum_{rm} nIntakeAdj_{rmn}}{NDays}$ | $\forall\, n$ | (**NC8**) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall\, nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall\, nmacro$ | (B9) |
| $avgIntake_n \leq \text{UL}_n$ | $\forall\, nmicro$ | (B10) |
| $avgIntake_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall\, nmicro$ | (B11) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall\, nmicro$ | (B12) |
| $absNhFe_{rm}, avgIntake_n, belowRDA_n, maxBelowRDA, nIntake_{rmn}, nIntakeAdj_{rmn}, nIntakeMeal_{rmn} \geq 0$ | $\forall\, r, m, n$ | (**NC9**) |
| $X_{rm} \in \mathbb{N}$ | $\forall\, r, m$ | (B14) |
| $Z_{rm} \in \{0, 1\}$ | $\forall\, r, m$ | (**NC10**) |
| $\min \left\{ maxBelowRDA + \varepsilon_1 \sum_n belowRDA_n - \varepsilon_2 \cdot avgIntake_{absFe} \right\}$ | | (**NA8**) |

**Table 3D.7:** Constraints mixed integer nonlinear programming model using Miller's equation to estimate absorbable zinc intake.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall r; m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall r; m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall i; m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall r, m, n$ | (B6) |
| $avgIntake_n = \frac{\sum_{rm} nIntake_{rmn}}{NDays}$ | $\forall n$ | (B7) |
| $avgAbsZnIntake =$ | | |
| $65.4 \cdot 0.5 \cdot \Bigg[ \quad 0.13 + \frac{avgIntake_{Zn}}{65.4} + 0.10 \left(1 + \frac{avgIntake_{Phy}/660}{1.2}\right)$ | | |
| $- \sqrt{\left(0.13 + \frac{avgIntake_{Zn}}{65.4} + 0.10 \left(1 + \frac{avgIntake_{Phy}/660}{1.2}\right)\right)^2 - 4 \cdot 0.13 \cdot \frac{avgIntake_{Zn}}{65.4}} \quad \Bigg]$ | | (**NM1**) |
| $avgIntakeAdj_{Zn} = avgAbsZnIntake$ | | (**NM2**) |
| $avgIntakeAdj_n = avgIntake_n$ | $\forall n$ not `Zn` | (**NM3**) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall nmacro$ | (B9) |
| $avgIntakeAdj_n \leq \text{UL}_n$ | $\forall nmicro$ | (**NM4**) |
| $avgIntakeAdj_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall nmicro$ | (**NM5**) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall nmicro$ | (B12) |
| $avgAbsZnIntake, avgIntake_n, avgIntakeAdj_n,$ $belowRDA_n, maxBelowRDA, nIntake_{rmn} \geq 0$ | $\forall r, m, n$ | (**NM6**) |
| $X_{rm} \in \mathbb{N}$ | $\forall r, m$ | (B14) |
| $\min \left\{ maxBelowRDA + \varepsilon_1 \sum_n belowRDA_n - \varepsilon_2 \cdot avgAbsZnIntake \right\}$ | | (**NM7**) |

### 3.D.4 Piecewise linear aproximation

*3D.4.1 Armah*

Armah's nonlinear programming model is linearized through the application of univariate piecewise linear approximation to the natural logarithm function for nonheme iron, absorbable nonheme iron, and all nutrients affecting nonheme iron absorption (PA1-PA4) (Table 3D.8). Section 3C.1.1 provides a detailed exposition of this method. Constraint PA5 (Equation 3A.2) computes the natural logarithm of nonheme iron absorption, while Constraint PA6 computes the natural logarithm of absorbable nonheme iron intake (Equation 3A.3) . Constraint PA7 is employed to determine the total absorbable iron intake. Note that the term $\ln(\max\{\text{dietary factor}, 1\})$ is not included in Constraint PA5, but is accounted for in the first line segment of the piecewise linear approximation. Furthermore, the natural logarithm of the dietary factors is calculated (PA1-PA4) starting from the average intake (B7) (Figure 3D.1A), while the average absorbable iron intake is calculated (PA1-PA4) starting from the natural logarithm of absorbable iron (PA5-PA6) (Figure 3D.1B). Both steps are interconnected, as the logarithm of absorbable iron is calculated based on the logarithms of the dietary factors (PA5).



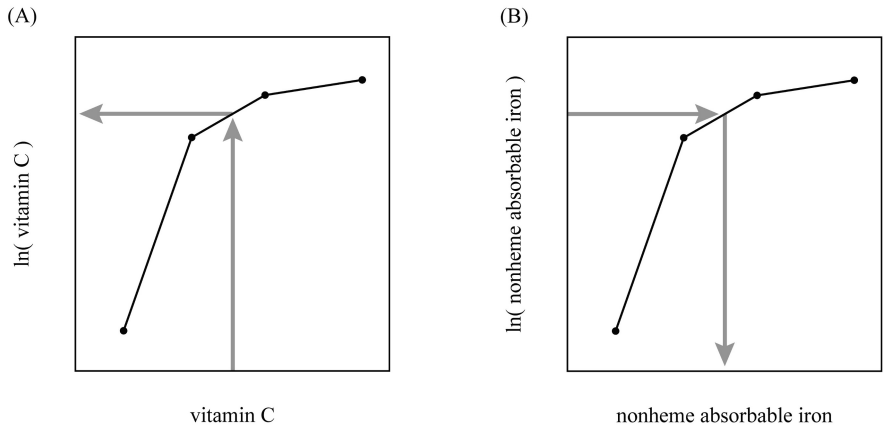



**Figure 3D.1:** Example of how piecewise linear approximation is applied: (A) to transition from vitamin C intake to the natural logarithm of vitamin C intake (likewise for all other dietary factors affecting nonheme iron absorption), and (B) to transition from the natural logarithm of nonheme absorbable iron intake to nonheme absorbable iron intake. Steps A and B are interconnected, as the logarithm of nonheme absorbable iron is calculated based on the logarithms of the dietary factors.

### *3D.4.2 Conway*

The nonlinear programming model using Conway's equation is linearized through the application of univariate piecewise linear approximation to the base 10 logarithm function for nonheme iron absorption and absorbable nonheme iron intake (PC1-PC4) (Table 3D.9). For each meal, Constraint PC5 computes the base 10 logarithm of nonheme iron absorption (Equation 3A.4), while Constraint PA6 computes the base 10 logarithm of absorbable nonheme iron intake (Equation 3A.5). Constraints PC7-8 ensure that the absorbable nonheme iron intake equals 0 when the meal is not selected. Constraint PA9 is employed to determine the total absorbable iron intake per meal. Note that the base 10 logarithm of nonheme iron is calculated (PC1-PC4) starting from nonheme iron content of a meal (NC3), while the absorbable nonheme iron content of a meal is calculated (PC1-PC4) starting from the base 10 logarithm of absorbable iron (PC5-PC6).

### *3D.4.3 Miller*

Miller's nonlinear programming model is linearized through the application of bivariate piecewise linear approximation. Section 3C.1.2 provides a detailed exposition of this method. Constraints PM1-PM6 are applied to estimate the absorbable zinc intake from the diet (Table 3D.10).

**Table 3D.8:** Constraints piecewise linear approximation model based on Armah's equation to estimate nonheme iron absorption.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall r; m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall r; m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall i; m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall r, m, n$ | (B6) |
| $avgIntake_n = \frac{\sum_{rm} nIntake_{rmn}}{NDays}$ | $\forall n$ | (B7) |
| $avgIntake_n = c^x + \sum_s d_s \cdot \delta_{ns}$ | $\forall nabs$ | (**PA1**) |
| $lnIntake_n = c^y + \sum_s a_s d_s \cdot \delta_{ns}$ | $\forall nabs$ | (**PA2**) |
| $Y_{ns} \leq \delta_{ns}$ | $\forall nabs, s$ | (**PA3**) |
| $Y_{ns} \geq \delta_{n,s+1}$ | $\forall nabs, s < S$ | (**PA4**) |
| $lnAbsNhFe = 6.294 - 0.709 \cdot lnSF + 0.119 \cdot lnIntake_{VitC} + 0.006 \cdot lnIntake_{MFP} - 0.055 \cdot lnIntake_{Tea} - 0.247 \cdot lnIntake_{Phy} - 0.137 \cdot lnIntake_{Ca} - 0.083 \cdot lnIntake_{nhFe}$ | | (**PA5**) |
| $lnIntake_{absNhFe} = lnIntake_{nhFe} + lnAbsNhFe - \ln 100$ | | (**PA6**) |
| $avgIntake_{absFe} = 0.25 \cdot avgIntake_{hFe} + avgIntake_{absNhFe}$ | | (**PA7**) |
| $avgIntakeAdj_{Fe} = avgAbsFeIntake$ | | (NA3) |
| $avgIntakeAdj_n = avgIntake_n$ | $\forall n$ not `Fe` | (NA4) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall nmacro$ | (B9) |
| $avgIntakeAdj_n \leq \text{UL}_n$ | $\forall nmicro$ | (NA5) |
| $avgIntakeAdj_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall nmicro$ | (NA6) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall nmicro$ | (B12) |
| $absNhFe,\ avgIntake_n,\ avgIntakeAdj_n,\ belowRDA_n,\ lnAbsNhFe,\ lnIntake_n,\ maxBelowRDA,\ nIntake_{rmn} \geq 0$ | $\forall r, m, n$ | (**PA8**) |
| $X_{rm} \in \mathbb{N}$ | $\forall r, m$ | (B14) |
| $\delta_{ns} \in [0, 1]$ | $\forall n, s$ | (**PA9**) |
| $Y_{ns} \in \{0, 1\}$ | $\forall n, s$ | (**PA10**) |
| $\min \left\{ maxBelowRDA + \varepsilon_1 \sum_n belowRDA_n - \varepsilon_2 \cdot avgAbsZnIntake \right\}$ | | (B15) |

**Table 3D.9:** Piecewise linear approximation model based on Conway's equation.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall\, m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall\, r; m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall\, r; m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall\, i; m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall\, r, m, n$ | (B6) |
| $Z_{rm} \cdot \frac{NDays}{14} \geq X_{rm}$ | $\forall\, r, m$ | (NC1) |
| $Z_{rm} \leq X_{rm}$ | $\forall\, r, m$ | (NC2) |
| $nIntakeMeal_{rmn} = \text{nContent}_{rmn} \cdot Z_{rm}$ | $\forall\, r, m, n$ | (NC3) |
| $nIntake_{rmn} = c^x + \sum_s d_{rms} \cdot \delta_{rmns}$ | $\forall\, r; m; s; n =$ `absNhFe`, `nhFe` | (**PC1**) |
| $logIntake_{rmn} = c^y + \sum_s a_{rms} d_{rms} \cdot \delta_{rmns}$ | $\forall\, r; m; s; n =$ `absNhFe`, `nhFe` | (**PC2**) |
| $Y_{rmns} \leq \delta_{rmns}$ | $\forall\, r; m; s; n =$ `absNhFe`, `nhFe` | (**PC3**) |
| $Y_{rmns} \geq \delta_{rmn,s+1}$ | $\forall\, r; m; s < S; n =$ `absNhFe`, `nhFe` | (**PC4**) |
| $logAbsNhFe_{rm} = 0.9385 \cdot Z_{rm} + 0.001 \cdot nIntakeMeal_{r,m,Fruit}$ $+ 0.0009 \cdot nIntakeMeal_{r,m,MFP} - 0.0052 \cdot nIntakeMeal_{r,m,Beans}$ $- 0.0016 \cdot nIntakeMeal_{r,m,Cereal} - 0.002 \cdot nIntakeMeal_{r,m,Tea}$ $- 0.0012 \cdot nIntakeMeal_{r,m,Dairy} - 0.0041 \cdot nIntakeMeal_{r,m,Cheese}$ $- 0.0032 \cdot nIntakeMeal_{r,m,Eggs} - 0.0062 \cdot nIntakeMeal_{r,m,Soy}$ $- 0.0087 \cdot nIntakeMeal_{r,m,Nuts}$ | $\forall\, r, m$ | (**PC5**) |
| $logIntake_{r,m,absNhFe} = logAbsNhFe_{rm} + (eqSF - \log 100) \cdot Z_{rm} + logIntake_{r,m,nhFe}$ | $\forall\, r, m$ | (**PC6**) |
| $absNhFeIntake_{rm} \leq nIntake_{r,m,nhFe}$ | $\forall\, r, m$ | (**PC7**) |
| $absNhFeIntake_{rm} \leq nIntake_{r,m,absNhFe}$ | $\forall\, r, m$ | (**PC8**) |
| $nIntake_{r,m,absFe} = 0.25 \cdot nIntake_{r,m,hFe} + absNhFeIntake_{rm}$ | $\forall\, r, m$ | (**PC9**) |
| $nIntakeAdj_{r,m,Fe} = nIntake_{r,m,absFe}$ | $\forall\, r, m$ | (NC6) |
| $nIntakeAdj_{rmn} = nIntake_{rmn}$ | $\forall\, r; m; n$ not `Fe` | (NC7) |
| $avgIntake_n = \frac{\sum_{rm} nIntakeAdj_{rmn}}{NDays}$ | $\forall\, n$ | (NC8) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall\, nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall\, nmacro$ | (B9) |
| $avgIntake_n \leq \text{UL}_n$ | $\forall\, nmicro$ | (B10) |
| $avgIntake_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall\, nmicro$ | (B11) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall\, nmicro$ | (B12) |
| $avgIntake_n, belowRDA_n, maxBelowRDA, nIntake_{rmn},$ $nIntakeAdj_{rmn}, nIntakeMeal_{rmn} \geq 0$ | $\forall\, r, m, n$ | (**PA8**) |
| $X_{rm} \in \mathbb{N}$ | $\forall\, r, m$ | (B14) |
| $Z_{rm}, Y_{rmns} \in \{0, 1\}$ | $\forall\, r, m, n, s$ | (**PC11**) |
| $\delta_{rmns} \in [0, 1]$ | $\forall\, r, m, n, s$ | (**PC12**) |
| $logIntake_{rmn}, logAbsNhFe_{rm} \in \mathbb{R}$ | $\forall\, r, m, n$ | (**PC13**) |
| $\min \left\{ maxBelowRDA + \varepsilon \sum_n belowRDA_n \right\}$ | | (B15) |

**Table 3D.10:** Constraints piecewise linear approximation model based on Miller's equation to estimate absorbable zinc intake.

| Constraints | | |
|---|---|---|
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = NDays$ | $\forall\, m$ not `snack` | (B1) |
| $\sum_r (X_{rm} \cdot \text{mealType}_{rm}) = 2 \cdot NDays$ | $m =$ `snack` | (B2) |
| $X_{rm} \leq \frac{NDays}{14}$ | $\forall\, r; m =$ `dinner`, `lunch` | (B3) |
| $X_{rm} \leq \frac{NDays}{7}$ | $\forall\, r; m =$ `breakfast`, `snack` | (B4) |
| $\sum_r (X_{rm} \cdot \text{mealFoodItems}_{rmi}) \leq \text{fDiversity}_i$ | $\forall\, i; m =$ `dinner`, `lunch` | (B5) |
| $nIntake_{rmn} = \text{nContent}_{rmn} \cdot X_{rm}$ | $\forall\, r, m, n$ | (B6) |
| $avgIntake_n = \frac{\sum_{rm} nIntake_{rmn}}{NDays}$ | $\forall\, n$ | (B7) |
| $\sum_z \sum_p (T_{zp} + B_{zp}) = 1$ | | (**PM1**) |
| $\sum_z \sum_p \lambda_{zp} = 1$ | | (**PM2**) |
| $\lambda_{zp} \leq T_{zp} + T_{z+1,p} + T_{z,p-1} + B_{zp} + B_{z-1,p} + B_{z,p+1}$ | $\forall\, z, p$ | (**PM3**) |
| $avgIntake_{Zn} = 65.4 \sum_z \sum_p x_z \lambda_{zp}$ | | (**PM4**) |
| $avgIntake_{Phy} = 660 \sum_z \sum_p y_p \lambda_{zp}$ | | (**PM5**) |
| $avgAbsZnIntake = 65.4 \sum_z \sum_p f_{zp} \lambda_{zp}$ | | (**PM6**) |
| $avgIntakeAdj_{Zn} = avgAbsZnIntake$ | | (NM2) |
| $avgIntakeAdj_n = avgIntake_n$ | $\forall\, n$ not `Zn` | (NM3) |
| $avgIntake_n \geq \text{LB}_n$ | $\forall\, nmacro$ | (B8) |
| $avgIntake_n \leq \text{UB}_n$ | $\forall\, nmacro$ | (B9) |
| $avgIntakeAdj_n \leq \text{UL}_n$ | $\forall\, nmicro$ | (NM4) |
| $avgIntakeAdj_n + \text{RDA}_n \cdot belowRDA_n \geq \text{RDA}_n$ | $\forall\, nmicro$ | (NM5) |
| $maxBelowRDA \geq belowRDA_n$ | $\forall\, nmicro$ | (B12) |
| $avgAbsZnIntake, avgIntake_n, avgIntakeAdj_n,$ $belowRDA_n, maxBelowRDA, nIntake_{rmn} \geq 0$ | $\forall\, r, m, n$ | (**PM7**) |
| $X_{rm} \in \mathbb{N}$ | $\forall\, r, m$ | (B14) |
| $\lambda_{zp} \in [0, 1]$ | $\forall\, z, p$ | (**PM8**) |
| $T_{zp} \in \{0, 1\}$ | $\forall\, z = 2 \dots Z, \quad p = 1 \dots P - 1$ | (**PM9**) |
| $B_{zp} \in \{0, 1\}$ | $\forall\, z = 1 \dots Z - 1, \quad p = 2 \dots P$ | (**PM10**) |
| $\min \left\{ maxBelowRDA + \varepsilon \sum_n belowRDA_n \right\}$ | | (B15) |

### 3.D.5 Modeling scenarios

**Table 3D.11:** Overview of modeling scenarios.[1]

| S | Model | Diet plan | Equation | Method | Solver | Time limit | Startpoint | Multistart |
|---|---|---|---|---|---|---|---|---|
| 1 | Mixed-integer | Vegan | Armah | PLA | CPLEX | 1h | | |
| 2 | Mixed-integer | Vegan | Armah | NLP | LINDO | 1h | | |
| 3 | Mixed-integer | Vegan | Armah | NLP | LINDO | 1h | ✓ | |
| 4 | Mixed-integer | Vegan | Armah | NLP | LINDO | 1h | ✓ | ✓ |
| 5 | Mixed-integer | Vegan | Armah | NLP | SCIP | 1h | | |
| 6 | Mixed-integer | Vegan | Armah | NLP | SCIP | 1h | ✓ | |
| 7 | Mixed-integer | Vegan | Conway | LP | CPLEX | 1h | | |
| 8 | Mixed-integer | Vegan | Conway | PLA | CPLEX | 1h | | |
| 9 | Mixed-integer | Vegan | Conway | NLP | LINDO | 1h | | |
| 10 | Mixed-integer | Vegan | Conway | NLP | LINDO | 1h | ✓ | |
| 11 | Mixed-integer | Vegan | Conway | NLP | LINDO | 1h | ✓ | ✓ |
| 12 | Mixed-integer | Vegan | Conway | NLP | SCIP | 1h | | |
| 13 | Mixed-integer | Vegan | Conway | NLP | SCIP | 1h | ✓ | |
| 14 | Mixed-integer | Vegetarian | Armah | PLA | CPLEX | 1h | | |
| 15 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 1h | | |
| 16 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 1h | ✓ | |
| 17 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 1h | ✓ | ✓ |
| 18 | Mixed-integer | Vegetarian | Armah | NLP | SCIP | 1h | | |
| 19 | Mixed-integer | Vegetarian | Armah | NLP | SCIP | 1h | ✓ | |
| 20 | Mixed-integer | Vegetarian | Conway | LP | CPLEX | 1h | | |
| 21 | Mixed-integer | Vegetarian | Conway | PLA | CPLEX | 1h | | |
| 22 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 1h | | |
| 23 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 1h | ✓ | |
| 24 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 1h | ✓ | ✓ |
| 25 | Mixed-integer | Vegetarian | Conway | NLP | SCIP | 1h | | |
| 26 | Mixed-integer | Vegetarian | Conway | NLP | SCIP | 1h | ✓ | |
| 27 | Mixed-integer | Omnivorous | Armah | PLA | CPLEX | 1h | | |
| 28 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 1h | | |
| 29 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 1h | ✓ | |
| 30 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 1h | ✓ | ✓ |
| 31 | Mixed-integer | Omnivorous | Armah | NLP | SCIP | 1h | | |
| 32 | Mixed-integer | Omnivorous | Armah | NLP | SCIP | 1h | ✓ | |
| 33 | Mixed-integer | Omnivorous | Conway | LP | CPLEX | 1h | | |
| 34 | Mixed-integer | Omnivorous | Conway | PLA | CPLEX | 1h | | |
| 35 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 1h | | |
| 36 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 1h | ✓ | |
| 37 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 1h | ✓ | ✓ |
| 38 | Mixed-integer | Omnivorous | Conway | NLP | SCIP | 1h | | |
| 39 | Mixed-integer | Omnivorous | Conway | NLP | SCIP | 1h | ✓ | |
| 40 | Mixed-integer | Vegan | Armah | PLA | CPLEX | 30min | | |
| 41 | Mixed-integer | Vegan | Armah | NLP | LINDO | 30min | | |
| 42 | Mixed-integer | Vegan | Armah | NLP | LINDO | 30min | ✓ | |
| 43 | Mixed-integer | Vegan | Armah | NLP | LINDO | 30min | ✓ | ✓ |
| 44 | Mixed-integer | Vegan | Armah | NLP | SCIP | 30min | | |
| 45 | Mixed-integer | Vegan | Armah | NLP | SCIP | 30min | ✓ | |
| 46 | Mixed-integer | Vegan | Conway | LP | CPLEX | 30min | | |
| 47 | Mixed-integer | Vegan | Conway | PLA | CPLEX | 30min | | |
| 48 | Mixed-integer | Vegan | Conway | NLP | LINDO | 30min | | |
| 49 | Mixed-integer | Vegan | Conway | NLP | LINDO | 30min | ✓ | |
| 50 | Mixed-integer | Vegan | Conway | NLP | LINDO | 30min | ✓ | ✓ |
| 51 | Mixed-integer | Vegan | Conway | NLP | SCIP | 30min | | |
| 52 | Mixed-integer | Vegan | Conway | NLP | SCIP | 30min | ✓ | |
| 53 | Mixed-integer | Vegetarian | Armah | PLA | CPLEX | 30min | | |
| 54 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 30min | | |
| 55 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 30min | ✓ | |
| 56 | Mixed-integer | Vegetarian | Armah | NLP | LINDO | 30min | ✓ | ✓ |
| 57 | Mixed-integer | Vegetarian | Armah | NLP | SCIP | 30min | | |
| 58 | Mixed-integer | Vegetarian | Armah | NLP | SCIP | 30min | ✓ | |
| 59 | Mixed-integer | Vegetarian | Conway | LP | CPLEX | 30min | | |

| S | Model | Diet plan | Equation | Method | Solver | Time limit | Startpoint | Multistart |
|---|---|---|---|---|---|---|---|---|
| 60 | Mixed-integer | Vegetarian | Conway | PLA | CPLEX | 30min | | |
| 61 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 30min | | |
| 62 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 30min | ✓ | |
| 63 | Mixed-integer | Vegetarian | Conway | NLP | LINDO | 30min | ✓ | ✓ |
| 64 | Mixed-integer | Vegetarian | Conway | NLP | SCIP | 30min | | |
| 65 | Mixed-integer | Vegetarian | Conway | NLP | SCIP | 30min | ✓ | |
| 66 | Mixed-integer | Omnivorous | Armah | PLA | CPLEX | 30min | | |
| 67 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 30min | | |
| 68 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 30min | ✓ | |
| 69 | Mixed-integer | Omnivorous | Armah | NLP | LINDO | 30min | ✓ | ✓ |
| 70 | Mixed-integer | Omnivorous | Armah | NLP | SCIP | 30min | | |
| 71 | Mixed-integer | Omnivorous | Armah | NLP | SCIP | 30min | ✓ | |
| 72 | Mixed-integer | Omnivorous | Conway | LP | CPLEX | 30min | | |
| 73 | Mixed-integer | Omnivorous | Conway | PLA | CPLEX | 30min | | |
| 74 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 30min | | |
| 75 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 30min | ✓ | |
| 76 | Mixed-integer | Omnivorous | Conway | NLP | LINDO | 30min | ✓ | ✓ |
| 77 | Mixed-integer | Omnivorous | Conway | NLP | SCIP | 30min | | |
| 78 | Mixed-integer | Omnivorous | Conway | NLP | SCIP | 30min | ✓ | |
| 79 | Mixed-integer | Vegan | Miller | PLA | CPLEX | 1h | | |
| 80 | Mixed-integer | Vegan | Miller | NLP | LINDO | 1h | | |
| 81 | Mixed-integer | Vegan | Miller | NLP | SCIP | 1h | | |
| 82 | Mixed-integer | Vegetarian | Miller | PLA | CPLEX | 1h | | |
| 83 | Mixed-integer | Vegetarian | Miller | NLP | LINDO | 1h | | |
| 84 | Mixed-integer | Vegetarian | Miller | NLP | SCIP | 1h | | |
| 85 | Mixed-integer | Omnivorous | Miller | PLA | CPLEX | 1h | | |
| 86 | Mixed-integer | Omnivorous | Miller | NLP | LINDO | 1h | | |
| 87 | Mixed-integer | Omnivorous | Miller | NLP | SCIP | 1h | | |
| 88 | Mixed-integer | Vegan | Miller | PLA | CPLEX | 30min | | |
| 89 | Mixed-integer | Vegan | Miller | NLP | LINDO | 30min | | |
| 90 | Mixed-integer | Vegan | Miller | NLP | SCIP | 30min | | |
| 91 | Mixed-integer | Vegetarian | Miller | PLA | CPLEX | 30min | | |
| 92 | Mixed-integer | Vegetarian | Miller | NLP | LINDO | 30min | | |
| 93 | Mixed-integer | Vegetarian | Miller | NLP | SCIP | 30min | | |
| 94 | Mixed-integer | Omnivorous | Miller | PLA | CPLEX | 30min | | |
| 95 | Mixed-integer | Omnivorous | Miller | NLP | LINDO | 30min | | |
| 96 | Mixed-integer | Omnivorous | Miller | NLP | SCIP | 30min | | |
| 97 | Continuous | I50-G40 | Armah | PLA | CPLEX | 1h | | |
| 98 | Continuous | I50-G40 | Armah | NLP | LINDO | 1h | | |
| 99 | Continuous | I50-G40 | Armah | NLP | SCIP | 1h | | |
| 100 | Continuous | I75-G50 | Armah | PLA | CPLEX | 1h | | |
| 101 | Continuous | I75-G50 | Armah | NLP | LINDO | 1h | | |
| 102 | Continuous | I75-G50 | Armah | NLP | SCIP | 1h | | |
| 103 | Continuous | I100-G60 | Armah | PLA | CPLEX | 1h | | |
| 104 | Continuous | I100-G60 | Armah | NLP | LINDO | 1h | | |
| 105 | Continuous | I100-G60 | Armah | NLP | SCIP | 1h | | |
| 106 | Continuous | I50-G40 | Miller | PLA | CPLEX | 1h | | |
| 107 | Continuous | I50-G40 | Miller | NLP | LINDO | 1h | | |
| 108 | Continuous | I50-G40 | Miller | NLP | SCIP | 1h | | |
| 109 | Continuous | I75-G50 | Miller | PLA | CPLEX | 1h | | |
| 110 | Continuous | I75-G50 | Miller | NLP | LINDO | 1h | | |
| 111 | Continuous | I75-G50 | Miller | NLP | SCIP | 1h | | |
| 112 | Continuous | I100-G60 | Miller | PLA | CPLEX | 1h | | |
| 113 | Continuous | I100-G60 | Miller | NLP | LINDO | 1h | | |
| 114 | Continuous | I100-G60 | Miller | NLP | SCIP | 1h | | |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); I, limit food item changes; G, limit food group changes; Miller, Hambidge et al. (2010); NLP, nonlinear programming; PLA; piecewise linear approximation.

# 3.E Appendix - Additional results

## 3.E.1 Accuracy piecewise linear approximation

**Table 3E.1:** Accuracy of absorbable nonheme iron and zinc estimates based on the nonlinear equation (NLP) and piecewise linear approximation (PLA). For Conway, the absorbable nonheme iron content was computed for each meal in the dataset with a nonheme iron content greater than 0. For Armah and Miller, the absorbable nonheme iron and zinc content was calculated for 1000 simulated two-week menu plans (Section 3C.3).

| | Nonheme Iron | | Zinc |
|---|---|---|---|
| | Armah (daily avg) | Conway (single meal) | Miller (daily avg) |
| Mean total content (mg) | 23.1 | 2.4 | 28.8 |
| Mean absorbable content NLP (mg) | 1.3 | 0.1 | 4.5 |
| Mean absorbable content PLA (mg) | 1.3 | 0.1 | 4.5 |
| Mean absolute difference NLP and PLA (mg) | 0.1 | 0.0 | 0.0 |
| Mean absolute difference NLP and PLA (%) | 5.1 | 3.9 | 0.4 |
| Max absolute difference NLP and PLA (mg) | 0.2 | 0.1 | 0.1 |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); Miller, Hambidge et al. (2010); NLP, nonlinear programming; PLA; piecewise linear approximation.

### 3.E.2 Performance mixed-integer diet model with 30 min time limit

**Table 3E.2:** Comparison of model runtime and average absorbable iron and zinc content (mg/day) based on Armah's, Conway's, and Miller's equations for NLP and PLA models.[1]

| | | Vegan | | Vegetarian | | Omnivorous | |
|---|---|---|---|---|---|---|---|
| Equation | Model | Runtime[2] | Absorbed[3] | Runtime | Absorbed | Runtime | Absorbed |
| **Iron** | | | | | | | |
| Armah | PLA[4] | 00:25 | **1.6** | 0:00:19 | **2.2** | 0:01:27 | **3.2** |
| | NLP | | | | | | |
| | LINDO | 06:03 | **1.6** | 0:13:35 | **2.2** | 0:07:19 | **3.2** |
| | LINDO+initial | 09:46 | **1.6** | 0:15:42 | **2.2** | 0:30:12 | 2.8 |
| | LINDO+initial+multistart | 03:03 | **1.6** | 0:03:47 | **2.2** | 0:28:13 | **3.2** |
| | SCIP | 30:26 | **1.6** | 0:30:07 | 1.7 | 0:30:11 | 3.1 |
| | SCIP+initial | 30:24 | **1.6** | 0:30:06 | **2.2** | 0:30:11 | 3.1 |
| Conway | PLA | 00:25 | **1.8** | 0:01:48 | **2.0** | 0:03:58 | **3.1** |
| | NLP | | | | | | |
| | LINDO | 30:18 | - | 0:30:16 | - | 0:30:18 | - |
| | LINDO+initial | 30:18 | - | 0:33:59 | 0.5 | 0:49:18 | 0.9 |
| | LINDO+initial+multistart | 30:26 | 0.8 | 0:34:04 | 0.5 | 0:49:17 | 0.9 |
| | SCIP | 01:18 | **1.8** | 0:30:11 | - | 0:31:57 | 3.0 |
| | SCIP+initial | 01:32 | **1.8** | 0:30:21 | 0.5 | 0:30:38 | 0.9 |
| **Zinc** | | | | | | | |
| Miller | PLA | 00:33 | **2.9** | 01:47 | **4.2** | 02:57 | **5.1** |
| | NLP | | | | | | |
| | LINDO | 01:30 | **2.9** | 03:26 | 4.1 | 06:59 | 5.0 |
| | SCIP | 30:12 | 2.7 | 30:25 | **4.2** | 30:58 | **5.1** |

[1] Abbreviations: Armah, Armah et al. (2013); Conway, Conway et al. (2007); NLP, nonlinear programming; LINDO, Linear Interactive and Discrete Optimizer; Miller, Hambidge et al. (2010); PLA, piecewise linear approximation; SCIP, Solving Constraint Integer Programs.

[2] Runtime is displayed in h:mm:ss. The limit for the model's runtime was set to 30 minutes.

[3] Per type of menu plan, highest absorbable iron and zinc intake (mg/day) estimates are highlighted in bold.

[4] For the menu plans obtained by the PLA model, absorbable iron and zinc intake estimates based on the original absorption equations were calculated retrospectively.

### 3.E.3 Micronutrient content mixed-integer menu plans

Menu plans with similar absorbable iron or zinc content generally had similar nutrient profiles, though some exceptions were observed. For example, in Armah's mixed-integer vegan menu plans, all diets had similar absorbable iron content, but the vitamin A content in the LINDO + startpoint + multistart model was 155 $\mu$g lower than in the PLA model (Figure 3E.1). Similarly, in Conway's vegan menu plans, both the PLA and SCIP (+ startpoint) models had similar absorbable iron content, yet the vitamin A content in PLA was 105 $\mu$g lower than in SCIP. The reason for this lies in the formulation of the models' objective function, which aims to maximize the bottleneck nutrient. Although menu plans had similar absorbable iron content when rounded to one decimal place, differences existed at higher precision. A small improvement in the bottleneck nutrient can lead to a relatively large reduction in other nutrients. Please note that in evaluating the performance of NLP and PLA, our goal was to obtain modeling solutions with the highest absorbable iron or zinc content. However, for practical applications, a different objective function formulation may be preferable to achieve a more balanced outcome. Gerdessen & De Vries (2015) provide a comprehensive explanation of this matter.

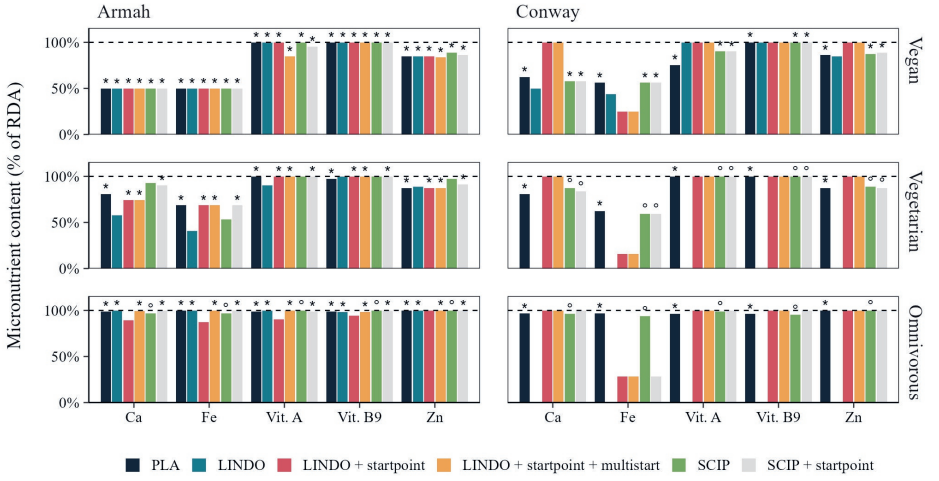



**Figure 3E.1:** Average micronutrient content of menu plans optimized for absorbable iron intake, shown as a percentage of the Recommended Daily Allowance (RDA). Only micronutrients with values below the RDA are displayed, with nutrient content capped at 100%. A * above the bars indicates that the mixed-integer diet model identified the solution with the highest absorbable iron content, whereas a ° denotes solutions where absorbable iron content was only 0.1 mg lower than the best observed solution.

# Part 2

Food Substitution

# Chapter 4

## Within-food-group optimization improves nutritional adequacy, sustainability, and acceptability of modeled diets



This chapter is published as:

van Wonderen, D., Gerdessen, J. C., Biesbroek, S. & Melse-Boonstra, A. (2025). Within-food-group optimization improves nutritional adequacy, sustainability, and acceptability of modeled diets. *Frontiers in Nutrition*, 12, 1648055.

# Abstract

Food production is a major contributor to global greenhouse gas emissions (GHGE). To mitigate this impact, researchers have developed methods for designing healthy and sustainable diets by modifying existing consumption patterns through dietary changes between food groups. However, the nutrient and emission profiles within these food groups can differ greatly. The aim of this study was to investigate the extent to which the nutritional adequacy, sustainability, and acceptability of diets can be improved through dietary changes within food groups. To analyze the potential of within-food-group optimization, we investigated several diet modeling strategies and scenarios to optimize nutrient intake while minimizing GHGE and dietary change. The diets used as input for the diet model were derived from the U.S. National Health and Nutrition Examination Survey (NHANES) 2017-2018 consumption dataset. By adjusting food quantities only within food groups, macro- and micronutrient recommendations could be met while achieving a 15% to 36% reduction in GHGE. When foods were optimized both within- and between food groups, only half the dietary change (23%) was required to achieve a 30% GHGE reduction, compared to optimizing between food groups alone (44%). This may improve consumer acceptance, assuming smaller dietary shifts are perceived as more acceptable. To conclude, within-food-group optimization increases opportunities to improve the nutritional adequacy, sustainability, and acceptability of diets.

## 4.1 Introduction

Food production for human consumption is estimated to contribute one-third of global greenhouse gas emissions (GHGE) caused by human activities (Tubiello et al., 2021). To reduce this footprint, various methodologies have been developed to design diets that are both environmentally and nutritionally sustainable. This research area is commonly referred to as diet modeling or diet optimization. Initial studies focused on defining nutritionally adequate diets within set cost limits, but from around 2010 onward, they began to incorporate environmental considerations (Gazan et al., 2018b; Perignon et al., 2016; van Dooren, 2018; Leonard & Kiely, 2024).

Multiple studies have shown that substantial dietary changes are necessary to improve the sustainability of diets, though the precise extent varies between studies (Table 4.1). For example, Vieux et al. (2018) examined the changes in food consumption needed to reduce GHGE in several European countries while complying with nutrient guidelines. Based on their results, it was computed that achieving a 30% reduction in GHGE required food quantity changes ranging from 40% to 65%. Similarly, it could be calculated from the study by Rocabois et al. (2022) that a 30% reduction in GHGE would involve food quantity changes of up to 69% in the French diet. Other studies have reported a lower level of required dietary change. For instance, Nordman et al. (2024) found that a 30% change in the Danish diet could achieve a 31% reduction in GHGE.



The differences in modeling results regarding GHGE reductions and required dietary changes can be explained by several factors. These include variations in the current consumption patterns of the target group under investigation, as well as in the GHGE and nutrient content of available foods, which may differ between countries. Results are also influenced by modeling decisions, such as the nutrient constraints applied, food quantity limits set to ensure the acceptability of modeled diets, and the level of detail at which foods are represented.

Diet modeling studies often model dietary changes at the level of food groups or subgroups, both referred to as food groups in this study. The number of food groups considered varies widely, ranging from 11 (Fu et al., 2024) to 402 (Perignon et al., 2016) (Table 4.1). When modeling at food group level, quantities of food groups can be adjusted, but the distribution of individual food items within each group remains unchanged, or an average food is used as proxy. For instance, while the overall quantity of the food group "vegetables" can be increased, it is not possible to adjust the distribution of specific food items, such as increasing "carrots" while decreasing "cucumber".

However, when modeling at food group level, the variability in nutrient composition (Figure 4.1) and GHGE profiles (Figure 4.2) within food groups (Rippin et al., 2021) is not taken into account, leaving opportunities to further improve the nutritional adequacy and sustainability of diets by optimizing foods within those food groups (Bolo

**Table 4.1:** Summary of diet modeling studies showing the trade-off between greenhouse gas emissions (GHGE) and total dietary change, as well as detailing the number of food groups used and the number of individual food items aggregated within those groups.

| Study | Country[1] | Trade-off[2,3] | | Number of food groups[4] | Number of individual food items |
|---|---|---|---|---|---|
| | | GHGE reduction (%) | Dietary change (%)[5] | | |
| Green et al. (2015) | UK | 30-60 | 68-94 | 148 | 2,980 |
| Horgan et al. (2016) | UK | 15 | 50[6,7] | 134 | 1,491 |
| Perignon et al. (2016) | FR | 10-60 | 5-506 | 402 | 1,342 |
| Gazan et al. (2018a) | FR | — | — | 212 | 1,342 |
| Vieux et al. (2018) | FR, UK, IT, FI, SE | 30 | 40-65 | 151 | 1,708–4,079 |
| Reynolds et al. (2019) | UK | 57 | 54 | 101 | 653 |
| Mariotti et al. (2021) | FR | — | — | 32 | 2,800 |
| Dussiot et al. (2022a) | FR | — | — | 45 | 1,533 |
| Rocabois et al. (2022) | FR | 30 | 69 | 207 | 1,342 |
| Tompa et al. (2022) | HU | — | — | 35 | 857 |
| Heerschop et al. (2023) | NL | 16 | 33 | 28 | 2,389 |
| Verly Jr et al. (2023) | BR | — | — | 85 | 1,591 |
| Fu et al. (2024) | CN | — | — | 11 | — |
| Nordman et al. (2024) | DK | 31 | 30 | 50 | 434 |
| Bashiri et al. (2025) | EE | — | — | 14 | 74 |
| Fouillet et al. (2025) | FR | — | — | 45 | 1,533 |
| Kesse-Guyot et al. (2025) | FR | 77 | 110 | 47 | 265 |

[1] Country of the consumption data set used for modelling.

[2] Results on GHGE reductions and total dietary change should be interpreted with caution across studies, as different modeling approaches were applied, limiting direct comparability.

[3] A '—' indicates that GHGE reductions or total dietary change were not reported or could not be calculated from the study's results.

[4] Some studies refer to food groups as food items. For example, Vieux et al. (2018) model diets using 151 food items such as root vegetables and citrus fruits. In this study, we classify these as food groups, with items like carrots and oranges representing individual foods within those groups.

[5] See Appendix 4A for the calculation of total dietary change.

[6] Average dietary change calculated at the food group level, not at the total diet level.

[7] Complete removal or addition of food groups is not included in this number.

et al., 2024). In addition, within-food-group optimization could improve the acceptance of generated diets for two reasons. First, foods within food groups are typically more similar than those between food groups, which may make such substitutions more preferable (Arens & Hamilton, 2018; Huh et al., 2016). Second, consumer acceptance is often linked to the extent of dietary change, with smaller changes generally considered more achievable (Hill, 2009; van Dooren, 2018). By allowing changes within food groups, rather than

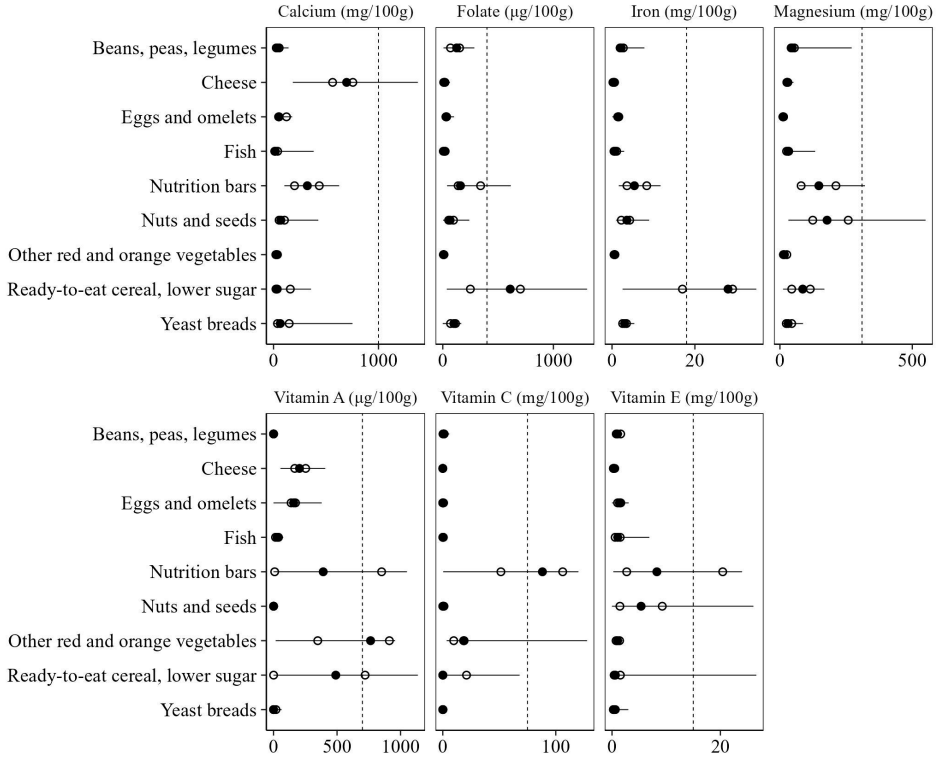



**Figure 4.1:** Variation in nutrient content within a selection of What We Eat In America (WWEIA) food subgroups that display considerable differences. The horizontal line shows the range of variation, with the black dot marking the average content and the white dots representing the $25^{th}$ and $75^{th}$ percentiles. The vertical line indicates the Recommended Daily Allowance (RDA).

only between food groups, the total dietary change required could be reduced, thereby increasing acceptance.

The impact of whether foods are optimized within food groups on the nutritional adequacy, environmental sustainability, and acceptability of diets has not yet been quantified. Accordingly, this study aims to demonstrate the potential benefits of within-group optimization, which may help future studies determine the most suitable level at which to model foods. To achieve this, we developed a diet model that optimizes nutrient intake while minimizing GHGE and dietary change by adjusting food quantities reported in the U.S. National Health and Nutrition Examination Survey 2017-2018 (NHANES) (CDC & NCHS, 2018). By exploring different modeling strategies, we evaluated and compared

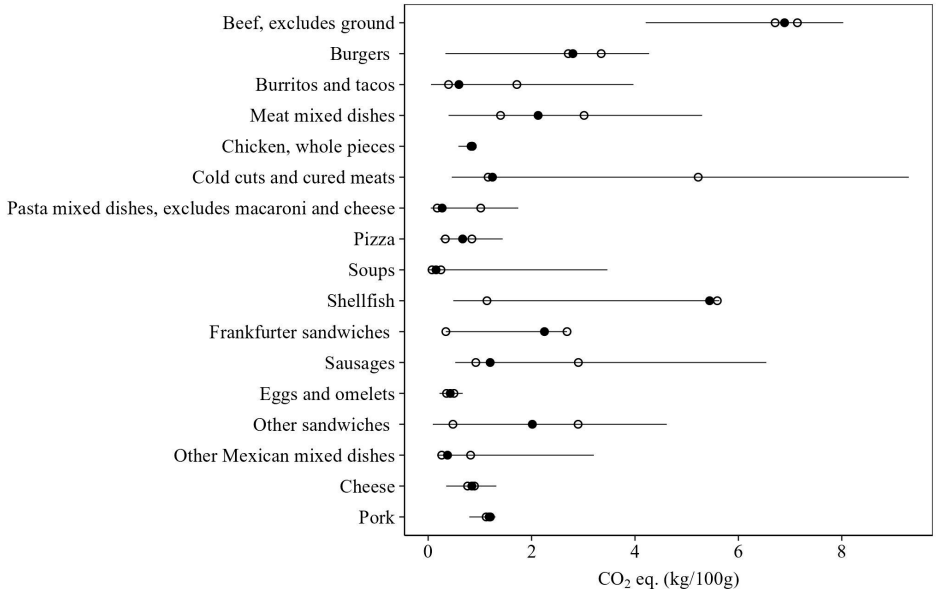



**Figure 4.2:** Variation in greenhouse gas emissions across the What We Eat In America (WWEIA) food subgroups that contribute most to observed diet-related CO2 eq. emissions, limited to groups with three or more food items. The horizontal line shows the range of variation, with the black dot marking the average emission and the white dots representing the $25^{th}$ and $75^{th}$ percentiles.

the performance of between-food-group and within-food-group optimization, along with the trade-offs among nutritional adequacy, environmental sustainability, and acceptability.

## 4.2 Materials and Methods

To demonstrate the impact of within-food-group optimization on the nutritional adequacy, environmental sustainability, and acceptability of diets, we started from observed consumption data and applied various diet modeling strategies to optimize nutrient intake while minimizing greenhouse gas emissions (GHGE) and dietary changes for both females and males.

### 4.2.1 Data

#### *4.2.1.1 Observed diet*

U.S. food consumption data, comprising two 24-hour dietary recalls, were retrieved from NHANES 2017-2018 (CDC & NCHS, 2018), and corresponding nutrient intakes from the Food and Nutrient Database for Dietary Studies (FNDDS) 2017-2018 (USDA & ARS, 2020a). For the purpose of this study, which is to compare methodological approaches, we focused on a single target group: adults aged 18 to 65, as they represent a relatively homogeneous group in terms of nutritional needs. Individuals were excluded if they reported a low energy intake (less than 1,200 kcal for women and 1,800 kcal for men) or a high energy intake (more than 3,000 kcal for women and 3,600 kcal for men). This left 1,738 female respondents and 1,428 male respondents, with 3,166 respondents in total. The consumption data were then summarized by calculating the average daily intake per food item (g/day) separately for females and males, which is referred to as the observed diets.

#### *4.2.1.2 Food group classification*

The food group classifications used for diet modeling were based on the What We Eat In America (WWEIA) (USDA & ARS, 2020d) and FNDDS (USDA & ARS, 2020a) subgroup classifications, comprising 153 and 46 groups, respectively. In addition, a custom food group classification was created using the method as described by Perignon et al. (2016) (Perignon), consisting of 345 groups (Appendix 4B). Separate model runs were conducted for each food group classification.

Of the 4,257 reported unique food items, the intake of 2,734 food items was considered for optimization. Foods excluded from optimization, i.e. for which quantities were kept identical to those in the observed diets, included items classified as "other" according to the WWEIA food group classification (e.g., nutritional powder mix and raw oats). Furthermore, food items consumed three times or less were excluded from optimization.

#### *4.2.1.3 Greenhouse gas emissions*

GHGE of NHANES (composite) foods, expressed in CO2 equivalents, were estimated using GHGE data for corresponding primary food products from the dataFIELD database and associated loss factors from the Loss-Adjusted Food Availability (LAFA) database (USDA ARS, 2022). LAFA provides the percentage of food weight lost throughout the supply chain (from farm to retail) as well as losses at the consumer level (e.g., inedible portions, cooking losses, and uneaten food). GHGE for each NHANES food was calculated as:

$$\text{GHGE} = \sum_i \text{Weight}_i \cdot \text{GHGE}_i \cdot \frac{100}{100 - \text{Food loss (\%)}_i} \Rightarrow \left(\frac{\text{CO}_2\,\text{eq.}}{100g}\right) = \left(\frac{g}{100g}\right) \cdot \left(\frac{\text{CO}_2\,\text{eq.}}{g}\right)$$

where $i$ represents the primary food products of the NHANES food. The weights of primary food products for each NHANES food were based on the Food Commodity Intake Database (FCID) (US EPA, 2001-2010), extended by Conrad et al. (2022) for more recent NHANES cycles (2011-2018). Both the weight and GHGE data align with FCID's classification of primary food products. For the loss factors, LAFA food descriptions were mapped to FCID food codes by Conrad et al. (2018). Using this approach, the calculated GHGE for the observed diets was 5.1 kg CO2 equivalents per day for females and 7.9 kg CO2 equivalents per day for males.

### 4.2.2 Diet modeling

The diet model optimized nutrient intake while minimizing GHGE and dietary change by adjusting food quantities. Different strategies for achieving this were explored to compare the performance of between- and within-food-group optimization, as well as to compute the trade-offs between nutritional adequacy, environmental sustainability, and acceptability. Full model definitions are provided in Appendix 4C Tables 4C.1 - 4C.3, with the applied nutritional guidelines outlined in Table 4C.4.

#### *4.2.2.1 Within-food-group optimization*

In the first modeling experiment, we examined the potential to improve the nutritional adequacy and GHGE of observed diets by adjusting food item quantities within food groups only. The model permitted changes only to the distribution of food items within each group, keeping the overall quantity of each food group similar to the observed diet. In the objective function of the model, the highest weight was assigned to minimizing the largest deviation from recommended macro- and micronutrient intake levels (Recommended Daily Allowances; RDA), then GHGE reduction, and finally, minimizing dietary change. A simplified version of the objective function is given by:

$$\min \left\{ D^{\text{macro}} + D^{\text{rda}} + \varepsilon_1 \cdot E + \varepsilon_2 \cdot C^{\text{within}} \right\}$$

where $D^{macro}$ and $D^{rda}$ represent deviations from macronutrient and micronutrient (RDA) guidelines, $E$ represents GHGE, $C^{within}$ represents within-food-group changes, and $\epsilon_1$ and $\epsilon_2$ are small weighting values such that $\epsilon_1 > \epsilon_2$. The full formulation is provided in Table 4C.3 Equation 4C.30.

Furthermore, we evaluated model outcomes using three different food group classifications (Perignon with 345 groups, WWEIA with 153 groups, and FNDDS with 46 groups) across

varying levels of within-food-group changes allowed (0%, 25%, 50%, 75%, and 100%). When no change was allowed (0% change), the diet remained identical to the observed diet. In the least restrictive scenario (100% change), all food item quantities within each food group could be adjusted freely. For intermediate scenarios, we constrained the absolute change within each food group to x% or less. Additionally, no food item could exceed the quantity of the highest-consumed item in that food group. Note that these upper bounds for food item quantities were based on daily average intakes from the consumption dataset and do not reflect the highest intakes reported by individuals. The unrestricted scenario may favor a narrow selection of the healthiest and most sustainable foods, whereas the restricted scenarios lead to more diverse and balanced diets, which we assumed are more acceptable. This allowed us to examine the trade-off between acceptability, nutritional adequacy, and sustainability.

#### *4.2.2.2 Lowered nutrient goals*

In the second modeling experiment, which builds on the first, we investigated to what extent lowering nutrient goals can further reduce GHGE. This provides a sharper picture on the trade-off between nutritional adequacy and sustainability. Here, macronutrient bounds were extended by 20% (e.g. the protein lower and upper bounds were relaxed from 10-35 E% to 8-42 E%), and micronutrient goals were lowered from the RDA to the Estimated Average Requirements (EAR):

$$\min \left\{ D^{\text{macro}\pm 20\%} + D^{\text{ear}} + \varepsilon_1 \cdot E + \varepsilon_2 \cdot C^{\text{within}} \right\}$$

For the full formulation, see Table 4C.3 Equation 4C.31.

#### *4.2.2.3 Between-food-group vs. Between-and-within-food-group optimization*

In the last modeling experiment, we investigated the minimal amount of dietary change required to achieve at least a 30% reduction in GHGE while complying with macronutrient and micronutrient (RDA) guidelines:

$$\min \left\{ C^{\text{total}} + \varepsilon \cdot E \right\}$$

where $C^{\text{total}}$ represents total dietary change and $D^{\text{macro}} = D^{\text{rda}} = 0$ (Table 4C.3 Equation 4C.37). We compared the required dietary change under two scenarios: one where only total food group quantities may be adjusted while distribution of food items within each group remains fixed (between-food-group optimization), and another where food item quantities can be changed without restriction (between-and-within-food-group optimization). This comparison demonstrates how within-food-group optimization can

improve the acceptability of designed diets. Acceptability was evaluated based on total dietary change (Appendix 4A) and food diversity. We assumed that diets requiring smaller deviations from observed consumption patterns are more acceptable (Hill, 2009; van Dooren, 2018). Food diversity was measured by the number of unique food items and the average share of the most prevalent food item in each food group, where diets with a higher number of unique food items and a lower average share were considered more diverse and balanced, and therefore more acceptable (Embling et al., 2021; Zandstra et al., 2000).

## 4.3 Results

To demonstrate the added value of within-food-group optimization on the nutritional adequacy, environmental sustainability, and acceptability of diets, we conducted various modeling experiments, for which the results are displayed below.

### 4.3.1 Within-food-group optimization

In the first modeling experiment, we examined to what extent the nutritional adequacy and GHGE of observed diets could be improved by adjusting food item quantities within food groups only. Figure 4.3 shows how deviations from macro- and micronutrient guidelines and GHGE were reduced, with greater reductions observed when changes within groups were larger. When no restrictions were placed on the magnitude of change (allowed within food group change of 100% and no constraint on the maximum quantity of food items), deviations from dietary guidelines were either absent or negligible, and GHGE was reduced by 15% to 36%, when using the WWEIA or FNDDS food group classification. However, when using the Perignon classification, the largest deviations were still 36% for macronutrients and 1% for micronutrients, with a maximum GHGE reduction of just 3%.

Figures 4.4 and 4.5 offer a closer look at the micro- and macro-nutrient content of observed and optimized diets. For females, the average micronutrient content of the observed diets was below the RDA for calcium, folate, iron, magnesium, vitamin A, and vitamin E. Increasing the intake of iron and vitamin E, followed by calcium and folate, proved to be the most difficult, as shown by the Perignon and WWEIA food group classifications where the allowed within food group change was 50%. For males, the bottleneck nutrients were magnesium, vitamin A, and vitamin E. In terms of macronutrient and sodium content, the observed diets for both females and males failed to meet the dietary guidelines for fiber, total fat, saturated fatty acids (SFA), and sodium. Optimizing these nutrients proved challenging, with only the scenarios allowing 100% within food group changes, using the WWEIA and FNDDS classifications, meeting or nearly meeting the dietary guidelines.

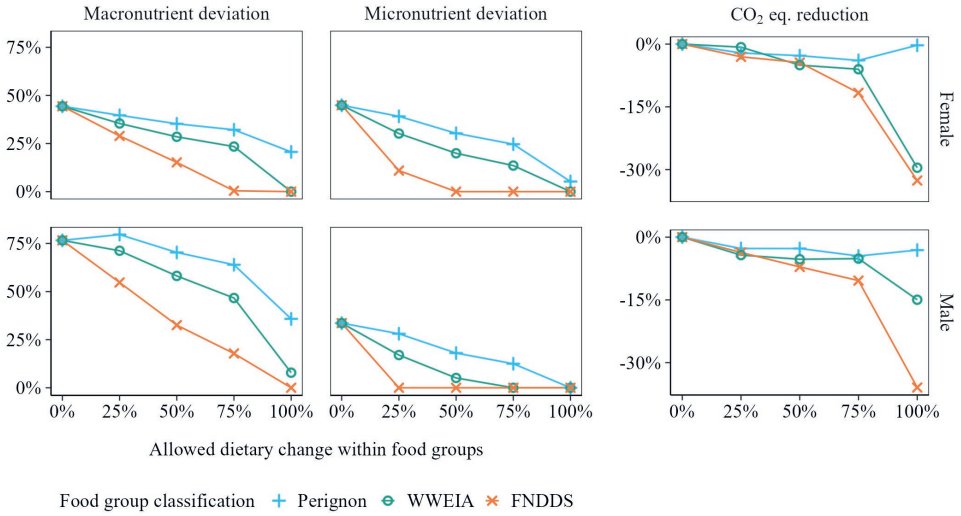



**Figure 4.3:** Deviations from dietary guidelines for the most limiting macro- and micronutrient, along with reductions in greenhouse gas emissions (GHGE), for observed and optimized diets in which food item quantities were adjusted within food groups only. As the diet model prioritizes optimizing nutritional adequacy over sustainability (Appendix 4C), a reduction in macro- and micronutrient deviations may come at the cost of increased GHGE, as demonstrated by the scenario where a dietary change of 100% was allowed for the Perignon food group classification.

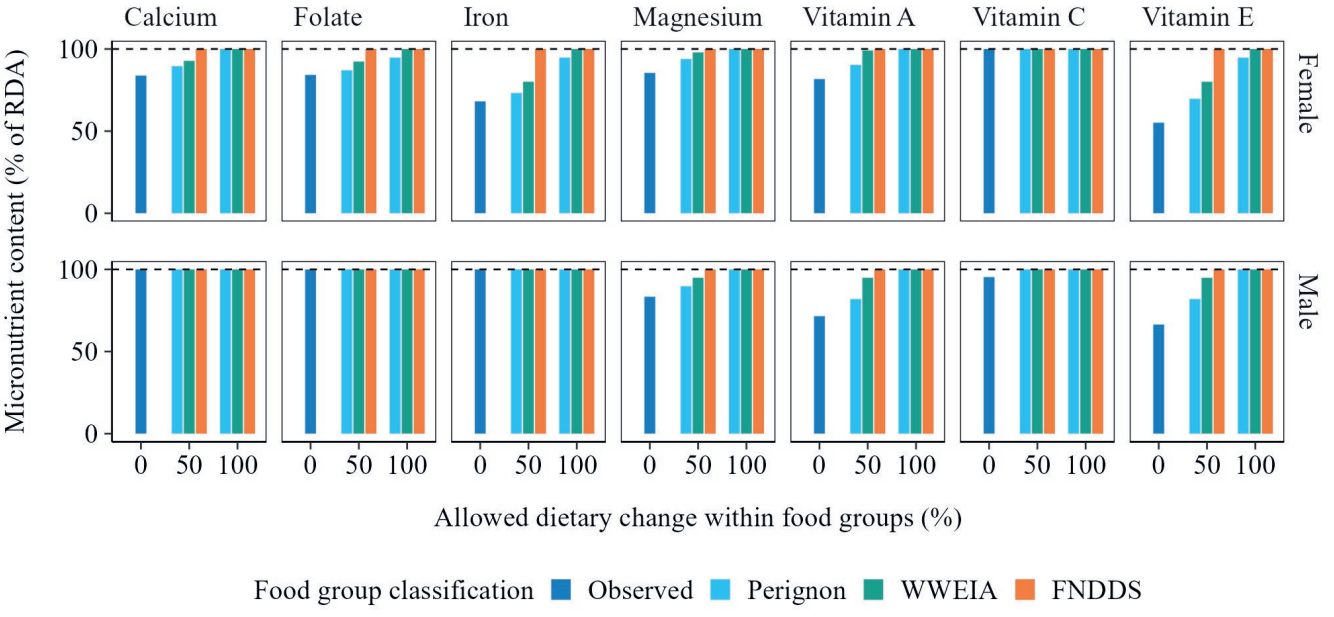



**Figure 4.4:** Average micronutrient content of the observed and optimized diets in which food item quantities were adjusted within food groups only. The micronutrient content is presented as a percentage of the Recommended Daily Allowance (RDA). Note that only micronutrients with values below the RDA are displayed. Additionally, nutrient intake is capped at 100%.

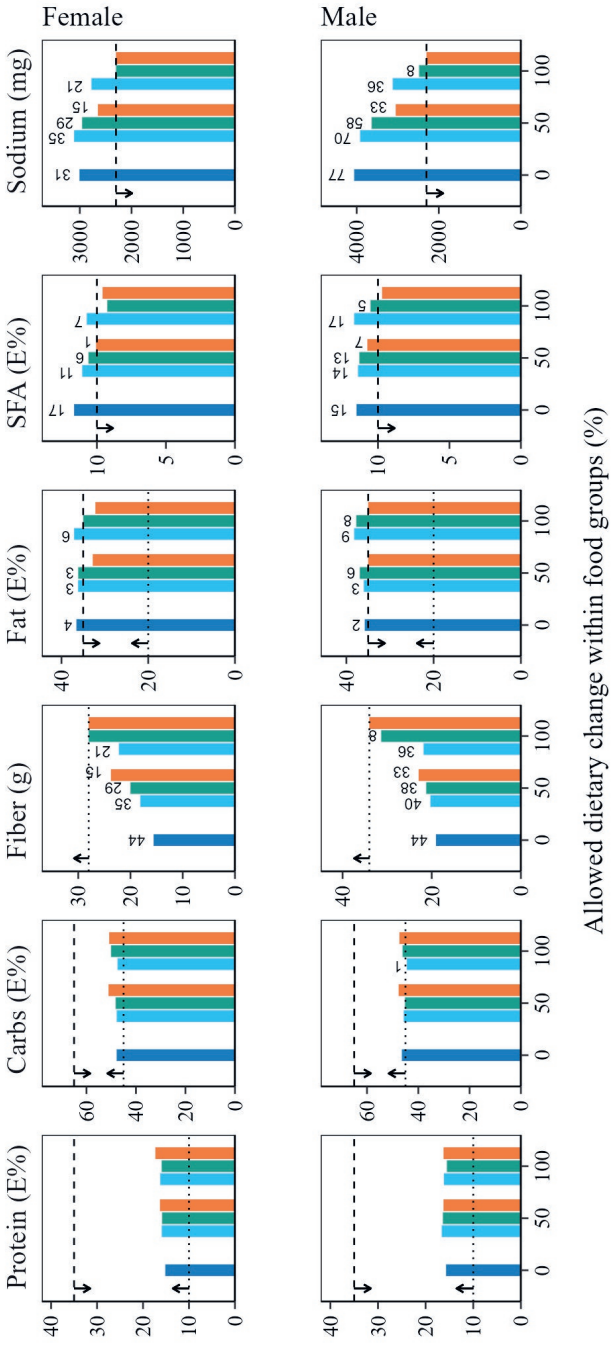



**Figure 4.5:** Macronutrient and sodium content of the observed and optimized diets in which food item quantities were adjusted within food groups only. For protein, carbohydrate, total fat, and saturated fatty acids (SFA), the nutrient content is expressed as a percentage of energy intake. The dotted and dashed horizontal lines display the lower and upper bounds of recommended intake, respectively. The labels above the bars display the deviation from dietary guidelines (%).

An example of how within-food-group optimization improved the nutrient content of diets is shown in Figure 4.6. For females, iron intake substantially increased through the selection of iron-rich food items in the scenario where the allowed change within food groups was 100%. For instance, in the "Yeast breads" group, the model favored "Bread, whole grain white, toasted" (5.4 mg/100g) over less iron-rich breads like "Bread, white" (3.4 mg/100g) and "Bread, wheat or cracked wheat" (3.6 mg/100g), the most consumed breads in the observed diet (Figure 4.7). In contrast, when dietary change was limited to 50%, a smaller increase in iron was observed, as it was constrained to stay closer to observed consumption patterns.

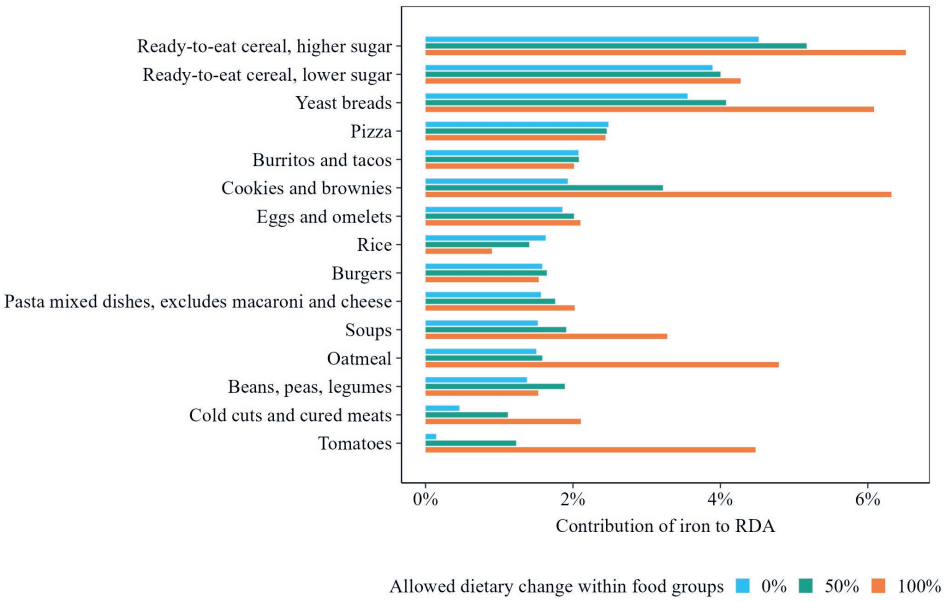



**Figure 4.6:** Iron content of WWEIA groups contributing most to iron intake in observed and optimized diets for females, presented as a percentage of the Recommended Daily Allowance (RDA).

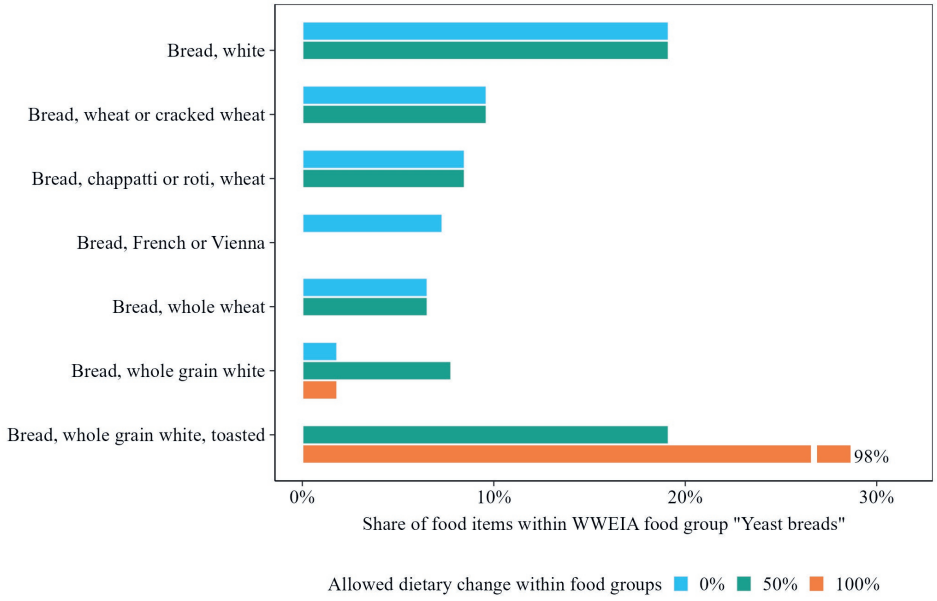



**Figure 4.7:** Distribution of foods with the highest share consumed in WWEIA group 'Yeast breads' in observed and optimized diets for females. Note that in the scenario where dietary change within food groups is limited to 50%, also no food item can exceed the highest observed quantity within its group.

### 4.3.2 Lowered nutrient goals

When nutrient goals were lowered by extending macronutrient bounds by 20% and lowering micronutrient goals from the RDA to the EAR, larger reductions in GHGE were achievable (Figure 4.8). This greater reduction was most evident when 100% of the diet could be adjusted. In this scenario, GHGE reductions exceeding 60% were possible, representing a twofold improvement compared to the full nutrient goals. Larger reductions in GHGE were also observed in the scenario where 75% of the diet could be altered, when the WWEIA and FNDDS classifications were used. However, in the other scenarios, lowering the nutrient goals did not result in a notable GHGE reductions. In these cases, the nutrient goals remained unachieved, leaving limited opportunity for further reductions in GHGE.

Examples of GHGE reductions for a selection of WWEIA food groups are shown in Figure 4.9. Reductions in groups such as 'Burgers' and 'Meat mixed dishes' were achieved by replacing beef with turkey or chicken. For groups such as 'Burritos and tacos' and 'Pasta dishes,' switching to vegetarian options lowered GHGE. Notably, for males, no

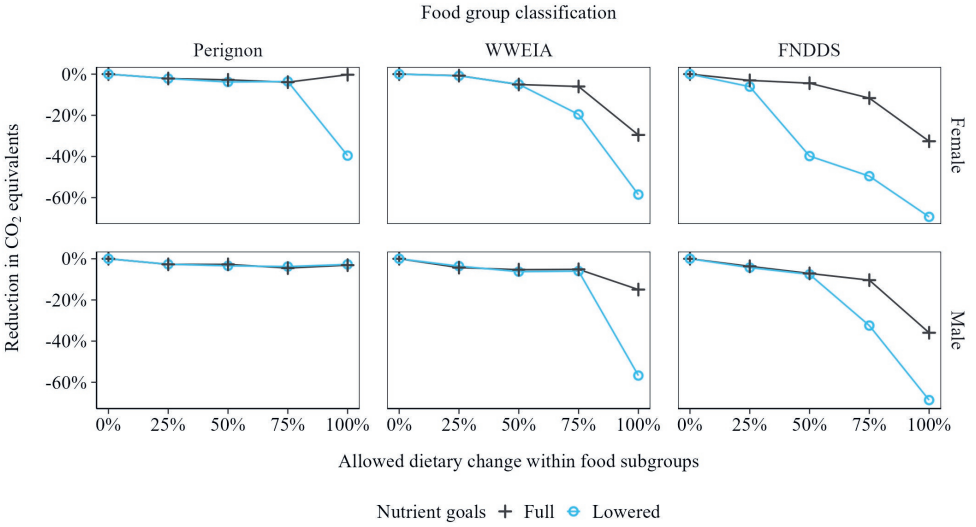



**Figure 4.8:** Greenhouse gas emission reductions of the optimized diets, relative to the observed diets, under full and lowered nutrient goals. For the lowered nutrient goals, macronutrient bounds were extended by 20%, and micronutrient goals were lowered from the Recommended Dietary Allowance (RDA) to the Estimated Average Requirements (EAR).

reductions were observed in groups such as 'Beef' or 'Meat mixed dishes' when full nutrient goals were applied. This was due to males' relatively high salt intake, which led the model to favor foods low in sodium over those with low GHGE.

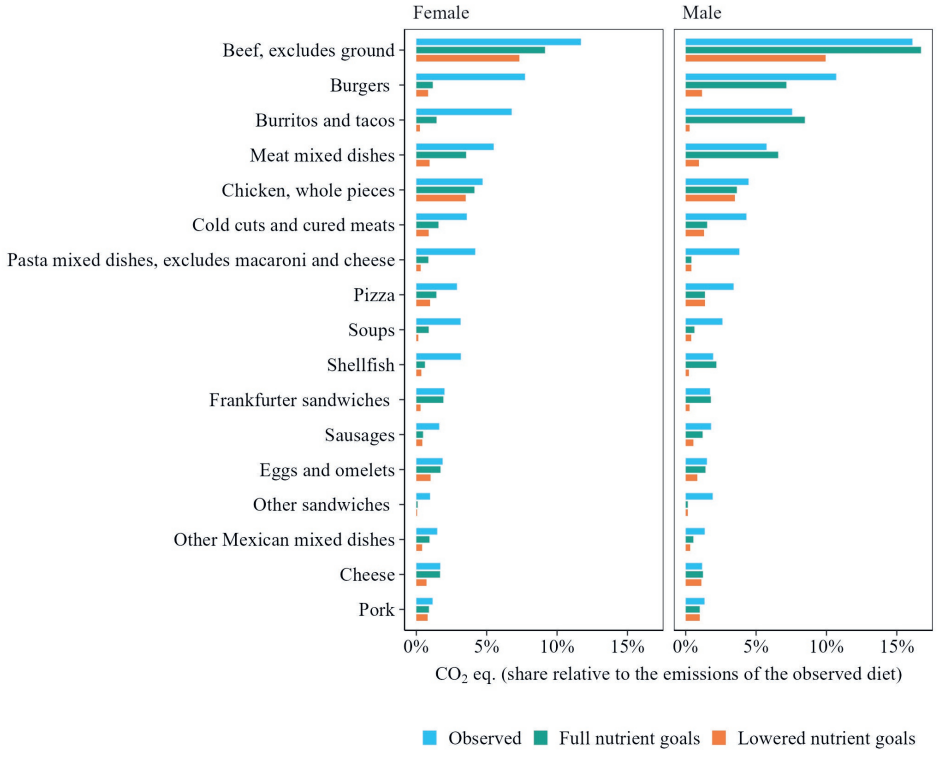



**Figure 4.9:** Greenhouse gas emissions from WWEIA food groups (≥3 items) with the highest contributions in observed and optimized diets, under full and lowered nutrient goals. For the lowered nutrient goals, macronutrient bounds were extended by 20%, and micronutrient goals were lowered from the Recommended Dietary Allowance (RDA) to the Estimated Average Requirements (EAR).

### 4.3.3 Between-food-group vs. Between-and-within-food-group optimization

In the previous sections, we presented improvements in nutritional adequacy and GHGE of diets when optimizing food item quantities within food groups only. In this section, we will examine the dietary changes required to achieve a 30% reduction in GHGE while complying with macro- and micronutrient (RDA) guidelines. We compared two modeling approaches: between-food-group optimization, where only total food group quantities can be adjusted while the distribution of food items within each group remains fixed, as well as between-and-within food-group optimization, where the quantities of individual food items can be adjusted freely (100% allowed dietary change).

For between-food-group optimization, the total dietary change ranged from 36% to 58%, depending on the food classification used (Figure 4.10). In contrast, for between-and-within-food-group optimization, the average required dietary change was 23%, which is half of that for between-food-group optimization (44%). Furthermore, the diets obtained by between-and-within-food-group optimization contained a higher number of unique food items. Here, only 38% of food items were removed on average, compared to 53% when applying between-food-group optimization, during which certain food groups were entirely removed. Nonetheless, the latter resulted in more balanced diets, with the average maximum food item share within a food group at 36%, similar to the observed diets, compared to 40% for between-and-within-food-group optimization. Note that the share is relatively higher for food group classifications with fewer food items within each food group.

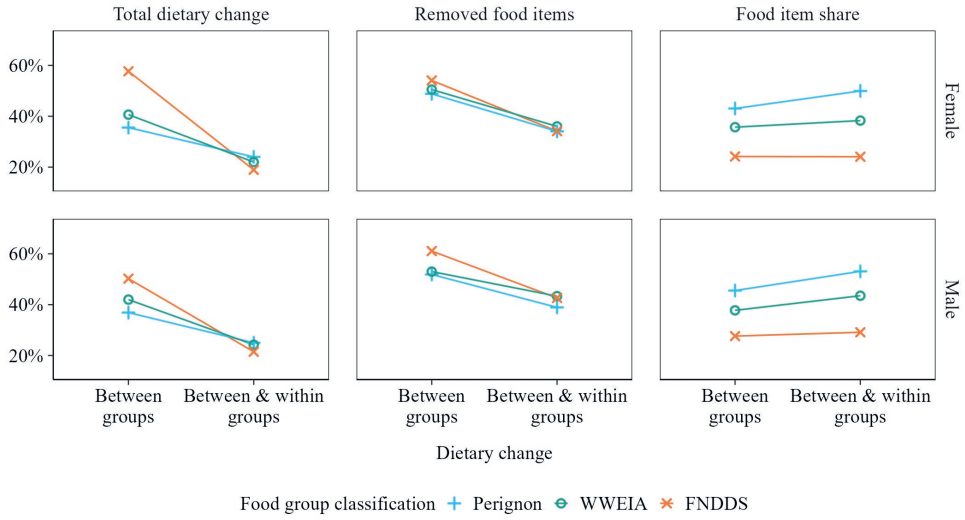



**Figure 4.10:** Comparison of between- versus between-and-within-food-group optimization for various acceptability measures. The calculation of total dietary change is detailed in Appendix 4A. The percentage of removed food items and the share of food items reflect diet diversity, with food item share being the average share of the most prevalent food item in each food group.

## 4.4 Discussion

This study demonstrated that between-and-within-food-group optimization facilitates greater improvements in the nutritional adequacy, environmental sustainability, and acceptability of diets than between-food-group optimization alone. By adjusting food item quantities within food groups only, macro- and micronutrient recommendations could be

met while achieving up to a 36% reduction in GHGE. A further reduction in GHGE, exceeding 60%, could be attained by lowering macro- (by 20%) and micronutrient (RDA to EAR) goals. To achieve a 30% reduction in GHGE while complying with macro- and micronutrient (RDA) guidelines, between-and-within-food-group optimization required, on average, only half the total dietary change (23%) compared to between-food-group optimization (44%). This 44% dietary change in U.S. diets for between-food-group optimization aligns with findings by Vieux et al. (2018) for France and Finland (45% on average). It is lower than the values reported for the United Kingdom and Sweden (62%) (Vieux et al., 2018) and by Green et al. (2015) for the United Kingdom (68%), and exceeds that of Nordman et al. (2024) for Denmark (30%) (Table 4.1). In summary, this study quantified the positive impact of the modeling decision to optimize food item quantities within food groups, as opposed to not doing so, on the nutritional adequacy, environmental sustainability, and acceptability of diets. Note that these results only provide an indication of the potential improvements achieved by optimizing foods within food groups, compared to not doing so. In different study settings, such as when applied to another target population or when additional nutritional constraints are included, the benefits of within-food-group optimization may vary in magnitude. In the remainder of this section, we will also discuss possible drawbacks of within-food-group optimization, as well as briefly address other modeling decisions, including handling missing data and outliers, selecting sustainability metrics, choosing dietary reference values, and quantifying consumer acceptability.



A potential drawback of optimizing food item quantities within food groups is the concern that only a narrow range of the healthiest and most sustainable foods is selected, making modeled diets less diverse and more sensitive to inaccuracies in nutrient (Ahuja & Perloff, 2008) and GHGE (Lansche et al., 2013) data. However, rather than avoiding within-food-group optimization altogether, this risk can be managed by applying constraints on food item quantities to help preserve dietary diversity, as demonstrated by our modeling scenarios where the percentage of allowed dietary change within food groups and the maximum quantity of food items were restricted. Additionally, the impact of data inaccuracies can be mitigated through outlier handling techniques, such as winsorization (Abuzaid & Alkronz, 2024).

Furthermore, within-food-group optimization requires GHGE data on food item level, which is not always available. When different data sources do not share the same level of detail, Gazan et al. (2018a) proposed to aggregate more detailed data (e.g. nutrient composition) to match the least detailed data source (e.g. GHGE). Instead, we recommend modeling at the available levels, where GHGE is optimized at group level and nutrient content at food item level. This approach makes it easier to achieve nutrient goals, thereby requiring less dietary change to achieve the same GHGE reduction. It offers a temporary solution until GHGE data becomes available for a wider range of foods, a need also highlighted by other researchers (Rippin et al., 2021). Alongside GHGE data, it is

equally important to collect and model data on other sometimes divergent sustainability mid-point metrics, such as blue water use (or water scarcity), or end-point metrics such as biodiversity, in order to support a more multidimensional approach to designing sustainable diets.

Regarding the modeling decision of which dietary reference values to apply, there may be a trade-off between nutritional adequacy and environmental sustainability. Lowering nutritional goals can allow for greater reductions in GHGE, as shown by Perignon et al. (2016) and this study (Figure 4.8). A question that remains is the extent to which improving sustainability may come at the expense of nutritional adequacy? For example, the Recommended Daily Allowances (RDA) serve as guidelines for micronutrient intake to meet the needs of 97.5% of the population and are often used in individual diet counseling. Many individuals, however, can remain healthy with lower intake levels, such as the Estimated Average Requirements (EAR), which meet the needs of 50% of the population (IOM, 2000; USDA & US HHS, 2020). Alternatively, increased use of dietary supplements and food fortification could help improve the sustainability of diets (Fatemi et al., 2023; Grasso et al., 2023).

Among all modeling decisions, determining how to quantify consumer acceptability may be the most technically challenging. In this study, we assumed that diets with greater diversity and fewer changes from the observed diet would be more acceptable. However, there is still limited understanding of which types of dietary change are most acceptable to consumers (Gazan et al., 2018b; van Dooren, 2018). For instance, are small adjustments across a variety of foods more acceptable than large changes to a few? Is it easier for consumers to adapt to the addition of new foods or to the elimination of familiar ones? Are modifications within food groups preferred over changes between groups? Although modifications within food groups may be preferred because the alternatives are likely more similar in taste (Kourouniotis et al., 2016), a potential downside is that dietary advice at individual food level can be more challenging to recall and follow than general food group recommendations (Mata et al., 2010). As for food diversity, its association with acceptability is not necessarily linear. For example, a diet with many different food items could become less acceptable if the preparation of such meals is deemed too effortful by consumers. Conversely, diets with less variety, like having the same breakfast every day, may be acceptable for consumers who prefer routine. Further research is needed to address these questions to ensure optimized diets are not only healthy and sustainable but also acceptable to individuals. This involves both understanding the behavioral drivers of consumer acceptance as well as developing methods to quantify and integrate those drivers into diet modeling.

## 4.5 Conclusion

This study demonstrated that within-food-group optimization increases opportunities to improve the nutritional adequacy, sustainability, and acceptability of diets. As such, we recommend optimizing both between and within food groups.

# 4.A Appendix - Calculation of dietary change

Among studies, different metrics are used to calculate dietary change. For example, Vieux et al. (2018) reported the mean absolute quantity variation (%) across 151 food items, referred to here as food groups, using the following equation:

$$\frac{1}{G}\sum_{g=1}^{G}\frac{\left|Q_g^{\text{optimized}} - Q_g^{\text{observed}}\right|}{Q_g^{\text{observed}}}$$

where $Q_g$ represents the quantity of food group $g$ in the observed and optimized diets, and $G$ is the total number of food groups.

Green et al. (2015) displayed the sum of squared percentage deviations (%) for 148 food groups:

$$\sum_{g=1}^{G}\left(\frac{Q_g^{\text{optimized}} - Q_g^{\text{observed}}}{Q_g^{\text{observed}}}\right)^2$$

In this study, we report the total dietary change (%) across 2,734 food items $i$:

$$\frac{\sum_i \left|Q_i^{\text{optimized}} - Q_i^{\text{observed}}\right|}{\sum_i Q_i^{\text{observed}}}$$

This metric was chosen because it is more stable than Vieux et al.'s (2018), where small denominators can cause large percentage differences when calculated at food item level. Additionally, it is easier to interpret than Green et al.'s (2015) metric, as squared changes can be less intuitive.

To compare our modeling results with the studies of Vieux et al. (2018), Green et al. (2015), Reynolds et al. (2019), Nordman et al. (2024), and Kesse-Guyot et al. (2025) we calculated total dietary change for their studies using the food group quantities reported in their manuscript or supplementary materials:

$$\frac{\sum_g \left|Q_g^{\text{optimized}} - Q_g^{\text{observed}}\right|}{\sum_g Q_g^{\text{observed}}}$$

These outcomes are reported in the introduction. For the studies by Perignon et al. (2016), Horgan et al. (2016), Rocabois et al. (2022), and Heerschop et al. (2023), this was not possible, so we reported the outcomes as presented by the original authors.

## 4.B Appendix - Custom food group classification

Following the methodology of Perignon et al. (2016) and Vieux et al. (2013), we began by selecting the most frequently consumed foods in each of the 153 WWEIA food groups (USDA & ARS, 2020a). Selection was based on ranking foods by the number of respondents who consumed them, with the most popular food ranked first. Foods were included if their rank fell between 1 and a threshold calculated as 10% of the total number of items in the food group. In this way, a greater number of foods were selected for groups containing more items. For example, the WWEIA group 'apple juice' contains only three food items. Since 10% of 3 equals 0.3, which rounds up to 1, only the item ranked first was selected. Similarly, in the larger group 'beans, peas, legumes', which contains 61 food items, the threshold allowed for the selection of items ranked between 1 and 7. Overall, this approach led to the selection of 345 foods, which were used as custom groups for assigning the remaining food items to the foods with most similar nutrient profiles, based on their scaled Euclidean distance. The nutrients considered in this process are listed in Table 4C.4. Of the 345 custom groups formed, 32 contained a single food item, while the largest group consisted of 45 items.

# 4.C Appendix - Diet model

## 4.C.1 Notation

**Table 4C.1:** Indices, parameters, and variables.

| Indices | |
|---|---|
| $i$ | Food item $i = 1 \dots I$ |
| $g$ | Food group $g = 1 \dots G$ |
| $n$ | Nutrient $n = 1 \dots N$ |
| $n.macro$ | Subset of macronutrients |
| $n.macro.e$ | Subset of macronutrients where bounds are expressed as percentage of energy intake |
| $n.macro.w$ | Subset of macronutrients where bounds are expressed in weight |
| $n.micro$ | Subset of micronutrients |

| Parameters | |
|---|---|
| $\varepsilon$ | Arbitrary small value |
| $b_{ig}$ | Binary value indicating food item $i$ belongs to food group $g$ (yes: 1 ; no: 0) |
| $c_{in}$ | Nutrient content for food item $i$ and nutrient $n$ |
| $e_i$ | Emission for food item $i$ |
| $e^f$ | Emission of food items that may not be substituted (fixed) |
| $e_g^{fg}$ | Average emission for food group $g$ |
| $ear_n$ | Estimated Average Requirement (EAR) for each micronutrient $n$ |
| $f^b$ | Factor to extend macronutrient bounds |
| $f^d$ | Fraction of allowed quantity deviation per food group |
| $f^e$ | Energy conversion factor |
| $f^r$ | Greenhouse gas emission (GHGE) reduction factor |
| $lb_n$ | Lower bound on intake level for each macronutrient $n$ |
| $o_i$ | Observed food intake for food item $i$ |
| $o^{en}$ | Observed energy intake |
| $o^{em}$ | Observed emission |
| $o_g^{fg}$ | Observed food intake for food group $g$ |
| $o_g^{fg.max}$ | Maximum observed food intake within food group $g$ |
| $o_{ig}^{fg.sh}$ | Observed food intake share of food item $i$ within food group $g$ |
| $rda_n$ | Recommended Daily Allowance (RDA) for each micronutrient $n$ |
| $t_n^f$ | Nutrient intake of food items that may not be substituted (fixed) for nutrient $n$ |
| $ub_n$ | Upper bound on intake level for each macronutrient $n$ |
| $ul_n$ | Tolerable upper intake level for each micronutrient $n$ |

| Variables | |
|---|---|
| $D_n^{ear}$ | Normalized deviation below EAR for micronutrient $n$ |
| $D^{max.ear}$ | Maximum micronutrient deviation from the EAR |
| $D^{max.macro}$ | Maximum macronutrient deviation from dietary guidelines |
| $D^{max.q}$ | Maximum absolute deviation from observed quantity |

| Variables | |
|---|---|
| $D^{max.rda}$ | Maximum micronutrient deviation from the RDA |
| $D_n^{lb}$ | Normalized deviation below lower bound for macronutrient $n$ |
| $D_i^q$ | Absolute deviation from observed quantity for food item $i$ |
| $D_i^{q-}$ | Negative deviation from observed quantity for food item $i$ |
| $D_i^{q+}$ | Positive deviation from observed quantity for food item $i$ |
| $D_n^{rda}$ | Normalized deviation below RDA for micronutrient $n$ |
| $D^{sum.ear}$ | Sum of micronutrient deviations from the EAR |
| $D^{sum.macro}$ | Sum of macronutrient deviations from dietary guidelines |
| $D^{q.sum}$ | Sum of absolute deviations from observed quantity |
| $D^{sum.rda}$ | Sum of micronutrient deviations from the RDA |
| $D_n^{ub}$ | Deviation above upper bound for macronutrient $n$ |
| $E$ | Emission |
| $Q_i$ | Quantity for food item $i$ |
| $Q_g^{fg}$ | Quantity for food group $g$ |
| $T_n$ | Nutrient intake for nutrient $n$ |

### 4.C.2 Base model

The diet model optimizes nutrient and greenhouse gas emission (GHGE) goals by changing food quantities between and within food groups. Different strategies for achieving this are explored through various modeling experiments. This section outlines the constraints of the base model, which are applied across all experiments (Table 4C.2). The following section discusses the additional constraints and objective functions specific to each modeling experiment (Table 4C.3).

Constraints 4C.1 and 4C.2 measure the absolute change in quantity relative to observed consumption. Constraint 4C.3 then calculates the total deviation in food quantities compared to the observed diet, and Constraint 4C.4 determines the maximum deviation, normalized by observed food group quantities. Food group quantities are calculated with Constraint 4C.5.

Next, the nutrient intake of the optimized diet is calculated (Constraint 4C.6), comprising two components: intake from optimized food items and intake from foods excluded from optimization. Micronutrient intake is constrained by the tolerable upper intake levels (Constraint 4C.7), and goals are established according to the Estimated Average Requirement (EAR) (Constraint 4C.8) and Recommended Daily Allowances (RDA) (Constraint 4C.9) (Table 4C.4). Deviations from these goals are normalized to a 0-1 scale for comparison. Constraints 4C.10 and 4C.11 then evaluate the maximum deviations below the EAR and RDA, and Constraints 4C.12 and 4C.13 calculate the total sum of these deviations.

Furthermore, Constraint 4C.14 guarantees that the optimized diet maintains an energy intake similar to the observed diet. Constraints 4C.15 - 4C.18 set the macronutrient goals

in alignment with dietary guidelines, distinguishing between goals expressed as energy percentages (Constraints 4C.15 - 4C.16) and those expressed in weight (Constraints 4C.17 - 4C.18). To ensure model linearity for the energy percentage goals, deviations are normalized using the energy intake of the observed diet. The macronutrient bounds can be relaxed by a factor to make them easier to meet. Constraints 4C.19 and 4C.20 calculate the maximum deviation and total deviations for macronutrient intake.

GHGE of the observed diet are calculated using Constraint 4C.21. As with nutrient intake, emissions are divided into two components: those from optimized food items and those from foods excluded from optimization. Constraint 4C.22 ensures that the emissions from the optimized diet are less than or equal to those of the observed diet.

Finally, Constraints 4C.23 - 4C.26 establish the variable domains.

**Table 4C.2:** Constraints of base diet model.

| Constraints | | |
|---|---|---|
| $Q_i - o_i = D_i^{q+} - D_i^{q-}$ | $\forall i$ | (4C.1) |
| $D_i^q = D_i^{q+} + D_i^{q-}$ | $\forall i$ | (4C.2) |
| $D^{sum.q} = \sum_i D_i^q$ | | (4C.3) |
| $D^{max.q} \geq D_i^q \cdot b_{ig} / o_g^{fg}$ | $\forall i, g$ | (4C.4) |
| $Q_g^{fg} = \sum_i (Q_i \cdot b_{ig})$ | $\forall g$ | (4C.5) |
| $T_n = \sum_i (Q_i \cdot c_{in}) + t_n^f$ | $\forall n$ | (4C.6) |
| $T_n \leq ub_n$ | $\forall n \in n.micro$ | (4C.7) |
| $T_n + ear_n \cdot D_n^{ear} \geq ear_n$ | $\forall n \in n.micro$ | (4C.8) |
| $T_n + rda_n \cdot D_n^{rda} \geq rda_n$ | $\forall n \in n.micro$ | (4C.9) |
| $D^{max.ear} \geq D_n^{ear}$ | $\forall n \in n.micro$ | (4C.10) |
| $D^{max.rda} \geq D_n^{rda}$ | $\forall n \in n.micro$ | (4C.11) |
| $D^{sum.ear} = \sum_n D_n^{ear}$ | | (4C.12) |
| $D^{sum.rda} = \sum_n D_n^{rda}$ | | (4C.13) |
| $T_{Energy} = o^{en}$ | | (4C.14) |
| $T_n + D_n^{lb} \cdot lb_n \cdot (1 - f^b) \cdot o^{en} / f^e \geq lb_n \cdot (1 - f^b) \cdot T_{Energy} / f^e$ | $\forall n \in n.macro.e$ | (4C.15) |
| $T_n - D_n^{ub} \cdot ub_n \cdot (1 + f^b) \cdot o^{en} / f^e \leq ub_n \cdot (1 + f^b) \cdot T_{Energy} / f^e$ | $\forall n \in n.macro.e$ | (4C.16) |
| $T_n + D_n^{lb} \cdot lb_n \cdot (1 - f^b) \geq lb_n \cdot (1 - f^b)$ | $\forall n \in n.macro.w$ | (4C.17) |
| $T_n - D_n^{ub} \cdot ub_n \cdot (1 + f^b) \geq ub_n \cdot (1 + f^b)$ | $\forall n \in n.macro.w$ | (4C.18) |
| $D^{max.macro} \geq D_n^{lb} + D_n^{ub}$ | $\forall n \in n.macro$ | (4C.19) |
| $D^{sum.macro} = \sum_n \left( D_n^{lb} + D_n^{ub} \right)$ | | (4C.20) |
| $E = \sum_i (Q_i \cdot e_i) + e^f$ | | (4C.21) |
| $E \leq o^{em}$ | | (4C.22) |
| $D^{max.ear}, D^{max.macro}, D^{max.rda}, D^{max.q}, D^{sum.ear}, D^{sum.macro}, D^{sum.rda}, D^{sum.q}, E \in \mathbb{R}_{\geq 0}$ | | (4C.23) |
| $D_n^{ear}, D_n^{lb}, D_n^{rda}, D_n^{ub}, T_n \in \mathbb{R}_{\geq 0}$ | $\forall n$ | (4C.24) |
| $D_i^q, D_i^{q-}, D_i^{q+}, Q_i \in \mathbb{R}_{\geq 0}$ | $\forall i$ | (4C.25) |
| $Q_g^{fg} \in \mathbb{R}_{\geq 0}$ | $\forall g$ | (4C.26) |

### 4.C.3 Modeling experiments

*Within-food-group optimization*

In the first modeling experiment, we examine the potential to improve the nutritional adequacy and greenhouse gas emissions (GHGE) of observed diets by adjusting food item quantities within food groups only. This means that the total quantity of each food group remains fixed, as in the observed diet (Constraint 4C.27, Table 4C.3). We evaluate this under varying levels of allowed dietary change. In the most restrictive scenario (0% change), the diet remains identical to the observed diet. In the least restrictive scenario (100% change), all food item quantities within each food group may be freely adjusted. For intermediate scenarios, we constrain the absolute change within each food group to x% ($f^d$) or less (Constraint 4C.28). Additionally, no food item may exceed the highest quantity observed within its respective food group (Constraint 4C.29). In the objective function, different goals are weighted, with the highest weight assigned to minimizing the largest deviation from recommended macro- and micronutrient intake levels (RDA), followed by the total sum of nutrient deviations, then GHGE reduction, and finally, minimizing overall dietary change (Constraint 4C.30).

*Lowered nutrient goals*

In the second modeling experiment, which builds on the first, we investigate to what extent lowering nutrient goals can further reduce GHGE. Here, macronutrient bounds are relaxed by a specified factor ($f^b$, Constraints 4C.15 - 4C.18 , Table 4C.2), and in the objective function, the micronutrient goals are lowered from the RDA to the EAR (Constraint 4C.31, Table 4C.3).

*Between-food-group vs. Between-and-within-food-group optimization*

In the last modeling experiment, we investigate the minimal amount of dietary change required to achieve specified nutrient (Constraints Constraints 4C.32 - 4C.33, Table 4C.3) and GHGE (Constraint 4C.34) goals. We compare the required dietary change under two scenarios: one where only total food group quantities may be adjusted—keeping the composition of food items within each group unchanged (Constraint 4C.35)—and another where food item quantities can be changed without restriction. Additionally, we compare the effects of calculating GHGE at the food group level (Constraint 4C.36) versus the food item level (Constraint 4C.21, Table 4C.2). In the objective function, the highest priority is given to minimizing dietary change, followed by further reducing GHGE (Constraint 4C.37, Table 4C.3).

The diet models were run in Python 3.11.4 with Gurobi 11.0.1, while all other data preparation and analyses were conducted in R 4.2.1.

**Table 4C.3:** Additional constraints and objective function for each modeling experiment.

Within-food-group optimization

$$Q_g^{fg} = o_g^{fg} \qquad \forall g \quad \text{(4C.27)}$$

$$\sum_i (D_i^q \cdot b_{ig}) \leq f^d \cdot o_g^{fg} \qquad \forall g \quad \text{(4C.28)}$$

$$\sum_i (D_i^q \cdot b_{ig}) \leq o_g^{fg.max} \qquad \forall i, g \quad \text{(4C.29)}$$

$$\min \Big\{ D^{max.macro} + D^{max.rda} + \varepsilon_1 \left( D^{sum.macro} + D^{sum.rda} \right) + \varepsilon_2 \cdot E + \varepsilon_3 \cdot D^{sum.q} \Big\} \quad \text{(4C.30)}$$

Lowered nutrient goals

$$\min \Big\{ D^{max.macro} + D^{max.ear} + \varepsilon_1 \left( D^{sum.macro} + D^{sum.ear} \right) + \varepsilon_2 \cdot E + \varepsilon_3 \left( D^{max.rda} + D^{sum.rda} \right) + \varepsilon_4 \cdot D^{sum.q} \Big\} \quad \text{(4C.31)}$$

Between-food-group vs. Between-and-within-food-group optimization

$$D^{max.macro} \leq 0 \quad \text{(4C.32)}$$

$$D^{max.rda} \leq 0 \quad \text{(4C.33)}$$

$$1 - E/o^{em} \geq f^r \quad \text{(4C.34)}$$

$$Q_i = \sum_g \left( Q_g^{fg} \cdot o_{ig}^{fg.sh} \right) \qquad \forall i \quad \text{(4C.35)}$$

$$E = \sum_g \sum_i \left( Q_i \cdot b_{ig} \cdot e_g^{fg} \right) + e^f \quad \text{(4C.36)}$$

$$\min \{ D^{max.q} + 0.001 \cdot D^{sum.q} + \varepsilon \cdot E \} \quad \text{(4C.37)}$$

### 4.C.4 Parameter input

**Table 4C.4:** Dietary reference intakes (women and men 19–30y) for nutrients considered in the diet models (NIH, 2024; USDA & US HHS, 2020).[1]

| | | LB | | UB | |
|---|---|---|---|---|---|
| Macronutrient | Unit | Female | Male | Female | Male |
| Energy[3] | kcal/day | 1800 | 2400 | 2400 | 3000 |
| Protein | E% | 10 | 10 | 35 | 35 |
| Total carbohydrates | E% | 45 | 45 | 65 | 65 |
| Dietary fibre | g | 25 | 38 | — | — |
| Total fat | E% | 20 | 20 | 35 | 35 |
| Total saturated fatty acids | E% | — | — | 10 | 10 |
| Sodium | mg | — | — | 2300 | 2300 |

| | | EAR | | RDA | | |
|---|---|---|---|---|---|---|
| Micronutrient[3] | Unit | Female | Male | Female | Male | UL |
| Calcium | mg | 800 | 800 | 1000 | 1000 | 2500 |
| Copper | mg | 0.7 | 0.7 | 0.9 | 0.9 | 10 |
| Iron | mg | 8.1 | 6 | 18 | 8 | 45 |
| Magnesium | mg | 255 | 330 | 310 | 400 | — |
| Selenium | $\mu$g | 45 | 45 | 55 | 55 | 400 |
| Zinc | mg | 6.8 | 9.4 | 8 | 11 | 40 |
| Vitamin A | $\mu$g | 500 | 625 | 700 | 900 | 3000 |
| Vitamin B1 | mg | 0.9 | 1 | 1.1 | 1.2 | — |
| Vitamin B2 | mg | 0.9 | 1.1 | 1.1 | 1.3 | — |
| Vitamin B3 | mg | 11 | 12 | 14 | 16 | — |
| Vitamin B6 | mg | 1.1 | 1.1 | 1.3 | 1.3 | 100 |
| Vitamin B9 | $\mu$g | 320 | 320 | 400 | 400 | 1000 |
| Vitamin B12 | $\mu$g | 2 | 2 | 2.4 | 2.4 | — |
| Vitamin C | mg | 60 | 75 | 75 | 90 | 2000 |
| Vitamin E | mg | 12 | 12 | 15 | 15 | — |

[1] Abbreviations: LB, Lower Bound; UB, Upper Bound; EAR, Estimated Average Requirement; RDA, Recommended Daily Allowance; UL, tolerable Upper Level.

[2] Energy reference intakes for a sedentary (LB) and active (UB) lifestyle.

[3] The diet model does not optimize vitamin D intake because guidelines are based on both food intake and sunlight exposure.

# Chapter 5

## Using diet optimization and machine learning for the design of healthy and acceptable menu plans



This chapter is published as:

van Wonderen, D., Gerdessen, J. C., Melse-Boonstra, A., & Onwezen, M. C. (2026). Using diet optimization and machine learning for the design of healthy and acceptable menu plans. *European Journal of Operational Research*, 328(2), 668-679.

# Abstract

The success of dietary plans relies on understanding and modelling consumer acceptance, yet quantifying this poses a challenge due to the complexity of individual preferences. Recent research is focused on deriving acceptability constraints directly from data, as demonstrated by its application in designing food baskets with a limited number of commodities. In this study, we applied diet optimization with machine learning to the more complex task of menu planning. This involved considering hundreds of potential food alternatives and assessing their compatibility within a meal using a recipe completion algorithm. Compared to the traditional diet modelling approach of food group filtering, the recipe completion model delivered diets with either higher nutritional adequacy or greater substitute acceptability, depending on the number of food groups used in the traditional method. While more research is needed to further improve the acceptability of substitutions, combining diet optimization with recipe completion presents a promising approach to enhance the nutritional adequacy of individual diets while maintaining the acceptability of food combinations within meals.

## 5.1 Introduction

Consumer acceptance is crucial in the shift from current to sustainable healthy diets because the success of any desired change in dietary behavior ultimately depends on the willingness of individuals to adopt and maintain those new dietary habits (Bos et al., 2013). However, including consumer acceptance into diet modelling — the process of creating mathematical models to design and optimize dietary plans — has posed a challenge due to the difficulty to quantify and translate the inherent variability in individuals' preferences into specific constraints (van Dooren, 2018). When underlying mechanisms are hard to capture, there is the possibility to derive constraints directly from data. Using data-driven methods to formulate consumer acceptability constraints has the advantage of enabling greater personalization while reducing the need for manual input (Fajemisin et al., 2024).

Data-driven methods to model consumer acceptability are rarely applied in diet modelling, despite their widespread use in related domains such as food recommender systems (Tran et al., 2021), food substitution algorithms (Shirai et al., 2021), and recipe completion algorithms (Lei et al., 2020). Within the field of diet modelling, one of the few studies that adopted a data-driven approach is the work by Fajemisin et al. (2024), who demonstrated, through a theoretical example, how integrating diet modelling with machine learning techniques could be applied to estimate the acceptability of daily food baskets (Maragno et al., 2021). However, in this example, the machine learning model was trained on simulated acceptability scores of food baskets containing only a limited number of commodities. The effectiveness of this approach has not yet been tested in a real-world setting, particularly in the more complex context of menu planning, which requires accounting for the compatibility of food items within meals.

The objective of this study is to contribute to the development of data-driven consumer acceptance modelling by integrating the domain of diet modelling with related food recommendation domains, and by evaluating this approach in a real-world setting through the task of menu planning. Concretely, we apply the recipe completion algorithm described by De Clercq et al. (2016) for the following reasons: recipe completion algorithms are specifically designed to generate food suggestions that maintain coherence with other items in a meal (Lei et al., 2020); recipe completion algorithms support multiple use cases, namely the substitution of food items as well as the addition of new items (in this study, we focus only on food substitution); the linear recipe completion algorithm proposed by De Clercq et al. (2016) can be easily implemented within diet modelling and has demonstrated competitive performance compared to more complex approaches (Cueto et al., 2019; Chen et al., 2020). Furthermore, we compare this data-driven approach to more conventional diet modelling strategies, in which food alternatives are selected based on food group filtering and item popularity. As a case study, we design menu plans for vegetarian women of reproductive age using data from the National Health and Nutrition

Examination Survey 2013-2018 (NHANES) (CDC & NCHS, 2018). This demographic group was selected because they are at a relatively higher risk of nutritional deficiencies (Phillips, 2005; Bailey et al., 2015; Stevens et al., 2022).

## 5.2 Background

In this section, we review the literature on addressing consumer acceptance in diet modelling (Section 5.2.1). Additionally, we discuss various algorithms developed for recipe completion (Section 5.2.2). To maintain conciseness, we limit our review to the research fields directly covered in this study and do not elaborate on food recommender systems or food substitution algorithms.

### 5.2.1 Optimizing consumer acceptance in diet modelling

Diet modelling is a widely employed technique for the design of healthy, affordable, and sustainable diets. However, without considering consumer acceptance, model outcomes are unlikely to be deemed optimal in the eyes of those they aim to serve. Over the years, literature has presented several strategies to enhance the acceptability of designed diets (van Dooren, 2018). For example, Masset et al. (2009) modelled consumer acceptance by minimizing differences in food consumption between observed and optimized diets and by imposing constraints on allowed ingredient quantities within different food groups. Similarly, Maillot et al. (2010) penalized deviations from the observed diet and applied restrictions on food intake, aligned with the 95th percentile of national consumption data. Dussiot et al. (2022b) employed a multi-objective programming approach to maximize nutritional adequacy and to minimize deviations from the current diet. Also in this case, acceptability constraints were applied to each food group, ensuring that food intake remained within the observed consumption range across various subpopulations. An alternative approach involves modelling with recipes, combined with constraints for food groups and items (Leung et al., 1995; Benvenuti & De Santis, 2020).

Recent research is focused on directly deriving acceptability constraints from data. One data-driven approach presented in literature is inverse optimization. In contrast to standard optimization, which begins with the formulation of an objective function and constraints, inverse optimization starts with an observed solution to deduce the corresponding objective function or constraints. Ghobadi & Mahmoudzadeh (2021) used inverse optimization to learn constraints on consumer preferences based on past dietary patterns considering twenty-six food items. Incorporating the constraints identified through inverse optimization resulted in food baskets that were more closely aligned with past consumption patterns. Conversely, baskets designed without these constraints demonstrated less food variety and included a higher proportion of food items not regularly consumed. Ahmadi et al. (2020) proposed a method, denoted as “inverse learning”, to identify the cost function for individuals’ food preferences. They concluded that the designed diets,

which allowed for the selection of thirty-eight different food groups, could be well-tailored to personal preferences.

Another data-driven approach presented in literature is optimization with machine learning. To the best of our knowledge, the study by Fajemisin et al. (2024) is the only one that utilized optimization with machine learning to quantify the acceptability of food baskets. The acceptability of food baskets was estimated using various machine learning algorithms that were trained on a dataset consisting of twenty-five food commodities and simulated acceptability scores (Maragno et al., 2021).

Although literature provides a number of promising data-driven techniques for modelling consumer acceptance, these methods have only been evaluated in scenarios with a limited number of food groups or items, and without considering the coherence of selected food items within a meal. Therefore, in our study, we combine optimization with machine learning to design healthy menu plans, considering hundreds of food items, where we use recipe completion to quantify the acceptance of food alternatives.

### 5.2.2 Recipe completion algorithms

Recipe completion algorithms are designed to generate food suggestions, considering the compatibility of these suggestions with other food items in a meal (Lei et al., 2020). One frequently cited study in this research domain is the research conducted by De Clercq et al. (2016). They applied two relatively straightforward machine learning methods for recipe completion: non-negative matrix factorization and two-step regularized least squares (Pahikkala et al., 2014). Their methodology involved leveraging data on both flavor components and the co-occurrence of ingredients within recipes. Most favorable outcomes were obtained for the two-step regularized least squares method, particularly when relying solely on co-occurrence data. When using a single dataset, this method can be considered as linear kernel ridge regression.

Subsequent to the research conducted by De Clercq et al. (2016), there have been several other research attempts to improve modelling results. Cueto et al. (2019) used memory-based collaborative filtering, a technique originating from the field of recommender systems, for the completion of partial recipes. While their results demonstrated lower performance compared to the findings of De Clercq et al. (2016), the authors expressed confidence that this performance gap could be narrowed through enhancements to the dataset's quality. Chen et al. (2020) developed a new embedding-based ingredient predictor to design nutritious recipes. Their approach outperformed non-negative matrix factorization, although it did not surpass the performance of the two-step regularized least squares method as showcased by De Clercq et al. (2016). Lei et al. (2020) tested the same techniques as those used by De Clercq et al. (2016) but took a different approach to the data and their representation in model training. Firstly, instead of flavor data, they employed nutrient data. Secondly, rather than utilizing a binary representation to

indicate ingredient presence or absence, they transformed the recipe data using an entropy density function. This entropy density function was used to express the role of each ingredient within a recipe, determined by its nutrient profile. The introduction of this new data representation led to noticeable improvements in results for the different recipe datasets tested by Lei et al. (2020). However, it did not outperform the results achieved by De Clercq et al. (2016) through the application of linear kernel ridge regression. It is worth noting though that this difference may be attributed to the datasets employed, as they significantly influenced the final performance measures (Lei et al., 2020).

All in all, the recipe completion method proposed by De Clercq et al. (2016) stands as a pivotal reference in the literature, notable for its straightforward design and robust performance compared to the alternative approaches. For these reasons, we have chosen to apply linear kernel ridge regression to quantify the acceptance of food alternatives. By integrating recipe completion with diet optimization, this study aims to contribute to the development of data-driven consumer acceptance modelling.

## 5.3 Methodology

For the design of healthy and acceptable vegetarian menu plans, we developed a novel approach that integrates recipe completion with diet optimization. We used recipe completion to identify substitutes that go well together with the remaining food items within a meal (Section 5.3.1) and diet modelling to replace food items with healthier alternatives (Section 5.3.2). The data used for this research is described in Section 5.3.3.

### 5.3.1 Recipe completion model

Following the example of De Clercq et al. (2016), we applied linear kernel ridge regression to quantify the compatibility of potential substitutes with other food items within a meal. First, we describe how these substitution scores were calculated, followed by an explanation of how we evaluated the model's performance.

#### *5.3.1.1 Calculation of food subsitutes scores*

Food substitutes scores were calculated as follows:

$$\hat{Y} = ZM \tag{5.1}$$

where matrix $\hat{Y}$ with entries $\hat{y}_{mi}$ contains the substitution scores: for each meal $m$, score $\hat{y}_{mi}$ indicates how well food item $i$ goes together with the food items already present in meal $m$, denoted by binary meal matrix $Z$ with entries $z_{mi}$ ($z_{mi} = 1$ if meal $m$ contains food item $i$). A positive substitution score $\hat{y}_{mi}$ implies that most ingredients are frequently consumed together, while a negative score implies that the ingredients are not

or not frequently consumed in the same meal. The symmetric coefficient matrix $M$, with entries $\mu_{ij}$ $(i = j)$, contains the scores of all food item pairings. For a given food item $i$, substitution score $\hat{y}_{mi}$ equals the sum of $\mu_{ij}$ over all food items $j$ present in meal $z_{mj}$. Matrix $M$ is estimated by:

$$M = X^T(XX^T + \lambda I)^{-1}Y \tag{5.2}$$

where $X$ and $Y$ are the input matrix and the response matrix respectively with entries $x_{mi}$ and $y_{mi}$, and $\lambda$ is a regularization term used to prevent overfitting. Both $X$ and $Y$ are binary meal matrices used for training the recipe completion model, and are in this case identical to each other.

To give the prevention of uncommon food combinations more weight, we transformed $M$ by multiplying the negative elements with a positive factor $\phi$:

$$M' = \begin{cases} \phi \cdot \mu_{ij}, & \text{if } \mu_{ij} < 0 \\ \mu_{ij}, & \text{if } \mu_{ij} \geq 0 \end{cases} \tag{5.3}$$

*5.3.1.2 Performance evaluation*

To evaluate the model performance, the meal dataset was split into training (80%) and test (20%) datasets (5.4). K-fold cross validation (K=5) was applied on the training set to determine the optimal values for $\lambda$ (5.5) and $\phi$ (5.3).

$$\hat{Y}_{test} = X_{test}X_{train}^T(X_{train}X_{train}^T + \lambda_{opt}I)^{-1}X_{train} \tag{5.4}$$

$$\hat{Y}_K = X_K X_{(train-K)}^T(X_{(train-K)}X_{(train-K)}^T + \lambda I)^{-1}X_{(train-K)} \tag{5.5}$$

Model performance was measured based on its capability of retrieving a food item that was removed from a meal (De Clercq et al., 2016). Therefore, for each meal in the test dataset, we randomly removed one food item. The model, trained on the training dataset with $\lambda_{opt}$ and $\phi_{opt}$, returned a list of substitution scores for all food items. Substitution scores for food items that were already present in the meal were ignored. The rank of the removed food item within the ordered list of substitute candidates was used as an indicator of how well the model is able to complete a meal. The median of the rankings of the removed items was also used to select $\lambda_{opt}$ and $\phi_{opt}$ during the K-fold cross validation step. Additionally, the percentage of meals for which the removed food item was ranked within the top 10 was used as a measure for model performance.

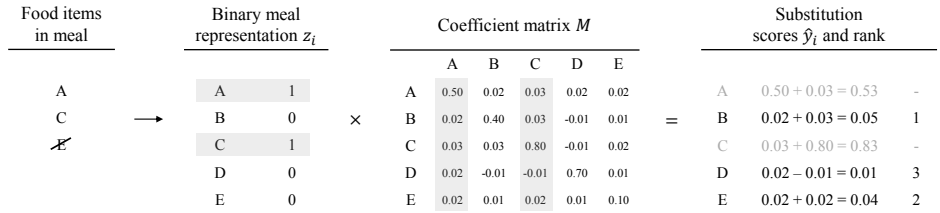



**Figure 5.1:** Numerical example of recipe completion for a single meal containing food items A,C, and E, with item E removed. The ranking of the removed item is used to measure model performance. The better the performance, the higher the ranking (with 1 being the best). In this example, removed food item E was ranked second-best.

Finally, the recipe completion model that was integrated into the diet model was trained on both training and test datasets with $\lambda_{opt}$ (5.1-5.2) and $\phi_{opt}$ (5.3). Figure 5.1 provides an overview of the functionality of the recipe completion model.

### 5.3.2 Diet modelling

Two diet models were developed to improve the healthiness of meals by substituting food items. Our approach replaced food items using the recipe completion algorithm (RC), while the other model followed the traditional approach by selecting substitutes based on food group filtering and food popularity (FGF). Notation and model formulations are presented below.

#### *5.3.2.1 Notation*

| Sets | |
|---|---|
| $c \in \mathcal{C}$ | Classes for the number of food items within a meal |
| $g \in \mathcal{G}$ | Food groups |
| $i, j \in \mathcal{I}$ | Food items |
| $m \in \mathcal{M}$ | Meals |
| $n \in \mathcal{N}$ | Nutrients and energy |
| $n \in \mathcal{N}^1$ | Macronutrients |
| $n \in \mathcal{N}^{1'}$ | Macronutrients expressed as percentage of energy intake |
| $n \in \mathcal{N}^{1''}$ | Other macronutrients |
| $n \in \mathcal{N}^2$ | Micronutrients |
| $s \in \mathcal{S}$ | Substitution round |

Parameters

| | |
|---|---|
| $b^c_{mc}$ | Binary value indicating meal $m$ belongs to number of food item class $c$ (1) or not (0) |
| $b^g_{gi}$ | Binary value indicating food item $i$ belongs to food group $g$ (1) or not (0) |
| $b^m_{mi}$ | Binary value indicating food item $i$ belongs to meal $m$ (1) or not (0) |
| $c_{in}$ | Nutrient content (mg or $\mu$g per g) for micronutrient $n$ and food item $i$ |
| $c^e_n$ | Total nutrient content of food items excluded from substitution (Section 5.3.3.2) |
| $e$ | Average observed energy intake (kcal/day) of respondent |
| $f^e_n$ | Factor to convert macronutrient $n$ intake (g) to energy intake (kcal) |
| $f^s$ | Factor to normalise the substitution scores based on the number of food items per meal |
| $lb_n$ | Lower bound (g) for macronutrient $n$ |
| $lb^s$ | Lower bound for substitution score |
| $n^d$ | Number of consumption days |
| $n^f_i$ | Food popularity score based on the count of food item $i$ divided by the total number of food items in the consumption dataset. Food items are ranked per food group. |
| $n^i_m$ | Number of food items per meal $m$ |
| $n^s_m$ | Maximum number of allowed substitutions per meal $m$ |
| $q^s_{ic}$ | Substitution quantity of food item $i$ |
| $q^o_{mi}$ | Quantity (g) of food item $i$ in observed meal $m$ |
| $rcM_{ii}$ | Recipe completion model (equation 5.3) |
| $rda_n$ | Recommended Daily Allowance (RDA) for micronutrient $n$ (mg or $\mu$g) |
| $ub_n$ | Upper bound (g) for macronutrient $n$ |
| $ul_n$ | Tolerable Upper Level (UL) for micronutrient $n$ (mg or $\mu$g) |
| $w$ | Weight of health objective |

Variables

| | |
|---|---|
| $A_{msi}$ | Binary value indicating food item $i$ will be added to meal $m$ in substitution round $s$ (1) or not (0) |
| $D^{macro}$ | Maximum deviation from macronutrient guidelines, $\max_n\{D^{lb}_n + D^{ub}_n\}$ |
| $D^{micro}$ | Maximum deviation below the RDA, $\max_n\{D^{rda}_n\}$ |
| $D^{lb}_n$ | Normalized deviation below macronutrient lower bound $n$ |
| $D^{rda}_n$ | Normalized deviation from the RDA for micronutrient $n$. A positive value means the average intake is below the RDA. |
| $D^{ub}_n$ | Normalized deviation above macronutrient upper bound $n$ |
| $F_{mi}$ | Binary value indicating the final meal $m$ contains food item $i$ (1) or not (0) |

| Variables | |
|---|---|
| $I_n$ | Average intake (mg or $\mu$g per day) of micronutrient $n$ |
| $K_{msi}$ | Binary value indicating food item $i$ will be kept in meal $m$ for substitution round $s$ (1) or not (0) ($Z$ in equation 5.1) |
| $K^o_{mi}$ | Binary value indicating food item $i$ of observed meal $m$ will be kept (1) or not (0) |
| $L_{ms}$ | Binary value indicating $S^a$ is larger than $N$ for meal $m$ and substitution round $s$ |
| $N_{ms}$ | Binary value equal to 1 if no food items have been substituted in meal $m$ in substitution round $s$ |
| $Q_{msi}$ | Quantity (g) of food item $i$ added to meal $m$ in substitution round $s$. Food items that are not selected receive a quantity of 0. |
| $Q^f_{mi}$ | Quantity (g) of all food items $i$ in final meal $m$ |
| $R_{msi}$ | Binary value indicating food item $i$ is replaced in meal $m$ in substitution round $s$ |
| $S_{msi}$ | Score that indicates how well food item $i$ goes together with the remaining food items in meal $m$ for substitution round $s$ ($Y$ in equation 5.1) |
| $S^a_{msi}$ | Score of food item $i$ added to meal $m$ in substitution round $s$ ($\hat{Y}$ in equation 5.1). Food items that are not selected receive a score of 0. |
| $S^{max}_{ms}$ | Maximum value of $S^a$ and $N$ for meal $m$ and substitution round $s$, $\max_{ms}\{\sum_i S^a_{msi}, N_{msi}\}$ |
| $S^{min}$ | Lowest substitution score of $S^{max}$ for all meals $m$ and substitution rounds $s$, $\min_{ms}\{S^{max}_{ms}\}$ |
| $S^n_{ms}$ | Normalised substitution score of meal $m$ in substitution round $s$ |

#### *5.3.2.2 Diet model based on recipe completion (RC)*

The diet model has two objectives: 1) maximizing the nutritional value of the respondent's diet, and 2) maximizing consumer acceptance. The objective function (5.6) maximizes a weighted sum of these objectives by minimizing the largest deviation from macro- and micronutrient guidelines, and by maximizing the lowest substitution score.

$$max\left\{ -w \cdot (D^{macro} + D^{micro}) + (1-w) \cdot S^{min} \right\} \tag{5.6}$$

Constraints 5.7-5.10 set the rules which food items may be replaced by which substitutes within a single meal. Only food items present in the observed meal can be substituted (5.7). For modelling purposes, substitution takes place in rounds. Per substitution round, at most one food item of the observed meal can be replaced (5.8). The number of substitution rounds depends on the number of allowed substitutions for each meal. For example, when 50% of the food items may be replaced, this implies that the number of substitution

rounds is 2 for a meal with 4 food items and 3 for a meal with 7 food items. If a food item is removed in substitution round $s$, there should also be a substitute and vice versa (5.9). Furthermore, food item $i$ can only be added once per meal, and the addition of a food item that is already present in the observed meal is not allowed (5.10).

$$\sum_{s=1}^{S} R_{msi} \leq b_{mi}^{m} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \qquad (5.7)$$

$$\sum_{i=1}^{I} R_{msi} \leq 1 \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n_m^s\} \qquad (5.8)$$

$$\sum_{i=1}^{I} R_{msi} = \sum_{i=1}^{I} A_{msi} \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n_m^s\} \qquad (5.9)$$

$$b_{mi}^{m} + \sum_{s=1}^{S} A_{msi} \leq 1 \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \qquad (5.10)$$

How well each potential substitute goes together with the remaining food items is calculated using the recipe completion model (5.11-5.13). Constraints 5.11 and 5.12 are used to identify the remaining food items, while the corresponding substitution scores are calculated in Constraint 5.13. Next, substitution scores are normalized since the scores are positively correlated with the number of food items within a meal (5.14). For the objective function, only the normalized score of the selected substitute is considered (5.15-5.17). Constraint 5.18 ensures that the score of the substitute exceeds a certain lower bound to prevent the selection of inappropriate substitutes. If no substitute is selected, the score equals 1, which is the highest score possible (5.19-5.23). Consequently, the diet model tends to preserve the observed diet with as few changes as possible. Constraint 5.24 calculates the lowest substitution score among all substitutions. Constraint 5.25 returns the final meal.

$$K_{m1i} = b_{mi}^{m} - R_{m1i} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \qquad (5.11)$$

$$K_{msi} = K_{m,s-1,i} + A_{m,s-1,i} - R_{msi} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{2, ..., n_m^s\} \qquad (5.12)$$

$$S_{msi} = \sum_{j=1}^{I} (K_{msj} \cdot rcM_{ji}) \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \qquad (5.13)$$

$$S_{msi}^{n} = S_{msi} - n_m^i \cdot f^s + lb^s \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \qquad (5.14)$$

$$S_{msi}^{a} \leq A_{msi} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \qquad (5.15)$$

$$S_{msi}^{a} \geq S_{msi}^{n} - (1 - A_{msi}) \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \qquad (5.16)$$

$$S_{msi}^{a} \leq S_{msi}^{n} + (1 - A_{msi}) \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \qquad (5.17)$$

$$S^a_{msi} \geq lb^s - (1 - A_{msi}) \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n^s_m\} \quad (5.18)$$

$$\sum_{i=1}^{I} A_{msi} + N_{ms} = 1 \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.19)$$

$$S^{max}_{ms} \geq \sum_{i=1}^{I} S^a_{msi} \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.20)$$

$$S^{max}_{ms} \geq N_{ms} \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.21)$$

$$S^{max}_{ms} \leq \sum_{i=1}^{I} S^a_{msi} + (1 - L_{ms}) \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.22)$$

$$S^{max}_{ms} \leq N_{ms} + L_{ms} \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.23)$$

$$S^{min} \leq S^{max}_{ms} \qquad \forall m \in \mathcal{M}, s \in \{1, ..., n^s_m\} \quad (5.24)$$

$$F_{ms} = K_{m,n^s_m,i} + A_{m,n^s_m,i} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \quad (5.25)$$

The quantity of added food items is determined based on the average intake per meal, as observed in the meal dataset, taking into account the number of items in each meal (5.26). This adjustment is made because some foods are consumed in smaller quantities when meals contain a larger number of food items. The quantity of food items that are not replaced (5.27-5.28) remains unchanged. Constraint 5.29 gives the quantity of all food items included in the final meal.

$$Q_{msi} = \sum_{c=1}^{C} (q^s_{ic} \cdot b^c_{mc} \cdot R_{msi}) \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n^s_m\} \quad (5.26)$$

$$b^m_{mi} + F_{mi} - 2 \cdot K^o_{mi} \geq 0 \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \quad (5.27)$$

$$b^m_{mi} + F_{mi} - 2 \cdot K^o_{mi} \leq 1 \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \quad (5.28)$$

$$Q^f_{mi} = q^o_{mi} \cdot K^o_{mi} + \sum_{s=1}^{S} Q_{msi} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I} \quad (5.29)$$

The nutrient intake is calculated by summing the nutrient content of the final meals and the nutrient content of that part of observed consumption that was considered to be beyond the scope of food substitution in this research (e.g. mixed dishes, snacks, and drinks, see Section 5.3.3.1). The average nutrient intake per day is calculated by dividing the total nutrient intake by the number of days of observed consumption (5.30).

$$I_n = \sum_{m=1}^{M} \sum_{i=1}^{I} (Q_{mi}^f \cdot c_{in} + c_n^e) \,/\, n^d \qquad \forall n \in \mathcal{N} \quad (5.30)$$

In the objective function, the highest normalized deviations from macronutrient and micronutrient guidelines are minimized (5.6). For the macronutrient guidelines that are expressed as a percentage of daily energy intake (Table 5.2), the bounds are converted to grams, as the average nutrient intake is measured in grams (5.31-5.32). To maintain linearity in the model, these deviations are normalized using the daily energy intake of the original diet. Constraints (5.33-5.34) establish goals for the other macronutrients, while deviations below the Recommended Daily Allowance (RDA) for micronutrients are managed in Constraint (5.35). Constraints (5.36) and (5.37) then identify the largest deviation from dietary guidelines. Finally, to prevent unhealthily high intake of micronutrients, nutrient intakes are limited by the tolerable Upper Levels (UL) (5.38).

$$I_n + D_n^{lb} \cdot lb_n \cdot e \,/\, f_n^e \geq lb_n \cdot I_{Energy} \,/\, f_n^e \qquad \forall n \in \mathcal{N}^{1'} \quad (5.31)$$

$$I_n - D_n^{ub} \cdot ub_n \cdot e \,/\, f_n^e \leq ub_n \cdot I_{Energy} \,/\, f_n^e \qquad \forall n \in \mathcal{N}^{1'} \quad (5.32)$$

$$I_n + D_n^{lb} \cdot lb_n \geq lb_n \qquad \forall n \in \mathcal{N}^{1''} \quad (5.33)$$

$$I_n - D_n^{ub} \cdot ub_n \leq ub_n \qquad \forall n \in \mathcal{N}^{1''} \quad (5.34)$$

$$I_n = rda_n(1 - D_n^{rda}) \qquad \forall n \in \mathcal{N}^2 \quad (5.35)$$

$$D^{macro} \geq D_n^{lb} + D_n^{ub} \qquad \forall n \in \mathcal{N}^1 \quad (5.36)$$

$$D^{micro} \geq D_n^{rda} \qquad \forall n \in \mathcal{N}^2 \quad (5.37)$$

$$I_n \leq ul_n \qquad \forall n \in \mathcal{N}^{\in} \quad (5.38)$$

The last constraints represent the variable domains (5.39-5.43).

$$A_{msi}, F_{mi}, K_{msi}, K_{msi}^o, L_{ms}, N_{ms}, R_{msi} \in \{1, 0\} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \mathcal{S} \quad (5.39)$$

$$S_{msi}, S_{ms}^n \in [-1, 1] \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \mathcal{S} \quad (5.40)$$

$$S_{msi}^a, S_{ms}^{max}, S^{min} \in [0, 1] \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \mathcal{S} \quad (5.41)$$

$$D^{macro}, D^{micro}, I_n, Q_{msi}, Q_{mi}^f \in \mathbb{R}_{\geq 0} \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \mathcal{S}, n \in \mathcal{N} \quad (5.42)$$

$$D_n^{lb}, D_n^{rda}, D_n^{ub} \in \mathbb{R} \qquad \forall n \in \mathcal{N} \quad (5.43)$$

#### *5.3.2.3 Diet model based on food group filtering and food popularity (FGF)*

The formulation of the FGF diet model is similar to that of the RC model, except for the constraints used to calculate and bound the substitution score (5.13-5.18). These constraints are replaced by 5.44-5.45. Only food items from the same food group may substitute each other (5.44), and the substitution score is based on the frequency that food items are present in the meal dataset (5.45). Food items outside the top 30 of most frequently consumed items per food group received a negative score to prevent these food items from being selected as substitutes.

$$\sum_{i=1}^{I}(R_{msi} \cdot b_{gi}^{g}) = \sum_{i=1}^{I}(A_{msi} \cdot b_{gi}^{g}) \qquad \forall m \in \mathcal{N}, s \in \{1, ..., n_m^s\}, g \in \mathcal{G} \quad (5.44)$$

$$S_{msi}^{a} = A_{msi} \cdot n_i^f \qquad \forall m \in \mathcal{M}, i \in \mathcal{I}, s \in \{1, ..., n_m^s\} \quad (5.45)$$

We implemented this diet model two times. In the first instance, we utilized the main food groups from the Food and Nutrient Database for Dietary Studies (FNDDS) (USDA & ARS, 2020b), which consists of eight categories that include the food items considered in this research (Section 5.3.3.1). In the second instance, we applied FNDDS's food subgroup classification, which includes thirty-one categories. These two approaches will be referred to as the FGF-M (main) and FGF-S (sub) diet models, respectively.

#### *5.3.2.4 Performance evaluation*

The RC and FGF diet models were evaluated based on nutritional adequacy and acceptability. Nutritional adequacy was measured by the extent to which the modeled diets met macro- and micronutrient guidelines. Acceptability was assessed qualitatively by two nutrition researchers who compared the substitutes selected by all diet models and determined which model designed more appropriate meals (Appendix 5A). Additionally, acceptability was measured with two quantitative metrics: the co-occurrence of food items and diversity of substitutes, where higher values indicate greater acceptability. Note that the qualitative method plays the leading role in evaluation, as the quantitative metrics function as indirect approximations of specific facets of consumer acceptability and consequently have lower validity.

### 5.3.3 Data

Dietary intake data from the National Health and Nutrition Examination Survey 2013-2018 (NHANES) (CDC & NCHS, 2014, 2016, 2018) was utilized to establish two datasets.

The first dataset was employed to train the recipe completion algorithm (Section 5.3.3.1), while the second dataset comprised the diets to be optimized by the diet models (Section 5.3.3.2). The meal and intake datasets vary only in terms of the filters that were applied. Nutrient values and food classifications of the food items were retrieved from the FNDDS (USDA & ARS, 2020b). Details regarding the parameter settings of the diet models are provided in Section 5.3.3.3.

#### *5.3.3.1 Meal data*

The meals used to fit the recipe completion model were created from consumption data aggregated by respondent, interview day and time of eating occasion. The meal dataset was filtered to align with the specific use case of designing culturally acceptable and nutritionally adequate diets for vegetarian women of reproductive age. Subsequent filters were then employed to refine the dataset further, making it suitable for recipe completion. The following steps outline the exact filtering procedure: 1) Only the consumption of individuals aged eighteen years and older was considered. 2) Meals that contained meat, fish, or poultry were filtered out. 3) Only food items belonging to main courses were selected. For example, food items such as drinks, snacks, fruit, and desserts were not considered. 4) Food items classified as mixed dishes (e.g. rice with carrots and tomato-based sauce) were excluded, as the recipe completion model requires separate food items. 5) Only meals consisting of three or more food items were retained. The assumption here is that the recipe completion model needs at least two food items to give a good recommendation for food substitution. This finally resulted in a meal dataset of 8,328 unique meals.



Furthermore, to avoid substitution of food items that are very similar, food items were clustered manually based on their description. For example, all food items described as 'black beans' (NFS, from canned, from dried, with or without added fat) were considered as the food cluster 'black beans'. This resulted in 398 unique food clusters. The nutrient content of each food cluster was calculated by taking the average content of corresponding food items weighted by the number of occurrences in the dataset.

#### *5.3.3.2 Intake data*

The following criteria were used to select the respondents whose diets were optimized: 1) Only vegetarian women of reproductive age (19-50 years) were selected. 2) Respondents that satisfied all Recommended Daily Allowances (RDA) and macronutrient guidelines were excluded, as their diets were already considered optimal. 3) Respondents that had no meals with 3 or more substitutable food items were omitted. This resulted in a intake dataset of 112 respondents. For each respondent and meal, indication was made of the food items allowed for replacement ($b^m_{mi}$) and those considered beyond the scope of food substitution ($c^e_n$). The criteria for exclusion included classification as a mixed dish, not

part of the main course, or consisting of fewer than three food items (Section 5.3.3.1). In total, 42% of the food items were allowed to be replaced.

#### *5.3.3.3 Parameter settings*

The parameter settings used for running the recipe completion model and the RC and FGF diet models are displayed in Table 5.1 and Table 5.2. The recipe completion model was trained using R Statistical Software v4.2.1. The diet models were run in Python 3.9.13 using Gurobi's 10.0.1 MIP solver with a time limit of 10 minutes per run.

**Table 5.1:** Parameter settings for running diet models.

| Parameter settings | |
|---|---|
| $\lambda$ | The following lambda values have been tested for the recipe completion model: 10; 100; 300; 500; 800; 1,000; 3,000; 5,000; 8,000; 10,000; 1e+05; 1e+06. |
| $\lambda_{opt}$ | The optimal lambda value used by the recipe completion model equalled 500. |
| $\phi$ | The following phi values have been tested for the recipe completion model: 1, 2, 4, 6, 8, 10, 12, 14. |
| $\phi_{opt}$ | The optimal phi value used by the recipe completion model equalled 6. |
| $c$ | The classes for the number of food items within a meal were 3-4, 5-7, and $\geq$ 8 items. |
| $f$ | The factor used to normalize the substitution score (0.00302) was derived from the slope of a regression model trained on $X \cdot M'$, where the number of food items was the independent variable and the dependent variable was the average score of the top 30 substitution options. |
| $G$ | Food groups in set $\mathcal{G}$ were based on the main (8) and sub (31) food group classifications from the FNDDS (USDA & ARS, 2020b). |
| $lb_n$ | Lower Bound (LB) on macronutrient $n$, see Table 5.2. |
| $lb_n^s$ | We based the lower bound of the substitution score on the average score of the top 30 substitution options for all meals containing three food items (0.01796). |
| $n^d$ | The meal dataset contained 2 days of consumption. |
| $n_m^s$ | Per meal, 50% of the food items considered for substitution were allowed to be replaced. The resulting number was rounded down. |

| Parameter settings | |
|---|---|
| $q^s_{ic}$ | The substitution quantity of food item $i$ is based on the average quantity in the consumption dataset for food item class $c$, normalized to a daily caloric intake of 2,000 kcal. |
| $rda_n$ | Recommended Daily Allowance (RDA) of micronutrient $n$, see Table 5.2. |
| $ub_n$ | Upper Bound (UB) on macronutrient $n$, see Table 5.2. |
| $ul_n$ | Tolerable Upper Level (UL) on micronutrient $n$, see Table 5.2. |
| $w$ | The following weights for the health objective have been applied for the RC model: 0, 0.75, 0.8, 0.84, 0.86, 0.88, 0.9, 1. The weights for the FGF models were: 0, 0.75, 0.8, 0.84, 0.86, 0.89, 0.91, 0.94, 1. These weights were selected manually based on visual inspection of shifts in the optimal solution. |

**Table 5.2:** Upper (UB) and Lower Bounds (LB) for macronutrients, and Recommended Daily Allowances (RDA) and Tolerable Upper Levels (UL) for micronutrients, as considered by the diet models for adult women of reproductive age (19-50 years) (Meyers et al., 2006).

| Macronutrient | Unit | LB | UB |
|---|---|---|---|
| Protein | E%[1] | 10 | 35 |
| Carbohydrates | E% | 45 | 65 |
| Dietary fiber | g | 25 | - |
| Total fat | E% | 20 | 35 |
| Total saturated fatty acids | E% | - | 10 |
| Sodium | mg | - | 2300 |
| **Micronutrient** | **Unit** | **RDA** | **UL** |
| Calcium | mg | 1000 | 2500 |
| Iron | mg | 18 | 45 |
| Zinc | mg | 8 | 40 |
| Vitamin A | $\mu$g | 700 | 3000 |
| Thiamin | mg | 1.1 | - |
| Riboflavin | mg | 1.1 | - |
| Niacin | mg | 14 | - |
| Vitamin B6 | mg | 1.3 | 100 |
| Folate | $\mu$g | 400 | 1000 |
| Vitamin C | mg | 75 | 2000 |

[1] E% = Energy percentage.

## 5.4 Results and Discussion

In this section, we present and discuss our new approach to data-driven consumer acceptance modelling, in which recipe completion is applied to substitute food items within meals. We begin by evaluating the performance of the recipe completion algorithm (Section 5.4.1). Next, we assess the performance of our Recipe Completion (RC) diet model and compare it to traditional Food Group Filtering (FGF) diet models (Section 5.4.2), focusing on two key aspects: the nutritional adequacy of the generated diets and the acceptability of the proposed food substitutions.

### 5.4.1 Performance recipe completion model

Performance of the recipe completion model was measured based on its capability of retrieving a food item that was removed from a meal. The median rank of the removed food item and the percentage of meals for which the removed food item was ranked within the top 10 are displayed in Table 5.3. Our results are comparable to those reported by De Clercq et al. (2016), which is reasonable since the same methodological approach was employed.

However, during initial experiments, we found that the food clustering step had a notable impact on performance. When we clustered food items by their nutrient profiles instead of their description, the top 10 performance dropped by 14%. The influence of the training dataset on model performance has also been noted by Cueto et al. (2019) and Lei et al. (2020) (Section 5.2.2). This suggests that, for effective recipe completion, data preparation is as critical as the algorithm itself. Unfortunately, the data cleaning process can be both complex and time-consuming, which might offset the time benefits of a data-driven approach compared to knowledge-driven diet modelling.

In addition, better model performance may be achieved by employing more advanced modelling techniques. For instance, in the field of food substitution, state-of-the-art approaches rely on large language models like BERT (Morales-Garzón et al., 2025) or knowledge graph embeddings (Shirai et al., 2021). While these methods have demonstrated superior performance compared to simpler techniques, their complexity presents greater challenges for implementation within a diet modelling framework, including non-linearity and the substantial increase in model size.

When we examine the rank of removed food items within the list of substitutes, it can be observed that food items that frequently occur in the dataset were often in the top of substitution recommendations (Figure 5.2A). Although the recommendation of popular items is desirable to some extent, this popularity bias also has several negative effects, including the homogeneity of recommendations (Abdollahpouri, 2020; Hong et al., 2020; Kunaver & Požrl, 2017; Chen et al., 2023) and the promotion of popular yet unhealthy food items (Trattner & Elsweiler, 2017; Musto et al., 2021). While the latter issue is not a

**Table 5.3:** Performance of recipe completion model compared to other studies. Performance is measured by the model's ability to retrieve a food item that was removed from a meal. The better the model performs, the higher the rank (with 1 being the best) of the removed food item and the higher the percentage of meals for which the removed food item was ranked within the top 10.

| Source | Nr. of food items | Nr. of recipes | Median rank | Top 10 (%) |
|---|---|---|---|---|
| De Clercq et al. (2016) | 381 | 55,001 | 6 | 61.6 |
| Cueto et al. (2019) | 267 | 37,340 | 15 | 40.3 |
| Chen et al. (2020) | 465 | 36,492 | 8 | 52.6 |
| Lei et al. (2020) | 1,463 | 188,446 | - | 60.9 |
| This research | 398 | 8,328 | 5 | 61.4 |

major concern in our approach, since the diet model will not replace a food item with an alternative that has a less beneficial nutrient profile, the former issue remains a concern and will be further discussed in Section 5.4.2.2.

Figure 5.2A also displays some food items that are popular but have a relatively low ranking. This is because the recipe completion model does not score the popularity of food items directly but the popularity of food item combinations (Figure 5.2B). As an example, one meal in the training dataset included spaghetti and spaghetti sauce, with eggs being removed. Despite the frequent consumption of eggs (1,271 occasions), they ranked low (position 373) because they were rarely consumed with spaghetti (0 occasions) or spaghetti sauce (7 occasions). Instead, the model suggested alternatives like cheese, lentils, and mixed vegetables, which were more commonly paired with these ingredients. This shows that a low item ranking does not necessarily imply that the model's top suggestions are inappropriate. Therefore, when assessing model performance, it may be more meaningful to evaluate the quality of the top substitution recommendations.

Evaluating the quality of top substitution options is challenging since there is no generally acknowledged dataset of valid substitutions (Shirai et al., 2021). As such, many studies have relied on human evaluation to assess the quality of food substitutes (Morales-Garzón et al., 2021; Eftimov et al., 2020; Pellegrini et al., 2021). However, human evaluation can be subjective and is time-consuming. A more quantitative alternative would be to evaluate substitutions based on how well they satisfy specific constraints. By examining examples of both suitable and unsuitable food substitutions (Table 5.4), necessary criteria for model evaluation can be established. For instance, the completeness of a meal and the diversity of food items could serve as quality indicators. For Meals 1 and 2, the addition of a staple food and dressing are beneficial substitutions as they enhance the meal's completeness by providing a carbohydrate source and condiment, respectively. On the other hand, for Meals 3 and 4, the addition of staple foods is less suitable, as these meals already contain sufficient carbohydrate sources, which reduces the diversity of food items.

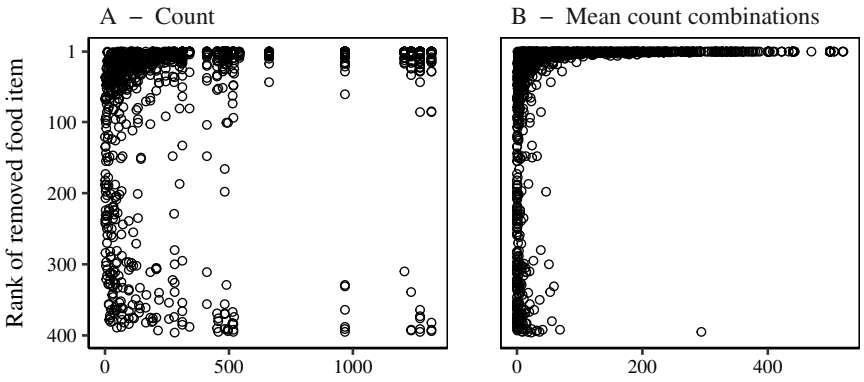



**Figure 5.2:** Figure A displays the correlation between the rank of the removed food item from the test dataset and the count of that food item in the training dataset. Figure B illustrates the average frequency of combinations between removed and remaining food items in the training dataset. For example, a meal contains food items A, B, and C. Food item A is the removed item and items B and C are the remaining items. For each combination of removed and remaining food items, it is counted how often this combination occurs in the training dataset (e.g. A-B: 10, A-C: 4). Then the average frequency equals 7 for this meal.

**Table 5.4:** Examples of top 5 substitution options.

| Meal | Remaining food items | Removed food item (Rank) | Top 5 substitutes |
|---|---|---|---|
| 1 | Eggs, Spinach | Bread-pita-whole grain (149) | Butter, Bread, Bread-whole grain, Tortillas, Rice-white |
| 2 | Carrots-raw, Cheese, Cucumber-raw, Romaine lettuce-raw, Cabbage-red-raw | Caesar dressing (1) | Caesar dressing, Italian dressing, Tomatoes-raw, Pepper-raw, Spinach-raw |
| 3 | Rice-white, Kidney beans, Queso, Sour cream | Avocado (13) | Tortillas, Soy sauce, Salsa, Potato, Lettuce-raw |
| 4 | Rice-brown, Asparagus, String beans | Summer squash (2) | Soy sauce, Summer squash, Rice-white, Salsa, Mushrooms |

In summary, measures are needed to quantify meal completeness and the diversity of food items and food groups. Using these measures, along with the ranking of the food item to be retrieved, may allow for better validation of different recipe completion algorithms. Moreover, these measures, in combination with the co-occurence of food items within a meal, could be used to train the recipe completion algorithm, potentially improving its overall performance. Alternatively, more advanced algorithms may be considered.

### 5.4.2 Comparison of the RC and FGF diet models

In order to assess the performance of our Recipe Completion diet model (RC) and the diet models based on Food Group Filtering (FGF), we examined the nutritional adequacy of modelled diets and the acceptability of food substitutes.

#### *5.4.2.1 Nutritional adequacy*

The trade-off between nutritional adequacy and consumer acceptability is illustrated in Figure 5.3. We assumed that fewer dietary adjustments lead to higher acceptability. Under this assumption, nutritional adequacy and acceptability were negatively correlated. Our RC diet model outperformed the FGF diet models in terms of macronutrient adequacy. This is because our RC model is not limited to selecting replacements from the same food group as the original item. However, when evaluating micronutrient adequacy, the results differed: while our RC model performed better than the FGF-S model (31 subgroups), it did not surpass the FGF-M model (8 main groups). The advantage of the FGF-M model lies in its smaller number of food groups, with more items within each category, increasing the likelihood of finding a healthier alternative.

Figure 5.4 provides a closer examination of the nutritional adequacy of the modelled diets for the scenario where 10% of the food items has been substituted. The average nutrient content of observed diets fell below recommended levels for calcium, folate, iron, niacin, vitamin A, vitamin B6, and zinc. Diet modelling improved the intake of these nutrients, resulting in the satisfaction of the Recommended Daily Allowance (RDA) for folate, vitamin A, and vitamin B6. The RC diet also satisfied the RDA for niacin.



The macronutrient content of modelled diets is displayed in Figure 5.5. For the observed diets, the average intake of fiber and saturated fat did not meet dietary guidelines. While the RC diet managed to meet the guideline for saturated fat, it fell short on fiber intake, though it still performed slightly better than the FGF diets.

To further improve the nutritional adequacy of the diets, the diet models could be extended to facilitate not only food substitutions but also the addition or removal of food items as well as the adjustment of portion sizes. For our RC diet model, the implementation of adding extra food items is straightforward, as recipe completion algorithms are specifically designed for this purpose. Furthermore, the diet models could be extended to allow for replacement of food items that can be consumed alone, such as beverages, fruits, and snacks, which were excluded in this study (Section 5.3.3.1). For such a task, food substitution algorithms could be used, which are designed to find similar but healthier food items. Integrating both recipe completion and food substitution algorithms with diet optimization would further broaden its applicability for adapting meals.

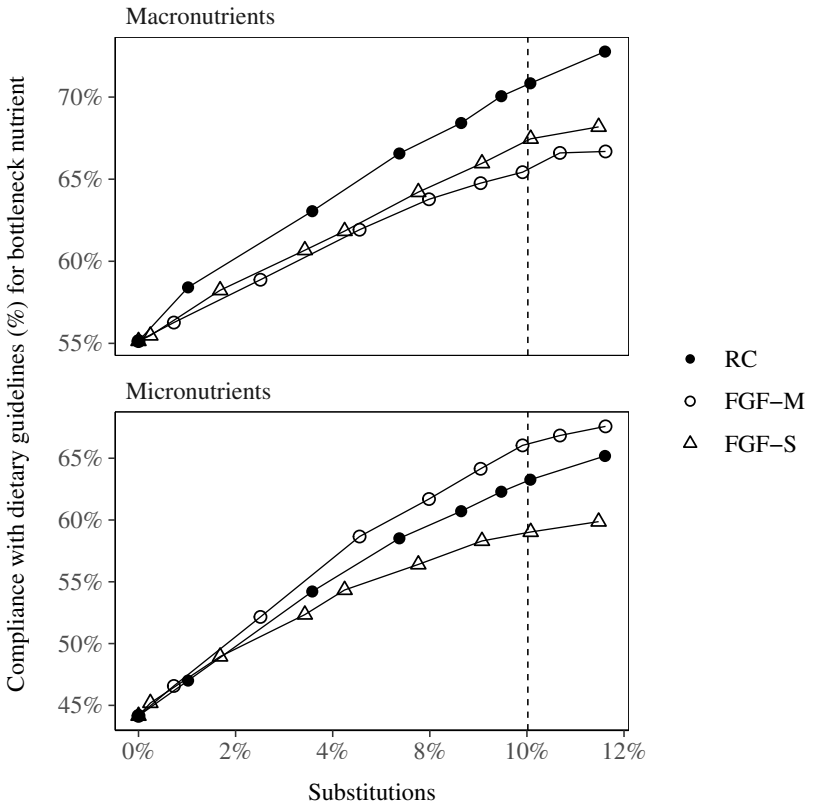



**Figure 5.3:** Compliance of the bottleneck nutrient with dietary guidelines for macro- and micronutrients plotted against the average percentage of substitutions for the Recipe Completion (RC) and Food Group Filtering (FGF) diet models (M: 8 main groups, S: 31 subgroups). Note that the bottleneck nutrient can be different for each respondent. The percentage of substitutions is calculated by dividing the number of substitutions by the total number of food items. Within a meal, a maximum of 50% of the food items considered for food substitution were allowed to be replaced (e.g. for a meal with 10 food items of which 7 items are considered for food substitution, maximum 3 items may be replaced). The dashed vertical line highlights the model scenario in which the percentage of substitutions equals 10%, a reference point utilized in other figures and tables.

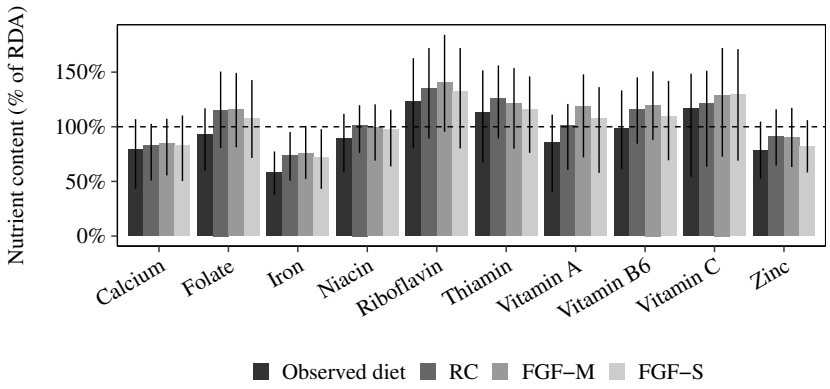



**Figure 5.4:** Average micronutrient content (model scenario = 10% substitutions) of the observed diets and the diets designed by the Recipe Completion (RC) and Food Group Filtering (FGF) models (M: 8 main groups, S: 31 subgroups). The micronutrient content is presented as a percentage of the Recommended Daily Allowance (RDA). The error bars, derived from the interquartile range, display the range of variation in nutrient content between the diets for each respondent.

#### *5.4.2.2 Substitute acceptability*

The acceptability of substitutes was evaluated qualitatively by two nutrition researchers. Complementary to that, acceptability was also approximated using two quantitative measures: the co-occurrence of substitutes with other items in the meal dataset, and the diversity of food substitutes.

*Qualitative evaluation*

A few examples of how meals were adapted by the RC and FGF diet models are given in Table 5.5. In the first meal, our RC model outperformed the FGF models by creating a more complete meal by replacing one of the two carbohydrate sources with a legume-based product. This was also the case in the second meal, where the RC model created a more balanced meal by introducing a carbohydrate source. Meals 3 and 4, however, demonstrate cases where the FGF models made better substitutions than the RC model. In Meal 3, the RC model removed all condiments, reducing meal variety. In Meal 4, it introduced an additional carbohydrate source, even though two were already present.

In a similar way, all meals were evaluated qualitatively by nutrition researchers (model scenario = 10% substitutions). When comparing the RC and FGF-M models, our RC model outperformed the FGF-M model in 39% of cases, while the FGF-M model produced better results in 19% of cases (equal score = 42%). As for the FGF-S model, the FGF-S

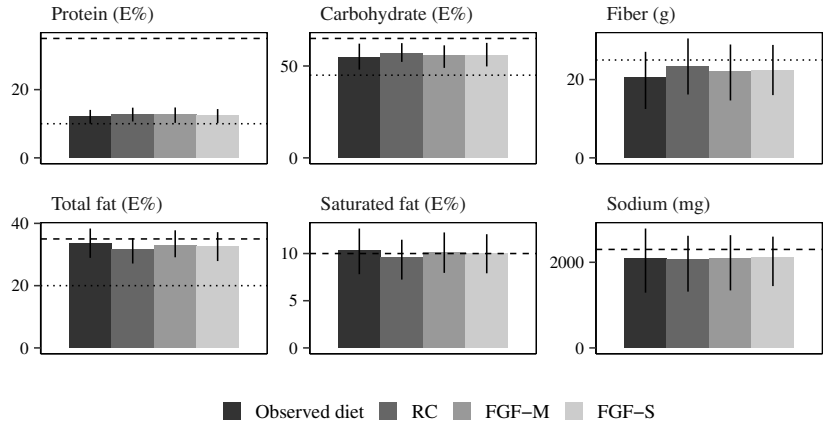



**Figure 5.5:** Average macronutrient content (model scenario = 10% substitutions) of the observed diet and the diets designed by the Recipe Completion (RC) and Food Group Filtering (FGF) models (M: 8 main groups, S: 31 subgroups). For protein, carbohydrate, and fat, the nutrient content is expressed as a percentage of the energy intake. The error bars, derived from the interquartile range, display the range of variation in nutrient content between the diets for each respondent. The dotted and dashed horizontal lines display the lower and upper bounds of recommended intake.

model showed better results in 42% of cases, with the RC model leading in 25% (equal score = 33%). This pattern reveals that the FGF model achieved better results in terms of substitute acceptability when a larger number of food groups was used. With many food groups, the variety of food items within each group is smaller, leading to substitutes that are more similar to the items being replaced. This increases the likelihood that a substitute fits well within the current meal context. However, because the substitutes are more similar, the potential to improve the nutritional value of meals is also reduced, as illustrated in Figure 5.3. In contrast, the RC model demonstrated greater creativity in food substitution, as it was not restricted by food group constraints. This flexibility enabled the model to achieve higher nutritional adequacy, but also led to a higher incidence of inappropriate substitutions.

*Co-occurence of food items*

A quantitative measure for approximating acceptability is the co-occurrence of substituted food items with other components of the meal. A higher frequency of co-occurrence correlates with a greater likelihood that the substitute is acceptable. The average frequencies of food item combinations (model scenario = 10% substitutions) were 26, 8, and

**Table 5.5:** Example of meal adaptations (model scenario = 10% substitutions) by the Recipe Completion (RC) and Food Group Filtering (FGF) diet models (M: 8 main groups, S: 31 subgroups).

| Meal | Food items | Substitutes RC | Substitutes FGF-M | Substitutes FGF-S |
|---|---|---|---|---|
| 1 | Tortillas-whole grain | Soybean curd | Cereal | Bread-nut |
| | Rice-white | | | |
| | Carrots | | | |
| 2 | Black beans | | | |
| | Cheese | Rice-white | Yogurt-Greek | Cheese-goat |
| | Sour cream | | | |
| | Salsa | Spinach | Spinach | Tomatoes-pickled |
| 3 | Potato patty | | | Potato-baked-peel-cheese |
| | Yogurt | Bread-chappatti or roti | | |
| | Spaghetti sauce | Lentils | Broccoli | |
| | Onions-raw | | Spinach | Mixed vegetables |
| 4 | Tortillas | | | |
| | Potato | | Broccoli | Plantain |
| | Sour cream | Rice-white | Yogurt-Greek | Dip |
| | Cheese-cottage | Broccoli | | |
| | Pinto beans | | | |

3 for the RC, FGF-M, and FGF-S diet models respectively (Figure 5.6). In the FGF-M and FGF-S diets, 14% and 43% of substitutions involved food item combinations that appeared, on average, less than once in the meal dataset, compared to 0% in the RC diet.



*Diversity of substitutes*

Another quantitative measure is the percentage of distinct substitutes, under the assumption that increased diversity indicates more tailored substitutions for specific meals. Without substitution rules, the same food items may be repeatedly selected. In contrast, when substitutions are meal-specific, a wider variety of food items is likely to be introduced across different meals, resulting in a more diverse diet. In this evaluation, our RC model (19%) outperformed the FGF-M model (16%) but was surpassed by the FGF-S model (28%). An overview of the most frequently selected substitutes is given in Table 5.6.

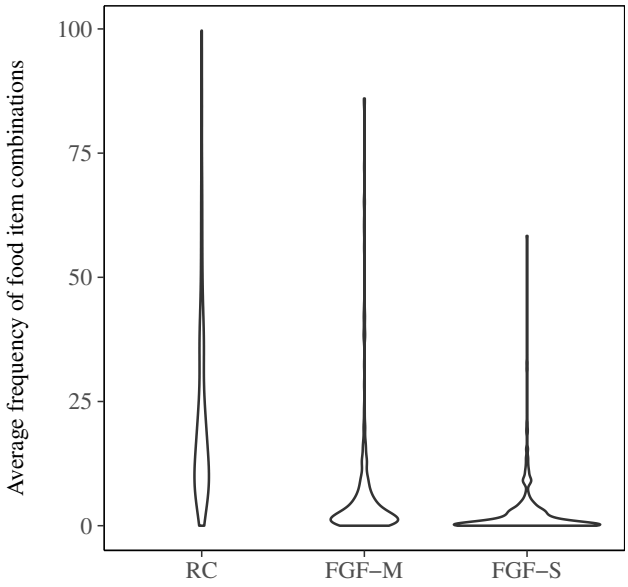



**Figure 5.6:** Distribution of average frequencies of food item combinations (model scenario = 10% substitutions) for the Recipe Completion (RC) and Food Group Filtering (FGF) models (M: 8 main groups, S: 31 subgroups). For example, a meal contains substituted item A and other food items B and C. For each combination of substituted and other food items, it is counted how often this combination occurs in the meal dataset (e.g. A-B: 8, A-C: 2). In this example, the average frequency equals 5. Frequencies above 100 are not displayed.

**Table 5.6:** Substitutes selected by the Recipe Completion (RC) and Food Group Filtering (FGF) diet models (M: 8 main groups, S: 31 subgroups) ordered by frequency (model scenario = 10% substitutions).

| | RC | | | FGF-M | | | FGF-S | |
|---|---|---|---|---|---|---|---|---|
| 1 | Cereal | 12.7 % | 1 | Spinach | 20.8 % | 1 | Mixed vegetables | 9.5 % |
| 2 | Lentils | 8.6 % | 2 | Cereal | 16.0 % | 2 | Spinach | 6.3 % |
| 3 | Spinach | 7.3 % | 3 | Oatmeal | 8.0 % | 3 | Pepper-sweet-red-raw | 6.0 % |
| 4 | Bread-whole grain | 5.1 % | 4 | Broccoli | 7.3 % | 4 | Whole wheat cereal | 5.7 % |
| 5 | Oatmeal | 4.4 % | 5 | Soybean curd | 4.2 % | 5 | Peas-green | 4.7 % |
| | ... | | | ... | | | ... | |
| 59 | Sweet potato | 0.3 % | 50 | String beans-green | 0.3 % | 89 | Winter squash | 0.3 % |

The relatively low diversity of food substitutes suggested by the RC diet model is due to popular food items frequently appearing at the top of substitution options. A simple

way to increase diversity is by introducing constraints that limit the selection of similar food items. A more fundamental solution would be to refine the recipe completion algorithm itself by, for example, rebalancing the sample data, integrating diversity metrics into the training process, or incorporating additional inputs on food item attributes (Abdollahpouri, 2020; Hong et al., 2020; Kunaver & Požrl, 2017; Chen et al., 2023).

In summary, the results demonstrated a trade-off between nutritional adequacy and substitute acceptability. The FGF-M model outperformed the FGF-S model in nutritional adequacy but lagged behind in substitute acceptability. The RC model showed performance that was intermediate between the two. The key strengths of our RC diet model are its ability to enhance the nutritional value of diets by allowing substitutions across different food groups, and the high co-occurrence of substituted items with the rest of the meal. Nonetheless, further improvement is needed to increase the acceptability of the substitutes.

*Portion size*

One aspect of acceptability that has not yet been addressed is the portion size of substitutes. In this study, we added substitutes using their average quantities as observed in the meal dataset, with adjustments based on the number of food items in a meal. However, this approach can lead to meals with disproportionate portion sizes. The literature presents several algorithms that predict ingredient quantities in recipes (Chen et al., 2020; Li et al., 2021; Choi et al., 2023). Embedding such an algorithm could improve the acceptability of substitution quantities.



#### *5.4.2.3 Sustainability and affordability*

Lastly, sustainability and affordability are two aspects not yet considered in this study. Although vegetarian diets typically contain fewer animal-based products, known to burden the environment (Chai et al., 2019), it is important to recognize that they may also include plant-based foods with low sustainability. To further improve the sustainability of vegetarian diets, such foods could be penalized within the model. The Food Impacts on the Environment for Linking to Diets (dataFIELD) database provides greenhouse gas emission data that can be linked to the food codes used in the National Health and Nutrition Examination Survey (NHANES) (Conrad et al., 2023). To account for affordability, a cost constraint could be introduced to ensure that meals remain within the same price range as before. The Purchase to Plate data can be used to connect price information with NHANES (USDA ERS, 2023).

## 5.5 Conclusion

The objective of this study was to contribute to the development of data-driven consumer acceptance modelling by integrating diet optimization with recipe completion, and comparing this approach to traditional diet modelling methods, which rely on food group filtering and item popularity. While broader food group categories in the traditional approach facilitated the creation of healthier diets, they often resulted in lower acceptability of substitutes, and vice versa. Our recipe completion diet model struck a balance by demonstrating greater acceptability than the traditional approach when broader food groups were utilized, while achieving higher nutritional adequacy with narrower food groups. To enhance the performance of the recipe completion diet model, the recipe completion algorithm needs further refinement to identify acceptable substitutes for designing complete and diverse meals. To further improve nutritional adequacy, model enhancements include adding extra food items to meals, replacing individual food items (no meal context), and adjusting portion sizes.

In conclusion, while more research is needed to further improve the acceptability of substitutions, combining diet optimization with recipe completion presents a promising approach to enhance the nutritional adequacy of individual diets while maintaining the acceptability of food item combinations within meals.

# 5.A Appendix - Qualitative evaluation

To evaluate the performance of food substitutions by the Recipe Completion (RC) and Food Group Filtering (FGF) models—where FGF-M contained 8 main food groups and FGF-S 31 subgroups—the designed diets (model scenario = 10% substitutions) were assessed by two nutrition researchers. For each meal, they determined whether the substitutions between RC and FGF-M, and RC and FGF-S, were better or comparable. If one model made a substitution and the other did not, that meal was excluded from evaluation. In this evaluation, only the acceptability of food substitutions was considered, not the nutritional adequacy or the appropriateness of substitution quantities. The acceptability of substitutions was evaluated based on the following criteria:

- Coherence: Does the substitute pair well with the other food items in the meal? For example, cereal might not be suitable for dinner but could work well for breakfast. Similarly, raw vegetables fit better in a salad than in a warm meal.
- Completeness: Does the substitute contribute to a balanced meal? For instance, a dinner lacking a protein source, carbohydrate source, or condiment may be considered incomplete.
- Diversity: Is the substitute sufficiently different from the other food items in the meal? For example, adding rice to a meal that already contains pasta may be less appropriate.



The following form was used to collect the responses from the nutritionists:

| Meal | Original food items | Substitutes RC | Substitutes FGF-M | Substitutes FGF-S | Best* RC vs FGF-M | Best* RC vs FGF-S | Comments |
|---|---|---|---|---|---|---|---|
| 1 | A | E | F | | RC | NA | |
| | B | | | | | | |
| | C | G | H | | | | |
| | D | | | | | | |
| 2 | A | | | | Equal | FGF-S | |
| | E | B | C | D | | | |
| | F | B | C | D | | | |
| ... | ... | ... | ... | ... | ... | ... | ... |

* Answer options: RC, FGF, Equal, NA (Not Applicable)

A summary of the collected responses is given below:

| | |
|---|---|
| Total number of evaluated meals | 157 |
| – Evaluated meals RC vs FGF-M | 150 |
| – Evaluated meals RC vs FGF-S | 151 |
| Average number of food items per meal | 4.4 |
| Average number of substitutions per meal | 1.9 |

Evaluation counts per nutritionist:

| | Best (RC vs FGF-M) | | | Best (RC vs FGF-S) | | |
|---|---|---|---|---|---|---|
| Nutritionist | RC | FGF-M | Equal | RC | FGF-S | Equal |
| 1 | 59 (39%) | 23 (15%) | 68 (45%) | 36 (24%) | 57 (38%) | 58 (38%) |
| 2 | 58 (39%) | 35 (23%) | 57 (38%) | 39 (26%) | 69 (46%) | 43 (29%) |
| Total | 117 (39%) | 58 (19%) | 125 (42%) | 75 (25%) | 126 (42%) | 101 (33%) |

Examples of comments given by the nutritionists:

- Why is plantain_raw in the list of substitution options, as this is never appropriate.
- Aggregated food item cereal is too broad, as oat flakes could be appropriate for a salad, but sweet breakfast cereals are not. Currently, cereal is almost never appropriate. Could be excluded from the list of substitution options.
- Food item mixed_vegetables is difficult to evaluate.
- Unexpected substitution: how can croissant be more healthy than whole_grain_bread?
- RC adds carb source making it more complete, which seems also a good alternative for a salad, although it removes a condiment.
- Both not appropriate: RC removes condiment, FGF-M adds carrots twice.
- RC more like a hamburger, FGF more like a salad.
- RC more diverse as it replaces one of the two condiments, but it changes the type of dish.
- Difficult to judge, because original meal is already not acceptable.
- RC is more complete, the others are more comparable to the original.
- RC adds another carb, so FGF-M is better.

# Chapter 6

General discussion



## 6.1 Positioning within existing knowledge

The aim of this thesis was to improve the nutritional adequacy, sustainability, and acceptability of modeled diets by evaluating current approaches and developing new methods for modeling iron absorption and food substitution. With respect to iron absorption, this focus was motivated by the fact that many diet modeling studies have relied on constant absorption factors rather than incorporating the diet-dependent equations available in the literature. The impact of this simplification on the assessment of iron adequacy in modeled diets had not yet been investigated. Our research demonstrated that diet models with constant absorption factors tend to overestimate the absorbable nonheme iron content of menu plans compared to models using diet-dependent absorption equations (Chapter 2, Table 6.1). To increase the estimated absorbable nonheme iron content, equations are more effective than constant factors, as they allow for the optimization of both enhancers and inhibitors of iron absorption. This is especially relevant for improving the nutritional adequacy of vegetarian and vegan diets, as nonheme iron from plant sources is less readily absorbed than heme iron from animal products (Hunt, 2002).

To identify the most effective method for implementing these nonlinear absorption equations in diet modeling, we compared the use of nonlinear programming and piecewise linear approximation. In diet modeling, piecewise linear approximation had not yet been applied, despite its potential performance advantages over nonlinear programming (Geißler et al., 2011). Our results showed that, in mixed-integer diet models, piecewise linear approximation outperformed nonlinear programming in terms of solution quality and runtime (Chapter 3, Table 6.1). Both methods performed equally well in continuous diet models and in modeling zinc absorption. As such, this study provides a practical example of how to implement piecewise linear programming and highlights the importance of validating solution quality when using nonlinear programming. This may inform future researchers developing diet models with nonlinear equations.

With regard to modeling food substitution, many studies have modeled diets at food-group level, allowing substitutions only between food groups. We demonstrated that modeling at food-item level, allowing both between- and within-food-group substitutions, leads to greater improvements in nutritional adequacy, environmental sustainability, and acceptability than relying solely on between-food-group substitutions (Chapter 4, Table 6.1). Within-food-group substitution allowed the selection of more nutrient-dense and sustainable foods, requiring less dietary change and thereby improving the acceptability of modeled diets, under the assumption that smaller changes are more acceptable. These findings underscore the relevance of considering both between- and within-food-group substitutions in future diet modeling studies.

In addition to food-group-based substitution, we also investigated data-driven food substitution. Data-driven methods have rarely been applied in diet modeling and have only

been tested in experimental settings. We developed a novel data-driven approach to substitute food items using a recipe completion algorithm and applied it to improve the nutritional adequacy of self-reported meals (Chapter 5, Table 6.1). Compared with traditional food group substitution, our recipe completion diet model achieved higher acceptability, as evaluated by nutritionists, than broad food group substitutions. While it did not exceed the acceptability of narrow food-group substitutions, it resulted in greater nutritional adequacy. As such, data-driven food substitution offers a promising strategy for personalizing food substitutions and thereby improving their acceptability.

**Table 6.1:** Overview of main research findings per chapter.

| Chapter | Main findings |
|---|---|
| | **Iron absorption** |
| 2 | For omnivorous, vegetarian, and vegan diets, constant absorption factors overestimated the absorbable nonheme iron content compared to diet-dependent absorption equations. Unlike constant absorption factors, the equations enable the optimization of food components that either enhance or inhibit nonheme iron absorption, resulting in more realistic estimates of absorbable iron. |
| 3 | Piecewise linear approximation provided accurate solutions for modeling iron absorption in a mixed-integer diet model, outperforming nonlinear programming in terms of both solution quality and runtime. For zinc, both the piecewise linear and nonlinear programming models produced accurate solutions, although the nonlinear programming model had longer runtimes. In the continuous diet model, nonlinear programming and piecewise linear approximation showed comparable performance for both iron and zinc. |
| | **Food substitution** |
| 4 | Allowing both within- and between-food-group substitutions notably improved the nutritional adequacy and sustainability of diets compared with between-food-group substitutions alone, achieving the same outcomes with only half the dietary change. |
| 5 | We introduced a new data-driven method for food substitution by combining the research domains of diet modeling and recipe completion, and compared its performance with traditional food group substitution. Traditional substitution using broad food groups, which include many food items, often resulted in healthier diets but was less acceptable to consumers, as evaluated by nutritionists. Conversely, substitution using narrow food groups, with fewer food items, resulted in higher acceptability but led to smaller nutritional benefits. Our recipe completion diet model balanced these trade-offs, achieving better acceptability than broad food group substitutions and greater nutritional adequacy than narrow food group substitutions. |



## 6.2 Societal relevance

Having discussed our main research findings, we now relate these outcomes to their societal relevance, corresponding to the aim of this thesis: improving the nutritional adequacy, environmental sustainability, and acceptability of modeled diets. To improve the envi-

ronmental sustainability of dietary intake, increasing the share of plant-based proteins is widely promoted as a strategy (Aiking & de Boer, 2020; Graça et al., 2019; Hemler & Hu, 2019; Sabate & Soret, 2014; Vinnari & Vinnari, 2014; Willett et al., 2019). This approach presents two key challenges: maintaining nutritional adequacy and ensuring that the resulting dietary changes remain acceptable to consumers.

Regarding nutritional adequacy, iron is particularly important, as iron from plant-based foods is less readily absorbed than iron from animal sources (Hunt, 2002). To ensure adequate levels of absorbable iron, diet-dependent iron absorption equations offer a more effective modeling approach than constant absorption factors, as they allow the optimization of both enhancers and inhibitors of iron absorption (Chapter 2). In mixed-integer diet models, these nonlinear absorption equations can be implemented most accurately and with shortest runtimes using piecewise linear approximation. For continuous diet models, both piecewise linear approximation and nonlinear programming provide reliable outcomes (Chapter 3). Furthermore, allowing the selection of nutrient-dense foods is essential for improving the nutritional adequacy of modeled diets. Including both between- and within-food-group substitutions creates more opportunities to incorporate these foods than between-food-group substitution alone (Chapter 4).

For people to adopt healthier and more sustainable diets, these plans must be acceptable to consumers (Bos et al., 2013). In diet modeling, acceptability is often maintained by minimizing dietary changes (van Dooren, 2018). Allowing both between- and within-food-group substitution requires fewer dietary changes to achieve the same nutritional and sustainability goals compared to relying only on between-food-group substitution (Chapter 4). Therefore, within-food-group substitution is an important approach for improving the acceptability of modeled diets. Alternatively, data-driven food substitution can be applied. This approach offers greater personalization than food-group-based substitutions (Fajemisin et al., 2024), which can improve the acceptability of modeled diets. We developed a novel, data-driven method that substitutes foods using a recipe-completion algorithm (Chapter 5). Our results showed that the recipe completion diet model achieved higher acceptability than food-group-based substitution using broad food groups, but it did not surpass the acceptability of substitution based on narrow food groups. Nonetheless, data-driven food substitution remains a promising strategy to personalize food substitutions and improve acceptability in the future.

Overall, the methods presented in this thesis for modeling iron absorption and food substitution, which we recommend applying in combination, offer effective support in designing diets that are healthy, environmentally sustainable, and acceptable. While most applicable to (individual) menu planning, they could also contribute to the refinement of national and international dietary guidelines. These methods are especially relevant for future studies focused on improving the sustainability of diets through reduced con-

sumption of animal-based products, while maintaining their nutritional adequacy and acceptability.

## 6.3 Remaining challenges and future research

In this section, we reflect on the remaining challenges, which can be viewed as the limitations of this thesis, and outline opportunities for future research. Each opportunity is numbered in the text and summarized in Table 6.2. Our discussion starts with the modeling of iron absorption and then turns to food substitution.

### 6.3.1 Iron absorption

This thesis examined whether to use constant absorption factors or diet-dependent absorption equations, and for the latter, whether to implement them through nonlinear programming or piecewise linear approximations. However, some of the other modelling decisions we made still require further investigation, including the choice of a specific nonheme iron absorption equation, the modeling of heme iron absorption, iron absorption from fortified foods, how to account for non-dietary factors affecting iron absorption, and the modeling of absorption for nutrients other than iron. These modeling choices will be discussed in more detail below.

#### *6.3.1.1 Nonheme iron absorption equation*

The literature provides several diet-dependent nonheme iron absorption equations, which can be categorized into two types: single-meal equations, which estimate absorption for each meal separately, and complete diet equations, which estimate absorption for the entire diet (De Carli et al., 2018). In our research, we applied the equations proposed by Hallberg & Hulthén (2000) (single-meal equation), Conway et al. (2007) (single-meal equation), and Armah et al. (2013) (complete diet equation). Although we found no statistically significant differences in absorption estimates between these equations (Chapter 2), the differences may still be relevant in practice. In the literature, the debate over whether single-meal or complete-diet equations yield more accurate estimates has not yet been resolved (De Carli et al., 2018; Beard et al., 2007). Consequently, further research is needed to determine which absorption equation provides the most accurate estimates of nonheme iron absorption and is therefore the most appropriate for use in diet modeling (1).

Additionally, new nonheme iron absorption equations could be developed. Given that prior studies are generally based on small datasets and rely on relatively simple regression techniques, the use of larger datasets combined with more sophisticated machine learning methods may enhance the accuracy of future absorption estimates (Khorraminezhad et al., 2020) (2).

Apart from accuracy, other factors that influence the choice of an absorption equation include ease of implementation and data availability. For example, to evaluate the performance of piecewise approximation (Chapter 3), we tested only the iron absorption equations of Armah et al. (2013) and Conway et al. (2007), as these allowed a univariate approximation. However, the equation by Hallberg & Hulthén (2000) requires multivariate piecewise linear approximation, which is more challenging to implement. Additionally, the performance of multivariate approximation has not yet been investigated. Further research is required to determine whether the most accurate absorption equation also produces the most accurate estimates for piecewise linear approximation, and whether a trade-off exists between accuracy and model runtimes (3).

Regarding data availability, some equations rely on information that is not readily accessible. For instance, the equations of Armah et al. (2013) and Hallberg & Hulthén (2000) require data on compounds like phytate and tannic acid, which are often missing from standard food composition databases. These compounds can strongly affect equation outcomes, as shown by our sensitivity analysis in Chapter 2. When phytate or tannic acid values cannot be accurately estimated, the equation of Conway et al. (2007) may offer a more practical alternative, as it relies not on nutrient component data but only on food group quantities. Future research could focus on expanding nutrient composition tables to include these compounds, so that data availability will not be a limiting factor in the choice of a particular absorption equation (4).

#### *6.3.1.2 Absorption of heme iron*

Heme iron absorption is often modeled using a constant absorption factor of 25% (Armah et al. 2015b, Chapter 3), and occasionally with the equation by Hallberg & Hulthén (2000) (Barré et al. 2018, Dussiot et al. 2022b, Chapter 2). However, the mechanisms underlying heme iron absorption remain incompletely understood (West & Oates, 2008). Given the variation in heme iron content estimates across studies (Schönfeldt & Hall, 2011; Archundia-Herrera et al., 2024), further research is needed to assess whether current estimation methods yield reliable values for heme iron absorption (5).

#### *6.3.1.3 Iron absorption from fortified foods*

Fortified foods are a major contributor to total iron intake in Western countries (Huma et al., 2007). However, existing iron absorption algorithms do not account for food fortification, despite evidence that the type and concentration of the iron fortificant, as well as the food processing techniques applied, can significantly influence absorption rates (Huma et al., 2007; Shubham et al., 2020; Hurrell, 2002). Further research is necessary to address this gap (6).

#### *6.3.1.4 Non-dietary factors affecting iron absorption*

Host-related factors, particularly an individual's iron status, strongly affect iron absorption (Gibson, 2007; Armah et al., 2013). We explored this by modeling with two different serum ferritin concentrations, the most commonly used indicator of iron status (Daru et al., 2017). The resulting estimates of absorbable iron differed substantially between the two scenarios (Chapter 2). Nevertheless, no universally accepted standard exists regarding which concentration should be used in diet modeling when individual iron status is unknown (Daru et al., 2017), highlighting the need for further research to define appropriate serum ferritin cutoffs (7).

#### *6.3.1.5 Absorption of other nutrients*

Lastly, this thesis focused on modeling the absorption of nonheme iron (Chapters 2 and 3), with zinc included to a lesser extent (Chapter 3). However, more precise modeling of the absorption of other nutrients is also important, particularly in the context of promoting a transition from animal-based to plant-based diets. Several essential nutrients are either predominantly found in animal products or are absorbed less efficiently from plant-based foods. In addition to iron and zinc, vegetarians and vegans may be at increased risk of deficiencies in vitamin A, vitamin B12, vitamin D, and calcium. For macronutrients, protein and essential fatty acids may require additional attention (Neufingerl & Eilander, 2021; Plotnikoff et al., 2023). Future studies could investigate methods to accurately model and optimize the absorbable intake of these nutrients (8).

### 6.3.2 Food substitution

The main challenge in modeling food substitution is ensuring the acceptability of substitutes (van Dooren, 2018). This thesis examined acceptability for both food-group- (Chapter 4) and data-driven substitution (Chapter 5). Acceptability was approximated through food similarity, assuming that more similar substitutes are generally more acceptable, and through food quantity adjustments, assuming smaller changes are preferable. This section discusses these approximations in more detail and highlights opportunities for future research.

#### *6.3.2.1 Food similarity*

Common metrics for approximating the acceptability of food substitutes based on similarity include food classification, recipe context, flavor, texture, and nutrient profile (Shirai et al., 2021). We first address food classification and recipe context, as applied in Chapters 4 and 5, followed by the other metrics.

*Food similarity - Food classification*

Food classification captures similarity by assuming that foods within the same group are more similar than those in different groups. In Chapter 4 we therefore assumed that within-food-group substitution improved the acceptability of diets by reducing the need for between-group substitutions. However, we did not model this explicitly, nor do most other diet modeling studies. Future research could incorporate this distinction by, for example, assigning smaller penalties to within-food-group substitutions than to substitutions between different food groups. A conceptually related approach was proposed by (Horgan et al., 2016), who implemented stepwise dietary changes in which adjustments to the quantities of foods already consumed were favored over the addition of new foods or the removal of existing ones. Differentiating and assigning weights to various types of dietary changes could improve the acceptability of modeled diets (9).

When relying on food classification to measure food similarity, it is important to consider the type of classification used. Since existing classifications were not necessarily designed for substitution purposes, the assumed similarity of foods can be arbitrary. For example, the food subgroup "dark-green vegetables" in the Food and Nutrient Database for Dietary Studies (FNDDS) (USDA & ARS, 2020a) includes both raw and cooked spinach. However, these items may not be suitable substitutes within the context of a meal. For raw spinach, a more acceptable substitute might be raw lettuce, which belongs to a different subgroup. This suggests that a vegetable classification based on preparation method may be more appropriate than one based on characteristics such as color. Developing a standardized food classification system for food substitution and making it publicly available could improve the acceptability of food group-based substitutions and prevent duplication of effort among researchers (10).

*Food similarity - Recipe context*

Recipe context defines food similarity based on the principle that foods occurring in similar recipes are more alike than those in different recipes. For instance, tomatoes and onions show high similarity because they often appear in Italian recipes. In contrast, tomatoes and Brussels sprouts show low similarity because the latter appears mostly in Dutch cuisine. In our recipe completion diet model (Chapter 5), food similarity was measured this way, using data on how frequently food items co-occurred in meals. However, we observed that a substitute with a high overall co-occurrence score with the other foods in the meal is not necessarily an acceptable substitute. One reason for this is that co-occurrence scores can compensate for each other. For example, consider a meal consisting of pasta and various vegetables where we want to replace one of the vegetables. Rice may have a high co-occurrence score with the other vegetables and, therefore, be selected as a substitute, even though the meal already contains pasta. The co-occurrence score between rice and pasta is very low, but this can be ignored because the high co-occurrence scores

between rice and the vegetables compensate for it. To prevent such compensation effects, training a recipe completion model from a reversed perspective, focusing on which food items should not be combined, could be an interesting research opportunity to improve the performance of recipe-context-based substitution (11).

Another situation where recipe-context-based substitution can fail is when the type of dish is not taken into account. For example, consider a breakfast dish with toast, eggs, tomatoes, and onions. Replacing the toast with pasta, which has a high co-occurrence score with tomatoes and onions, might result in a suitable dinner recipe but would likely be inappropriate for breakfast. Future research could examine whether training machine learning models with dish type data improves the acceptability of suggested substitutes (12).

Completeness of meals presents a further challenge for recipe-context-based substitutions. In our study (Chapter 5), we observed that although some food suggestions aligned well with the meal, important components such as condiments or carbohydrate sources were still missing. Future research could leverage information on the role of a food item within a meal to improve the quality of substitutions generated by machine learning models. Alternatively, a hybrid approach could be employed, where the machine learning model scores the acceptability of food combinations, while additional constraints in the diet model ensure that each meal contains a complete set of components (13).

The performance of recipe-context-based substitution could also be improved by applying more advanced machine learning methods. In our study (Chapter 5), we used kernel ridge regression, as this linear model can be readily integrated into a linear diet model. However, state-of-the-art approaches rely on large language models such as Bidirectional Encoder Representations from Transformers (BERT) (Morales-Garzón et al., 2025), or on knowledge graph embeddings (Shirai et al., 2021). While these advanced methods offer improved accuracy in selecting acceptable food alternatives, their implementation within diet modeling remains challenging due to non-linearity and high computational cost. Future research may explore strategies to integrate complex machine learning models within the constraints of diet modeling (14).

Furthermore, the quality of the data used to train the model affects the appropriateness of substitutes. The consumption data we used for model training regularly included food combinations that most consumers would find unacceptable, at times resulting in inappropriate suggestions (Chapter 5). Using recipe data rather than consumption data may improve results. Future studies could train the recipe completion model on recipe data, which could then be applied to optimize observed consumption. However, using recipe data for model training also presents challenges, as foods are often named differently across databases. Matching recipe and consumption data can therefore be difficult, since one-to-one correspondences are not always available and identifying the most appropriate alternative may be ambiguous. Additionally, when performed manually, this matching

process is time-consuming. The research domain that focuses on automating this process is known as entity linking (Feher et al., 2023). In future studies, this technique could be applied to match recipe data with food consumption data (15).

The use of consumer preference data could also improve the quality of food substitutes. Although such data is rarely freely accessible and thus not commonly used in diet modeling, it is frequently employed in food recommender systems (Tran et al., 2021). Future research could assess whether including this data improves the acceptability of substitutions in diet models (16).

Lastly, rather than substituting individual foods within a recipe, future studies could explore substituting entire meals based on recipe similarity. This may offer a simpler approach to designing acceptable menu plans. For this, we may also take inspiration from food recommender systems, where both individual food items and recipes can be suggested (Mahajan & Kaur, 2024). However, a key prerequisite is that recipe databases include a sufficient number of healthy alternatives, which is not always the case at present (Trattner & Elsweiler, 2017) (17).

*Food similarity - Flavor, Texture, and Nutrient profile*

Flavor is occasionally used as a measure of food similarity in food recommender systems, food substitution algorithms, and recipe completion algorithms (Abdool et al., 2015; Gim et al., 2022; De Clercq et al., 2016). The underlying idea is that the more flavor compounds food items share, the more similar they are. Additionally, flavor information can also be used to reveal complementary foods and flavor combinations (Park et al., 2021). To the best of our knowledge, flavor information has not yet been incorporated into diet modeling. Investigating its potential to improve the suitability of food substitutions could be a promising area for future research (18).

Texture is not commonly used as a food similarity metric, possibly due to the lack of dedicated food texture databases. One example of a food recommender system that incorporated texture data collected this information through a user survey (Abdool et al., 2015). Nonetheless, given that texture is an important driver of food preferences (Jeltema et al., 2016), it may be worthwhile to investigate how texture data can be used to improve the acceptability of food substitutions (19).

Lastly, nutrient profiles can be used to measure food similarity, but their relevance for food substitution in diet modeling remains uncertain. Since the objective is often to find healthier alternatives, it might be preferable to select substitutes with substantially different nutrient profiles. On the other hand, there may be situations where minor changes suffice, such as substituting one type of bread with another that has less salt but comparable levels of other nutrients. Further research could clarify whether nutrient profiles serve as a useful similarity metric for selecting acceptable substitutes (20).

#### *6.3.2.2 Food quantities*

Apart from food similarity, the second main factor influencing the acceptability of food substitutions is food quantity (Section 6.3.2). In diet modeling, acceptability is often approximated by the magnitude of food quantity changes, with smaller changes associated with higher acceptability (van Dooren, 2018), as was also assumed in Chapters 4 and 5. Additionally, quantities of food groups and items are commonly constrained by lower and upper bounds, such as the 5th and 95th percentiles of observed intakes (van Dooren, 2018). However, applying such constraints becomes more challenging in the context of meals, as appropriate portion sizes depend not only on the type of food, but also on the number of items in the meal and the type of recipe. Alternatively, a data-driven approach could be employed, in which machine learning is used to estimate appropriate food quantities, as demonstrated by (Chen et al., 2020; Li et al., 2021; Choi et al., 2023). This method has not yet been applied in diet modeling, and future research could investigate its feasibility and potential benefits over conventional rule-based approaches (21).

#### *6.3.2.3 Trade-offs*

Acceptability of food substitution thus depends on both the similarity of food substitutes and the extent of food quantity changes. However, since multiple metrics can be used to quantify acceptability, an important question is how to combine these measures and address potential trade-offs. For example, there can be a trade-off between similarity metrics such as food classification and recipe context. A food substitute may be considered similar because it belongs to the same food group, yet dissimilar due to its low co-occurrence with the other items in the meal. Another example of a trade-off exists between food similarity and food quantity changes. Foods that are more similar may have smaller differences in nutrient profiles and therefore require larger quantity adjustments to meet nutritional and environmental goals. This trade-off was also demonstrated in Chapter 5, where using broader food groups, with items that are less similar, enabled greater improvements in nutritional adequacy and sustainability of diets, but led to less acceptable substitutions. Further exploration of these trade-offs across the different aspects of nutrition, sustainability, and acceptability could provide valuable insights. In doing so, future diet modeling studies may draw inspiration from other research fields. For instance, epidemiology may offer a scientific basis for modeling nutritional trade-offs, such as prioritizing specific macro- or micronutrient deficiencies when it is not possible to meet all dietary guidelines. Likewise, consumer research on food preferences and choices could provide valuable insights for modeling trade-offs in acceptability. Here, it is also important to consider, in addition to food similarity and changes in food quantities, other factors that influence acceptability, such as food prices (Fantechi et al., 2025) (22).

Insights from environmental sciences are essential for modeling trade-offs in sustainability. In this thesis (Chapter 4), we focused solely on greenhouse gas emissions, as they are the

most commonly used indicator of dietary sustainability (Wickramasinghe et al., 2013) and because our study specifically addressed the methodology of food substitution rather than providing a full evaluation of the environmental impacts of dietary changes. Future research aiming to assess the latter could benefit from a more integrated approach to designing sustainable diets by incorporating additional environmental metrics (23).

Finally, alongside assessing various trade-offs, it is essential to examine methods for engaging consumers in making these trade-offs, since they ultimately determine their dietary intake. For example, consumers could be presented with different menu plan options illustrating various strategies for improving the nutritional adequacy and sustainability of their diet. This approach allows consumers to select the types of dietary changes they prefer. A comparable approach has been investigated in food recommender systems (Starke et al., 2023) (24).

**Table 6.2:** Summary of research opportunities.

| | |
|---|---|
| **Iron absorption** | |
| 1 | Investigate which nonheme iron absorption equation is most accurate |
| 2 | Develop new nonheme iron absorption equation |
| 3 | Investigate performance of multivariate piecewise linear approximation |
| 4 | Expand nutrient composition tables with compounds that influence iron absorption |
| 5 | Assess accuracy of heme iron absorption estimates |
| 6 | Account for iron absorption from fortified foods |
| 7 | Determine appropriate serum ferritin cutoffs |
| 8 | Investigate modeling the absorption of other important nutrients |
| **Food substitution** | |
| 9 | Differentiate between various types of dietary changes |
| 10 | Develop a standardized food classification system for food substitution |
| 11 | Train substitution model on uncommon food combinations |
| 12 | Train substitution model on dish type data |
| 13 | Account for meal completeness |
| 14 | Apply more advanced machine learning models for food substitution |
| 15 | Train substitution model on recipe data |
| 16 | Train substitution model on consumer preference data |
| 17 | Substitute meals instead of individual food items |
| 18 | Train substitution model on flavor data |
| 19 | Train substitution model on texture data |
| 20 | Train substitution model on nutrient data |
| 21 | Train model to estimate acceptable food quantities |
| 22 | Explore trade-offs between nutrition, sustainability, and acceptability |
| 23 | Include additional environmental metrics to model sustainability |
| 24 | Investigate methods for engaging consumers in making trade-offs |

## 6.4 Conclusion

In this thesis, we discussed various approaches for modeling iron absorption and food substitution to support the development of diets that are healthy, sustainable, and acceptable. Modeling iron absorption using diet-dependent absorption equations, rather than constant absorption factors, allows for more precise optimization of absorbable iron content in diet plans. These nonlinear absorption equations can be integrated into diet modeling through nonlinear programming or piecewise linear approximation, with the latter recommended for mixed-integer diet models. When substituting food items, it is recommended to exchange items both between- and within food groups, as this requires fewer dietary changes to achieve the same nutritional and sustainability goals compared with substitutions limited to exchanges between food groups alone. As an alternative to such a rule-based approach, machine learning can be applied, as demonstrated by the recipe completion diet model we developed, which offers a promising strategy for personalizing food substitutions and thereby improving their acceptability.

To conclude, the methods presented for modeling iron absorption and food substitution effectively support the design of diets that are healthy, environmentally sustainable, and acceptable. They are particularly valuable for future research aimed at promoting sustainable diets with reduced animal-based products, while ensuring nutritional adequacy and acceptability.

# References


Abdollahpouri, H. (2020). *Popularity bias in recommendation: a multi-stakeholder perspective*. Ph.D. thesis University of Colorado at Boulder.

Abdool, H., Pooransingh, A., & Li, Y. (2015). Recommend my dish: A multi-sensory food recommender. In *2015 IEEE Pacific Rim conference on communications, computers and signal processing (PACRIM)* (pp. 240–245). IEEE.

Abuzaid, A., & Alkronz, E. (2024). A comparative study on univariate outlier winsorization methods in data science context. *Italian Journal of Applied Statistics*, *36*, 85–99.

Ahmadi, F., Ganjkhanloo, F., & Ghobadi, K. (2020). Inverse learning: Solving partially known models using inverse optimization. ArXiv preprint arXiv:2011.03038.

Ahuja, J. K., & Perloff, B. P. (2008). Quality control procedures for the usda food and nutrient database for dietary studies nutrient values. *Journal of Food Composition and Analysis*, *21*, S119–S124.

Aiking, H., & de Boer, J. (2020). The next protein transition. *Trends in Food Science & Technology*, *105*, 515–522.

Andrewski, E., Cheng, K., & Vanderpool, C. (2022). Nutritional deficiencies in vegetarian, gluten-free, and ketogenic diets. *Pediatrics In Review*, *43*, 61–70.

Archundia-Herrera, M. C., Nunes, F., Barrios, I. D., Park, C. Y., Bell, R. C., & O'Brien, K. O. (2024). Development of a database for the estimation of heme iron and nonheme iron content of animal-based foods. *Current Developments in Nutrition*, *8*, 102130.

Arens, Z. G., & Hamilton, R. W. (2018). The substitution strategy dilemma: Substitute selection versus substitute effectiveness. *Journal of the Academy of Marketing Science*, *46*, 130–146.

Armah, S. M., Boy, E., Chen, D., Candal, P., & Reddy, M. B. (2015a). Regular consumption of a high-phytate diet reduces the inhibitory effect of phytate on nonheme-iron absorption in women with suboptimal iron stores. *The Journal of nutrition*, *145*, 1735–1739.

Armah, S. M., Carriquiry, A., Sullivan, D., Cook, J. D., & Reddy, M. B. (2013). A complete diet-based algorithm for predicting nonheme iron absorption in adults. *The*

*Journal of nutrition*, *143*, 1136–1140.

Armah, S. M., Carriquiry, A. L., & Reddy, M. B. (2015b). Total iron bioavailability from the us diet is lower than the current estimate. *The Journal of Nutrition*, *145*, 2617–2621.

Babayev, D. A. (1997). Piece-wise linear approximation of functions of two variables. *Journal of Heuristics*, *2*, 313–320.

Bailey, R. L., West Jr, K. P., & Black, R. E. (2015). The epidemiology of global micronutrient deficiencies. *Annals of nutrition and metabolism*, *66*, 22–33.

Barré, T., Perignon, M., Gazan, R., Vieux, F., Micard, V., Amiot, M.-J., & Darmon, N. (2018). Integrating nutrient bioavailability and co-production links when identifying sustainable diets: how low should we reduce meat consumption? *PLoS One*, *13*, e0191767.

Bashiri, B., Kaleda, A., Gavrilova, O., & Vilu, R. (2025). A culturally acceptable shift in diet to reduce land footprint: an optimization study for estonia. *Environmental Modeling and Assessment*, *30*, 125–139.

Beard, J. L., Murray-Kolb, L. E., Haas, J. D., & Lawrence, F. (2007). Iron absorption prediction equations lack agreement and underestimate iron absorption. *The Journal of nutrition*, *137*, 1741–1746.

Benvenuti, L., & De Santis, A. (2020). Making a sustainable diet acceptable: an emerging programming model with applications to schools and nursing homes menus. *Frontiers in Nutrition*, *7*, 562833.

Bolo, A., Verger, E., Fouillet, H., & Mariotti, F. (2024). Exploring multidimensional and within-food group diversity for diet quality and long-term health in high-income countries. *Advances in Nutrition*, *15*, 100278.

Bos, C., Van der Lans, I. A., Van Rijnsoever, F. J., & Van Trijp, H. C. (2013). Understanding consumer acceptance of intervention strategies for healthy food choices: a qualitative study. *BMC public health*, *13*, 1–14.

Bowman, S. A., Clemens, J. C., Friday, J. E., Lynch, K. L., & Moshfegh, A. J. (2022). Food patterns equivalents database 2017-2018: Methodology and user guide [online]. `http://www.ars.usda.gov/nea/bhnrc/fsrg`.

Broekema, R., Tyszler, M., van't Veer, P., Kok, F. J., Martin, A., Lluch, A., & Blonk, H. T. J. (2020). Future-proof and sustainable healthy diets based on current eating patterns in the netherlands. *The American journal of clinical nutrition*, *112*, 1338–1347.

Burer, S., & Letchford, A. N. (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, *17*, 97–106.

CDC, & NCHS (2014). National health and nutrition examination survey questionnaire (or examination protocol, or laboratory protocol). https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013.

CDC, & NCHS (2016). National health and nutrition examination survey questionnaire (or examination protocol, or laboratory protocol). https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015.

CDC, & NCHS (2018). National health and nutrition examination survey questionnaire (or examination protocol, or laboratory protocol). https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017.

Chai, B. C., van der Voort, J. R., Grofelnik, K., Eliasdottir, H. G., Klöss, I., & Perez-Cueto, F. J. (2019). Which diet has the least environmental impact on our planet? a systematic review of vegan, vegetarian and omnivorous diets. *Sustainability*, *11*, 4110.

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, *41*, 1–39.

Chen, M., Jia, X., Gorbonos, E., Hoang, C. T., Yu, X., & Liu, Y. (2020). Eating healthier: Exploring nutrition information for healthier recipe recommendation. *Information Processing & Management*, *57*, 102051.

Choi, D., Gim, M., Badreddine, S., Kim, H., Park, D., & Kang, J. (2023). Kitchenscale: Learning to predict ingredient quantities from recipe contexts. *Expert Systems with Applications*, *224*, 120041.

Cioffi, C. E., Figueroa, J., & Welsh, J. A. (2018). Added sugar intake among pregnant women in the united states: National health and nutrition examination survey 2003-2012. *Journal of the Academy of Nutrition and Dietetics*, *118*, 886–895. e1.

Claassen, G., Hendriks, T. H., & Hendrix, E. M. (2007). *Decision Science: theory and applications*. Wageningen Academic Publishers.

Collings, R., Harvey, L. J., Hooper, L., Hurst, R., Brown, T. J., Ansett, J., King, M., & Fairweather-Tait, S. J. (2013). The absorption of iron from whole diets: a systematic review. *The American journal of clinical nutrition*, *98*, 65–81.

Conrad, Z., Cyril, A., Kowalski, C., Jackson, E., Hendrickx, B., Lan, J. J., McDowell, A., Salesses, M., Love, D. C., & Wiipongwii, T. (2022). Diet sustainability analyses can be improved with updates to the food commodity intake database. *Frontiers in Nutrition*, *9*, 868485.

Conrad, Z., Drewnowski, A., Belury, M. A., & Love, D. C. (2023). Greenhouse gas emissions, cost, and diet quality of specific diet patterns in the united states. *The American Journal of Clinical Nutrition*, *117*, 1186–1194.

Conrad, Z., Niles, M. T., Neher, D. A., Roy, E. D., Tichenor, N. E., & Jahns, L. (2018). Relationship between food waste, diet quality, and environmental sustainability. *PloS one*, *13*, e0195405.

Conway, R. E., Powell, J. J., & Geissler, C. A. (2007). A food-group based algorithm to predict non-heme iron absorption. *International Journal of Food Sciences and Nutrition*, *58*, 29–41.

Cueto, P. F., Roet, M., & Słowik, A. (2019). Completing partial recipes using item-based collaborative filtering to recommend ingredients. ArXiv preprint arXiv:1907.12380.

Dahl, G. (2010). An introduction to convexity.

Dainty, J. R., Berry, R., Lynch, S. R., Harvey, L. J., & Fairweather-Tait, S. J. (2014). Estimation of dietary iron bioavailability from food iron intake and iron status. *PLoS One*, *9*, e111824.

Daru, J., Colman, K., Stanworth, S. J., De La Salle, B., Wood, E. M., & Pasricha, S.-R. (2017). Serum ferritin as an indicator of iron status: what do we need to know? *The American journal of clinical nutrition*, *106*, 1634S–1639S.

De Carli, E., Dias, G. C., Morimoto, J. M., Marchioni, D. M. L., & Colli, C. (2018). Dietary iron bioavailability: agreement between estimation methods and association with serum ferritin concentrations in women of childbearing age. *Nutrients*, *10*, 650.

De Clercq, M., Stock, M., De Baets, B., & Waegeman, W. (2016). Data-driven recipe completion using machine learning methods. *Trends in Food Science & Technology*, *49*, 1–13.

Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv preprint arXiv:1810.04805.

van Dooren, C. (2018). A review of the use of linear programming to optimize diets, nutritiously, economically and environmentally. *Frontiers in nutrition*, *5*, 352128.

Dussiot, A., Fouillet, H., Perraud, E., Salomé, M., Huneau, J.-F., Kesse-Guyot, E., & Mariotti, F. (2022a). Nutritional issues and dietary levers during gradual meat reduction–a sequential diet optimization study to achieve progressively healthier diets. *Clinical Nutrition*, *41*, 2597–2606.

Dussiot, A., Fouillet, H., Wang, J., Salomé, M., Huneau, J.-F., Kesse-Guyot, E., & Mariotti, F. (2022b). Modeled healthy eating patterns are largely constrained by currently estimated requirements for bioavailable iron and zinc: a diet optimization study in french adults. *The American journal of clinical nutrition*, *115*, 958–969.

EFSA (2015). Scientific opinion on dietary reference values for iron. *EFSA Journal*, *13*, 4254. (NDA).

Eftimov, T., Popovski, G., Valenčič, E., & Seljak, B. K. (2020). Foodex2vec: New foods' representation for advanced food data analysis. *Food and Chemical Toxicology*, *138*,

111169.

Elorinne, A.-L., Alfthan, G., Erlund, I., Kivimäki, H., Paju, A., Salminen, I., Turpeinen, U., Voutilainen, S., & Laakso, J. (2016). Food and nutrient intake and nutritional status of finnish vegans and non-vegetarians. *PloS one*, *11*, e0148235.

Embling, R., Pink, A. E., Gatzemeier, J., Price, M., Lee, M. D., & Wilkinson, L. L. (2021). Effect of food variety on intake of a meal: a systematic review and meta-analysis. *The American journal of clinical nutrition*, *113*, 716–741.

Ericksen, P. J. (2008). Conceptualizing food systems for global environmental change research. *Global environmental change*, *18*, 234–245.

Eustachio Colombo, P., Elinder, L. S., Lindroos, A. K., & Parlesak, A. (2021). Designing nutritionally adequate and climate-friendly diets for omnivorous, pescatarian, vegetarian and vegan adolescents in sweden using linear optimization. *Nutrients*, *13*, 2507.

Fairweather-Tait, S. J., Jennings, A., Harvey, L. J., Berry, R., Walton, J., & Dainty, J. R. (2017). Modeling tool for calculating dietary iron bioavailability in iron-sufficient adults. *The American journal of clinical nutrition*, *105*, 1408–1414.

Fajemisin, A. O., Maragno, D., & den Hertog, D. (2024). Optimization with constraint learning: A framework and survey. *European Journal of Operational Research*, *314*, 1–14.

Fantechi, T., Contini, C., Casini, L., & Lähteenmäki, L. (2025). Dietary dilemmas: Navigating trade-offs in food choice for sustainability, health, naturalness, and price. *Food Quality and Preference*, *129*, 105497.

FAO, & WHO (2011). Fao's approach to nutrition-sensitive agricultural development. Food and Agriculture Organization of the United Nations. `http://www.fao.org/3/a-at543e.pdf`.

FAO, IFAD, UNICEF, WFP and WHO (2024). The state of food security and nutrition in the world 2024: Financing to end hunger, food insecurity and malnutrition in all its forms. Food and Agriculture Organization of the United Nations. `https://doi.org/10.4060/cd1254en`.

Fatemi, S. F., Irankhah, K., Kruger, J., Bruins, M. J., & Sobhani, S. R. (2023). Implementing micronutrient fortification programs as a potential practical contribution to achieving sustainable diets. *Nutrition Bulletin*, *48*, 411–424.

Feher, D., Ibrahim, F., Cheng, Z., Schlegel, V., Maidment, T., Bagshaw, J., & Batista-Navarro, R. (2023). Few-shot entity linking of food names. *Information Processing & Management*, *60*, 103463.

Forde, C. G., & Decker, E. A. (2022). The importance of food processing and eating behavior in promoting healthy and sustainable diets. *Annual review of nutrition*, *42*,

377–399.

Fouillet, H., Huneau, J., Perraud, E., Dussiot, A., Wang, J., Kesse-Guyot, E., & Mariotti, F. (2025). Plant to animal protein ratio in the diet of the elderly: potential for increase and impacts on nutrient adequacy and long-term health - a diet optimization study. *The American Journal of Clinical Nutrition*, .

Fu, H., Li, Y., Jiang, P., Zhou, S., & Liao, C. (2024). Transition towards sustainable diets: Multi-objective optimization of dietary pattern in china. *Sustainable Production and Consumption*, *48*, 14–28.

Galetti, V., Stoffel, N. U., Sieber, C., Zeder, C., Moretti, D., & Zimmermann, M. B. (2021). Threshold ferritin and hepcidin concentrations indicating early iron deficiency in young women based on upregulation of iron absorption. *EClinicalMedicine*, *39*, 101052.

GAMS (2024a). Cplex. `https://www.gams.com/latest/docs/S_CPLEX.html`.

GAMS (2024b). LINDO and LINDOGlobal. `https://www.gams.com/latest/docs/S_LINDO.html`.

Gazan, R., Barré, T., Perignon, M., Maillot, M., Darmon, N., & Vieux, F. (2018a). A methodology to compile food metrics related to diet sustainability into a single food database: application to the french case. *Food chemistry*, *238*, 125–133.

Gazan, R., Brouzes, C. M., Vieux, F., Maillot, M., Lluch, A., & Darmon, N. (2018b). Mathematical optimization to explore tomorrow's sustainable diets: a narrative review. *Advances in Nutrition*, *9*, 602–616.

Geißler, B., Martin, A., Morsi, A., & Schewe, L. (2011). Using piecewise linear functions for solving minlps. In *Mixed integer nonlinear programming* (pp. 287–314). Springer.

Gerdessen, J., & De Vries, J. (2015). Diet models with linear goal programming: impact of achievement functions. *European journal of clinical nutrition*, *69*, 1272–1278.

Ghobadi, K., & Mahmoudzadeh, H. (2021). Inferring linear feasible regions using inverse optimization. *European Journal of Operational Research*, *290*, 829–843.

Gibson, R. S. (2007). The role of diet-and host-related factors in nutrient bioavailability and thus in nutrient-based dietary requirement estimates. *Food and Nutrition Bulletin*, *28*, S77–S100.

Gim, M., Choi, D., Maruyama, K., Choi, J., Kim, H., Park, D., & Kang, J. (2022). Recipemind: guiding ingredient choices from food pairing to recipe completion using cascaded set transformer. In *Proceedings of the 31st ACM international conference on information & knowledge management* (pp. 3092–3102).

Gómez, M. I., Barrett, C. B., Raney, T., Pinstrup-Andersen, P., Meerman, J., Croppenstedt, A., Carisma, B., & Thompson, B. (2013). Post-green revolution food systems and the triple burden of malnutrition. *Food Policy*, *42*, 129–138.

Grasso, A. C., Besselink, J. J., Tyszler, M., & Bruins, M. J. (2023). The potential of food fortification as an enabler of more environmentally sustainable, nutritionally adequate diets. *Nutrients*, *15*, 2473.

Graça, J., Godinho, C. A., & Truninger, M. (2019). Reducing meat consumption and following plant-based diets: Current evidence and future directions to inform integrated transitions. *Trends in Food Science & Technology*, *91*, 380–390.

Green, R., Milner, J., Dangour, A. D., Haines, A., Chalabi, Z., Markandya, A., Spadaro, J., & Wilkinson, P. (2015). The potential to reduce greenhouse gas emissions in the uk through healthy and realistic dietary change. *Climatic Change*, *129*, 253–265.

György, A., & Kocsis, L. (2011). Efficient multi-start strategies for local search algorithms. *Journal of Artificial Intelligence Research*, *41*, 407–444.

Haider, L. M., Schwingshackl, L., Hoffmann, G., & Ekmekcioglu, C. (2018). The effect of vegetarian diets on iron status in adults: A systematic review and meta-analysis. *Critical reviews in food science and nutrition*, *58*, 1359–1374.

Hallberg, L., & Hulthén, L. (2000). Prediction of dietary iron absorption: an algorithm for calculating absorption and bioavailability of dietary iron. *The American journal of clinical nutrition*, *71*, 1147–1160.

Hallberg, L., & Hulthén, L. (2000b). Erratum to hallberg & hulthén (2000). *The American Journal of Clinical Nutrition*, *72*, 1242–1242.

Hallberg, L., & Hulthén, L. (2006). Errata to hallberg & hulthén (2000). *The American Journal of Clinical Nutrition*, *84*, 1253–1253.

Hambidge, K. M., Miller, L. V., Westcott, J. E., Sheng, X., & Krebs, N. F. (2010). Zinc bioavailability and homeostasis. *The American journal of clinical nutrition*, *91*, 1478S–1483S.

Heerschop, S. N., Kanellopoulos, A., Biesbroek, S., & van 't Veer, P. (2023). Shifting towards optimized healthy and sustainable dutch diets: impact on protein quality. *European Journal of Nutrition*, *62*, 2115–2128.

Heller, M. C., Willits-Smith, A., Meyer, R., Keoleian, G. A., & Rose, D. (2018). Greenhouse gas emissions and energy use associated with production of individual self-selected us diets. *Environmental Research Letters*, *13*, 044004.

Hemler, E. C., & Hu, F. B. (2019). Plant-based diets for personal, population, and planetary health. *Advances in Nutrition*, *10*, S275–S283.

Hill, J. O. (2009). Can a small-changes approach help address the obesity epidemic? a report of the joint task force of the american society for nutrition, institute of food technologists, and international food information council. *The American journal of clinical nutrition*, *89*, 477–484.

Hong, C. et al. (2020). *An Approach to de-homogenizing recommender systems*. Ph.D. thesis Massachusetts Institute of Technology.

Hoos, H. H., & Tsang, E. (2006). Local search methods. In *Foundations of Artificial Intelligence* (pp. 135–167). Elsevier volume 2.

Horgan, G. W., Perrin, A., Whybrow, S., & Macdiarmid, J. I. (2016). Achieving dietary recommendations and reducing greenhouse gas emissions: modelling diets to minimise the change from current intakes. *International Journal of Behavioral Nutrition and Physical Activity*, *13*, 1–11.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2016). More similar but less satisfying: Comparing preferences for and the efficacy of within-and cross-category substitutes for food. *Psychological Science*, *27*, 894–903.

Huma, N., Salim-Ur-Rehman, Anjum, F. M., Murtaza, M. A., & Sheikh, M. A. (2007). Food fortification strategy - preventing iron deficiency anemia: a review. *Critical reviews in food science and nutrition*, *47*, 259–265.

Hunt, J. R. (2002). Moving toward a plant-based diet: are iron and zinc at risk? *Nutrition reviews*, *60*, 127–134.

Hunt, J. R. (2003). Bioavailability of iron, zinc, and other trace minerals from vegetarian diets. *The American journal of clinical nutrition*, *78*, 633S–639S.

Hurrell, R., & Egli, I. (2010). Iron bioavailability and dietary reference values. *The American journal of clinical nutrition*, *91*, 1461S–1467S.

Hurrell, R. F. (2002). Fortification: overcoming technical and practical barriers. *The Journal of nutrition*, *132*, 806S–812S.

Hurrell, R. F., Reddy, M. B., Juillerat, M., & Cook, J. D. (2006). Meat protein fractions enhance nonheme iron absorption in humans. *The Journal of nutrition*, *136*, 2808–2812.

IOM (2000). Dietary reference intakes: Applications in dietary assessment. Institute of Medicine.

IOM (2001). *Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc*. Washington (DC): National Academies Press (US); 2001. Institute of Medicine.

IOM (2006). *Dietary reference intakes: the essential guide to nutrient requirements*. National Academies Press. Institute Of Medicine.

IOM (2011). *Dietary reference intakes for calcium and vitamin D*. National Academies Press. Institute Of Medicine.

Jeltema, M., Beckley, J., & Vahalik, J. (2016). Food texture assessment and preference based on mouth behavior. *Food Quality and Preference*, *52*, 160–171.

Kazimipour, B., Li, X., & Qin, A. K. (2013). Initialization methods for large scale global optimization. In *2013 IEEE congress on evolutionary computation* (pp. 2750–2757). IEEE.

Kesse-Guyot, E., Baudry, J., Berlivet, J., Perraud, E., Julia, C., Touvier, M., Allès, B., Lairon, D., Hercberg, S., & Fouillet, H. (2025). To be climate-friendly, food-based dietary guidelines must include limits on total meat consumption - modeling from the case of france. *International Journal of Behavioral Nutrition and Physical Activity*, *22*, 95.

Khorraminezhad, L., Leclercq, M., Droit, A., Bilodeau, J.-F., & Rudkowska, I. (2020). Statistical and machine-learning analyses in nutritional genomics studies. *Nutrients*, *12*, 3140.

Kourouniotis, S., Keast, R. S., Riddell, L. J., Lacy, K., Thorpe, M. G., & Cicerale, S. (2016). The importance of taste on dietary choice, behaviour and intake in a group of young adults. *Appetite*, *103*, 1–7.

Kronqvist, J., Bernal, D. E., Lundell, A., & Grossmann, I. E. (2019). A review and comparison of solvers for convex minlp. *Optimization and Engineering*, *20*, 397–455.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems–a survey. *Knowledge-based systems*, *123*, 154–162.

Lansche, J., Gaillard, G., Nemecek, T., Mouron, P., Peano, L., Bengoa, X., Humbert, S., & Loerincik, Y. (2013). The world food lca database project: towards more accurate food datasets. In *The 6th International Conference on Life Cycle Management (LCM 2013), Gothenburg*.

Larvie, D. Y., & Armah, S. M. (2021). Estimating phytate intake from the us diet using the nhanes data. *Journal of Food Composition and Analysis*, *102*, 104050.

Lee, J., & Leyffer, S. (2011). *Mixed integer nonlinear programming* volume 154. Springer Science & Business Media.

Lei, Z., ul Haq, A., Dorraki, M., Zhang, D., & Abbott, D. (2020). Composing recipes based on nutrients in food in a machine learning context. *Neurocomputing*, *415*, 382–396.

Lenaerts, B., & Demont, M. (2021). The global burden of chronic and hidden hunger revisited: new panel data evidence spanning 1990–2017. *Global food security*, *28*, 100480.

Leonard, U. M., & Kiely, M. E. (2024). Can micronutrient requirements be met by diets from sustainable sources: outcomes of dietary modelling studies using diet optimization. *Annals of Medicine*, *56*, 2389295.

Leung, P., Wanitprapha, K., & Quinn, L. A. (1995). A recipe-based, diet-planning modelling system. *British Journal of Nutrition*, *74*, 151–162.

Li, K., Chen, Y., Li, H., Mu, X., Zhang, X., & Liu, X. (2021). Deeprecipes: Exploring massive online recipes and recovering food ingredient amounts. *IEEE Access*, *9*, 67859–

67873.

Lopez, A., Cacoub, P., Macdougall, I. C., & Peyrin-Biroulet, L. (2016). Iron deficiency anaemia. *The Lancet*, *387*, 907–916.

Lourenço, H. R., Martin, O. C., & Stützle, T. (2003). Iterated local search. In *Handbook of metaheuristics* (pp. 320–353). Springer.

Mahajan, P., & Kaur, P. D. (2024). A systematic literature review of food recommender systems. *SN Computer Science*, *5*, 174.

Maillot, M., Vieux, F., Amiot, M. J., & Darmon, N. (2010). Individual diet modeling translates nutrient recommendations into realistic and individual-specific food choices. *The American journal of clinical nutrition*, *91*, 421–430.

Maragno, D., Wiberg, H., Bertsimas, D., Birbil, S. I., Hertog, D. d., & Fajemisin, A. (2021). Mixed-integer optimization with constraint learning. ArXiv preprint arXiv:2111.04469.

Mariotti, F., Havard, S., Morise, A., Nadaud, P., Sirot, V., Wetzler, S., & Margaritis, I. (2021). Perspective: modeling healthy eating patterns for food-based dietary guidelines - scientific concepts, methodological processes, limitations, and lessons. *Advances in Nutrition*, *12*, 590–599.

Martınez-Navarrete, N., Camacho, M. M., Martınez-Lahuerta, J., Martınez-Monzó, J., & Fito, P. (2002). Iron deficiency and iron fortified foods—a review. *Food Research International*, *35*, 225–231.

Masset, G., Monsivais, P., Maillot, M., Darmon, N., & Drewnowski, A. (2009). Diet optimization methods can help translate dietary guidelines into a cancer prevention food plan. *The Journal of nutrition*, *139*, 1541–1548.

Mata, J., Todd, P. M., & Lippke, S. (2010). When weight management lasts. lower perceived rule complexity increases adherence. *Appetite*, *54*, 37–43.

McCullough, B. D., & Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, *93*, 873–892.

Meyers, L. D., Hellwig, J. P., Otten, J. J. et al. (2006). *Dietary reference intakes: the essential guide to nutrient requirements*. National Academies Press.

Miller, L. V., Krebs, N. F., & Hambidge, K. M. (2007). A mathematical model of zinc absorption in humans as a function of dietary zinc and phytate. *The Journal of nutrition*, *137*, 135–141.

Miller, L. V., Krebs, N. F., & Hambidge, K. M. (2013). Mathematical model of zinc absorption: effects of dietary calcium, protein and iron on zinc absorption. *British Journal of Nutrition*, *109*, 695–700.

Monsen, E. R. (1988). Iron nutrition and absorption: dietary factors which impact iron bioavailability. *Journal of the American Dietetic Association*, *88*, 786–790.

Morales-Garzón, A., Gómez-Romero, J., & Martin-Bautista, M. J. (2021). A word embedding-based method for unsupervised adaptation of cooking recipes. *IEEE Access*, *9*, 27389–27404.

Morales-Garzón, A., Gutiérrez-Batista, K., & Martin-Bautista, M. J. (2025). Adaptafood: an intelligent system to adapt recipes to specialised diets and healthy lifestyles. *Multimedia Systems*, *31*, 1–24.

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*, 3055–71.

Musto, C., Starke, A. D., Trattner, C., Rapp, A., & Semeraro, G. (2021). Exploring the effects of natural language justifications in food recommender systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 147–157).

Neubert, K., & Brunner, E. (2007). A studentized permutation test for the non-parametric behrens–fisher problem. *Computational Statistics & Data Analysis*, *51*, 5192–5204.

Neufingerl, N., & Eilander, A. (2021). Nutrient intake and status in adults consuming plant-based diets compared to meat-eaters: a systematic review. *Nutrients*, *14*, 29.

NIH (2024). Nutrient recommendations and databases. National Institutes of Health. `https://ods.od.nih.gov/HealthInformation/nutrientrecommendations.aspx`.

Nordman, M., Stockmarr, A., Lassen, A. D., & Trolle, E. (2024). Low-carbon diets across diverse dietary patterns: Addressing population heterogeneity under constrained optimization. *Science of the Total Environment*, *953*, 176155.

Pahikkala, T., Stock, M., Airola, A., Aittokallio, T., De Baets, B., & Waegeman, W. (2014). A two-step learning approach for solving full and almost full cold start problems in dyadic prediction. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part II 14* (pp. 517–532). Springer.

Park, D., Kim, K., Kim, S., Spranger, M., & Kang, J. (2021). Flavorgraph: a large-scale food-chemical graph for generating food representations and recommending food pairings. *Scientific reports*, *11*, 931.

Pasricha, S.-R., Tye-Din, J., Muckenthaler, M. U., & Swinkels, D. W. (2021). Iron deficiency. *The Lancet*, *397*, 233–248.

Pellegrini, C., Özsoy, E., Wintergerst, M., & Groh, G. (2021). Exploiting food embeddings for ingredient substitution. In *HEALTHINF* (pp. 67–77).

Perignon, M., Masset, G., Ferrari, G., Barré, T., Vieux, F., Maillot, M., Amiot, M.-J., & Darmon, N. (2016). How low can dietary greenhouse gas emissions be reduced without

impairing nutritional adequacy, affordability and acceptability of the diet? a modelling study to guide sustainable food choices. *Public health nutrition*, *19*, 2662–2674.

Phillips, F. (2005). Vegetarian nutrition. *Nutrition Bulletin*, *30*, 132–167.

Ploskas, N., & Sahinidis, N. V. (2022). Review and comparison of algorithms and software for mixed-integer derivative-free optimization. *Journal of Global Optimization*, (pp. 1–30).

Plotnikoff, G. A., Dobberstein, L., & Raatz, S. (2023). Nutritional assessment of the symptomatic patient on a plant-based diet: seven key questions. *Nutrients*, *15*, 1387.

Reddy, M. B., Hurrell, R. F., & Cook, J. D. (2000). Estimation of nonheme-iron bioavailability from meal composition. *The American journal of clinical nutrition*, *71*, 937–943.

Reynolds, C. J., Horgan, G. W., Whybrow, S., & Macdiarmid, J. I. (2019). Healthy and sustainable diets that meet greenhouse gas emission reduction targets and are affordable for different income groups in the uk. *Public health nutrition*, *22*, 1503–1517.

Rippin, H. L., Cade, J. E., Berrang-Ford, L., Benton, T. G., Hancock, N., & Greenwood, D. C. (2021). Variations in greenhouse gas emissions of individual diets: Associations between the greenhouse gas emissions and nutrient intake in the united kingdom. *Plos one*, *16*, e0259418.

Rocabois, A., Tompa, O., Vieux, F., Maillot, M., & Gazan, R. (2022). Diet optimization for sustainability: Indigoo, an innovative multilevel model combining individual and population objectives. *Sustainability*, *14*, 12667.

Rust, N. A., Ridding, L., Ward, C., Clark, B., Kehoe, L., Dora, M., Whittingham, M. J., McGowan, P., Chaudhary, A., Reynolds, C. J. et al. (2020). How to transition to reduced-meat diets that benefit people and the planet. *Science of the Total Environment*, *718*, 137208.

Sabate, J., & Soret, S. (2014). Sustainability of plant-based diets: back to the future. *The American journal of clinical nutrition*, *100*, 476S–482S.

Sanders, T. A. B. (1999). The nutritional adequacy of plant-based diets. *Proceedings of the Nutrition Society*, *58*, 265–269.

Schönfeldt, H., & Hall, N. (2011). Determining iron bio-availability with a constant heme iron value. *Journal of Food Composition and Analysis*, *24*, 738–740.

Shirai, S. S., Seneviratne, O., Gordon, M. E., Chen, C.-H., & McGuinness, D. L. (2021). Identifying ingredient substitutions using a knowledge graph of food. *Frontiers in Artificial Intelligence*, *3*, 621766.

Shubham, K., Anukiruthika, T., Dutta, S., Kashyap, A., Moses, J. A., & Anandharamakrishnan, C. (2020). Iron deficiency anemia: A comprehensive review on iron absorption, bioavailability and emerging food fortification approaches. *Trends in Food Science & Technology*, *99*, 58–75.

Singh, P., & Prasad, S. (2023). A review on iron, zinc and calcium biological significance and factors affecting their absorption and bioavailability. *Journal of Food Composition and Analysis*, *123*, 105529.

Spronk, I., Kullen, C., Burdon, C., & O'Connor, H. (2014). Relationship between nutrition knowledge and dietary intake. *British journal of nutrition*, *111*, 1713–1726.

Starke, A. D., Asotic, E., Trattner, C., & Van Loo, E. J. (2023). Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices. *ACM Transactions on Recommender Systems*, *1*, 1–31.

Stevens, G. A., Beal, T., Mbuya, M. N., Luo, H., Neufeld, L. M., Addo, O. Y., Adu-Afarwuah, S., Alayón, S., Bhutta, Z., Brown, K. H. et al. (2022). Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: a pooled analysis of individual-level data from population-representative surveys. *The Lancet Global Health*, *10*, e1590–e1599.

Tompa, O., Kiss, A., Maillot, M., Sarkadi Nagy, E., Temesi, Á., & Lakner, Z. (2022). Sustainable diet optimization targeting dietary water footprint reduction: a country-specific study. *Sustainability*, *14*, 2309.

Tran, T. N. T., Felfernig, A., Trattner, C., & Holzinger, A. (2021). Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems*, *57*, 171–201.

Trattner, C., & Elsweiler, D. (2017). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web* (pp. 489–498).

Tubiello, F. N., Rosenzweig, C., Conchedda, G., Karl, K., Gütschow, J., Xueyao, P., Obli-Laryea, G., Wanner, N., Qiu, S. Y., & De Barros, J. (2021). Greenhouse gas emissions from food systems: building the evidence base. *Environmental Research Letters*, *16*, 065007.

UN (2016). Sdg goal 2: End hunger, achieve food security and improved nutrition, and promote sustainable agriculture. United Nations Department of Economic and Social Affairs Statistics Division. `https://unstats.un.org/sdgs/report/2016/goal-02/`.

UNDP (2025). Goal 2: Zero hunger. United Nations Development Programme. `https://www.undp.org/sustainable-development-goals/zero-hunger`.

US EPA (2001-2010). Food Commodity Intake Database (FCID). US Environmental Protection Agency. `http://fcid.foodrisk.org/`.

USDA, & ARS (2020a). USDA Food and Nutrient Database for Dietary Studies 2013-2018. U.S. Department of Agriculture and Agricultural Research Service.

USDA, & ARS (2020b). USDA food and nutrient database for dietary studies 2013-2018. `http://www.ars.usda.gov/nea/bhnrc/fsrg`.

USDA, & ARS (2020c). What we eat in america food categories 2017-2018. http://www.ars.usda.gov/nea/bhnrc/fsrg.

USDA, & ARS (2020d). What We Eat in America Food Categories 2017-2018. U.S. Department of Agriculture (USDA) and Agricultural Research Service (ARS). http://www.ars.usda.gov/nea/bhnrc/fsrg.

USDA, & US HHS (2020). Dietary guidelines for americans, 2020-2025. U.S. Department of Agriculture and U.S. Department of Health and Human Services. https://www.dietaryguidelines.gov/sites/default/files/2021-03/Dietary_Guidelines_for_Americans-2020-2025.pdf.

USDA ARS (2022). Food Availability (Per Capita) Data System. U.S. Department of Agriculture and Agricultural Research Service.

USDA DGA (2025). Scientific report of the 2025 dietary guidelines advisory committee. https://www.dietaryguidelines.gov/2025-advisory-committee-report.

USDA ERS (2023). Purchase to Plate. https://www.ers.usda.gov/data-products/purchase-to-plate.

Verly Jr, E., Machado, I. E., Meireles, A. L., & Nilson, E. A. (2023). Avoidable diet-related deaths and cost-of-illness with culturally optimized modifications in diet: The case of brazil. *PLoS One*, *18*, e0288471.

Vieux, F., Perignon, M., Gazan, R., & Darmon, N. (2018). Dietary changes needed to improve diet sustainability: are they similar across europe? *European journal of clinical nutrition*, *72*, 951–960.

Vieux, F., Soler, L.-G., Touazi, D., & Darmon, N. (2013). High nutritional quality is not associated with low greenhouse gas emissions in self-selected diets of french adults. *The American journal of clinical nutrition*, *97*, 569–583.

Vigerske, S., & Gleixner, A. (2018). Scip: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optimization Methods and Software*, *33*, 563–593.

Vinnari, M., & Vinnari, E. (2014). A framework for sustainability transition: The case of plant-based diets. *Journal of agricultural and environmental ethics*, *27*, 369–396.

West, A. R., & Oates, P. S. (2008). Mechanisms of heme iron absorption: current questions and controversies. *World journal of gastroenterology: WJG*, *14*, 4101.

Wickramasinghe, K., Scarborough, P., Goldacre, M., & Rayner, M. (2013). Defining sustainable diets by comparing greenhouse gas emissions from different food groups: a systematic review. *The Lancet*, *382*, S104.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., & Wood, A. (2019). Food in the anthropocene: the eat–lancet commission on healthy diets from sustainable food systems. *The lancet*, *393*,

447–492.

van Wonderen, D., Gerdessen, J. C., Biesbroek, S., & Melse-Boonstra, A. (2025a). Within-food-group optimization improves nutritional adequacy, sustainability, and acceptability of modeled diets. *Frontiers in Nutrition*, *12*, 1648055.

van Wonderen, D., Gerdessen, J. C., Kirst, P., & Melse-Boonstra, A. (2025b). Diet optimization: modeling iron and zinc absorption by nonlinear programming and piecewise linear approximation using nhanes. *The American Journal of Clinical Nutrition*, *122*, 835–858.

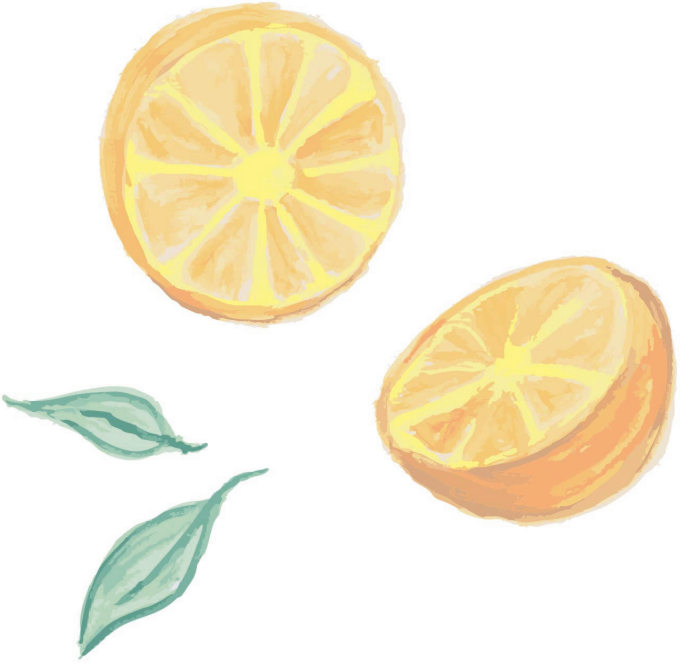
van Wonderen, D., Gerdessen, J. C., Melse-Boonstra, A., & Onwezen, M. C. (2026). Using diet optimization and machine learning for the design of healthy and acceptable menu plans. *European Journal of Operational Research*, *328*, 668–679.

van Wonderen, D., Melse-Boonstra, A., & Gerdessen, J. C. (2023). Iron bioavailability should be considered when modeling omnivorous, vegetarian, and vegan diets. *The Journal of nutrition*, *153*, 2125–2132.

World Obesity Federation (2023). World obesity atlas 2023. `https://data.worldobesity.org/publications/?cat=19`.

Zandstra, E. H., de Graaf, C., & van Trijp, H. C. (2000). Effects of variety and repeated in-home consumption on product acceptance. *Appetite*, *35*, 113–119.

Zimmermann, M. B., & Hurrell, R. F. (2007). Nutritional iron deficiency. *The lancet*, *370*, 511–520.



# Summary

Diet modeling is applied to support the design of diets that are healthy, environmentally sustainable, and acceptable. In this thesis, we evaluated different approaches to modeling iron absorption and food substitution.

Iron deficiency is a major contributor to the global disease burden, particularly affecting children, premenopausal women, and individuals in low- and middle-income countries. People following vegetarian or vegan diets are also at risk, as nonheme iron from plant sources is less readily absorbed than heme iron from animal products. Although the literature provides several equations for estimating nonheme iron absorption from diets, many diet modeling studies still apply a constant absorption factor instead. In Chapter 2 of this thesis, we evaluated the differences in absorbable nonheme iron content when modeling with constant absorption factors versus diet-dependent absorption equations across omnivorous, vegetarian, and vegan diets. For all diet types, constant absorption factors tend to overestimate the absorbable nonheme iron content compared to diet-dependent equations. Unlike constant absorption factors, the equations enable the optimization of food components that either enhance or inhibit nonheme iron absorption, resulting in more realistic estimates of absorbable iron.

These iron absorption equations are nonlinear and they can be integrated into diet modeling using either nonlinear programming or piecewise linear approximation. In Chapter 3, we compared both methods within a mixed-integer and a continuous diet model. For the mixed-integer model, piecewise linear approximation produced accurate results and outperformed nonlinear programming in terms of solution quality and computation time. For the continuous model, the performance of both methods was comparable.

Food substitution involves replacing specific foods with alternatives to create healthier and more environmentally sustainable diets. One of the primary challenges in modeling food substitution is ensuring the acceptability of these changes. In diet modeling, strategies to maintain acceptability include minimizing the differences between observed and optimized diets - based on the assumption that smaller dietary changes are more acceptable - and imposing constraints on substitutions both between- and within food groups. However, many studies have modeled diets only at food group level, overlooking the benefits of substitutions within food groups, where variation in nutrient content and

environmental impact can be substantial. In Chapter 4, we investigated how allowing substitutions within food groups can enhance the nutritional adequacy, environmental sustainability, and acceptability of modeled diets. We found that enabling substitutions within food groups led to notable improvements in nutritional adequacy and sustainability. Consequently, achieving the same nutrition and sustainability goals required only half as much dietary change when both between- and within-food-group substitutions were permitted, compared to using only between-food-group substitutions.

In addition to food-group-based substitutions, machine learning offers an alternative approach for identifying acceptable substitutes. Since it is challenging to quantify individual preferences and translate them into explicit constraints, recent research has focused on learning these preferences directly from data. In Chapter 5, we introduced a novel data-driven method for food substitution by integrating the fields of diet modeling and recipe completion. Recipe completion algorithms aim to identify food items that are commonly consumed together and can complement each other within a meal. In our approach, we used such an algorithm to suggest acceptable food substitutes that pair well with the other food items within a meal and compared its performance to traditional food-group-based substitutions. Traditional substitution using broad food groups, which include many food items, tended to be less acceptable to consumers as evaluated by nutritionists. In contrast, substitution within narrow food groups, containing fewer items, resulted in higher acceptability but offered smaller nutritional improvements. Our recipe completion diet model balanced these trade-offs, achieving greater acceptability than substitutions based on broad food groups and improved nutritional adequacy compared to substitutions based on narrow food groups.

In conclusion, the methods presented for modeling iron absorption and food substitution have proven effective in supporting the design of diets that are healthy, environmentally sustainable, and acceptable.



# Acknowledgements

Writing this PhD thesis has truly been a team effort, and I would like to express my sincere gratitude for that. First and foremost, I would like to thank my daily supervisors, Joke van Lemmen and Alida Melse. Joke, as the diet modeling expert of the Operations Research and Logistics group (ORL), provided invaluable support in guiding me through and reviewing the mathematical modeling aspects of this thesis. Alida, our team's nutrition expert from the Human Nutrition and Health group (HNH), provided essential support in translating technical results into insights that are meaningful for a nutrition-focused audience. I have truly enjoyed working with both of you, not only because of your timely and insightful feedback, but also because of the thoughtful way in which you provided it. You always made me feel encouraged and gave me the confidence to freely discuss the best way forward. I believe you are not only gifted researchers but also truly dedicated and talented teachers and supervisors.

Secondly, I would like to sincerely thank my co-authors Peter Kirst, Sander Biesbroek, and Marleen Onwezen. Peter, a highly technically skilled colleague from ORL, provided invaluable support in helping me understand the theory behind nonlinear programming and piecewise linear approximation. Sander, from HNH, joined our team to contribute expertise on the sustainability aspects of our within-food-group optimization study. And Marleen, our consumer behavior expert from Wageningen Social and Economic Research (WSER) and the Marketing and Consumer Behavior group (MCB), really helped 'market' the added value of our study on integrating diet modeling with machine learning for food substitution. To all three of you, I truly appreciate the time and effort you invested, as well as the thoughtful and constructive feedback you provided.

I would also like to thank the people who made my PhD appointment possible: my promotor Sander de Leeuw from ORL, my former promotor Edith Feskens from HNH, and Bert Vermeer and Quirijn van der Goes, my former managers at WSER, who allowed me to continue working at WSER while being part-time seconded to ORL and HNH. Thank you, Sander, Edith, Bert and Quirijn, for your support, encouragement, and the confidence you placed in me throughout this journey. I also sincerely thank everyone else who helped make this arrangement work, including those involved in securing the necessary funding.

Furthermore, I want to thank all my other colleagues at ORL. I always felt very welcome in the group, even though I was not present as often as most of you due to being a part-time PhD candidate. I truly enjoyed the days I spent working in the Leeuwenborch, not so much because of the building itself but because of your great company. I am also grateful to my colleagues at HNH, who were equally kind and helpful, even though I spent even less time in Helix. My sincere thanks go as well to my WESR colleagues, who frequently asked how my PhD was progressing and showed great flexibility with our projects when I needed to focus on meeting paper revision deadlines. Finally, I would like to acknowledge colleagues throughout the Wageningen University and Research (WUR) community. The conversations I had with several of you, especially at the start of my PhD, offered valuable inspiration and insight for my research.

Last but not least, I am immensely thankful for my family, who support me in everything I do. For them, it makes no difference whether I earn a doctorate or hold no title at all; what matters most is that I enjoy what I do. And that I have; I have truly enjoyed my PhD journey, but nothing compares to the joy I find in spending time with all of you.

# Authorship statement

PhD candidate: Dominique van Wonderen
First promotor: Prof. Dr Sander de Leeuw
Title: Iron Absorption and Food Substitution in Diet Modeling
Date of public defence: 17 September 2026

*Chapter 1 General introduction*
I wrote the first draft and revised it after receiving feedback from Sander de Leeuw, Alida Melse, and Joke van Lemmen.

*Chapter 2 Modeling iron absorption*
The research was designed by me, under the supervision of Alida Melse and Joke van Lemmen. Then I carried out the data preprocessing, conducted the diet modeling, and interpreted the results, with advisory input from my supervisors. I also wrote the first draft and revised it based on their feedback. The manuscript underwent peer review before publication.

*Chapter 3 Nonlinear absorption equations*
I designed the research in collaboration with co-authors Joke van Lemmen, Peter Kirst, and Alida Melse. I was primarily responsible for conducting the diet modeling and interpreting the results, with advisory input from the co-authors. Furthermore, I drafted the manuscript, and the co-authors provided critical feedback. The manuscript was peer-reviewed and revised prior to publication.

*Chapter 4 Within-food-group substitution*
I, together with co-authors Joke van Lemmen, Sander Biesbroek, and Alida Melse, designed the research. I subsequently performed data preprocessing, modeling, and result analysis, with validation provided by the co-authors. The manuscript was drafted by me and revised in response to feedback from the co-authors. The manuscript underwent peer review before publication.

*Chapter 5 Food substitution by machine learning*
The research was designed by me together with co-authors Joke van Lemmen, Alida Melse, and Marleen Onwezen. I performed data preprocessing, modeling, and analysis, with validation from the co-authors. Alida Melse and Karin Borgonjen conducted the qualitative evaluation of the modeling results. I drafted the manuscript, with critical input and feedback from the co-authors. The manuscript was peer-reviewed and revised prior to publication.

*Chapter 6 General discussion*
I prepared the first draft and refined it following input from Sander de Leeuw, Alida Melse, and Joke van Lemmen.

*Generative Artificial Intelligence*
In all chapters except Chapter 2, I have used ChatGPT to provide language support, and to enhance the clarity of the text. I am fully responsible for the final content and interpretation of the research. I have used Open AI. (2024-2025). ChatGPT [Large language model]. https://chat.openai.com/.



Signature PhD candidate



Signature first promotor

# Completed training and supervision plan

Dominique van Wonderen
Wageningen School of Social Sciences (WASS)

| Learning activity | Institute[1] | Year | ECTS |
|---|---|---|---|
| **A1. Managing a research project** | | | |
| WASS PhD introduction | WASS | 2022 | 1.0 |
| Writing research proposal | ORL, WASS | 2022 | 6.0 |
| Scientific writing | Wagingeningen in'to languages | 2022 | 1.8 |
| Scientific publishing | WGS | 2022 | 0.3 |
| Course on presenting | The song works | 2023 | 0.6 |
| Poster presentation[2] | MNF, Den Haag | 2023 | 1.0 |
| Conference presentation[3] | GP-OML, Soesterberg | 2023 | 1.0 |
| **A2. Integrating research in the corresponding discipline** | | | |
| Food Digestion: Nutrient Breakdown & Absorption | HNH, WUR | 2022 | 6.0 |
| Assessment of Dietary Intake | HNH, WUR | 2022 | 3.0 |
| Optimization and Machine Learning | LNMB | 2023 | 4.0 |
| **B1. Placing research in a broader scientific context** | | | |
| Learn Python 3 | Codecademy | 2022 | 0.9 |
| Learn intermediate Python 3 | Codecademy | 2022 | 0.5 |
| Natural Language Processing | Codecademy | 2023 | 1.1 |
| Text generation | Codecademy | 2023 | 0.1 |
| Learn Git & GitHub | Codecademy | 2023 | 0.5 |
| Introduction to LaTeX | PE&RC | 2022 | 0.1 |
| Rmarkdown (Quarto) | WGS | 2022 | 0.6 |
| Bayesian Statistics | PE&RC | 2024 | 1.2 |
| Risk Analysis & Risk Management in Agriculture | WASS | 2024 | 3.0 |
| Big Food Hackaton | RIVM | 2025 | 0.7 |
| **B2. Placing research in a societal context** | | | |
| Guest lecture at high school[4] | Graafschap College, Doetinchem | 2022 | 0.5 |
| **C1. Employing transferable skills in different domains** | | | |
| Start to supervise BSc & MSc thesis students | WUR, Education Support Centre | 2022 | 0.1 |
| Supervising thesis students | ORL, WUR | 2022–2025 | 2.0 |
| Student assistant Decision Science 2 | ORL, WUR | 2022–2025 | 2.0 |
| Scientific Artwork | WUR Library | 2023 | 1.2 |
| **Total credits** | | | **39.2** |

1 Abbreviations: Graduate Program Operations Management & Logistics (GP-OML); Micronutrient Forum (MNF); National Network for Operations Research (LNMB); National Institute for Public Health and the Environment (RIVM)
2 Iron bioavailability should be considered when modelling omnivorous, vegetarian, and vegan diets
3 Food substitution by diet optimization with recipe completion
4 Wiskunde in de praktijk

This thesis was financially supported by the Operations Research and Logistics group, the Human Nutrition and Health group, and Wageningen Social and Economic Research at Wageningen University.

Cover design and illustrations by Dominique van Wonderen

Printed by ProefschriftMaken