Multi-objective decision-making for dietary assessment and advice

J.C. van Lemmen-Gerdessen

Thesis committee

Promotors

Prof. Dr J.G.A.J. van der Vorst
Professor of Operations Research and Logistics
Wageningen University & Research

Prof. Dr P. van ’t Veer
Professor of Nutrition, Public Health and Sustainability
Wageningen University & Research

Co-promotor

Dr G.D.H. Claassen
Associate professor, Operations Research and Logistics Group
Wageningen University & Research

Other members

Prof. Dr E.J.M. Feskens, Wageningen University & Research
Dr M.C. Ocké, National Institute for Public Health and the Environment (RIVM), Bilthoven
Dr E.L. Ferguson, London School of Hygiene and Tropical Medicine, United Kingdom
Prof. Dr C. Romero, Technical University of Madrid, Spain

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

Multi-objective decision-making for dietary assessment and advice

J.C. van Lemmen-Gerdessen

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 15 December 2017
at 4 p.m. in the Aula.

J.C. van Lemmen-Gerdessen
Multi-objective decision-making for dietary assessment and advice,
138 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)
With references, with summary in English

ISBN 978-94-6343-707-3
DOI 10.18174/425370

Contents

Chapter 1

Introduction

1.1 Background

Importance of healthy diets

Poor quality of diet is a major cause of mortality and disability worldwide (Lim et al., 2012, Imamura et al., 2015, WHO, 2013b). In international food programmes, most attention has been paid to food security and micronutrient deficiency, but the diet-related health burdens due to non-communicable chronic diseases are now surpassing those due to undernutrition in nearly every region of the world (Imamura et al., 2015).

Non-communicable chronic diseases (NCDs), also known as chronic diseases, are of long duration and generally slow progression (WHO, 2017). WHO (2014) points out NCDs as the leading cause of death globally: of the world's 56 million deaths in 2012, 38 million (68%) were attributable to NCDs. The number of NCD deaths has increased from 31 million in 2000, to the 38 million in 2012, and is projected to reach 52 million by 2030 (WHO, 2014, Mathers and Loncar, 2006). Worldwide, more than 40% of the NCD deaths were premature deaths under age 70 years. These were distributed unequally: in low- and middle-income countries, NCDs accounted for 82% of the premature deaths, which is regarded to act as key barrier to poverty reduction and sustainable development (WHO, 2014). The cumulative economic losses due to NCDs under a "business as usual" scenario in low- and middle-income countries during 2011-2025 have been estimated at US$ 7 trillion (WHO, 2014). So, NCDs have a negative impact on the quality of life and well-being of the individual and of society as a whole, and put a high burden on health systems and the economy (WHO, 2013b).

Worldwide, in 2012, four main NCDs were responsible for 82% of NCD deaths: cardiovascular diseases (17.5 million), cancers (8.2 million), chronic respiratory diseases (4 million), and diabetes (1.5 million) (WHO, 2014). These four main NCDs share four behavioural risk factors: tobacco use, unhealthy diet, physical inactivity and harmful use of alcohol (WHO, 2013a). With respect to diets, Lim et al. (2012) point out the individual dietary risk factors with the largest attributable burden to NCDs in 2010, see Table 1.1. So, healthy diets can contribute to reducing NCDs.

Due to their potential to reduce NCDs, healthy diets are a global priority, as addressed in e.g. the Global Nutrition Report (International Food Policy Research

Institute, 2016). In order to decrease the burden of NCDs, the WHO (2013a) presented a Global Action Plan, which includes a policy action to "Strengthen the scientific basis for decision-making through non-communicable disease-related research and its translation to enhance the knowledge base for ongoing national action". In the context of healthy diets, two research fields are particularly relevant to study aetiology, monitor the status quo, and develop prevention strategies: dietary assessment and dietary advice.

Table 1.1 Individual dietary risk factors with the largest attributable burden to NCDs in 2010, expressed in millions of deaths and in percentage of global DALYs (Lim et al., 2012).

Dietary risk factor	**Disease burden**[1]	
Diets that are	in 10^6 deaths	in %DALY[2]
low in fruits	4.9 (3.8 – 5.9)	4.2 (3.3 – 5.0)
high in sodium	4.0 (3.4 – 4.6)	2.5 (1.7 – 3.3)
low in nuts and seeds	2.5 (1.6 – 3.2)	2.1 (1.3 – 2.7)
low in whole grains	1.7 (1.3 – 2.1)	1.6 (1.3 – 1.9)
low in vegetables	1.8 (1.2 – 2.3)	1.5 (1.0 – 2.1)
low in seafood omega-3 fatty acids	1.4 (1.0 – 1.8)	1.1 (0.8 – 1.5)

[1]Note that the joint effect of multiple risk factors is not a simple addition of the individual effects and is often smaller than the sum of the individual effects.
[2]Disability adjusted life years, that is: sum of years lived with disability and years of life lost.

Dietary assessment and advice

Dietary assessment contributes to NCD reduction by assessing the food and nutrient intake of target groups and individuals in order to investigate the relation between diet and disease. It helps to point out which foods and nutrients critically contribute to the health status of consumers, and to formulate food and nutrient recommendations. Dietary assessment, for instance, asks respondents to fill in a questionnaire on their use of certain foods products. Nowadays, supermarkets sell thousands of food products, which contribute in varying degrees to the intake of a wide range of nutrients that the dietary expert might want to assess. However, one cannot expect respondents to fill in a comprehensive questionnaire on all the foods in their consumption patterns. The challenge for dietary experts is to compose a questionnaire that is short enough to be acceptable for the respondent, and yet sufficiently long to cover all nutrients that the dietary expert wants to assess.

Dietary advice contributes to NCD reduction by translating food and nutrient recommendations into realistic food choices. Here, dietary experts face another challenge: from the range of thousands of products that contain multiple nutrients, how to compose a dietary pattern that complies with all nutritional constraints, and is acceptable for the consumer?

So, both dietary assessment and dietary advice give rise to complex decision problems: which foods to include in dietary assessment or advice to pursue the multiple objectives of the researcher or fulfil the requirements of the consumer?

Operations Research and Multi-Criteria Decision-Making

Operations Research (OR) is a discipline that deals with the application of advanced analytical methods to help make better decisions (INFORMS, 2017). It employs techniques such as mathematical modelling and mathematical optimisation to arrive at optimal or near-optimal solutions to complex decision problems. It is often concerned with determining a maximum (such as nutritional adequacy of a diet) or minimum (such as the length of a food frequency questionnaire). Widely used techniques in OR are linear programming (LP) and mixed-integer linear programming (MILP). Within OR, multi-criteria decision-making (MCDM) approaches can be used to support the decision-maker in situations where one wants to pursue multiple maxima or minima at the same time (Romero and Rehman, 2003), for instance in situations where multiple nutrients are relevant. Commonly, there is a level of conflict between these maxima and minima.

MCDM problems can be classified in many ways, depending on characteristics of the problem. Based on the number of a problem's alternatives or solutions, a distinction is made between multi-attribute decision-making (MADM) and multi-objective decision-making (MODM). MADM deals with situations in which a (commonly relatively small) set of explicitly listed solutions has to be considered. MADM approaches then aim to rank these alternatives based on their performance with respect to multiple criteria and the decision-maker's preferences, or to point out a preferred alternative. MODM focusses on situations in which the set of solutions is implicitly described via a set of constraints, and often is very large. MODM approaches aim to find a 'best possible' solution, that is, they aim to find a feasible solution that has the best possible performance with respect to the decision-maker's objectives and preferences. For a rational decision-maker, a solution to an MODM problem should be Pareto-optimal, that is, it should have the property that none of its objectives can be improved without deteriorating one of the others. Basic methods for finding Pareto-optimal solutions for MODM problems are the weighting method (in which a weighted sum of the individual objectives is optimised) and the ε-constraint method (which optimises one of the objectives, and converts the others into constraints). Other common methods are compromise programming (which aims to find a solution that is as close to ideal as possible) and goal programming (which aims to minimise deviations from aspirations levels for the objectives) (Miettinen, 2008).

OR and MCDM have been used in many application areas, as is demonstrated in the variety of streams in recent conferences such as EURO2016 and MCDM2013: inventory management, production planning, vehicle routing, agricultural and forestry resources management, sustainable development, efficient use of energy, humanitarian logistics. So far, the number of applications in human nutrition is limited. This thesis demonstrates how MODM can contribute to two fields in human nutrition, viz. dietary assessment and dietary advice, and it shows how these applications contribute to model building and solving for OR.

1.2 Problem statement

People need nutrients but eat foods (Buttriss et al., 2014). The nutrient requirements vary between individuals, depending for instance on age, gender, activity and health status. Nutrients are not evenly distributed in foods. Moreover, there exists an ever-increasing variety of foods, and these foods are used within a broad spectrum of consumption patterns. Dietary assessment aims to assess habitual food intake, and uses food composition to calculate the resulting (observed) nutrient intake. Dietary advice works the other way around: starting from food-based dietary guidelines (FBDG) and nutrient requirements, it uses food composition to come to (advised) food intakes. In short: dietary assessment requires conversion of foods to nutrients, whereas dietary advice converts nutrients to foods, see Figure 1.1.

Figure 1.1 Dietary assessment uses food composition to convert food intake to nutrient intake, dietary advice uses food composition to convert nutrient intake to food intake.

Dietary assessment – food frequency questionnaires

For assessing habitual nutrient intake of subjects in a population, often food frequency questionnaires (FFQs) are used (Willett, 1998). FFQs consist of a limited number of questions on consumption of the foods of interest during a predefined period of time, such as the past month or year (Cade et al., 2002). On the one hand, an FFQ should include enough (questions on) food items to capture sufficient information on all nutrients of interest. On the other hand, an FFQ should be as short as possible, because long FFQs may bore respondents, and make them less motivated to fill out an FFQ accurately (Willett, 1998). The challenge is to develop short FFQs that provide sufficient information for each of the nutrients. Selection of food items to be included in an FFQ is based on their contribution to the nutrient intake of a population and its variance. The common selection procedure, usually based on stepwise regression (Molag et al., 2010, Willett, 1998), is time-consuming. Moreover, it is hard to select items in such a way that all nutrients of interest are sufficiently covered. As a result, the selection of items for FFQs strongly depends on intuition and (different) personal experiences of domain experts. Therefore, there is a need for quantitative decision support for designing short, informative FFQs in a way that is transparent and reproducible. As FFQs commonly aim for multiple nutrients, MODM approaches may contribute to FFQ development.

Dietary advice – Diet modelling

Dietary advice aims to translate nutrient recommendations into realistic food choices. Such translation occurs in several contexts, for instance in composing diets for randomised controlled trials, in individual advice by dieticians, or in developing guidelines for policy makers. In the past, dietary advice strongly depended on intuition and personal experiences of domain experts. Nowadays, diet modelling is considered a useful tool to help identify solutions to complex nutrition problems, such as the access to nutritionally adequate and affordable diets, and the development of dietary recommendations (Buttriss et al., 2014).

Diet modelling is defined as the use of mathematical techniques to formulate and optimise diets (Buttriss et al., 2014). Commonly, linear programming models are used. In such models, constraints ensure that the proposed diet meets requirements on nutrient content and consumer preference. If no diets exist that satisfy all constraints, the aim is to compose diets that violate the constraints as little as possible (Anderson and Earle, 1983, Ferguson et al., 2006, Fletcher et al., 1994). The resulting decision problem has multiple objectives. If diets do exist that satisfy all constraints, then commonly the aim is to compose diets that are as close as possible to the consumer's current diet. The objectives then are to minimise the differences between advised and current intakes of various foods (Darmon et al., 2002, Darmon et al., 2006, Maillot et al., 2010, Masset et al., 2009, Thompson et al., 2013). MODM approaches may contribute to diet modelling, as they can integrate conflicting and incomparable objectives, such as intakes of various nutrients, and show trade-offs between them.

For the quality of decision-making in MCDM problems, assumptions on the so-called preference structure of the decision-maker are of major importance (Tamiz et al., 1998, Romero, 2001, Romero, 2004, Jones, 2011). In diet models, the preference structure can for instance represent the way in which performance indicators for a diet's adequacy with respect to intake of multiple individual foods and nutrients are aggregated into a single performance indicator for the overall quality of the diet. All diet modelling references mentioned above use model formulations that require that the decision-maker has a preference structure in which the trade-offs between objectives are known and constant. This is a severe assumption. There is a need to investigate how other preference structures can be incorporated into diet models and how they affect the resulting solutions.

1.3 Research objective

As argued above, there is a need for MODM approaches to support decision-making for dietary assessment and advice. Therefore, the aim of this research is to investigate MODM approaches for dietary assessment and advice, thus contributing to formulating healthier diets. This is relevant not only to nutrition research as such, but also contributes to model building and solving in OR. This is summarised in Figure 1.2.

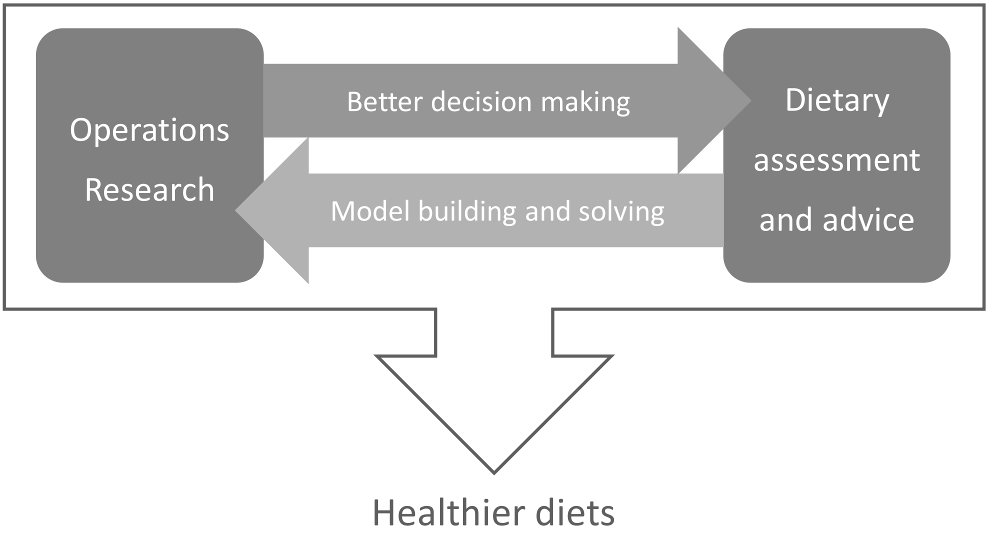

Figure 1.2 The aim of this thesis is to contribute to formulating healthier diets by investigating MODM approaches for dietary assessment and advice. For OR, this contributes to model building and solving.

1.4 Research challenges and research questions

Based on the previous sections, four MODM research challenges in nutrition research are identified. This section elaborates how each of these is translated into a research question for this thesis.

Research challenge 1. Use MODM for FFQ development

There is a need for quantitative decision support for designing short FFQs that provide sufficient information for each of the nutrients of interest, in a transparent and reproducible way. The various nutrients are incomparable entities (for instance, saturated fat and retinol). Therefore, the FFQ optimisation problem does not have one single objective, but one objective for each nutrient, which makes it a multi-objective problem. Literature review did not reveal any use of MODM approaches for development of FFQs. This gives rise to

Research question 1. *How can MODM support selection of items for FFQs?*

Research challenge 2. Solve fractional problems with conditional fractional terms

The basis of every FFQ is a tree structure in which all potential food items are ordered (Molag et al., 2010). Figure 1.3 shows an illustrative and simplified part of the food tree that comprises citrus fruit. An FFQ can use single items such as ORANGE and GRAPEFRUIT (Figure 1.3a), and/or aggregated items such as CITRUS (Figure 1.3b). Each of these items comprises exactly one node in the food tree, so their composition follows directly from the dataset. CITRUS, as well as ORANGE, is therefore called a *direct item*. An FFQ can also contain *indirect items*. In an indirect item, several direct items are grouped into one question that is characterised by the word “other”, for instance OTHER CITRUS in Figures 1.3c and 1.3d. The composition of OTHER CITRUS depends on the preceding direct items: in Figure 1.3c, OTHER CITRUS comprises TANGERINE, GRAPEFRUIT and LEMON, whereas in Figure 1.3d it only comprises GRAPEFRUIT and LEMON. So, the composition of OTHER CITRUS will only be known after optimisation, that is, after it is known which direct items are selected.

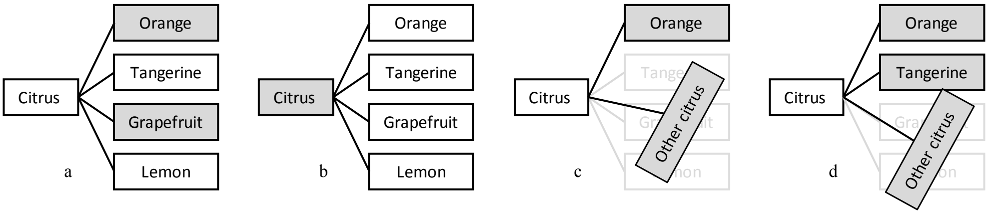

Figure 1.3 Food lists with direct and indirect items. Shaded items are selected for the food list.

Molag et al. (2010) identifies the inability to cope with indirect items as a major limitation of linear programming for FFQ development. Although there is debate, nutritionists assume that the contribution of indirect items to the quality of an FFQ can be approximated by the average contribution of its constituent items. This can be modelled by extending the FFQ model with fractional terms for the indirect items: the contribution of an indirect item is calculated as the summed contributions of its constituent items divided by the number of constituent items. These fractional terms, however, pose a problem: as long as the direct items are not selected, the composition of the indirect items is unknown. Phrased differently: the contribution of an indirect item is calculated as the average contribution of a set of items that is only known after optimisation. In case the set is empty, both the summed contribution and the number of items are zero, which makes the resulting fractional term undefined (that is, “zero divided by zero”). We refer to this problem as a general 0-1 fractional programming problem with conditional fractional terms, because the fractional terms are only defined on the condition that their denominator is non-zero. Of course, all undefined fractional terms should be excluded from the objective function. However, before optimisation it is not known which fractional terms will actually be undefined. Papers about fractional problems commonly start with the remark that only fractions with strictly positive or strictly negative denominators are considered (Stancu-Minasian, 1999, Stancu-Minasian, 2006), so existing literature does not provide a solution method for the problem at hand. This gives rise to

Research question 2. *How to solve general 0-1 fractional programming problems with conditional fractional terms?*

Research challenge 3. Explore preference structures in diet models

Existing diet models commonly use achievement functions that minimise a weighted sum of unwanted deviations from pre-set target levels on food and nutrient intakes. Implicitly, it is assumed that the decision-maker has a preference structure in which trade-offs between objectives are known and constant. In a diet model, this would mean that trade-offs between nutrients are constant and do not depend on intake level, for instance 10% extra deviation from a vitamin C target is considered equally serious as 5% extra deviation from a calcium target, no matter whether the vitamin C intake is almost adequate or dangerously low. A well-known drawback of achievement functions with weighted sums is that they may generate solutions in which some criteria completely meet their targets, whereas others are (very) far off (Romero, 2001). Moreover, they can be sensitive to changes in the used weights. Literature proposes several other achievement functions, representing other preference structures, that can generate solutions in which the unwanted deviations from targets are much more evenly spread, and that are less sensitive to weight changes. This leads to

Research question 3. *What is the impact of achievement functions in diet models?*

Research challenge 4. Find a compromise between total utility and lowest utility

The diet model addressed in research question 3 combines an achievement function in which a (weighted) sum of utilities is maximised and an achievement function in which the lowest value within a set of utilities is maximised. Combining these requires weighting again. The problem is how to justify and interpret weights (Korhonen et al., 2013, Hooker and Williams, 2012). Considering the key importance of preference structure (as implemented via achievement functions and weights) for the quality of decision-making in MCDM, it is crucial to keep an open mind for novel approaches. Hooker and Williams (2012) introduce an approach for balancing between total (sum of utilities) and lowest utility that does not require specification of a set of weights, but uses only one parameter. So far, the approach has not been evaluated and used in a practical context besides that of Hooker and Williams (2012). In order to assess its added value for diet modelling, it is important to gain insights into its properties and (dis)advantages. This leads to

Research question 4.	*What is the added value of a novel method for finding a compromise between total utility and lowest utility in the context of diet models?*

1.5 Outline of the thesis

This thesis includes a collection of five papers. The first two papers focus on research question 1. The other three papers focus on one research question each. Two research questions focus on dietary assessment, and two focus on dietary advice. From OR perspective, each dietary topic provides one application-oriented question and one methodological question (Table 1.2).

Table 1.2 Outline of the thesis.

Research challenge	Research question	Thesis chapter	Nutrition research	Operations Research
1. Use MODM for FFQ development	1	2	Assessment	Application
2. Solve fractional programming problem	2	3	Assessment	Method
3. Explore preference structures in diet models	3	4	Advice	Application
4. Combine total and lowest utility	4	5	Advice	Method

Chapter 2

Chapter 2, focusing on research question 1, describes how Operations Research can support the selection of food items for food frequency questionnaires by modelling the decision problem as a mixed integer linear programming problem. It is based on the following papers:

- Gerdessen, JC, Slegers, PM, Souverein, OW and De Vries, JHM (2012). Use of OR to design food frequency questionnaires in nutritional epidemiology. *Operations Research for Health Care,* 1**:** 30-33.
- Gerdessen, JC, Souverein, OW, Van 't Veer, P and De Vries, JHM (2015). Optimising the selection of food items for FFQs using Mixed Integer Linear Programming. *Public Health Nutrition,* 18**:** 68-74.

Chapter 3

Chapter 3, focusing on research question 2, describes how general 0-1 fractional programming problems with conditional fractional terms can be solved via a reformulation approach. It is based on the following paper:

- Gerdessen, JC, Claassen, GDH and Banasik, A (2013). General 0–1 fractional programming with conditional fractional terms for design of food frequency questionnaires. *Operations Research Letters,* 41**:** 7-11.

Chapter 4

Chapter 4, focusing on research question 3, describes how various achievement functions, representing various preference structures, can be incorporated into diet models. It is based on the following paper:

- Gerdessen, JC and De Vries, JHM (2015). Diet models with linear goal programming: impact of achievement functions. *European Journal of Clinical Nutrition,* 69**:** 1272-1278.

Chapter 5

Chapter 5, focusing on research question 4, studies the behaviour of a novel method for finding a compromise between the lowest and the total utility in the context of the diet model that was presented in Chapter 4. It is based on the following paper:

- Gerdessen, JC, Kanellopoulos, A, Claassen, GDH, 'Combining equity and utilitarianism' – Additional insights in a novel approach, *International Transactions in Operational Research* (DOI: 10.1111/itor.12415)

Chapter 2

Optimising the selection of food items for food frequency questionnaires

Based on:

Gerdessen, JC, Slegers, PM, Souverein, OW and de Vries, JHM (2012)
Use of OR to design food frequency questionnaires in nutritional epidemiology
Operations Research for Health Care, 1**:** 30-33

Gerdessen, JC, Souverein, OW, van 't Veer, P and de Vries, JHM (2015)
Optimising the selection of food items for FFQs using Mixed Integer Linear Programming
Public Health Nutrition, 18**:** 68-74

Abstract

In order to support dietary assessment, this chapter focusses on RQ1 “How can MCDM support selection of food items for FFQs?”. The challenge of selecting food items for FFQs in such a way that the amount of information on all relevant nutrients is maximised while the food list is as short as possible is modelled as a Mixed Integer Linear Programming (MILP) model. The model is demonstrated for an FFQ with interest in energy, total protein, total fat, saturated fat, monounsaturated fat, polyunsaturated fat, total carbohydrates, mono- and disaccharides, dietary fibre, and potassium. The food lists generated by the MILP model have good performance in terms of length, coverage, and R^2 of all nutrients. MILP-generated food lists were 32-40% shorter than a benchmark food list, whereas their quality in terms of R^2 was similar to that of the benchmark. These results suggest that the MILP model makes the selection process faster, more standardised and transparent, and is especially helpful in coping with multiple nutrients. The complexity of the method does not increase with increasing number of nutrients. The generated food lists appear either shorter or provide more information than a food list generated without the MILP model.

2.1 Introduction

FFQs are often used to assess usual long-term dietary intake of subjects in nutritional epidemiological studies, because they are easy to administer with relatively low costs (Thompson and Byers, 1994). However, it is widely acknowledged that they cannot estimate true usual intake of individuals without errors, and that these errors affect the estimated diet-disease relation. Nevertheless, FFQs may provide a more realistic instrument to assess long-term intake because they also capture infrequently consumed foods, whereas short-term instruments like 24-hour recalls have presumably less bias, but require many repeats. Indeed, a combination of FFQ with 24 hour recalls was shown to provide a superior assessment compared to either method alone for some foods and nutrients (Carroll et al., 2012). However, practical and financial constraints still often favour the use of FFQs.

The basis of any FFQ is a food list enumerating all food items on which respondents are questioned. The aim in developing FFQs is to select items for the food list such that as much information as possible is obtained for all nutrients of interest. However, the food list should not be too long in order to minimise the burden for respondents and the research costs. In the selection process decisions have to be taken on the level of aggregation of food items. Highly aggregated food items (such as ‘fresh fruit’) can capture a high coverage (i.e. the fraction of population intake that is covered by the items in the food list) in relatively few items, but they often are not suitable to capture the between-person variance in intake (Willett, 1998). Non-aggregated food items (single foods such as ‘apples’ or ‘oranges’) perform better with respect to capturing the between-person variance in intake. However, it takes many of these items to obtain sufficient coverage on all nutrients of interest.

To develop an FFQ, experts use standardised procedures. However, the process is time-consuming and the selection of food items strongly depends on the personal expertise of the expert. Molag et al. (2010) describe an automated procedure for selecting food items. In their selection procedure one nutrient at a time is taken into account. For that nutrient, all food items are ranked based on their contribution to the variance in the population. Then the highest ranked food items are added to the food list. Selection stops as soon as the selected items suffice to obtain a predefined level of R^2 (explained variance), for example 80%. Then the same procedure is followed for the next nutrient, until all nutrients have been taken into account.

Drawback of this procedure is that it only adds food items to the food list, and never removes items. As most food items contribute to R^2 of several nutrients it can be expected that food items added for one nutrient will also contribute to R^2 of another nutrient. Thus, the final food list will be unnecessarily long, and the selection of food items will depend on the order in which nutrients are taken into account. Therefore, this procedure does not suffice to minimise the number of questions in case the food list is targeted for multiple nutrients. Furthermore, the issue of choosing aggregation levels is not addressed. These drawbacks expose the need for a selection procedure that optimises the food list for multiple nutrients simultaneously, by taking into account the contribution of food items to both the level of intake and variance of multiple nutrients. The procedure should also include the selection of food items at the best aggregation level. Selecting the subset of food items that maximises R^2 resembles the variable selection problem in statistics. However, the FFQ problem distinguishes itself from the common variable selection problem by the large number of available food items, the issue of selecting aggregation levels, and the aim to optimise the food list for multiple nutrients simultaneously. This paper describes how selection of food items for the food list of an FFQ can be modelled and supported by Mixed Integer Linear Programming (MILP) models. It also compares MILP-generated food lists to a food list developed with the procedure described in Molag et al. (2010) in terms of length, coverage and R^2.

2.2 Methods

2.2.1 Data

To select food items to be included in the FFQ, food consumption data of the Dutch National Food Consumption Survey of 1997/1998 of the 3524 individuals in the age group of 25 up to 64 years of age (The Dutch Nutrition Centre, 1998) were used. The food consumption was assessed using a 2d food record, and converted into energy ($n = 1$), total protein ($n = 2$), total fat ($n = 3$), saturated fat ($n = 4$), monounsaturated fat ($n = 5$), polyunsaturated fat ($n = 6$), total carbohydrates ($n = 7$), mono- and disaccharides ($n = 8$), dietary fibre ($n = 9$), and potassium ($n = 10$) with the Dutch food composition database of 1996 (NEVO, 1996).

2.2.2 Aggregation level of food items; food tree

The food items are organised in a food tree with 5 levels, see Figure 2.1. Level 5 contains all items that can be found as “food codes” (single foods) in the NEVO food composition table (NEVO, 1996). Based on similarities in eating occasions, portion sizes and nutrient content these detailed food items are aggregated into more aggregated food items and food groups (Molag, 2010). The food groups at the highest aggregation level are the food groups as specified by NEVO (NEVO, 1996). The food tree contains 1697 items.

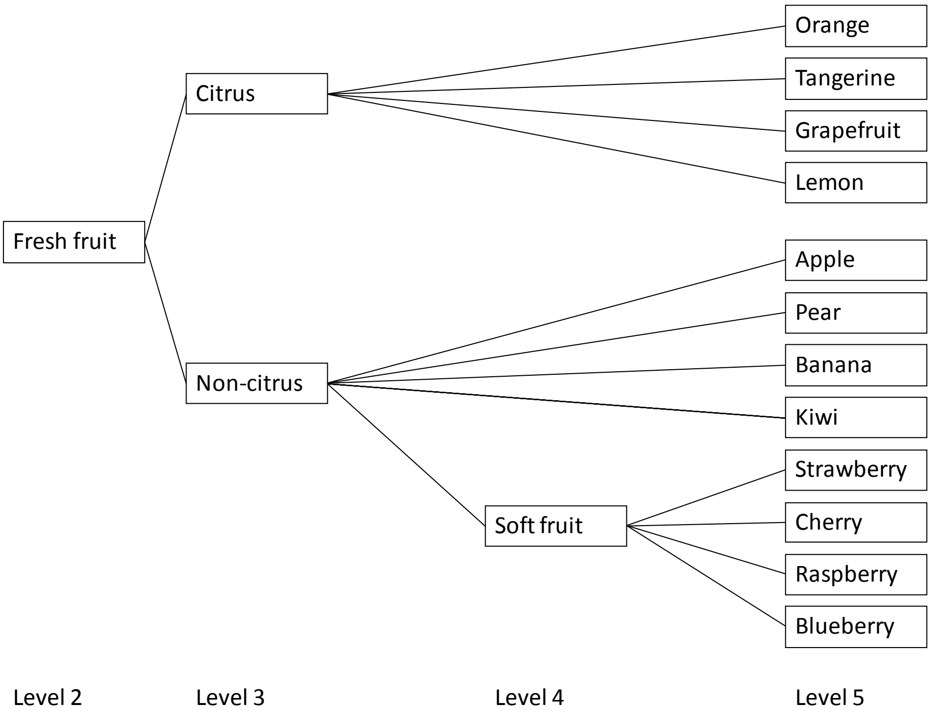

Figure 2.1 Simplified and illustrative part of the tree structure that comprises FRESH FRUIT. Item FRESH FRUIT (level 2) can be further aggregated into item FRUIT (level 1), which also contains items such as canned fruits and dried fruits.

2.2.3 Performance indicators for the quality of a food list

In this paper, we use three quantitative performance indicators to measure the quality of a food list: *length* of the food list, *coverage* of the level of nutrient intake, and explained variance (R^2). Performance indicator *length* counts the number of food items in the food list (i.e. the number of selected "boxes" in the tree of Figure 2.1). *Coverage* describes the fraction of population intake covered by the items in the food list. The R^2 of a nutrient is obtained from linear regression of total nutrient intake on nutrient intakes of all food items in the food list, see Appendix 2A.

FFQs that are targeted for multiple nutrients have a *coverage* and an R^2 for each nutrient n ($n = 1, \dots, N$), to be denoted as $coverage_n$ and R_n^2.

2.2.4 Optimising the performance indicators of a food list

There is a crucial difference between *length* and $coverage_n$ of a food list on the one hand, and R_n^2 on the other hand: the contribution of each individual food item to *length* and $coverage_n$ of a food list can be uniquely quantified, whereas the contribution of an item to R_n^2 depends on the set of other items in the list. This makes it impossible to calculate in a straightforward way which combination of items provides maximal R_n^2 on all nutrients. Therefore, a three-step procedure was employed to select a set of items (i.e. a food list) with high R_n^2:

1. A parameter $p_{j,n}$ is defined, which is a proxy for the contribution of item j to R_n^2 of a food list. P_n is defined as the sum of the $p_{j,n}$ of all selected items. It is therefore a proxy of the R_n^2 of the whole food list.
2. An MILP-model is used to select the optimal set of items with respect to *length*, $coverage_n$, and P_n.
3. Of the resulting set of items (i.e. food list) all R_n^2 are calculated.

The challenge is to find an effective proxy $p_{j,n}$, i.e. a proxy $p_{j,n}$ that demonstrates to be able to generate food lists with high R_n^2. As it is expected that items with high intake and high variance in a population might be good candidates for a food list (Willett, 1998), the following two implementations for proxy $p_{j,n}$ were tested:

- based on intake: $\mathrm{MOM1}_{j,n}$ is the percentage that food item j contributes to the coverage of nutrient n intake of the population (Mark et al., 1996). For items in level 5 we define $p_{j,n}$ as $\mathrm{MOM1}_{j,n}$. For the aggregated items in the other levels we define $p_{j,n}$ as 90% of the sum of the $p_{j,n}$ in their constituent items. We refer to the intake-based proxy as $\mathrm{MOM1}^*_{j,n}$. The 90% was chosen after experiments with several values (and calculating the R_n^2 of the resulting food lists) which showed that in general the value of 90% led to good results.
- based on variance: $\mathrm{MOM2}_{j,n}$ is defined as the percentage that food item j contributes to sum of the variances of nutrient n within a specific level of the food tree (Mark et al., 1996):

$$\text{MOM2}_{j,n} = \frac{\sum_{i=1}^{I}\left(F_{ij} - \overline{F}_{ij}\right)^2}{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(F_{ij} - \overline{F}_{ij}\right)^2}$$

where $j = 1, \ldots, J$ refers to all food items at a specific level in the food tree and F_{ij} individual i's intake of the nutrient from food list item j. We define $p_{j,n}$ as $\text{MOM2}_{j,n}$.

2.2.5 Mixed Integer Linear Programming (MILP) model

The basis for the MILP-model is the tree structure presented in Figure 2.1. For every food item j in the tree we define binary decision variable X_j:

> $X_j = 1$ denotes that we decide to include item j in the food list, and
> $X_j = 0$ denotes that we decide *not* to include item j in the food list.

We can express performance indicators *length*, $coverage_n$, and proxy P_n as linear functions of X_j:

$$length = \sum_{j=1}^{J} X_j$$

$$coverage_n = \sum_{j=1}^{J} \text{MOM1}_{j,n} \cdot X_j \qquad \forall n$$

$$P_n = \sum_{j=1}^{J} p_{j,n} \cdot X_j \qquad \forall n$$

Now we can formulate an MILP-model that optimises one of these performance indicators while keeping the others at user-specified levels and ensuring that only feasible food lists are generated. For instance, if we want to optimise P_n while the *length* of the food list is no more than a predefined level (here: 50 items) and the $coverage_n$ of each nutrient is at least 75% then an appropriate MILP-model is:

(1) maximize $\left\{ \sum_{n=1}^{N} P_n = \sum_{j=1}^{J}\sum_{n=1}^{N} p_{j,n} \cdot X_j \right\}$

subject to restrictions

(2) $length = \sum_{j=1}^{J} X_j \leq 50$

(3) $coverage_n = \sum_{j=1}^{J} \text{MOM1}_{j,n} X_j \geq 0.75 \qquad \forall n$

(4) The food list should not contain overlapping items.

Restriction (4) is added in order to assure the model will only generate feasible food lists. For example, if item ORANGE is selected then CITRUS and FRESH FRUIT cannot be included, and if NON-CITRUS is included then FRESH FRUIT cannot be included, and neither can all items on the right-hand side of NON-CITRUS. This can be modelled by adding one restriction for every item in level 5 of the food tree (see Figure 2.1). For example, for ORANGE and CHERRY we add:

$$X_{\text{FreshFruit}} + X_{\text{Citrus}} + X_{\text{Orange}} \leq 1$$
$$X_{\text{FreshFruit}} + X_{\text{Noncitrus}} + X_{\text{SoftFruit}} + X_{\text{Cherry}} \leq 1$$

Solving model (1)-(4) generates a feasible food list that has maximal value for average(P_n) among all food lists with at most 50 items and $coverage_n \geq 0.75$, provided that restrictions (2) and (3) are not conflicting. For example, if the user specifies that $length \leq 10$ and $coverage_n \geq 0.95$ then no food list is generated, because no such list exists.
An MILP-generated solution for the part of the food tree shown in Figure 2.1 might be $X_{\text{Citrus}} = X_{\text{Apple}} = X_{\text{Cherry}} = 1$, and $X_j = 0$ for all other food items. This should be interpreted as follows: three food items are included in the food list: CITRUS, APPLE, CHERRY. This implies that the model has chosen to aggregate ORANGE, GRAPEFRUIT, TANGERINE, and LEMON into one aggregated item at level 3.
By interchanging the performance indicators in (1), (2), and (3) various models can be obtained. Several examples are provided in Appendix 2B.

2.2.6 Experiments

With the MILP-model food lists have been generated of various lengths: 10, 20, …, 150 items. Two different proxies $p_{j,n}$ were tested: $\text{MOM1}^*_{j,n}$ and $\text{MOM2}_{j,n}$. Quality of the resulting food lists was measured in terms of their performance indicators *length*, $coverage_n$, and R^2_n. Standard MILP-software (Xpress-Mosel 7.0.1) was used to solve the models (i.e. generate the food lists). Runtime of the MILP-model was very small: for all instances a global optimal solution was found in less than 5s.

2.2.7 Comparison with ValNed

We compared *length*, $coverage_n$, and R^2_n of MILP-generated food lists with those of an actual FFQ, the so-called ValNed questionnaire (Molag, 2010). This questionnaire was developed for the same nutrient set, and with use of the same data source as the MILP-generated food lists. For constructing the food list of ValNed, the procedure of Molag et al. (2010) was used. The ValNed food list consisted of 117 items.

2.3 Results

2.3.1 Trade-off between length and R_n^2

For food lists of *length* = 10, 20, …, 150 items all R_n^2 (n = 1, …, N) were calculated. Figure 2.2 shows the trade-off between *length* and R_n^2. Figure 2.2 helps to weigh the amount of added information against the number of questions needed. For example, it took 40 items to obtain a food list in which all nutrients have $R_n^2 \geq 70\%$, 50 items to have all $R_n^2 \geq 80\%$, and 80 items to have all $R_n^2 \geq 85\%$.

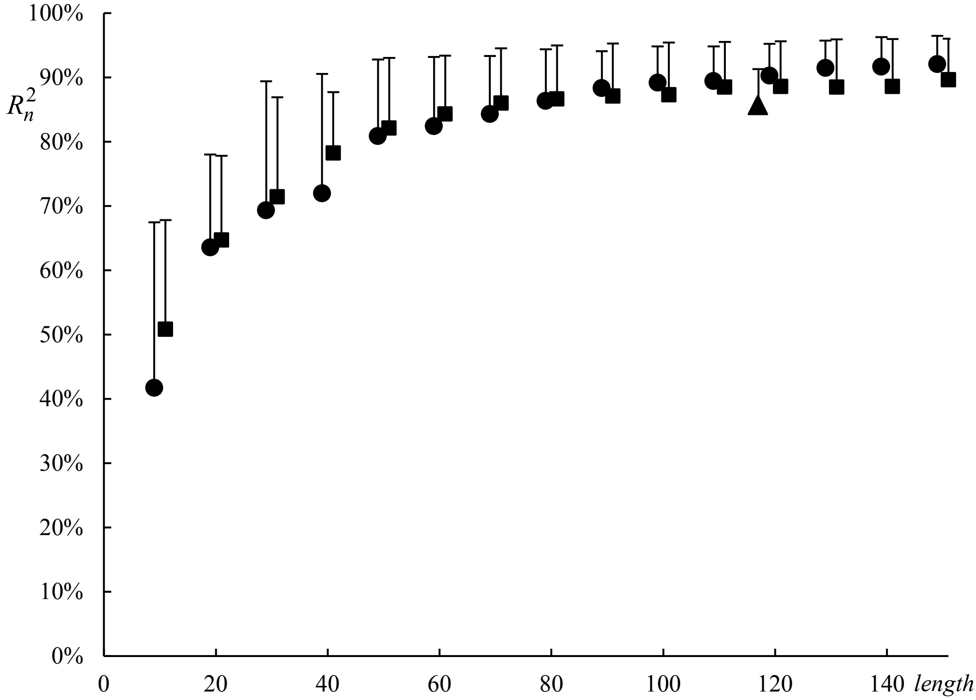

Figure 2.2 Trade-off between number of items in the food list (*length*) and R_n^2 for a coverage-based implementation and a variance-based implementation of proxy $p_{j,n}$ (food item j = 1, ..., J; nutrient n = 1, ..., N). For each food list the R_n^2 for all nutrients were calculated. The range of these R_n^2 is represented with a bar. The lowest among the R_n^2 of a food list is represented with a circle (●) for $p_{j,n} = \text{MOM1}^*_{j,n}$ and a square (■) for $p_{j,n} = \text{MOM2}_{j,n}$. As point of reference also the *length* and R_n^2 of ValNed (▲) are shown.

The choice for proxy $p_{j,n}$ (i.e. $\text{MOM1}^*_{j,n}$ or $\text{MOM2}_{j,n}$) had impact on the R_n^2 of the resulting food list. For food lists of up to 70 items $p_{j,n} = \text{MOM2}_{j,n}$ was the best proxy. For food lists of more than 90 items $p_{j,n} = \text{MOM1}^*_{j,n}$ was the best proxy.
Figure 2.3 shows the impact of *length* on the number of single food items selected for the FFQ. (All other selected items are aggregated items.) Both absolute and

relative number of single food items grows with growing *length*, because a longer food list allows selection of more detailed food items, and thus selection of relatively many single foods. ValNed uses fewer single foods than the MILP-generated food lists.

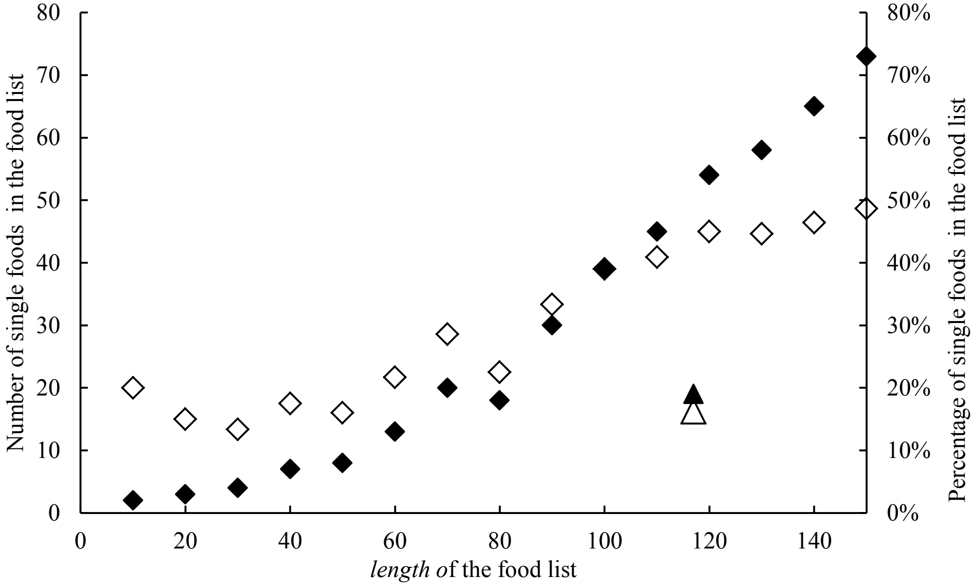

Figure 2.3 The impact of the *length* of the food list on the absolute and relative number of single foods selected for the food list with solid diamond (◆) for number of single foods in MILP food lists, open diamond (◇) for percentage of single foods in MILP food lists, solid triangle (▲) for number of single foods in ValNed food list, and open triangle (△) for percentage of single foods in the ValNed food list. For example, the MILP-generated food list of *length* 20 items contained 3 single foods (15%). The ValNed food list of *length* 117 items contained 19 single foods (16%).

2.3.2 Comparison with ValNed

Length, *coverage*$_n$, and R_n^2 of the MILP food lists were compared to those of the food list of ValNed. Figure 2.2 shows that the MILP-model obtained the same R_n^2 as ValNed in substantially fewer items, or vice versa obtained higher values for R_n^2 with the same number of items. This is further illustrated in Table 2.1, which shows the performance indicators of ValNed and of three food lists generated with the MILP-model (with proxy $p_{j,n}$ = $\mathrm{MOM1}^*_{j,n}$ and iterative improvement procedure as described in Appendix 2B).

Table 2.1 Performance indicators of the food list of ValNed and three MILP-generated lists.

Food list	*length*	$coverage_n$ (range, in %)	R_n^2 (range, in %)
ValNed	117	79.2; 89.1	85.7; 91.3
MILP117	117	96.0; 97.3	90.5; 95.2
MILP80	80	93.9; 97.3	86.4; 94.4
MILP70	70	91.4; 97.2	84.3; 93.4

Food list MILP117 was generated with an MILP-model that maximises P_n while putting an upper bound of 117 on the *length* of the generated food list. MILP117 has substantially higher values for $coverage_n$ and R_n^2 than ValNed, while *length* is the same. For generating MILP80 and MILP70 an upper bound of 80 resp 70 was put on the *length* of the food list. MILP80 has slightly higher values for R_n^2 than ValNed, and MILP70 has slightly lower values for R_n^2 than ValNed. Both lists have substantially higher $coverage_n$. In other words: the MILP-model generated lists that obtained the same R_n^2 as ValNed in substantially fewer (i.e. 32-40% less) items.

2.4 Discussion

This paper presents a methodology for optimising food lists when developing FFQs. The decision problem of selecting food items for food lists was formulated as a Mixed Integer Linear Programming (MILP) model with three performance indicators: *length*, *coverage*, and R^2. The MILP-model generated food lists with good performance in terms of *length*, *coverage*, and R^2 for all nutrients of interest. It supported the selection of the most informative aggregation level for food items, and optimised multiple nutrients at the same time. The generated food lists were either shorter or provided more information than a food list generated without the MILP-model.

The MILP-model chooses the most informative combination of food items from different aggregation levels fast and objectively. Also, food lists of various *length*s may be generated and the increase of *coverage* and R^2 obtained by adding more or other items to the food list may be investigated. The model provides objective information that can help to judge whether the extra information obtained by adding more food items justifies the additional burden for respondents and the additional research cost. With the MILP-model multiple nutrients can be optimised simultaneously. In contrast with a manual selection procedure the number of nutrients has no impact on the complexity of the model. The results of the MILP-based selection procedure are highly reproducible. In addition, the MILP-model can be included in a computer system.

In a typical MILP-supported selection procedure the MILP-model is used to generate an initial food list, which is scrutinised by the nutritionist. The nutritionist indicates which constraints must be added to the model, for example by specifying that some items should or should not be included in the food list to improve face validity. Then the MILP-model is re-run in order to generate a food list that takes into account this expert knowledge. This loop is repeated until the nutritionist is satisfied. The nutritionist then decides how the items are ordered in the actual FFQ. In this iterative procedure, the nutritionist is supported by the MILP-model. It combines the strong points of human insight and experience on the one hand and the efficiency and accuracy of quantitative optimisation techniques on the other hand (Claassen et al., 2007). It is complementary to the current practice of post-hoc validation studies and it can help to guide efficient design of future web-based and personal monitoring tools.

Some considerations have to be taken into account.

It is important to realise that the food tree used here is constructed based on expert knowledge, and that different choices in the structure of the food tree would have led to different food lists.

Figure 2.2 shows the trade-off between *length* of the food list and R_n^2 for lists of 10 to 150 items. The shortest of these lists were included for illustrative purposes; in practice, they would not be used for covering the set of nutrients.

The data used to calculate the MOM1 and MOM2 values that are used as input are based on intake data of a Dutch adult population. Other populations will require other intake data, resulting in different food lists. Major advantage of the MILP-model is that it generates new food lists fast and objectively when the user changes the input data, which facilitates easy adaptation to the characteristics and dietary habits of a population.

A limitation of the current dataset is that only two subsequent food record days were available for each subject. As a result, the between-person variance was artificially high since it contains part of day-to-day variation within persons

(Lambe et al., 2000). Also, the dataset included multiple persons from the same household, lowering between-person variance in food intake and increasing correlations between foods. In general, the data used to generate the food list will have to be able to estimate the intake and variance of the foods and food groups adequately for the target population. Therefore, especially for foods that are not consumed by many persons in the population, the sample size of the used survey will have to be large enough.

Table 2.1 indicates that for the used set of nutrients there is relatively little gain in either coverage or R^2 for food lists longer than 70 items. This might have been different for a different set of nutrients. If more (disperse) nutrients, such as vitamin C and carotenoids, would be added the food list would probably need to be longer. The optimal length of the food list in general depends on the (number of) nutrients, and on the dispersion of the nutrients through the available foods.

Even though these aspects may affect the resulting food lists, they have no effect on the methodology described in this paper in terms of speed, transparency and reproducibility. This makes the MILP-methodology useful for a large number of nutritionists worldwide who develop FFQs for a large variety of studies and different target populations (Bharathi et al., 2008, Cade et al., 2002, Wakai, 2009). It may help them to devise shorter questionnaires of which the performance is as good as that of FFQs with more food items, which generally provide the better results (Molag et al., 2007). Validation with for instance biomarkers will have to answer the question on how well FFQs generated with this methodology perform in different populations.

In conclusion, MILP-models can support development of food lists for FFQs. The results suggest that the MILP-model makes the selection process faster, more standardised and transparent, and is especially helpful in coping with multiple nutrients. The generated food lists appear either shorter or provide more information than a food list generated without the MILP-model.

Appendix 2A Calculation of R^2, C_j', and F_{ij}

The R^2 of a nutrient is obtained from the linear regression of total nutrient intake on the nutrient intakes of all food items in the food list:

$$Z_i = \beta_0 + \beta_1 F_{i1} + \beta_2 F_{i2} + \cdots + \beta_j F_{ij} + \varepsilon_i \, ,$$

where Z_i is the total nutrient intake of individual i (i = 1, …, I) and F_{ij} is individual i's intake of the nutrient from food list item j (j = 1, …, J). F_{ij} is calculated as the multiplication of the amount of the item that is reported with the nutrient content of that item. However, since a food list will contain single foods at level 5 and aggregated items at the other levels, F_{ij} may be either the nutrient intake of individual i from a single food or the nutrient intake from an aggregated item.

For single foods, the F_{ij} is calculated as the multiplication of the reported amount of the food (A_{ij}) and C_j, which is the nutrient content (gram per gram) as reported in the food composition table: $F_{ij} = A_{ij} \times C_j$.

The nutrient content of an aggregated item is not reported in the food composition table, but has to be calculated via the single foods that are taken into account by this aggregated item j. Therefore, for aggregated food items the weighted composition C_j' is used, with weights derived from the observed consumption:

$$C_j' = \sum_{k \in K_j} \left(C_k \times \frac{\sum_{i=1}^{I} A_{ik}}{\sum_{\kappa \in K_j} \sum_{i=1}^{I} A_{i\kappa}} \right)$$

where K_j is the set of single foods that are taken into account by aggregated item j, and $\sum_{i=1}^{I} A_{ik} / \sum_{\kappa \in K_j} \sum_{i=1}^{I} A_{i\kappa}$ is the fraction of the amount of aggregated item j attributable to single food k in the whole population. Then for an aggregated item j the F_{ij} is calculated as

$$F_{ij} = A_{ij} \times C_j'$$

The R^2 is then defined in the usual way:

$$R^2 = \frac{\sum_{i=1}^{I} \left(\hat{Z}_i - \overline{Z} \right)^2}{\sum_{i=1}^{I} \left(Z_i - \overline{Z} \right)^2} \, ,$$

with $\hat{Z}_i$ the predicted nutrient intake of individual i, $\overline{Z}$ the mean observed nutrient intake in the population and Z_i the observed nutrient intake of individual i.

Example of calculation of C_j' and F_{ij} of an aggregated food item

Suppose we have a food tree that contains an aggregated food item NON-CITRUS (referred to as j), constructed by aggregation of four single food items APPLE, PEAR, BANANA, KIWI (referred to as $k = 1,...,4$), see Figure 2A.1.

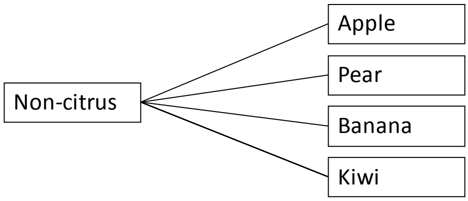

Figure 2A.1 Example of a food tree that contains one aggregated item NON-CITRUS, constructed by aggregation of four single foods APPLE, PEAR, BANANA, KIWI.

Suppose that we have a population of 3 individuals i. Table 2A.1 shows how much individual i has consumed of (single) food item k (A_{ik}), and it shows how much the food items k contain of the nutrient of interest (C_k). For example, individual i=1 consumed 300g of pear, and pear contains 30% of the nutrient of interest.

Table 2A.1 Example of calculation of the nutrient content of aggregated food item NON-CITRUS, where A_{ik} denotes how much individual i has consumed of (single) food item k, and C_k is the nutrient content of food item k.

		A_{ik}	i=1	i=2	i=3	$\sum_{i=1}^{I} A_{ik}$	C_k
NON-CITRUS:	k=1	APPLE	150	250	0	400	0.40
	k=2	PEAR	300	200	100	600	0.30
	k=3	BANANA	350	250	400	1000	0.25
	k=4	KIWI	200	100	200	500	0.20
		Consumption of NON-CITRUS:	1000	800	700	2500	

Total consumption of APPLE in the population can be calculated as

$$\sum_{i=1}^{I} A_{i1} = 150 + 250 + 0 = 400 .$$

Total consumption of NON-CITRUS in the population can be calculated as

$$\sum_{k \in Kj} \sum_{i=1}^{I} A_{ik} = (150 + 250 + 0) + ... + (200 + 100 + 200) = 2500$$

and in this population the fraction of NON-CITRUS consumption attributable to APPLE is

$$\frac{\sum_{i=1}^{I} A_{i1}}{\sum_{k \in K_j} \sum_{i=1}^{I} A_{ik}} = \frac{400}{2500}.$$

So in this population the nutrient content of aggregated food item NON-CITRUS is

$$C_j' = \sum_{k \in K_j} C_k \frac{\sum_{i=1}^{I} A_{ik}}{\sum_{\kappa \in K_j} \sum_{i=1}^{I} A_{i\kappa}} = 0.40 \cdot \tfrac{400}{2500} + 0.30 \cdot \tfrac{600}{2500} + 0.25 \cdot \tfrac{1000}{2500} + 0.20 \cdot \tfrac{500}{2500} = 0.276$$

and the nutrient intake of individual i via NON-CITRUS has to be calculated as

$$F_{ij} = A_{ij} \times C_j'.$$

For example, for individual 1 this is F_{ij} = (150+300+350+200)×0.276 = 276g.

Appendix 2B Mixed Integer Linear Programming models

This appendix shows how various types of optimisations can be performed by interchanging the performance indicators in (1), (2), (3). For example, the model

(2) minimise *length*
subject to restrictions
(1) $P_n \geq 0.80$ for all nutrients n
(3) $coverage_n \geq 0.80$ for all nutrients n
(4) The food list should not contain overlapping items.

generates the shortest food list with $coverage_n \geq 0.80$ and $P_n \geq 0.80$ for all nutrients.

In case the objective is to generate a list of at most 60 items of which the lowest among the N values of P_n is as high as possible the following model applies:

maximise $min\_P_n + \varepsilon \cdot \sum_n P_n$
subject to restrictions
(1a) $P_n \geq min\_P_n$ for all nutrients n
(2) $length \leq 60$ for all nutrients n
(4) The food list should not contain overlapping items

in which min_P_n is the lowest among the N values of P_n and ε a very small, positive number. This is a so-called maxmin model (Claassen et al., 2007).

In case the objective is to generate lists of which the minimum$_n(R_n^2)$ is as high as possible it can be useful to apply an iterative improvement procedure on the generated lists. This improvement procedure aims to improve the minimum$_n(R_n^2)$ by adding lower bound constraints with respect to the P_n of the nutrient that has lowest R_n^2.We will demonstrate this with an example in which food lists of *length* 20 are generated with the following model and $p_{j,n} = \text{MOM}^*_{j,n}$:

$$\text{maximise } (1)\left\{ \sum_{n=1}^{N} P_n = \sum_{j=1}^{J}\sum_{n=1}^{N} p_{j,n} \cdot X_j \right\}$$

subject to restrictions

$$(2) \quad length = \sum_{j=1}^{J} X_j \leq 20$$

(4) The food list should not contain overlapping items.

The R_n^2 of the resulting food list ranges from 60.5%-78.7%, so minimum$_n$(R_n^2) = 60.5%. The nutrient with lowest R_n^2 is polyunsaturated fat. In the food list the $P_{\text{polyunsaturatedfat}}$ is 0.586. Therefore, we add nutrient specific constraint (5a)

$$(5a)\ P_{\text{polyunsaturatedfat}} \geq 1.01 \cdot 0.586 = 0.592$$

to the model, generate a second food list, and calculate its values of R_n^2. It turns out that the R_n^2 of this second food list ranges from 61.0%-76.6%, so minimum$_n$(R_n^2) = 61.0%, which is higher than that of the initial food list. Again, the nutrient with lowest R_n^2 is polyunsaturated fat. In the second food list the $P_{\text{polyunsaturatedfat}}$ is 0.597. Therefore, we add nutrient specific constraint (5b)

$$(5b)\ P_{\text{polyunsaturatedfat}} \geq 1.01 \cdot 0.597 = 0.603$$

to the model, generate a third food list, and calculate its values of R_n^2. It turns out that the R_n^2 of this third food list ranges from 63.6%-78.0%, so minimum$_n$(R_n^2) = 63.6%, which is higher than that of the second food list. Now mono- and disaccharides has lowest R_n^2. In the third food list the $P_{\text{mono-anddisaccharides}} = 0.548$. Therefore, we add nutrient specific constraint (5c)

$$(5c)\ P_{\text{mono-anddisaccharides}} \geq 1.01 \cdot 0.548 = 0.553$$

to the model, generate a fourth food list, and calculate its values of R_n^2. It turns out that the minimum$_n$(R_n^2) of this fourth food list does not exceed 63.6%, and therefore we stop the iterative improvement procedure.

Chapter 3

General 0-1 fractional programming with conditional fractional terms for design of food frequency questionnaires

Based on:

Gerdessen, JC, Claassen, GDH and Banasik, A (2013)
General 0–1 fractional programming with conditional fractional terms
for design of food frequency questionnaires
Operations Research Letters, 41**:** 7-11

Abstract

Chapter 2 models selection of the most informative items to be included in the food list of FFQs as an MILP model. The MILP models choose the most informative combination of food items for different aggregation levels fast and objectively. The resulting food lists can contain single items such as APPLES and ORANGES and aggregated items such as CITRUS or FRUITS. The interpretation of these items follows directly from their names and (in case of aggregated items) the structure of the food tree. Therefore, we refer to them as *direct items.* In FFQs, also *indirect items* like OTHER CITRUS are used. The interpretation of such an indirect item depends on the direct items that are included in the food list. For instance, the interpretation of OTHER CITRUS depends on the citrus fruits that are added to the FFQ as direct items. Note the conceptual difference between CITRUS and OTHER CITRUS, which both are aggregated items: CITRUS is a direct item, because its interpretation follows directly from the food tree and thus is known *before* optimisation. OTHER CITRUS is an indirect item, because its interpretation depends on direct items that are included in the food list, and thus is known *after* optimisation.

Chapter 3 shows that the problem of selecting both direct and indirect food items to be included in an FFQ can be modelled as a general 0-1 fractional programming problem with more than 200 fractional terms. All fractional terms are conditional, that is, in every feasible solution only a subset of the fractional terms is actually defined. Existing literature does not provide a solution method. Chapter 3 therefore focusses on RQ3 “How to solve general 0-1 fractional programming problems with conditional fractional terms?”. We show how classical transformation principles can be combined and extended in order to eliminate the undefined fractional terms from the objective function. The resulting MILP model can be solved with standard software.

3.1 Introduction

Fractional programming problems can be classified as a separate entity within the area of non-linear programming. Several extensive reviews on fractional programming problems have been published, e.g. by Schaible (1995) and Stancu-Minasian (1999, 2006). Although there is a vast amount of literature on Fractional Programming, Schaible and Shi (2004) state that especially integer fractional programming is a somewhat neglected field. The bibliographies of Stancu-Minasian (1999, 2006) confirm the latter statement as only 5, respectively 7 percent of all manuscripts are devoted to integer Fractional Programming. A common assumption in fractional programming is that fractional terms are always defined (i.e. denominators can never become zero).

In development of tools for dietary assessment, a fractional programming problem with the following characteristics was encountered:

i. all decision variables are binary,
ii. there are more than 200 fractional terms in the objective function,
iii. all the fractional terms are conditional, i.e. in every feasible solution only a subset of the fractional terms is actually defined.

To the best of our knowledge, no method has been described that solves this problem. We exploit the special structure of the problem with conditional fractions to reformulate the 0-1 fractional programming problem such that it can be solved by standard MILP software to obtain an optimal solution.

In Section 3.2 we give a brief description of the background in nutritional epidemiology. Section 3.3 describes the resulting general 0-1 fractional programming problem. In Section 3.4 we present the reformulation approach that transforms the general 0-1 fractional programming problem into an equivalent MILP problem. With our reformulated model, we analysed a case with 219 conditional fractional terms (Section 3.5). Concluding remarks follow in Section 3.6.

3.2 Background

Epidemiological studies investigate the relationship between diet and disease. Their relevance can hardly be overestimated, e.g. in designing effective and efficient intervention studies related to obesity, diabetes or cancer (Willett, 1998). For example, the hypotheses that consumption of red and processed meat increases colorectal cancer risk while intake of fish decreases risk is strongly supported by the results of the European Prospective Investigation into Cancer and Nutrition study (Gonzalez and Riboli, 2010). This large epidemiological study investigates the relationships between diet and the incidence of cancer and other chronic diseases in European countries. For dietary assessment, large epidemiological studies commonly use food frequency questionnaires (FFQs). FFQs assess a population's habitual (nutrient) intake (Willett, 1998) by asking respondents about their consumption of several food items during a predefined time period. On the one hand an FFQ should include enough questions on food items to capture sufficient information on all nutrients of interest. On the other hand an FFQ should be as short as possible, because long FFQs are less cost and time efficient, may bore respondents and make them less motivated to fill out an FFQ accurately (Willett, 1998).

Selection of questions on food items to be included in an FFQ – from this point onwards simply referred to as "items" – is based on their contribution to the nutrient intake of a population. Aim of the selection procedure is to compose a set of items that captures as much information as possible. An upper bound is specified for the size of this set. Although the selection procedure is done by skilled nutritionists, it is neither standardised nor transparent. The common selection procedure, usually based on stepwise regression (Molag et al., 2010, Willett, 1998), is time-consuming. Moreover, it is hard to select items in such a way that all nutrients of interest are sufficiently covered. As a result, the selection of items strongly depends on intuition and (different) personal experiences of nutritionists. The aim of this study is to support, standardise and improve the selection procedure to obtain as much information as possible within a limited number of items.

The basis of every FFQ is a tree structure in which all potential items are ordered (Molag et al., 2010). Figure 3.1 shows an illustrative and simplified part of the tree structure that comprises fresh fruit. Level 5 contains the items that can be found as "food codes" in a food composition table. Based on similarities in eating occasions, portion sizes and nutrient content these detailed food items are aggregated into increasingly broad items in levels 4, 3, and 2. Items of level 2 can be further aggregated into items at level 1, but these are not suitable for being used as items in an FFQ (Molag et al., 2010). The process of aggregation causes a loss of information, e.g. an FFQ that asks for FRESH FRUIT will provide less detailed information than an FFQ that asks for CITRUS and NON-CITRUS.

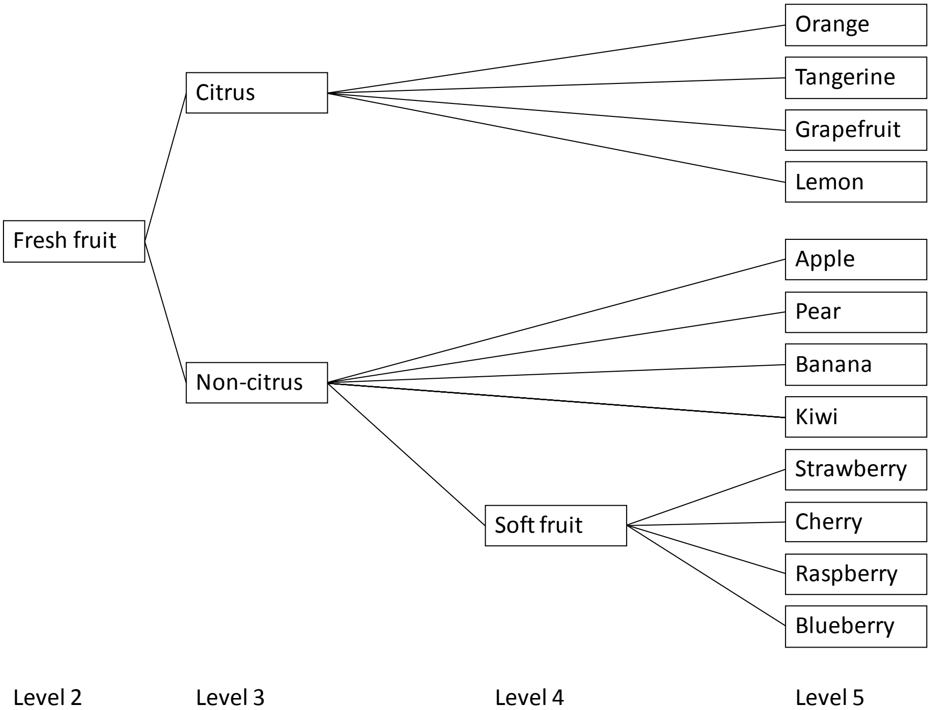

Figure 3.1 Simplified and illustrative part of the tree structure that comprises FRESH FRUIT.

In the tree structure, several paths can be seen, all starting at level 2, e.g. the path FRESH FRUIT – CITRUS – ORANGE and the path FRESH FRUIT – NON-CITRUS – SOFT FRUIT – CHERRY. The right-most item in a path is called a leaf. The leaves of the aforementioned paths are ORANGE respectively CHERRY. Items on the same path as leaf i are called predecessors of leaf i. The predecessors of CHERRY are SOFT FRUIT, NON-CITRUS, FRESH FRUIT. ORANGE has two predecessors: CITRUS and FRESH FRUIT. The item at the left-hand side of item i is called the parent of item i. All nodes that have item j as parent are called children of item j. E.g. SOFT FRUIT has parent NON-CITRUS and children {STRAWBERRY, CHERRY, RASPBERRY, BLUEBERRY}. To prevent overlap of information no more than one item of every path can be included in the FFQ.

Items that comprise exactly one node in the tree of Figure 3.1 are called *direct* items, e.g. FRESH FRUIT, CITRUS, CHERRY. It's also possible to use *indirect* items. In an indirect item several items are grouped into one question, e.g. item 2 in FFQ

"Do you eat 1. ORANGES, 2. OTHER CITRUS?" An indirect item is characterised by use of the word "other". Nutritionists assume that the information content of an indirect item is proportional to the average information content of its constituent items. However, there is no consensus on the latter assumption. The problem of weighing the information content of direct versus indirect items is still a topic of research in nutritional epidemiology (Molag et al., 2010).

For modelling purposes, we assume that respondents understand which items are implied in an indirect item in the FFQ. For example, respondents will interpret item 2 in FFQ "Do you eat 1. ORANGES, 2. OTHER CITRUS?" as "Do you eat LEMON, TANGERINE, GRAPEFRUIT?"

3.3 FFQ model

For modelling the FFQ problem we define the following index sets, decision variables and parameters:

Index sets

N	set of all nutrients
I	set of all items, i.e. all potential questions in the FFQ
L	set of all leaves
P_i	set of all predecessors of item i ($i \in L$)
p_i	parent of item i ($i \in I \mid i$ in level 3, 4, 5)
K_i	set of all children of item i ($i \in I \mid i$ in level 2, 3, 4)
F	set of all items that have at least 3 children

Decision variables

$xd_i = 1$ if item i is included in the FFQ as a direct item, and
$xd_i = 0$ otherwise ($i \in I$)

$y_f = 1$ if item f is included in the FFQ as an indirect item, and
$y_f = 0$ otherwise ($f \in F$)

$xg_i = 1$ if item i is grouped, and
$xg_i = 0$ otherwise $(i \in I \wedge p_i \in F)$

Parameters

$q_{i,n}$	information content of item i regarding nutrient n ($i \in I$, $n \in N$)
ub	upper bound on the number of items in the FFQ

Using an indirect item like e.g. "OTHER CITRUS" only makes sense if (at least) one direct item among the children of CITRUS is used, e.g. ORANGE. This implies that indirect items can only be considered for items that have at least three children, i.e. for the items in F. In the tree of Figure 3.1 the set F consists of three items: $F :=$ {CITRUS, NON-CITRUS, SOFT FRUIT}. For every child of the items in F the variable xg_i is defined, so for SOFT FRUIT three variables are defined: SOFT FRUIT can be used as direct item ($xd_{\text{softfruit}} = 1$), as indirect item OTHER SOFT FRUIT ($y_{\text{softfruit}} = 1$), or it can be grouped ($xg_{\text{softfruit}} = 1$).

An example of an FFQ of length 4 with only direct items may be: CITRUS, APPLE, PEAR, CHERRY. For this FFQ we have $xd_{\text{citrus}} = xd_{\text{apple}} = xd_{\text{pear}} = xd_{\text{cherry}} = 1$, all other $xd_i = 0$, all $y_i = 0$, and all $xg_i = 0$. The total information content of this list with respect to nutrient n is calculated as $q_{\text{citrus},n} + q_{\text{apple},n} + q_{\text{pear},n} + q_{\text{cherry},n}$.

Another example of a FFQ of length 4, now including 3 direct items and 1 indirect item, may be: ORANGE, OTHER CITRUS, APPLE, CHERRY. For this FFQ we have $xd_{\text{orange}} = xd_{\text{apple}} = xd_{\text{cherry}} = 1$, $y_{\text{citrus}} = 1$, $xg_{\text{tangerine}} = xg_{\text{grapefruit}} = xg_{\text{lemon}} = 1$, all other xd_i and y_i and xg_i are 0. The information content of this list of items for nutrient n is calculated as the sum of all information contributed by direct items and the average information content of indirect items:

$$q_{\text{orange},n} + \frac{q_{\text{tangerine},n} + q_{\text{grapefruit},n} + q_{\text{lemon},n}}{3} + q_{\text{apple},n} + q_{\text{cherry},n}$$

The problem of weighing the information contributed by direct items on the one hand and indirect items on the other hand, can be supported by the introduction of a weighing coefficient w. Now, the total weighted information on all nutrients $n \in N$, contributed by direct and indirect items, can be formulated as:

$$\sum_{n \in N} \left\{ \sum_{i \in I} q_{i,n} xd_i + w \left(\sum_{(f \in F | y_f = 1)} \frac{\sum_{i \in K_f} q_{i,n} xg_i}{\sum_{i \in K_f} xg_i} \right) \right\}$$

The first term sums the information contributed by all direct items. The second term adds up all information derived from indirect items. Each of the fractional

terms is conditional, i.e. fractional term f is only defined if indirect item f is chosen ($y_f = 1$). In case indirect item f is not chosen ($y_f = 0$), fraction f must be excluded from the objective function as both the numerator and the denominator of fraction f will be zero. In such cases the resulting problem is obviously not defined.
Now the FFQ model can be formulated as follows:

$$\text{maximize} \sum_{n \in N} \left\{ \sum_{i \in I} q_{i,n}\, xd_i + w \left(\sum_{(f \in F | y_f = 1)} \frac{\sum_{i \in K_f} q_{i,n} xg_i}{\sum_{i \in K_f} xg_i} \right) \right\} \tag{1}$$

subject to

$$\sum_{i \in I} xd_i + \sum_{f \in F} y_f \le ub \tag{2}$$

$$xd_i + xg_i + \sum_{j \in P_i} (xd_j + xg_j) \le 1 \qquad i \in L \tag{3}$$

$$xg_i \le y_f \qquad f \in F, i \in K_f \tag{4}$$

$$y_f \le xd_i + xg_i \qquad f \in F, i \in K_f \tag{5}$$

$$\sum_{i \in K_f} xg_i \le \left(\left| K_f \right| - 1 \right) \sum_{i \in K_f} xd_i \qquad f \in F, i \in K_f \tag{6}$$

$$xd_i\ (i \in I),\ xg_i\ (i \in I \wedge p_i \in F),\ y_f\ (f \in F) \text{ all binary} \tag{7}$$

Objective function (1) maximises the total amount of information derived from both direct (first term) and indirect items (second term). Constraint (2) ensures that the total number of items in the FFQ does not exceed the upper bound on the number of questions.

Constraints (3) ensure that for every path in the tree at most one item is included in the FFQ. Note that the actual number of terms in constraints (3) depends on the level of the leaf i.

Constraints (4) are classical feasibility constraints for connecting two types of variables (i.e. xg_i and y_f). They state that if an item i is grouped ($xg_i = 1$) then the related indirect item must exist ($y_f = 1$).

Constraints (5) denote that if indirect item f is chosen ($y_f = 1$) then each of its children should either be asked directly ($xd_i = 1$) or be grouped ($xg_i = 1$).
Constraints (6) ensure that items $i \in K_f$ can be grouped if and only if one of the children of f is used as a direct item.

The food tree that is used for designing FFQs contains 219 items for which an indirect question can be formulated, so the objective function of the FFQ problem contains 219 conditional fractional terms.

In the next section we elaborate a reformulation approach that transforms problem (1)-(7) into an MILP problem that can be solved with a branch-and-bound procedure, commonly available in standard mathematical programming software. As the reformulation approach is not affected by weighing coefficient w, we will omit it (i.e. choose $w = 1$) in Section 3.4.

3.4 Reformulation of the FFQ problem

Model (1)-(7) can be regarded as an extended version of the general 0-1 fractional programming problem (G-FP), described by Li (1994) and Wu (1997). The crucial difference between model (1)-(7) and the (G-FP) of Li (1994) and Wu (1997) refers to the conditional summation of fractional terms in (1).

If a G-FP problem contains only one fractional term in the objective function, the problem becomes a linear 0-1 fractional programming problem. For these problems, if it exists, at least one optimal solution is a vertex of the convex hull for the set of discrete solutions (Barros, 1998). The latter also holds for mixed integer linear programming (MILP) problems. Now the basic idea is to build on and extend known reformulation approaches in order to reformulate model (1)-(7) into an MILP model such that an optimal solution can be found using standard available MILP software.

The objective function (1) with conditional summation of the fractional terms can be reformulated into:

$$\text{maximize} \sum_{n \in N} \left(\sum_{i \in I} q_{i,n}\, xd_i + \sum_{(f \in F | y_f = 1)} y_f \cdot \frac{\sum_{i \in K_f} q_{i,n}\, xg_i}{\sum_{i \in K_f} xg_i} \right) \tag{1b}$$

In order to eliminate the fractional terms, parts of a classical transformation principle, initially suggested by Charnes and Cooper (1962), can be extended and applied to (1b). Suppose we define:

$$t_f = \frac{y_f}{\sum_{i \in K_f} xg_i} \qquad f \in F$$

Then objective function (1b) can be transformed into:

$$\text{maximize} \sum_{n \in N} \left(\sum_{i \in I} q_{i,n}\, xd_i + \sum_{(f \in F | y_f = 1)} \sum_{i \in K_f} q_{i,n} \cdot t_f \cdot xg_i \right) \tag{1c}$$

In which

$$\sum_{i \in K_f} t_f \cdot xg_i = y_f \qquad f \in F \tag{8}$$

Because of (4), all fractions with $y_f = 0$ have $\sum_{i \in K_f} t_f \cdot xg_i = 0$. Therefore, (1c) can be written as

$$\text{maximize} \sum_{n \in N} \left(\sum_{i \in I} q_{i,n}\, xd_i + \sum_{f \in F} \sum_{i \in K_f} q_{i,n} \cdot t_f \cdot xg_i \right) \tag{1d}$$

To eliminate the product terms of variables $t_f xg_i$ in (1d) and (8), a classical transformation principle can be applied (Glover, 1975) which is comparably used by Li (1994) and Wu (1997). According to Glover (1975) this transformation principle was initially suggested by Petersen and Clifford (1971). The transformation principle is suitable for linearising the product of a nonnegative continuous variable and a binary variable. It replaces the product terms $t_f xg_i$ by new continuous variables $z_{i,f} \geq 0$ ($f \in F$, $i \in K_f$) which are forced to satisfy the linear constraints (9)-(11) below.

Now, model (1)-(7) can be transformed to the following mixed integer linear programming model:

$$\text{maximize} \sum_{n \in N} \left(\sum_{i \in I} q_{i,n}\, xd_i + \sum_{f \in F} \sum_{i \in K_f} q_{i,n}\, z_{i,f} \right) \tag{1e}$$

subject to

$$\sum_{i \in I} xd_i + \sum_{f \in F} y_f \leq ub \tag{2}$$

$$xd_i + xg_i + \sum_{j \in P_i} (xd_j + xg_j) \leq 1 \qquad i \in L \tag{3}$$

$$xg_i \leq y_f \qquad f \in F, i \in K_f \tag{4}$$

$$y_f \leq xd_i + xg_i \qquad f \in F, i \in K_f \tag{5}$$

$$\sum_{i \in K_f} xg_i \leq \left(\left| K_f \right| - 1 \right) \sum_{i \in K_f} xd_i \qquad f \in F, i \in K_f \tag{6}$$

$$\sum_{i \in K_f} z_{i,f} = y_f \qquad f \in F \tag{8}$$

$$z_{i,f} \leq \tfrac{1}{2} xg_i \qquad f \in F, i \in K_f \tag{9}$$

$$z_{i,f} \geq t_f + \tfrac{1}{2}(xg_i - 1) \qquad f \in F, i \in K_f \tag{10}$$

$$z_{i,f} \leq t_f \qquad f \in F, i \in K_f \tag{11}$$

$$xd_i\ (i \in I),\ xg_i\ (i \in I \wedge p_i \in F),\ y_f\ (f \in F) \text{ all binary} \tag{12}$$

$$t_f\ (f \in F),\ z_{i,f}\ (f \in F, i \in K_f) \geq 0 \tag{13}$$

The factor '$^1/_2$' in (9) and (10) is the ceiling value of t_f, which is derived as follows: any indirect item contains two or more grouped items, so for any item with $y_f = 1$ the corresponding $\sum_{i \in K_f} xg_i \geq 2$. Therefore, $t_f = y_f / \sum_{i \in K_f} xg_i \leq 1/2$.

Model (1e) – (13) can be solved by standard MILP software to obtain an optimal solution.

3.5 Numerical analysis

3.5.1 Data

As basis for the experiments a food tree is used that contains 1697 items, of which 1340 are leaves. The tree contains 219 items with more than three children (set F). A dataset was obtained from food consumption data of the Dutch National Food

Consumption Survey of 1997/1998 (The Dutch Nutrition Centre, 1998) of the 3524 individuals in the age group of 25 up to 64 years of age. The food consumption was assessed using a 2-day food record. With the food composition database of 1996 (NEVO, 1996) the dataset was converted into the parameter values $q_{i,n}$ as relative variance in intake of nutrient *n* through item *i*. This type of data is used in studies that aim to rank respondents according to their intake of nutrients (Molag et al., 2010).

3.5.2 Experiments

The reformulated FFQ model (1e) – (13) is used to generate food lists for an FFQ with interest in 10 nutrients: energy, protein, total fat, saturated fat, monounsaturated fat, polyunsaturated fat, total carbohydrates, mono and disaccharides, dietary fibre, and potassium.

After reintroduction of the non-negative weight factor *w* in (1) for all indirect items, objective function (1e) can be replaced by:

$$\text{maximize} \sum_{n \in N} \left(\sum_{i \in I} q_{i,n}\, xd_i + w \left(\sum_{f \in F} \sum_{i \in K_f} q_{i,n}\ z_{i,f} \right) \right) \tag{1f}$$

Now, the impact of indirect items in an FFQ can be analysed numerically.

In Figure 3.2 the number of indirect items in the optimal solution is plotted against the weight *w* of the indirect items. It shows how the number of selected indirect items increases for increasing *w*.

Figure 3.3 shows for $w = 1$ how the normalised objective function increases with an increasing upper bound $ub = 1, 2, \ldots, 200$ to the number of items in the food list. The 100% value of the normalised objective function is calculated by setting *ub* to the maximum number of items (1697). E.g. it takes 11 items to capture 50% of the maximum amount of information.

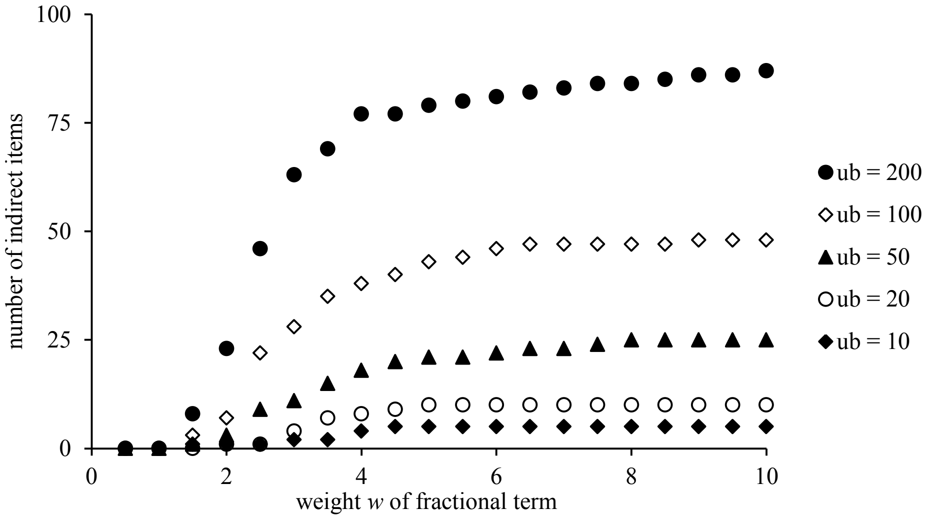

Figure 3.2 Number of indirect items in the optimal solution as function of the weight w of the fractional term in the objective function.

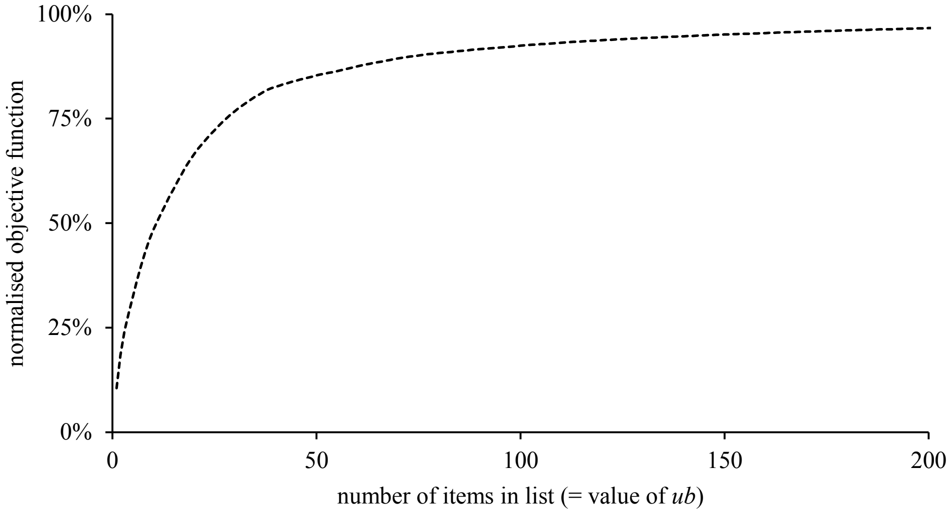

Figure 3.3 Normalised objective function value for increasing upper bound ub for the number of items in the food list.

Standard optimisation software was used (Xpress-Mosel 7.0.1). The calculation times for the 300 runs that were used to generate Figure 3.2 and Figure 3.3 ranged from 0.5 to 8 seconds (average 2 seconds).

3.6 Discussion and concluding remarks

The developed model helps to optimise the selection of items in any FFQ for dietary assessment in epidemiological studies. It provides an objective and fast starting point for developing an FFQ, but it can never replace nutritional expertise, because no MILP model can capture the many intangible factors that are critical in the cognitive interface between the respondent and the FFQ. For example, our assumption that respondents understand which items are implied by an indirect item might not always be valid. For example, it might not be apparent to everyone that “Soft fruit” means “Strawberry, Cherry, Raspberry, Blueberry”, thus making the answer to the question “Do you eat other soft fruit” hard to interpret. And when too many items are grouped into one indirect item the respondent might not think of all these items, unless a full enumeration is provided. In some cases, the nutritionist might decide to insert a question that clarifies another question. For example, asking a question about lettuce will help clarify to the respondent that a question about “greens” does not include lettuce, which would otherwise be interpreted variably. Thus, detailed knowledge of food use and culturally-related culinary practices will be essential in designing an FFQ. This knowledge is also crucial in designing the food tree. The way in which items are ordered in level 5, and the way in which they are aggregated into levels 4, 3, and 2, should reflect the food use and culinary practice of the respondents. For example, in the Netherlands potatoes are considered a starchy food, whereas in France they are considered a vegetable. In a typical MILP-supported selection procedure the MILP model generates a food list that is used as a starting point for developing an FFQ. Based on experience the nutritionist will indicate where problems might arise with respect to interpretation of the questions by respondents, and give suggestions for adding or removing items. These suggestions are formulated as extra constraints, and the model is re-run to generate an FFQ that takes into account the recently added expert knowledge. This iterative procedure, in which the expert is supported by the MILP model, is repeated until the nutritionist is satisfied. Such an iterative procedure combines the strong points of human insight and experience on the one hand and the efficiency and accuracy of quantitative optimisation techniques on the other hand (Claassen et al., 2007). It helps nutritionists to choose the most informative combination of direct and indirect items. Also, the FFQ model gives nutritionists a fast way to generate food lists of various lengths and to see how much extra information can be obtained by adding more or other items to the questionnaire. This helps nutritionists to weigh the amount of obtained information against the burden for respondents.

Chapter 4

Diet models with linear Goal Programming: impact of achievement functions

Based on:

Gerdessen, JC and De Vries, JHM (2015)
Diet models with linear goal programming: impact of achievement functions
European Journal of Clinical Nutrition, 69**:** 1272-1278

Abstract

Supporting decision-making related to dietary advice requires translating food and nutrient recommendations into realistic food choices. For this, linear programming-based diet models are a robust and flexible tool. Existing diet models commonly use achievement functions that minimise a weighted sum of unwanted deviations from pre-set target levels on food and nutrient intakes. Implicitly, it is assumed that the decision-maker has a preference structure in which trade-offs between objectives are known and constant. A well-known drawback of achievement functions with weighted sums is that they may generate solutions in which some criteria completely meet their targets, whereas other criteria are (very) far off. Moreover, they can be sensitive to changes in the used weights. MCDM literature, however, also proposes achievement functions that can generate solutions in which the differences among the deviations are much smaller, and that are less sensitive to weight changes.

In order to provide methodological insight into achievement functions, we formulated RQ3 "What is the impact of achievement functions in diet models?", which is investigated in Chapter 4. The chapter provides small numerical examples that illustrate the "mechanics" of achievement functions, and then uses them for a diet problem with 144 foods, 19 nutrients, and several types of palatability constraints, in which the nutritional constraints are modelled via nutrient adequacy curves. In this problem, no solution exists that satisfies all nutritional constraints. In order to find 'best possible' solutions, achievement functions are used that aim to minimise the (unwanted) deviations from the nutritional targets. It is demonstrated that using an achievement function that minimises a weighted sum of deviations generates a solution in which the total deviation is as small as possible. However, the solution is unbalanced: most nutrients are at their target level, but one is very far off. Using an achievement function that minimises the largest among the unwanted deviations generates a much more balanced solution: the largest deviation is as small as possible, so the differences between the individual deviations are much smaller. However, the total deviation is much larger than in the solution of the weighted sum achievement function. The extended goal programming achievement function minimises a convex combination of both previously mentioned achievement functions and therefore generates solutions of which the sum of deviations and the largest deviation are between the extremes provided by the other two achievement functions.

4.1 Introduction

Mathematical modelling of diets can be defined as the use of mathematical techniques to formulate and optimise diets (Buttriss et al., 2014). We refer to Buttriss et al. (2014) for a description of relevance, history and applications of diet (planning) models based on linear programming (LP).

LP-based diet models contain decision variables, an objective function, and a set of constraints. Commonly, the decision variables are defined as X_i = (proposed) daily intake of food i (i = 1, ..., I), $X_i \geq 0$. The objective function minimises (or maximises) a linear function of the decision variables, for example total cost or total energy content of the diet. The (linear) constraints ensure that the proposed diet meets requirements on e.g. nutrient content and consumer preference[1]. If the model contains integer variables it is called a Mixed Integer Linear Programming (MILP) model.

Classification of diet models

We distinguish two classes:

1. Single-objective problems
   Minimise (or maximise) one linear function of X_i : Minimise $\left\{\sum_{i=1}^{I} c_i X_i\right\}$. If c_i represents the cost of food i then the objective function minimises total diet cost, etc. (Darmon et al., 2002a, Briend et al., 2003, Thompson et al., 2013). Commonly, single objective problems are formulated as straightforward (MI)LP models.

2. Multi-objective problems
   Two types are observed in literature:
   2a. *With the set of available foods no diet can be planned that complies with all constraints* (Anderson and Earle, 1983, Ferguson et al., 2006, Fletcher et al., 1994). A common approach is to search for a “best possible” diet, which violates the constraints as little as possible. This is a problem with multiple objectives: Minimise{Violation of constraint 1}, Minimise{Violation of constraint 2}, …, Minimise{Violation of the last constraint}.

[1] In Chapters 4 and 5, constraints that model consumer preferences will be referred to as ‘palatability constraints’.

2b. *The decision-maker aims to plan a diet that complies with all constraints, and that differs as little as possible from the actual diet (which does not comply with all constraints)* (Darmon et al., 2002b, Darmon et al., 2006b, Maillot et al., 2010, Masset et al., 2009, Thompson et al., 2013). In other words, (s)he wants to minimise the differences between the proposed (optimised) diet X^P and the actual diet X^A:
$\text{minimise}\{|X_1^A - X_1^P|\}$, $\text{minimise}\{|X_2^A - X_2^P|\}$, …, $\text{minimise}\{|X_I^A - X_I^P|\}$.

This paper focuses on multi-objective diet models in Class 2a. For completeness, Class 2b is discussed in Appendix 4A.

Multi-objective problems – searching Pareto-optimal solutions

The objectives of a multi-objective problem are usually conflicting; commonly no solution exists that optimises all objectives at the same time. For instance, the diet with lowest violation of a constraint on iron intake might have a considerable violation of a constraint on intake of saturated fat, and vice versa. Methods for generating solutions to multi-objective problems therefore commonly focus on finding so-called Pareto-optimal solutions (also denoted as efficient solutions). Solution X is called *Pareto-optimal* if no other solution $X^{\#}$ exists that performs at least as good as X with respect to all objectives, and better with respect to at least one objective (Jones and Tamiz, 2010). In other words: the achieved value for one objective cannot be improved without worsening the level of another objective (Tamiz et al., 1998). The concept of Pareto-optimality is illustrated in Table 4.1, which shows four fictitious diets, and their violations v_{Fe} and v_{SF} of intake constraints on iron and saturated fat, respectively. No diet exists that has both lowest v_{Fe} and lowest v_{SF}. We therefore aim to identify the Pareto-optimal diets.

Table 4.1 Example of Pareto-optimality

	Violation of intake constraint on		
	Iron: v_{Fe}	Saturated fat: v_{SF}	Pareto-optimal or not?
Diet A	0	11	Yes
Diet B	4	3	Yes
Diet C	6	4	No: B has lower v_{Fe} and v_{SF}
Diet D	3	8	Yes

All references mentioned in Class 2 use linear Goal Programming (GP). GP uses the following steps (Jones and Tamiz, 2010) to find “best possible” (i.e. Pareto-optimal) diets for problems in Class 2a:

(GP1) quantify the extent to which a diet violates the constraints, and then

(GP2) minimise a function of these violations in order to obtain a diet that violates the constraints as little as possible. This function is called the *achievement function.*

Multi-criteria decision-making (MCDM) literature, in which GP is positioned, recognises achievement function selection, weight selection and weight space analysis as topics of major importance for the quality of decision-making (Tamiz et al., 1998, Romero, 2001, Romero, 2004, Jones, 2011), because any choice made in formulating the achievement function uses judgment of the modeller and implies assumptions on the preference structure of the decision-maker. If the election of the achievement function is wrong, then the decision-maker will probably not accept the solution (Romero, 2004). Results derived from GP-models usually are very sensitive to the type of achievement function that is chosen (Romero, 2004). All references mentioned in Class 2 use a weighted additive achievement function. GP literature offers other achievement functions as well (Romero, 2004).

Typically, with a weighted additive achievement function each set of data and weights results in a single solution. It would be useful to offer the decision-maker more than just this single solution and present a range of Pareto-optimal solutions. Offering multiple solutions allows choice of a solution that is most suitable for a specific decision problem, and that best meets non-quantifiable goals and preferences (Brill Jr, 1979, Makowski et al., 2000, Makowski et al., 2001).

This paper aims to provide methodological insight into several GP achievement functions: MinSum, MinMax, and Extended GP (EGP). It shows that the EGP achievement function is able to generate a solution that minimises the sum of all violations, as well as a solution that minimises the largest violation, and compromises between them. The EGP achievement function thus provides a way to obtain a range of solutions from one set of data and weights.

4.2 Methods

Goal Programming – Numerical example

In order to reveal the 'mechanics' of achievement functions we use a simplified diet model with two foods: 1 ~ bread and 2 ~ meat, with associated decision variables X_1 (X_2) = amount of bread (meat) in the diet. Three nutritional constraints are formulated: (1) an upper bound on salt intake restricts bread consumption to 3 units or less, (2) An upper bound on saturated fat intake restricts meat consumption to 2 units or less, (3) for sufficient iron intake the diet should contain at least 6 units of bread and/or meat:

$$\begin{array}{lllr} X_1 & & \leq 3 & (1) \\ & X_2 & \leq 2 & (2) \\ X_1 + & X_2 & \geq 6 & (3) \\ X_1, X_2 \geq 0 & & & \end{array}$$

Figure 4.1a shows that constraints (1-3) are conflicting: no diet (X_1, X_2) exists that complies with all constraints. The resulting model belongs to Class 2a.

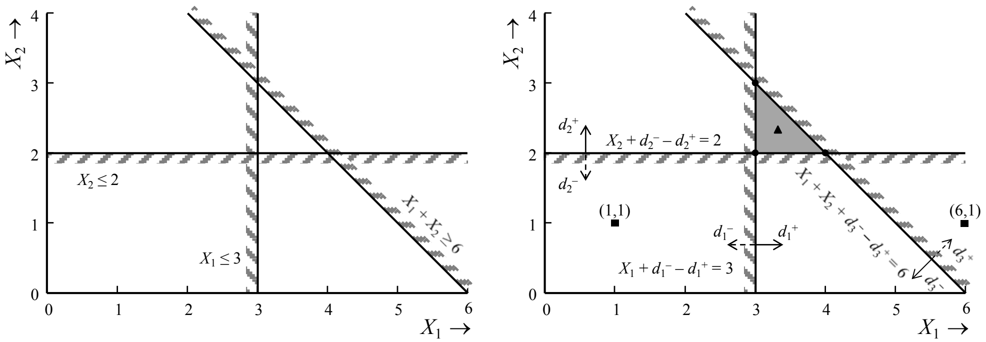

Figure 4.1a **b**

a. A diet model in Class 2a: no diet exists that complies with all constraints.

b. A GP-model uses deviational variables to quantify the deviations from the targets. Deviations d_1^-, d_2^-, d_3^+ (dashed line arrows) are allowed by the original constraints. The d_1^+, d_2^+, d_3^- (full line arrows) are unwanted deviations. A MinSum GP-model will propose a corner-point (•) of the shaded area, no matter which set of weights is used. A MinMax GP-model is able to propose solutions inside the shaded area, e.g. if all weights are equal then $(X_1,X_2) = (3\,^1/_3, 2\,^1/_3)$ (▲) is proposed.

Step (GP1) quantifies the violation of the nutritional constraints by adding deviational variables d_j^-, $d_j^+ \geq 0$, which represent the negative and positive deviation from the right-hand side value of nutritional constraint j:

$$\begin{array}{llcl} (1) & X_1 \leq 3 & \rightarrow & X_1 + d_1^- - d_1^+ = 3 \\ (2) & X_2 \leq 2 & \rightarrow & X_2 + d_2^- - d_2^+ = 2 \\ (3) & X_1 + X_2 \geq 6 & \rightarrow & X_1 + X_2 + d_3^- - d_3^+ = 6 \end{array}$$

For instance, $(X_1, X_2) = (1,1)$ implies $d_1^- = 2$, $d_2^- = 1$, $d_3^- = 4$, and $(X_1, X_2) = (6,1)$ implies $d_1^+ = 3$, $d_2^- = 1$, $d_3^+ = 1$.[2] As initial constraint (3) provides a lower bound for $X_1 + X_2$ it is allowed to have a positive deviation d_3^+. Any shortage d_3^-, however, violates nutritional constraint (3) and is therefore considered an *unwanted deviation*. Likewise, d_1^+ and d_2^+ are unwanted deviations, see Figure 4.1b.
Next, step (GP2) searches the "best possible" diet by formulating and minimising the achievement function.

MinSum achievement function

A weighted additive achievement function – which we refer to as *MinSum* achievement function – minimises a weighted sum of the unwanted deviations. For model (1-3) the unwanted deviations are d_1^+, d_2^+, d_3^-, which results in the following MinSum GP model:

$$\text{Minimise}\{D_{\text{sum}} = w_1^+ d_1^+ + w_2^+ d_2^+ + w_3^- d_3^-\} \tag{4}$$
$$X_1 + d_1^- - d_1^+ = 3 \tag{5}$$
$$X_2 + d_2^- - d_2^+ = 2 \tag{6}$$
$$X_1 + X_2 + d_3^- - d_3^+ = 6 \tag{7}$$
$$X_1, X_2, D_{\text{sum}}, d_1^-, d_1^+, d_2^-, d_2^+, d_3^-, d_3^+ \geq 0 \tag{8}$$

in which w_1^+, w_2^+, w_3^- are user-defined, non-negative weights. The optimal solution of problem (4-8) depends on the weights. Independent of the weights, it is always located in the shaded area of Figure 4.1b. For $w_1^+ = w_2^+ = w_3^-$ solutions in the shaded area have the same value of D_{sum}. As in LP-problems the optimal solution is always a corner-point (Jones and Tamiz, 2010), an LP-solver will generate a corner-point of the shaded area: (3,2) or (4,2) or (3,3). However, if the user specifies $w_1^+ = w_2^+ = w_3^-$ (s)he expresses that (s)he does not want to prioritise one nutrient, i.e. (s)he cannot justify assigning one deviation more importance than another. Hence, it would be natural to obtain a balanced solution that spreads the total unwanted deviation over all three deviational variables are (d_1^+, d_2^+, d_3^-) instead of an unbalanced solution that piles the unwanted deviation on only one of them. Table 4.2 shows that the MinSum model is sensitive to weight changes: slight weight changes make the optimal diet 'jump' from one corner-point to another. However, if the weights expressed by the user are similar, one would expect similar solutions.

[2] Note that in any optimal solution either one of the deviational variables (d_j^- or d_j^+) equals zero, or both (Jones and Tamiz, 2010).

Table 4.2 Solutions of the MinSum, MinMax and Extended GP model

Optimal solution of MinSum GP-model (4)-(8) for various choices of the weights

Weights (w_1^+, w_2^+, w_3^-)	Diet (X_1, X_2)	Deviations (d_1^+, d_2^+, d_3^-)	Weighted deviations $(w_1^+d_1^+, w_2^+d_2^+, w_3^-d_3^-)$	Objective function $D_{\text{sum}} = w_1^+d_1^+ + w_2^+d_2^+ + w_3^-d_3^-$
(0.9, 1, 1)	(4, 2)	(1, 0, 0)	(0.9, 0, 0)	0.9
(1, 0.9, 1)	(3, 3)	(0, 1, 0)	(0, 0.9, 0)	0.9
(1, 1, 0.9)	(3, 2)	(0, 0, 1)	(0, 0, 0.9)	0.9

Optimal solution of MinMax GP-model (10)-(17) for various choices of the weights

Weights (w_1^+, w_2^+, w_3^-)	Diet (X_1, X_2)	Deviations (d_1^+, d_2^+, d_3^-)	Weighted deviations $(w_1^+d_1^+, w_2^+d_2^+, w_3^-d_3^-)$	Objective function $D_{\max} = \max\left(w_1^+d_1^+; w_2^+d_2^+; w_3^-d_3^-\right)$
(1, 1, 1)	(3.333, 2.333)	(0.333, 0.333, 0.333)	(0.333, 0.333, 0.333)	0.333
(0.9, 1, 1)	(3.357, 2.321)	(0.357, 0.321, 0.321)	(0.321, 0.321, 0.321)	0.321
(1, 0.9, 1)	(3.321, 2.357)	(0.321, 0.357, 0.321)	(0.321, 0.321, 0.321)	0.321
(1, 1, 0.9)	(3.321, 2.321)	(0.321, 0.321, 0.357)	(0.321, 0.321, 0.321)	0.321

Optimal solution of Extended GP-model (11)-(18) for various choices of parameter λ.
Weights $(w_1^+, w_2^+, w_3^-) = (0.5;0.75;1)$ **were used.**

Parameter λ	Diet (X_1, X_2)	Deviations (d_1^+, d_2^+, d_3^-)	Weighted deviations $(w_1^+d_1^+, w_2^+d_2^+, w_3^-d_3^-)$	D_{sum}	$D_{\max}$	$D_{\text{ext}} = (1-\lambda)D_{\text{sum}} + \lambda D_{\max}$
0.00	(4, 2)	(1, 0, 0)	(0.50, 0, 0)	0.50	0.50	0.50
0.25	(4, 2)	(1, 0, 0)	(0.50, 0, 0)	0.50	0.50	0.50
0.50	(3.60, 2.40)	(0.60, 0.40, 0)	(0.30, 0.30, 0)	0.60	0.30	0.45
0.75	(3.46, 2.31)	(0.46, 0.31, 0.23)	(0.23, 0.23, 0.23)	0.69	0.23	0.35
1.00	(3.46, 2.31)	(0.46, 0.31, 0.23)	(0.23, 0.23, 0.23)	0.69	0.23	0.23

These imbalance and sensitivity are typical for an additive achievement function (Romero, 2001, Jones and Tamiz, 2010). Using an additive achievement function implies the assumption that all weighted unwanted deviations are additive and that nutritional adequacy of a diet is determined by the sum of its weighted unwanted deviations. This presupposes – implicitly – that

- deviations can compensate each other, e.g. an increase in the deviation from a vitamin C target can be compensated by a decrease in the deviation from a calcium target,
- trade-offs between the deviations are precisely known, e.g. 10% deviation from a vitamin C target is considered equally serious as 5% deviation from a calcium target,
- trade-offs are constant and do not depend on intake level, e.g. 10% extra deviation from a vitamin C target is considered equally serious as 5% extra deviation from a calcium target, no matter whether the vitamin C intake is almost adequate or dangerously low.

MinMax achievement function

A *MinMax* achievement function (also called Chebyshev achievement function) aims to minimise the largest among the weighted unwanted deviations (Romero, 2001, Romero, 2004, Romero et al., 1998):

$$\text{Minimise}\{D_{\max} = \max(w_1^+ d_1^+;\ w_2^+ d_2^+;\ w_3^- d_3^-)\} \tag{9}$$

In order to obtain a linear model that minimises this non-linear achievement function a constraint $D_{\max} \geq w_j^- d_j^-$ ($D_{\max} \geq w_j^+ d_j^+$) is added for every unwanted deviation d_j^- (d_j^+) (Claassen et al., 2007):

$$\text{Minimise}\{D_{\max}\} \tag{10}$$

$$X_1 + d_1^- - d_1^+ = 3 \tag{11}$$

$$X_2 + d_2^- - d_2^+ = 2 \tag{12}$$

$$X_1 + X_2 + d_3^- - d_3^+ = 6 \tag{13}$$

$$D_{\max} \geq w_1^+ d_1^+ \tag{14}$$

$$D_{\max} \geq w_2^+ d_2^+ \tag{15}$$

$$D_{\max} \geq w_3^- d_3^- \tag{16}$$

$$X_1, X_2, D_{\max}, d_1^-, d_1^+, d_2^-, d_2^+, d_3^-, d_3^+ \geq 0 \tag{17}$$

Achievement function (10) and constraints (14-16) together ensure that $D_{\max}$ will take the value $D_{\max} = \max(w_1^+ d_1^+; w_2^+ d_2^+; w_3^- d_3^-)$, and that $D_{\max}$ is as low as possible. Table 4.2 shows that using the MinMax achievement function for $w_1^+ = w_2^+ = w_3^- = 1$ results in a balanced solution: all unwanted deviations are equal. Moreover, the model is less sensitive to weight changes: slight weight changes cause minor shifts in (X_1, X_2). The MinMax achievement function attempts to spread the unwanted deviations as evenly as possible and it enables to identify non-cornerpoint solutions in the shaded area. However, possibly many unwanted deviations are found (Romero, 2001). With a MinMax achievement function the nutritional adequacy of a diet is mainly determined by its poorest nutrient.

Extended GP achievement function

The MinSum and MinMax achievement functions can be combined into the so-called *Extended GP achievement function* D_{ext} (Romero, 2001, Romero, 2004):

$$\text{Minimize } \{D_{\text{ext}} = (1-\lambda)\cdot D_{\text{sum}} + \lambda \cdot D_{\max}\} \tag{18}$$

where parameter $\lambda \in [0;1]$ weighs the importance attached to minimisation of D_{sum} and $D_{\max}$.

D_{ext} comprises both D_{sum} and $D_{\max}$: using $\lambda = 0$ implies $D_{\text{ext}} = D_{\text{sum}}$, so the model minimises the weighted sum of the unwanted deviations. Using $\lambda = 1$ implies $D_{\text{ext}} = D_{\max}$, so the model spreads the deviations and keeps the largest unwanted deviation as low as possible. For intermediate values of λ solutions are found that are compromises between the MinSum and MinMax solution. Thus, by varying λ the decision-maker can obtain a set of dietary suggestions ‘between’ the MinSum and MinMax diets, see also Table 4.2.

Normalisation to [0;1]-interval; interpretation as fuzzy sets

In order to deal with issues of scaling and incommensurability it is useful to formulate a GP-model in such a way that unwanted deviational variables are automatically normalised to a [0;1] interval (Tamiz et al., 1998). For instance, consider a diet model with two foods and two nutrients, see Table 4.3.

Table 4.3 Data for diet model

	Nutrient content		Nutritional constraints		
	(mg in 100g of food 1)	(mg in 100g of food 2)	Lower bound on intake (mg)	Target intake (mg)	Upper bound on intake (mg)
Nutrient 1	500	800	1000	[1500; 3000]	7000
Nutrient 2	0.3	0.1	0.6	[0.9; 2.4]	5

Define

dl_j^-, dl_j^+ ~ deviations from the *l*eft bound of target intake for nutrient *j*

dr_j^-, dr_j^+ ~ deviations from the *r*ight bound of target intake for nutrient *j*

The dl_j^- and dr_j^+ are the unwanted deviations. The nutritional constraints can be formulated as:

$$500X_1 + 800X_2 + 500dl_1^- - 500dl_1^+ = 1500 \quad (19)$$
$$dl_1^- \leq 1 \quad (20)$$
$$500X_1 + 800X_2 + 4000dr_1^- - 4000dr_1^+ = 3000 \quad (21)$$
$$dr_1^+ \leq 1 \quad (22)$$
$$0.3X_1 + 0.1X_2 + 0.3dl_2^- - 0.3dl_2^+ = 0.9 \quad (23)$$
$$dl_2^- \leq 1 \quad (24)$$
$$0.3X_1 + 0.1X_2 + 2.6dr_2^- - 2.6dr_2^+ = 2.4 \quad (25)$$
$$dr_2^+ \leq 1 \quad (26)$$

Thus, all unwanted deviational variables are normalised to [0;1], which facilitates judgment of trade-offs. Moreover, the normalised deviational variables can be used to incorporate fuzzy sets for nutrient intake.

Figure 4.2 shows a graph of intake I_1 of nutrient 1 versus $\mu_1 = 1 - dl_1^- - dr_1^+$.[3] For $1500 \leq I_1 \leq 3000$ the intake of nutrient 1 is considered fully adequate: both unwanted deviational variables dl_1^- and dr_1^+ are zero and $\mu_1 = 1$. Intake $I_1 = 1000$ is considered fully inadequate: $dl_1^- = 1$ and $\mu_1 = 0$. Likewise, $I_1 = 7000$ has $dl_1^+ = 1$ and $\mu_1 = 0$. So, an adequate intake has $\mu_1 = 1$ and an inadequate intake has $\mu_1 = 0$. If we assume the adequacy of intake I_1 increases linearly from 0 to 1 on interval

[3] In an optimal solution dl_j^- and dr_j^+ can never be both nonzero at the same time, so $\mu_1(I_1) = \min(1 - dl_1^-; 1 - dr_1^+) = 1 - dl_1^- - dr_1^+$.

[1000; 1500] we can use μ_1 as proxy for the adequacy of the diet with respect to nutrient 1. In case overall nutritional quality M of the diet is determined by the adequacy of its poorest nutrient we can calculate it as

$$M = \text{Minimum}\{\mu_1; \mu_2\} = 1 - \text{Maximum}\{dl_1^-; dr_1^+; dl_2^-; dr_2^+\} = 1 - D_{\text{max}} \qquad (27)$$

which implies the MinMax achievement function yields diets with maximal M.

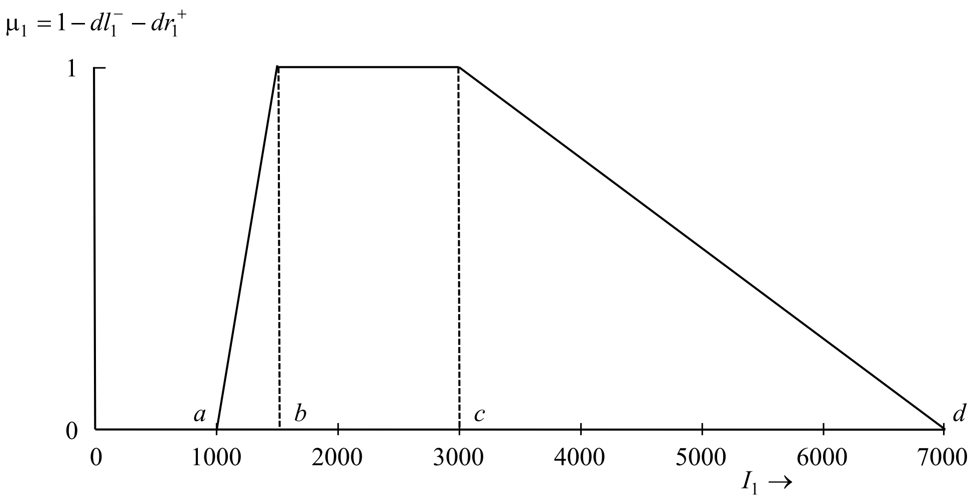

Figure 4.2 Adequacy curve for nutrient 1

The adequacy curve in Figure 4.2 can be interpreted as fuzzy set for the adequacy of intake I_1 with membership function $\mu_1(I_1) = 1 - dl_1^- - dr_1^+$. For more information on use of fuzzy sets for modelling intake adequacy we refer to Gedrich et al. (1999), Wirsam et al. (1997), Wirsam and Uthus (1996). More information on LP-formulations for fuzzy sets is found in Yaghoobi and Tamiz (2007). MILP-formulations of Dantzig (1960) can be used to construct curves with more than three intervals.

4.3 Diet model

This section specifies the diet model with which the results, described in Section 4.4, were obtained. The list of all used foods is provided in Appendix 4B.

Notation

The model uses the following indices, decision variables and data:

Indices

i ~ index for foods, $i = 1, \ldots, I$
j ~ index for nutrients, $j = 1, \ldots, J$

Decision variables

X_i ~ consumption of food i in grams per day
$intake_j$ ~ intake of nutrient j (in nutrient-specific unit: g, mg, µg, en%, or g/MJ)
dl_j^-, dl_j^+ ~ normalised deviation from *l*eft bound of target intake for nutrient j
dr_j^-, dr_j^+ ~ normalised deviation from *r*ight bound of target intake for nutrient j
μ_j ~ adequacy for intake of nutrient j

For sake of readability, many additional decision variables were defined, such as:
bread ~ consumption of bread

These variables are specified in Table 4.4. Moreover, Table 4.4 shows which additional decision variables were defined to indicate how the proposed amount of a certain food (group) is subdivided, for instance:

bread_cheese ~ amount of bread that is filled with cheese
bread_meat ~ amount of bread that is filled with meat
bread_sweetsav ~ amount of bread that is filled with sweet or savory filling

which indicate how the amount of *bread* is subdivided over the types of filling:

$bread = bread\_cheese + bread\_meat + bread\_sweetsav$

Parameters

$nc_{i,j}$ ~ nutrient content of food i for nutrient j (in g or mg per g food i)
$total\_energy$ ~ total energy in the diet: 11100 kJ
a_j, b_j, c_j, d_j ~ parameters for adequacy curves

Palatability constraints

Palatability constraints are used to assure that a diet is generated that is acceptable for the consumer. Some provide lower bounds (for instance allow 25g of snacks), others provide upper bounds (for instance no more than 300g of starch component). Moreover, several linking constraints are defined, such as a link between salad and dressing, a link between bread and spread. Table 4.4 presents each palatability constraint both verbally and as a formula.

Table 4.4 Palatability constraints

Bread	$bread = \sum_{i=11}^{i=15} X_i$
3-7 slices of 35g of bread.	$3 \cdot 35 \leq bread \leq 7 \cdot 35$
Bread can have cheese, meat, sweet/savory filling.	$bread = bread\_cheese + bread\_meat + bread\_swsav$
Use at least two types of filling.	$bread\_cheese \leq bread/2$; $bread\_meat \leq bread/2$ $bread\_sweetsav \leq bread/2$
Spread	$spread = \sum_{i=47}^{i=57} X_i + buttersalt\_spread + butter\_spread$
Per slice use 3-7g spread.	$3 \cdot bread\_total/35 \leq$ spread $\leq 7 \cdot bread\_total/35$
Butter can be used as spread and as cooking fat.	$X_{24} = buttersalt\_spread + buttersalt\_cooking$ $X_{25} = butter\_spread + butter\_cooking$
Cheese	$cheese = X_{20} + X_{21} + X_{22}$
Cheese can be used as filling or at dinner.	$cheese = cheese\_filling + cheese\_dinner$
On bread one cheese filling is 15-30g.	$15 \cdot bread\_cheese/35 \leq cheese\_filling \leq 30 \cdot bread\_cheese/35$
Meat product	$meat\_filling = \sum_{i=71}^{i=77} X_i$
One meat filling is 10-25g.	$10 \cdot bread\_meat/35 \leq meat\_filling \leq 25 \cdot bread\_meat/35$
Sweet and savory filling	$sweetsav\_filling = \sum_{i=123}^{i=135} X_i$
One sweet/savory filling is 10-20g.	$10 \cdot bread\_swsav/35 \leq swsav\_filling \leq 20 \cdot bread\_swsav/35$
Starch component	$starchcomp = \sum_{i=16}^{i=19} X_i + \sum_{i=63}^{i=65} X_i + \sum_{i=121}^{i=122} X_i$
No more than 300g.	$starchcomp \leq 300$
Pulses no more than twice a week.	$X_{63} + X_{64} + X_{65} \leq starchcomp \cdot 2/7$
Potatoes, rice, pasta no more than 3 times/wk.	$X_{16} + X_{17} \leq starchcomp \cdot 3/7$; $X_{18} + X_{19} \leq starchcomp \cdot 3/7$; $X_{121} + X_{122} \leq starchcomp \cdot 3/7$
Protein component (fish, meat, cheese, nuts, eggs)	$proteincomp = \sum_{i=66}^{i=70} X_i + \sum_{i=58}^{i=60} X_i +$ $cheese\_diner + nuts\_dinner + eggs\_dinner$
50-300g.	$50 \leq proteincomp \leq 300$
Nuts as snack or as protein component.	$X_{92} = nuts\_snack + nuts\_dinner$
Eggs with bread or as protein component.	$X_{23} = eggs\_bread + eggs\_dinner$
Fat component	$fatcomp = \sum_{i=26}^{i=32} X_i +$ $butter\_cooking + buttersalt\_cooking$
5-30g.	$5 \leq fatcomp \leq 30$
Vegetables[1]	
No more than 400g of cooked vegetables.	$X_{143} \leq 400$
No more than 250g of salad.	$X_{144} \leq 250$
Dressing	$dressing = \sum_{i=33}^{i=43} X_i$
10-30g per 100g salad	$0.1 \cdot X_{144} \leq dressing \leq 0.3 \cdot X_{144}$
Dessert	$dessert = \sum_{i=84}^{i=91} X_i$
150-300g	$150 \leq dessert \leq 300$
Fruit[1]	
No more than 300g	$X_{61} + X_{62} \leq 300$
Snacks	$snack = \sum_{i=93}^{i=107} X_i + nuts\_snack$
25-100g	$25 \leq snacks \leq 100$
Drinks	$drinks = \sum_{i=1}^{i=10} X_i + \sum_{i=78}^{i=83} X_i$
1500-3000g	$1500 \leq drinks \leq 3000$
Spread over categories	$X_1 + X_2 \leq drinks \cdot 0.4$; $X_3 + X_4 + X_5 + X_6 + X_7 \leq drinks \cdot 0.4$ $X_8 + X_9 \leq drinks \cdot 0.4$; $X_{10} \leq drinks \cdot 0.4$ $X_{78} + X_{79} + X_{80} + X_{81} + X_{82} + X_{83} \leq drinks \cdot 0.4$

[1]Note that lower bounds on intake of vegetables and fruits are provided via the nutritional constraints.

Administrative constraints

Administrative constraints are used to convert food intakes to nutrient intakes, expressed in units that match the norm to which they must comply. For instance, protein intake is expressed in energy%, Ca-intake is expressed in mg, and fibre intake is expressed in g/MJ.

For j = protein, carbohydrates, mono-di-saccharides:

$$intake_j = \left(\sum_{i \in \{1..I\}} nc_{i,j} X_i \cdot 17 \right) / total\_energy,$$

For j = total fat, SFA, MUFA, PUFA, linoleic acid:

$$intake_j = \left(\sum_{i \in \{1..I\}} nc_{i,j} X_i \cdot 37 \right) / total\_energy,$$

For j = EPA+DHA, cholesterol, Ca, Fe, K, Vit B_1, Vit B_2, Vit B_6, Vit B_{12}, Vit C, folate:

$$intake_j = \sum_{i \in \{1..I\}} nc_{i,j} X_i,$$

For j = dietary fibre:

$$intake_j = \left(\sum_{i \in \{1..I\}} nc_{i,j} X_i \cdot 1000 \right) / total\_energy,$$

Nutritional constraints

Nutritional constraints are used to calculate the adequacy of the intake of nutrients. In a similar way, the adequacy of the intake of vegetables and fruit is calculated. They are coded as nutrients j = 20 (vegetables) and j = 21 (fruits).

$$intake_j + (b_j - a_j) \cdot dl_j^- + dl_j^+ = b_j \qquad \forall j$$

$$intake_j + dr_j^- + (d_j - c_j) \cdot dr_j^+ = c_j \qquad \forall j$$

$$\mu_j + dl_j^- + dr_j^+ = 1 \qquad \forall j$$

The intake of nutrient 0 (= energy) is fixed to *total_energy*:

$$\sum_{i \in \{1..I\}} nc_{i,0} X_i = total\_energy$$

MCDM-related constraints

These constraints are used to calculate μ_j, μ_{min}, D_{sum}, D_{max}, D_{ext}. The ε-terms are used to ensure efficiency, with ε very small (e.g. 0.001).

$$D_{\text{sum}} = \sum_j \left(dl_j^- + dr_j^+ \right) - \varepsilon \cdot \mu_{\text{min}}$$

$$\mu_{\text{min}} \le \mu_j \qquad \forall j$$

$$D_{\text{min}} = 1 - \mu_{\text{min}} - \varepsilon \cdot \sum_j \mu_j$$

$$D_{\text{ext}} = (1 - \lambda) \cdot D_{\text{sum}} + \lambda \cdot D_{\text{max}} \qquad \text{with } 0 \le \lambda \le 1$$

Objective function

minimise $\{ D_{ext} \}$

4.4 Results

The Extended GP achievement function is used to obtain “best possible” solutions for a diet model for planning of diets for dietary controlled trials for men aged 19-30y as presented in Section 4.3. Palatability constraints define lower and upper bounds on intake of various foods (“at most 245g of bread”), and they link intakes of foods (“3-7g spread per slice”). Nutritional constraints are formulated via adequacy curves for 19 nutrients and for vegetables and fruits. The four characteristic points of each adequacy curve are defined in the following way: *a* is the Lower Intake Level below which an intake could lead to risk in most individuals; *b* is the average requirement, sufficient for virtually 50% of healthy people in a group, *c* is the Recommended Daily Intake which is sufficient for nearly all people; *d* is the Upper Intake Level that is unlikely to pose a risk of adverse health effects (Nordic council of Ministers, 2014). If no information was available a nutrition expert (JdV) made an estimate (Health Council of the Netherlands, 2006, Nordic council of Ministers, 2014). Energy intake was fixed to 100% of the estimated average requirements (EAR). Food compositions were obtained from RIVM (2012). The model was programmed in Fico Xpress 7.0.1 and calculations were done on an HP desktop with Intel I7 processor. The model generates the 11 diets in one second.

Table 4.5 provides reference values for the adequacy curves and summarises results for $\lambda \in \{0; 0.1; \ldots; 1\}$. Using $\lambda = 0$ implies $D_{ext} = D_{sum}$, so the model finds

Table 4.5 Reference values for nutritional adequacy curves and summary of results for $\lambda = 0, 0.1, \ldots, 1$.

Reference values for adequacy curves						Results for $\lambda = 0, 0.1, \ldots, 1$																
						0				0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1			
Nutrient	unit[1]	*a*	*b*	*c*	*d*	*I*	dl^-	dr^+	μ	μ									Intake	dl^-	dr^+	μ
Protein	en%	8%	10%	20%	25%	14.0%	0	0	1	1	1	1	1	1	1	1	1	1	19.2%	0	0	1
Total fat	en%	2%	25%	40%	45%	40.0%	0	0	1	1	1	1	1	1	1	1	1	1	40.8%	0	0.163	0.837
SFA[2]	en%	0%	0%	10%	15%	10.0%	0	0	1	1	1	1	1	1	1	1	1	1	10.8%	0	0.163	0.837
MUFA[3]	en%	8%	10%	20%	28%	17.6%	0	0	1	1	1	1	1	1	1	1	1	1	16.9%	0	0	1
PUFA[4]	en%	0%	3%	10%	12%	10.0%	0	0	1	1	1	1	1	1	1	1	1	1	10.3%	0	0.163	0.837
Linol	en%	0%	2%	9%	10%	8.82%	0	0	1	1	1	1	1	1	1	1	1	1	8.42%	0	0	1
EPA+DHA	mg	350	450	3000	4000	450	0	0	1	1	1	1	1	1	1	1	1	1	434	0.163	0	1
Cholesterol	mg	0	0	200	300	74.6	0	0	1	1	1	1	1	1	1	1	1	1	180	0	0	1
MDSacch[5]	en%	0%	0%	5%	10%	7.30%	0	0.460	0.540	0.691	0.692	0.700	0.702	0.703	0.704	0.729	0.782	0.820	5.82%	0	0.163	0.837
Fiber	gr/MJ	0	3	---[6]	---[6]	3.00	0	0	1	1	1	1	1	1	1	1	1	0.820	2.51	0.163	0	1
Ca	mg	400	500	800	2500	720	0	0	1	1	1	1	1	1	1	1	1	1	949	0	0.088	0.912
Fe	mg	7	9	23	60	12.1	0	0	1	1	1	1	1	1	1	1	1	1	12.4	0	0	1
K	mg	1600	3100	3500	10000	3227	0	0	1	1	1	1	1	1	1	1	1	1	3745	0	0.038	0.962
Vit B_1	mg	0.5	0.9	1.4	7	1.35	0	0	1	1	1	1	1	1	1	1	1	0.856	2.32	0	0.163	0.837
Vit B_2	mg	0.8	1.1	1.7	8.5	1.1	0	0	1	1	1	1	1	1	1	1	1	1	1.26	0	0	1
Vit B_6	mg	0.8	1	1.5	25	1.5	0	0	1	1	1	1	1	1	1	0.993	0.993	0.979	2.15	0	0.028	0.972
Vit B_{12}	μg	1	1.4	2	10	2.0	0	0	1	1	0.990	0.981	0.981	0.980	0.980	0.926	0.911	0.869	3.31	0	0.163	0.837
Vit C	mg	10	50	75	375	75	0	0	1	0.846	0.854	0.852	0.850	0.848	0.847	0.835	0.866	0.867	124	0	0.163	0.837
Folate	μg	100	200	400	2000	292	0	0	1	1	1	1	1	1	1	1	1	1	309	0	0	1
Vegetables	g	150	200	400[7]	400[7]	200	0	0	1	1	1	1	1	1	1	1	1	1	192	0.163	0	1
Fruits	g	100	200	300[7]	300[7]	200	0	0	1	1	1	1	1	1	1	1	0.782	0.820	184	0.163	0	1
						D_{sum}			0.460	0.463	0.463	0.466	0.468	0.469	0.469	0.518	0.666	0.969				1.95
						D_{max}			0.460	0.309	0.308	0.300	0.298	0.297	0.296	0.271	0.218	0.180				0.163
						D_{ext}			0.460	0.448	0.432	0.416	0.400	0.383	0.365	0.345	0.307	0.259				0.163

[1]These units apply to *a*, *b*, *c*, *d*, *I*. The dl^-, dr^+, μ are dimensionless.

[2]Saturated Fatty Acids, [3]Mono-Unsaturated Fatty Acids, [4]Poly-Unsaturated Fatty Acids, [5]Mono- and DiSaccharides

[6]No *c* and *d* are used, because within the feasible diets fibre intake will never be too high. [7]These are palatability constraints.

the diet with minimal sum of unwanted deviations. In this MinSum diet all nutrient intakes except mono- and disaccharides (MDSacch) are in their optimal range (at plateau [b,c] of the adequacy curve). However, this diet has dr^{+}_{MDSacch} = (7.30 – 5)/(10 – 5) = 0.460, which means that MDSacch intake is far too high and has a suboptimal adequacy μ_{MDSacch} = 1 – 0.460 = 0.540. As MDSacch is the only nutrient with suboptimal intake the MinSum diet has $D_{\max}$ = 0.460. Increasing λ implies that the model lowers $D_{\max}$ at the cost of increasing D_{sum}, thus spreading the unwanted deviations in order to find a more balanced solution: for λ = 0.1 the unwanted deviation of MDSacch decreases to dr^{+}_{MDSacch} = 0.309 (so μ_{MDSacch} = 1 – 0.309 = 0.691) whereas vitamin C intake becomes suboptimal (μ_{VitC} = 0.846). For λ = 0.9 the largest deviation has decreased to 0.180. However, seven intakes are then suboptimal. For $\lambda = 1$ the model strictly minimises the largest unwanted deviation, which results in the diet with lowest $D_{\max}$. This MinMax diet has 14 suboptimal intakes and the highest D_{sum}. It is entirely up to the nutritionist to judge whether a decrease in $D_{\max}$ is worth an increase in D_{sum}, and/or an increase in the number of violated constraints and to express a preference for any of the generated diets, based on the specific situation on hand. Table 4.6 shows which foods were chosen.

4.5 Discussion

This paper aims to provide methodological insight into several GP achievement functions: MinSum, MinMax, and Extended GP. A MinSum achievement function minimises the sum of the unwanted deviations from nutritional targets and is thus appropriate in situations where diet quality is determined by the sum of these unwanted deviations. It can, however, lead to solutions that are unbalanced and that are sensitive to changes in preferential weights. A MinMax achievement function minimises the largest among the unwanted deviations, and is thus appropriate when diet quality is mainly determined by the nutrient with the largest unwanted deviation. MinMax GP provides solutions that are as balanced as possible with respect to the unwanted deviational variables. However, possibly many unwanted deviations occur. An EGP achievement function is a compromise between MinSum and MinMax. It can – from one set of data and weights – obtain the MinSum solution, the MinMax solution, and a set of solutions ‘between’ the MinSum and MinMax solutions. Offering multiple solutions allows choice of a solution that is most suitable for a specific decision problem, and that best meets non-quantifiable goals and preferences (Brill Jr, 1979, Makowski et al., 2000, Makowski et al., 2001).

Table 4.6 Food intake for λ = 0, 0.1, …, 1 (in grams).

i	Food	λ										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
2	Tea prepared	600	600	600	600	600	600	600	600	600	600	600
6	Tomato juice	0	0	0	0	0	0	0	0	0	0	53
8	Beer alcohol free <0,1 vol%	151	172	171	193	199	205	208	300	300	300	247
10	Water 50-100 mg calcium per liter	600	600	600	600	600	600	600	600	600	600	600
15	Bread wholemeal average	197	205	205	201	200	198	198	203	208	145	142
16	Pasta wholemeal boiled	86	86	86	86	86	86	86	86	86	86	86
18	Rice brown boiled	72	40	38	46	45	45	45	0	0	0	0
19	Rice white boiled	56	88	91	83	83	84	84	129	129	129	129
21	Cheese 48+ less salt average	0	0	0	0	0	0	0	0	20	0	0
22	Cheese 30+ average	0	0	0	9	9	10	10	17	3	22	35
23	Eggs chicken boiled average	0	0	0	0	0	0	0	24	20	0	0
35	Salad dressing vinaigrette	33	31	30	25	22	20	19	9	5	24	13
44	Oil olive	27	26	26	29	30	29	28	25	25	30	22
46	Oil sunflower seed	3	4	4	1	0	1	2	5	5	0	8
48	Low fat margarine 40% fat <17 g sat	3	6	7	0	0	0	0	0	0	0	0
53	Margarine 80% fat > 24g sat unsalted	0	0	0	32	40	36	35	19	0	0	0
54	Margarine 80% fat >24 g saturates	36	35	34	8	0	0	0	0	0	0	0
55	Margarine 80% fat 17-24 g saturates	0	0	0	0	0	4	5	22	42	29	28
60	Fish fat > 10 g fat raw	14	14	14	14	14	14	14	14	14	14	13
61	Fruit fresh average excluding citrus	144	0	0	0	0	0	0	0	0	0	0
62	Fruit fresh citrus average	56	200	200	200	200	200	200	200	178	182	184
65	Lentils boiled	86	86	86	86	86	86	86	86	86	86	86
68	Pork <10% fat prepared	55	57	57	60	63	65	66	63	64	148	147
69	Pork >19% fat prepared	0	0	0	0	0	0	0	0	0	51	64
71	Processed meat prd <10g fat excl liver	17	17	0	0	0	0	0	0	0	0	0
73	Processed meat prd >30g fat excl liver	0	0	0	0	10	18	22	29	30	0	0
75	Processed meat products 10-20 g fat	11	13	29	23	13	3	0	0	0	21	20
83	Buttermilk	149	128	129	107	101	95	92	0	0	0	0
90	Yoghurt full fat	150	150	150	150	150	150	150	150	150	150	150
98	Crisps potato average	0	0	0	0	0	0	0	21	22	7	30
99	Japanese rice cracker mix wo peanuts	100	100	100	100	100	100	100	79	78	93	70
123	Peanut butter	28	29	29	29	29	28	28	17	14	6	0
143	Vegetables average boiled	89	98	102	117	125	133	135	170	185	120	148
144	Vegetables mixture raw	111	102	98	83	75	67	65	30	15	80	43

Extending a MinSum GP model to a MinMax GP model takes one additional variable (D_{max}) plus one constraint for every unwanted deviation. In terms of model size this can be considered as very small.
The Extended GP achievement function requires one extra model parameter (λ), which weighs the importance attached to minimising the total unwanted deviation versus the importance of minimising the largest unwanted deviation. No general rule can be given for setting the most useful value of λ. It seems most practical to let the model run for e.g. $\lambda \in \{0; 0.1; \ldots; 1\}$ and then judge the resulting diets and their nutritional adequacy. If desired, the model can be re-run with smaller step-sizes for λ in relevant sub-intervals, e.g. $\lambda \in \{0.71; 0.72; \ldots; 0.89\}$. In this way, the decision-maker is supported in finding his/her own trade-offs.

GP offers several other achievement functions that are worth exploring in diet modelling context. For instance, Lexicographic GP (Tamiz et al., 1998, Romero, 2004), in which the deviational variables are assigned to a number of priority levels that are minimised sequentially. In minimisation runs for lower-level deviational variables the higher-level deviational variables are fixed to their (previously obtained) optimal values. Lexicographic GP requires the decision-maker to provide a strict hierarchy of the unwanted deviations. Also, in GP it is possible to minimise the number of unmet nutritional constraints (Jones and Jimenez, 2013). This requires introducing binary variables indicating whether nutritional constraints have been met. It is useful in situations where unmet nutritional constraints incur costs, e.g. due to necessary fortification programs.

A key consideration in achievement function selection is the preference structure of the decision-maker (Romero, 2004). If the deviational variables can be classified into strict priority classes between which no finite trade-offs exist then Lexicographic GP should be considered. If finite trade-offs do exist between deviational variables then the decision-maker should consider Extended GP, which offers the opportunity either to minimise the total deviation or the largest deviation or a compromise between both. A decision-maker who wants to minimise the number of unmet nutritional constraints needs to introduce binary variables. It will depend on the type of nutrition question, for example whether it is aimed at the individual or population level, what the best modelling approach is. Further research is necessary to build a comprehensive framework that helps nutritionists to select the most suitable modelling approach for a wide range of diet problems.

This paper focuses on linear achievement functions. In literature also models are described with quadratic achievement functions (Carlson et al., 2007, Gao et al., 2006, Gedrich et al., 1999, Cleveland et al., 1993). In a quadratic achievement function the (weighted) sum of squared unwanted deviations is minimised, which means that large deviations are penalised more than small deviations (Cleveland et al., 1993). Quadratic achievement functions can find non-cornerpoint solutions. However, possibly local optima are generated (Bazaraa et al., 1993).

The quality of the solutions of a diet model depends on the quality and choice of data (Buttriss et al., 2014). For our experiment, we chose a limited number of foods and nutrients, and had to base adequacy curves partly on expert opinion, because not all needed information was available in literature. Also, we did our experiment only for a specific diet of one energy level. Other foods could have been used, as well as other bounds and other nutrients. However, it was not our intention to present a comprehensive diet model, but to demonstrate the impact of different achievement functions.

The presented diet model uses a continuous decision variable X_i to denote intake of food i. The model could therefore propose a diet with e.g. 2g apple (unrealistically low) or 17g margarine and 13g low-fat margarine (whereas a consumer would probably want to use either margarine or low-fat margarine but not both). Such issues can be overcome by extending the model with semi-continuous variables and binary variables.

This paper provides methodological insight into several GP achievement functions: MinSum, MinMax, and Extended GP. It shows that the EGP achievement function is able to generate a solution that minimises the sum of all violations, as well as a solution that minimises the largest violation, and compromises between them. The EGP achievement function thus provides a way to obtain a range of solutions from one set of data and weights.

Appendix 4A Model Class 2b: Minimise the differences between the actual diet and the proposed diet

Models in Class 2b search for diets that conform as closely as possible to an actual diet, while meeting constraints on e.g. nutrient recommendations, palatability and total cost.
The composition of the actual and proposed diet can be expressed in amounts consumed per food group (Maillot et al., 2010, Masset et al., 2009) or in energy provided per food group (Darmon et al., 2006b). These result in models of similar structure, and therefore we only elaborate models that express diets in amounts consumed per food group.

Notation, GP constraint

In this appendix, we use the following notation:

Q_i ~ Data: Intake of food group i in actual diet ($i = 1 \ldots I$)
X_i ~ Decision variable: Intake of food group i in proposed diet ($i = 1 \ldots I$)

with I the number of food groups. The deviation between the actual diet and the proposed diet can be calculated via a GP constraint:

$$X_i + Q_i d_i^- - Q_i d_i^+ = Q_i \qquad \text{for all food groups } i$$
$$d_i^-, d_i^+ \geq 0$$

Now $d_i = d_i^- + d_i^+$ is the absolute value of the relative deviation between the actual and the proposed diet[4]. The negative and positive deviations d_i^-, d_i^+ are normalised: $d_i^- = 1$ means that the consumption in food group i of the proposed diet is 100% lower than in the actual diet, and $d_i^+ = 1$ means that the consumption in food group i of the proposed diet is 100% higher than in the actual diet.

MinSum achievement function

As an example we consider a diet problem with two foods. Actual food intakes are $Q_1 = 3$ and $Q_2 = 2$, and a nutritional constraint $2X_1 + 3X_2 \geq 15$ applies. The

[4] Note that it is also possible to model absolute deviations: $X_i + d_i^- - d_i^+ = Q_i$ for all i. One should keep in mind that any choice made in modelling the deviation uses judgment from the decision-maker and implies assumptions on the underlying preference structure.

proposed diet should resemble the actual diet as much as possible. Therefore, all deviations d_i^-, d_i^+ from the actual diet are regarded as unwanted. A MinSum GP model could look as follows:

$$\text{Minimise}\{D_{\text{sum}} = w_1^- d_1^- + w_1^+ d_1^+ + w_2^- d_2^- + w_2^+ d_2^+ \} \tag{A1}$$
$$X_1 \quad + 3d_1^- - 3d_1^+ = 3 \tag{A2}$$
$$X_2 + 2d_2^- - 2d_2^+ = 2 \tag{A3}$$
$$2X_1 + 3X_2 \geq 15 \tag{A4}$$
$$X_1, X_2, D_{\text{sum}}, d_1^-, d_1^+, d_2^-, d_2^+ \geq 0$$

with w_1^-, w_1^+, w_2^-, w_2^+ the non-negative weights assigned by the user.

The shaded area in Figure 4A.1 contains all diets that comply with nutritional constraint (A4). Of these, we want to find the one that has minimal deviation from the actual diet. The set of Pareto-optimal solutions is indicated with a bold line segment connecting (3,3) and (4.5,2).

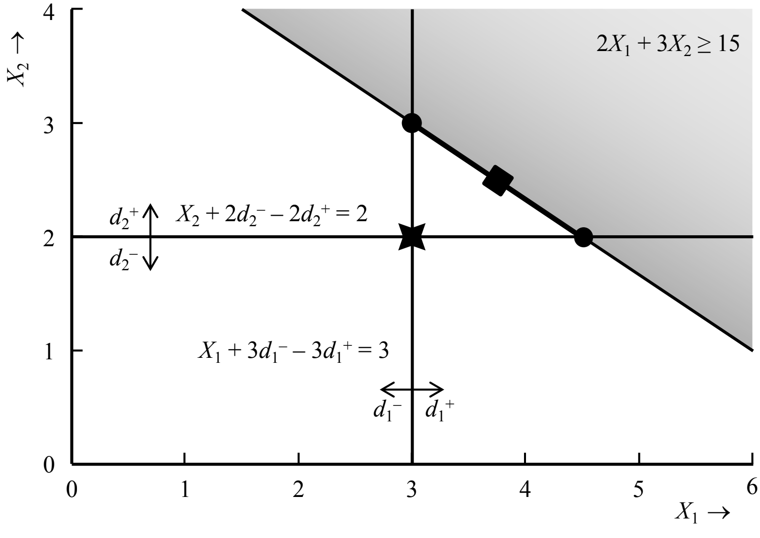

Figure 4A.1 Class 2b: The shaded area contains all diets that comply with nutritional constraint (A4). Of these we aim to find the diet with minimal deviation from the actual diet (✖). A MinSum GP model will yield diet (4.5, 2) or diet (3,3) (●). Diet (3.75, 2.5) (◆) is reachable via a MinMax GP model.

The optimal solution of model (A1)-(A4) depends upon the weights that are chosen. However, independent of the weights, it is always located within the Pareto set (i.e. the set of Pareto-optimal solutions). Note that all Pareto-optimal solutions have $d_1^- = d_2^- = 0$. For $w_1^+ = w_2^+ = 1$ all Pareto-optimal solutions have total weighted deviation $D_{\text{sum}} = 1$. That means that – according to the MinSum GP model with $w_1^+ = w_2^+ = 1$ – all Pareto-optimal diets are equally preferable. Due to the fact that in LP-problems the optimal solution is always found in a corner-point, an LP-solver will generate one of the corner-points of the efficient set: (4.5,2) or (3,3). However, if the user specifies that $w_1^+ = w_2^+$, (s)he expresses that (s)he cannot justify assigning one deviation more importance than another, and therefore it would be natural to obtain a balanced solution that spreads the deviations over both deviational variables (d_1^+, d_2^+) instead of an unbalanced solution that piles the unwanted deviation on only one of them.

Table 4A.1 Optimal solution of MinSum GP model (A1)-(A4) for various choices of the weights.

	Optimal solution			
Weights (w_1^+, w_2^+)	Diet (X_1, X_2)	Deviations (d_1^+, d_2^+)	Weighted deviations $(w_1^+ d_1^+, w_2^+ d_2^+)$	Achievement function $D_{\text{sum}} = w_1^+ d_1^+ + w_2^+ d_2^+$
(0.9, 1)	(4.5, 2)	(0.5, 0)	(0.45, 0)	0.45
(1, 1)	(4.5, 2) or (3, 3)†	(0.5, 0) or (0, 0.5)	(0.5, 0) or (0, 0.5)	0.50
(1, 0.9)	(3, 3)	(0, 0.5)	(0, 0.45)	0.45

† According to the MinSum achievement function all diets on line segment (4.5, 2)-(3, 3) are equally preferable, because they all have the same achievement function value: 0.50. However, an LP-solver will only generate a corner-point: (4.5, 2) or (3, 3).

Table 4A.1 shows the optimal solution of model (A1)-(A4) for various choices of the weights w_1^+, w_2^+. It shows that the MinSum model is sensitive to weight changes: slight changes in the weights make the optimal diet 'jump' from one corner-point to another.

MinMax achievement function

The MinMax GP version of model (A1)-(A4) is

$$
\begin{array}{llr}
\text{Minimise}\{D_{\max}\} & & \text{(A5)} \\
X_1 + 3d_1^- - 3d_1^+ = 3 & & \text{(A2)} \\
X_2 + 2d_2^- - 2d_2^+ = 2 & & \text{(A3)} \\
2X_1 + 3X_2 \geq 15 & & \text{(A4)} \\
D_{\max} \geq w_1^- d_1^- & & \text{(A6)} \\
D_{\max} \geq w_1^+ d_1^+ & & \text{(A7)} \\
D_{\max} \geq w_2^- d_2^- & & \text{(A8)} \\
D_{\max} \geq w_2^+ d_2^+ & & \text{(A9)} \\
X_1, X_2, D_{\max}, d_1^-, d_1^+, d_2^-, d_2^+ \geq 0 & &
\end{array}
$$

Table A2 shows the optimal solution of model (A2)-(A9) for various choices of the weights w_1^+, w_2^+. Now, for $w_1^+ = w_2^+ = 1$, a balanced solution is obtained, in which all unwanted deviations are equal. Moreover, the model is less sensitive to weight changes: slight changes in the weights cause minor shifts in the optimal solution. The MinMax GP model represents a situation in which the resemblance between two diets is determined by the food that differs the most.

Table 4A.2 Optimal solution of MinMax GP model (A2)-(A9) for various choices of the weights.

	Optimal solution			
Weights (w_1^+, w_2^+)	Diet (X_1, X_2)	Deviations (d_1^+, d_2^+)	Weighted deviations $(w_1^+ d_1^+, w_2^+ d_2^+)$	Achievement function $D_{\max}$
(0.9, 1)	(3.79, 2.47)	(0.263, 0.237)	(0.237, 0.237)	0.237
(1, 1)	(3.75, 2.50)	(0.250, 0.250)	(0.250, 0.250)	0.250
(1, 0.9)	(3.71, 2.53)	(0.237, 0.263)	(0.237, 0.237)	0.237

Extended GP achievement function

The achievement functions of the MinSum and MinMax GP model can be combined into the achievement function D_{ext} of a so-called *Extended GP model* (Romero, 2004, Romero et al., 1998):

$$\text{Minimize } \left\{ D_{\text{ext}} = (1-\lambda) \cdot D_{\text{sum}} + \lambda \cdot D_{\max} \right\} \tag{A10}$$

where parameter $\lambda \in [0;1]$ weighs the importance attached to the minimisation of the MinSum and MinMax achievement function. An Extended GP model can find solutions that are compromises between the MinSum and MinMax solutions. It can thus provide valuable dietary suggestions that are located ‘between’ the MinSum and MinMax diets. However, the numerical example is too small to contain any of these intermediate solutions.

Appendix 4B List of foods

Alc. and non-alc. drinks	Index *i*	Nevo code
Coffee prepared	1	644
Tea prepared	2	645
Juice grapefruit	3	664
Juice apple	4	383
Juice grape	5	396
Tomato juice	6	413
Juice orange pasteurised	7	410
Beer alcohol free <0,1 vol%	8	1519
Soft drink wo caffeine	9	400
Water 50-100 mg calcium per liter	10	599
Bread		
Bread brown wheat	11	236
Bread white water based	12	248
Bread white milk based	13	241
Bread rye average	14	1395
Bread wholemeal average	15	246
Cereal products		
Pasta wholemeal boiled	16	2157
Pasta white average boiled	17	659
Rice brown boiled	18	1014
Rice white boiled	19	658
Cheese		
Cheese Gouda 48+ average	20	513
Cheese 48+ less salt average	21	881
Cheese 30+ average	22	1382
Eggs		
Eggs chicken boiled average	23	84
Fats, oils and savoury sauces		
Butter salted	24	879
Butter unsalted	25	310
Cooking fat liq 97%fat <17g sat unsalted	26	2562
Cooking fat liquid 97% fat <17 g sat	27	2066
Cooking fat sol 97% fat>17g sat unsalted	28	2563
Cooking fat solid 80% fat >17 g sat	29	2073
Cooking fat solid 97% fat >17 g sat	30	2067
Margarine liq 80% fat < 17g sat unsalted	31	2558
Margarine liq 80% fat <17 g saturates	32	2077
Salad cream 25% oil	33	458
Salad dressing naturel without oil	34	844
Salad dressing vinaigrette	35	2466
Salad dressing./sauce approx 13% oil	36	2467
Mayonnaise	37	451
Mayonnaise low fat 40% oil	38	729
Mayonnaise product approx 35% oil	39	2471
Mayonnaise product w olive oil	40	2083
Sauce for chips 25% oil	41	465
Sauce for chips 35% oil	42	466
Sauce for chips 5% oil	43	2470
Oil olive	44	601
Oil peanut	45	308
Oil sunflower seed	46	317
Low fat marg prod 20-25% fat <10 g sat	47	2061

Low fat margarine 40% fat <17 g sat	48	2059
Low fat margarine prod 35% fat <10 g sat	49	2060
Low fat margarine prod Becel Omega3 Plus	50	2422
Low fat margarine prod Blue Band Idee	51	2423
Low fat spread Becel pro-activ	52	1956
Margarine 80% fat > 24 g sat unsalted	53	2557
Margarine 80% fat >24 g saturates	54	2063
Margarine 80% fat 17-24 g saturates	55	2062
Margarine product 60% fat <17 g sat	56	2072
Margarine product 70% fat >17 g sat	57	2065
Fish		
Fish lean 0-2 g fat raw	58	114
Fish medium fat > 2-10 g fat raw	59	115
Fish fat > 10 g fat raw	60	116
Fruit		
Fruit fresh average excluding citrus	61	173
Fruit fresh citrus average	62	172
Legumes		
Peas marrowfat legumes boiled	63	969
Beans white/brown boiled	64	968
Lentils boiled	65	970
Meat, meat products and poultry		
Beef <10% fat prepared	66	1665
Beef >10% fat prepared	67	1666
Pork <10% fat prepared	68	1670
Pork >19% fat prepared	69	1672
Pork 10-19% fat prepared	70	1671
Processed meat prod <10 g fat excl liver	71	1908
Processed meat prod 20-30 g fat ex liver	72	1910
Processed meat prod >30 g fat excl liver	73	1911
Processed meat products <10 g fat	74	1211
Processed meat products 10-20 g fat	75	1172
Processed meat products 20-30 g fat	76	1171
Processed meat products >30 g fat	77	1151
Milk en milkproducts		
Milk chocolate-flavoured low fat	78	273
Milk chocolate-flavoured full fat	79	272
Milk skimmed	80	294
Milk semi-skimmed	81	286
Milk whole	82	279
Buttermilk	83	289
Custard vanilla low fat	84	477
Custard half fat all flavours	85	2519
Custard several flavours full fat	86	1720
Yoghurt low fat with fruit	87	284
Yoghurt low fat	88	301
Yoghurt half fat	89	1502
Yoghurt full fat	90	278
Ice cream dairy cream based	91	303
Nuts, seeds and snacks		
Nuts mixed unsalted	92	207
Nuts mixed salted	93	1935
Mixed nuts and raisins	94	205
Peanuts salted	95	876
Peanuts unsalted	96	204

Peanuts coated	97	546
Crisps potato average	98	122
Japanese rice cracker mix wo peanuts	99	2147
Snack sausage roll puff pastry	100	266
Biscuit salted average	101	264
Salad egg	102	1499
Salad cucumber	103	1876
Salad meat	104	1877
Salad chicken curry	105	1498
Salad ham and leek	106	1497
Salad fish	107	1496
Pastry, cake and biscuits		
Biscuit brown/wholemeal	108	263
Biscuit muesli	109	636
Biscuit sweet	110	252
Biscuits averaged	111	258
Cake Dutch spiced Ontbijtkoek	112	240
Cake Dutch spiced Ontbijtkoek wholemeal	113	925
Cake sponge Dutch Eierkoek	114	254
Muesli bar	115	2239
Apple pie Dutch w shortbread w marg	116	251
Cake made with butter	117	1969
Flan with fruit filling	118	486
Gateau with butter-cream filling	119	256
Gateau with whipped cream	120	255
Potatoes		
Potatoes wo skins boiled average	121	982
Potatoes fried	122	1457
Savoury sandwich filling		
Peanut butter	123	455
Sandwich spread original	124	575
Sugar, confectionary, sweet prod.		
Chocolate flakes average	125	2531
Chocolate confetti averaged	126	1311
Coloured confetti fruit-flavoured	127	442
Coconut bread sweetened sliced	128	449
Honey	129	443
Jam without sugar	130	807
Jam reduced sugar	131	484
Jam	132	445
Spread chocolate hazelnut	133	436
Syrup Keukenstroop	134	378
Syrup apple	135	427
M&M's chocolate with peanuts	136	621
Candybar Mars	137	487
Chocolate bar milk with nuts	138	717
Chocolate plain	139	432
Chocolate milk	140	431
Sugar granulated	141	377
Boiled sweets	142	450
Vegetables		
Vegetables average boiled	143	1904
Vegetables mixture raw	144	127

Chapter 5

Combining equity and utilitarianism

Additional insights into a novel approach

Based on:

Gerdessen, JC, Kanellopoulos, A, Claassen, GDH
'Combining equity and utilitarianism' – Additional insights into a novel approach
DOI: 10.1111/itor.12415

Abstract

The diet model addressed Chapter 4 uses an extended goal programming (EGP) achievement function, which combines an achievement function in which a (weighted) sum of utilities is maximised and an achievement function in which the lowest value within a set of utilities is maximised. The combination requires weighting. The problem is how to justify and interpret weights. Considering the key importance of preference structure (as implemented via achievement functions and weights) for the quality of decision-making in MCDM, it is crucial to keep an open mind for novel approaches, investigate their behaviour, and assess their added value for practice. Chapter 5 explores such a novel approach (referred to as CEU) proposed by Hooker and Williams (2012) for *c*ombining *e*quity (comparable with the MaxMin achievement function in Chapter 4) and *u*tilitarianism (comparable with the MaxSum achievement function in Chapter 4) in a single model that does not require specification of a set of weights, but uses a single parameter. So far, CEU has not been evaluated and used in a practical context besides that of Hooker and Williams (2012). In order to assess its added value for diet modelling, it is important to gain insights into its properties and (dis)advantages.

Chapter 5 addresses research question 4 "What is the added value of a novel method for finding a compromise between total utility and lowest utility in the context of diet models?". It provides new insights into CEU and assesses its added value for practice by comparing it with EGP. Chapter 5 shows that CEU balances between equity and utilitarianism in a way that is basically different from using a convex combination of these two criteria, as is done in EGP. Moreover, CEU's parameter has an intuitive interpretation. The set of solutions generated by CEU is smaller and wider spaced than EGP's set of solutions, which can be an advantage for the decision-maker. CEU generates solutions on the Pareto front of the decision-maker's *n*-criteria problem. However, CEU's way of balancing equity and utilitarianism causes a (small) distance to the Pareto front of the associated bi-criteria problem on the aggregate criteria of lowest utility (representing equity) and total utility (representing utilitarianism). Reporting this distance will support the decision-maker to assess whether the achieved balance is worth its price.

Chapter 5 concludes that for applying CEU the investigated decision problem should have the following two characteristics: (i) summing individual utility values is meaningful, (ii) small increments of utilities within a predefined (commonly small) range from the lowest utility do not affect the decision-maker's perceived quality of a solution. The diet problems addressed in this thesis will often not have these characteristics, which limits the applicability of CEU for diet modelling.

5.1 Introduction

Hooker and Williams (2012) introduce a novel approach to combine the often conflicting criteria of equity and utilitarianism in a single mathematical model (from here on referred to as CEU). To the best of our knowledge, no follow-up research has been reported that compares CEU with a commonly used approach for combining equity and utilitarianism. Neither have any additional insights been generated regarding the properties and (dis)advantages of CEU. This paper provides additional insights into CEU by comparing it with a commonly used goal programming approach in a practical context.

Goal programming can be regarded as one of the most widely used multi-criteria decision-making techniques (Caballero et al., 2009), and is often cited as the "work horse" of multi objective optimisation (Romero, 2004, Charnes and Cooper, 1977). Romero (2001) introduces extended goal programming (EGP) to combine the often conflicting criteria of equity and utilitarianism. The achievement function of EGP balances between equity and utilitarianism by minimising a convex combination of the maximum deviation from a goal (equity) and the weighted sum of deviations from all goals (utilitarianism). A dimensionless parameter λ is used to control the importance of equity and utilitarianism in the achievement function. Romero (2004) states that the right choice of the achievement function is a key element for the success of a goal programming model; if the chosen achievement function is wrong then it is very likely that the decision-maker will not accept the solution. Romero (2004) also points out that determination of the precise value of parameter λ is a key question that is still open. The problem is how to justify and interpret any particular value of λ (Hooker and Williams, 2012). Jones and Jimenez (2013) recommend to carry out sensitivity or parametric analysis on λ. However, the problem of determining the appropriate value of λ remains unsolved.

Hooker and Williams (2012) introduce a novel approach to combine equity and utilitarianism (CEU) in a single mathematical model. For formulating equity, CEU uses the maximin principle of Rawls (1971), that is "one seeks to allocate goods so as to maximise the welfare of the worst off". Hooker and Williams (2012) argue that most people regard it as unreasonable to take a Rawlsian policy to its extreme. Hence, the authors propose to switch from a Rawlsian to a utilitarian criterion when inequality exceeds a threshold value Δ. This has the effect of adhering to a Rawlsian criterion unless the decrease in total utility is too great (Hooker and Williams, 2012). Hooker and Williams (2012) describe CEU's parameter Δ as the level of inequality at which efficiency considerations take over. So, both EGP and CEU are single-parameter approaches for combining the conflicting criteria of equity and utilitarianism. Therefore, we investigate CEU by comparing it with EGP.

This paper aims to contribute to the insight into the properties and (dis) advantages of CEU and to assess its added value for practice by comparing it with EGP. The comparison comprises the way of balancing equity and utilitarianism, method-specific parameters and their interpretation, location of the solutions, number of solutions, differences between neighbouring solutions, discrete versus continuous nature of the methods, distance to the Pareto front, and computational effort. As a practical case, we use a diet modelling problem. In order to provide a self-contained paper, Section 2 summarises CEU and Section 3 outlines the core of the diet modelling case.

5.2 Combing Equity and Utilitarianism (CEU)

Hooker and Williams (2012) suppose that a population consists of individuals (or classes of individuals), and that the decisions result in an allocation of utilities u_1, u_2, …, u_n to these individuals. The authors present a model that maximises the utility of the worst off – i.e. maximise($\min_i\{u_i\}$) – unless this takes too many resources from the others. For the two-person case the switch from a Rawlsian[1] to a utilitarian criterion takes place when inequality between u_1 and u_2 exceeds a threshold Δ: that is, when $|u_1 - u_2| \geq \Delta$. The contours of the aggregated welfare function are shown in Figure 5.1. When $|u_1 - u_2| \leq \Delta$, the contours reflect the Rawlsian criterion $\min\{u_1, u_2\}$. Otherwise, the contours reflect the utilitarian criterion $u_1 + u_2$.

The Rawlsian solution R (denoted with ○ in Figure 5.1) allocates equal utility to each person. From there, shifting resources from person 1 to person 2 would considerably improve u_2 at the expense of a slight decrease in u_1. Thus, switching to the utilitarian solution Q (denoted with ● in Figure 5.1), would improve total utility. Parameter Δ is the maximum sacrifice of total utility one is willing to make in order to maximise the utility of the worst off. It is the level of inequality at which efficiency considerations take over (Hooker and Williams, 2012). Figure 5.2

[1] Here, we follow the terminology of Hooker and Williams (2012), who denote $\min_i\{u_i\}$ as the Rawlsian criterion.

illustrates that for small values of Δ (Δ_1) the model will tend to select the utilitarian solution *Q*, whereas for large Δ (Δ_2) the Rawlsian solution *R* will be preferred.

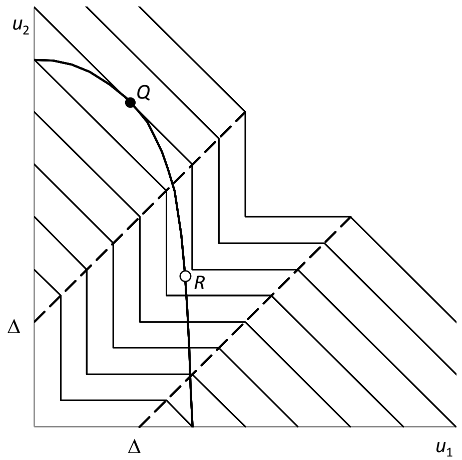

Figure 5.1 Contours of the aggregated welfare function. The diagonal part of the contours corresponds to a utilitarian criterion, and the L-shaped part corresponds to a Rawlsian criterion. The curve reflects resource limits (Hooker and Williams, 2012).

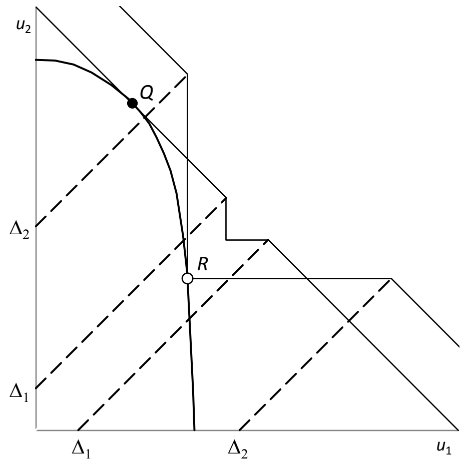

Figure 5.2 For small values of Δ (Δ_1) the utilitarian solution *Q* (●) is preferred, and for large values of Δ (Δ_2) the Rawlsian solution *R* (○) is preferred.

Two-person problem

The two-person problem aims to allocate utilities u_1 and u_2 ($u_1, u_2 \geq 0$) such that for $|u_1 - u_2| \leq \Delta$ the Rawlsian criterion $\min\{u_1, u_2\}$ is maximised, and for $|u_1 - u_2| \geq \Delta$ the utilitarian criterion $u_1 + u_2$. To ensure continuity of the welfare function, in the Rawlsian case $2\min\{u_1, u_2\} + \Delta$ is used. The optimisation problem is to maximise z subject to

$$z \leq \begin{cases} 2\min\{u_1, u_2\} + \Delta & \text{if } |u_1 - u_2| \leq \Delta, \\ u_1 + u_2 & \text{otherwise} \end{cases} \tag{1}$$

$$(u_1, u_2) \in U$$

in which U defines the feasible set. Welfare function (1) switches from a Rawlsian to a utilitarian criterion when the inequality between u_1 and u_2 exceeds the threshold Δ. Hooker and Williams (2012) present the following MILP formulation for problem (1):

$$\text{maximise } \{ z \} \tag{2}$$
s.t.
$$z \leq 2u_i + \Delta + (M - \Delta)\delta \tag{3}$$
$$z \leq u_1 + u_2 + \Delta(1 - \delta) \tag{4}$$
$$u_1 - u_2 \leq M \tag{5}$$
$$u_1 - u_2 \leq M \tag{6}$$
$$u_1, u_2 \geq 0, \quad \delta \in \{0, 1\} \tag{7}$$

If $|u_1 - u_2| \leq \Delta$ then $\delta = 0$ and if $|u_1 - u_2| \geq \Delta$ then $\delta = 1$.

In Figure 5.3, the Rawlsian solution R and the utilitarian solution Q are on the same contour, which implies that welfare function (1) is indifferent between R and Q:

$$2\min\{u_1^R; u_2^R\} + \Delta = u_1^Q + u_2^Q \tag{8}$$

As $u_1^R = u_2^R$, this can be written as

$$u_1^R + u_2^R + \Delta = u_1^Q + u_2^Q \tag{9}$$

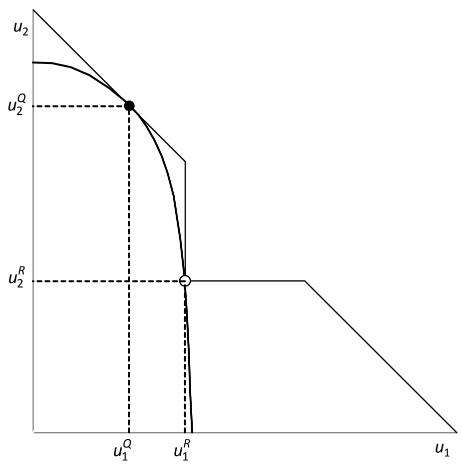

Figure 5.3 The Rawlsian solution R (○) and the utilitarian solution Q (●) are on the same contour of the welfare function.

So, switching from the utilitarian solution Q to the Rawlsian solution R incurs a loss of Δ in total utility. A decision-maker who is willing to make this sacrifice or more (that is, who uses a large value of Δ) will prefer the Rawlsian solution R, and a decision-maker who is not willing to make this sacrifice (that is, who uses a smaller value of Δ) will prefer the utilitarian solution Q, see Figure 5.2.

Many-person problem

In the many-person problem, utilities $u_1, u_2, \ldots, u_n \geq 0$ have to be allocated to n persons. For formulating the welfare function of the many-person problem, the two-person welfare function (1) is rewritten as

$$z \leq \Delta + 2u_{\min} + \max\{0, u_1 - u_{\min} - \Delta\} + \max\{0, u_2 - u_{\min} - \Delta\} \tag{10}$$

and then generalised for the many-person problem:

$$z \leq (n-1)\Delta + nu_{\min} + \sum_{i=1}^{n} \max\{0, u_i - u_{\min} - \Delta\} \tag{11}$$

with $u_{\min} = \min_i\{u_i\}$. Thus, person i makes a utilitarian contribution to the welfare function if $u_i - u_{\min} > \Delta$, and is otherwise represented by $u_{\min}$. We will refer to utilities that make a utilitarian contribution as utilitarian contributors, and to the others as Rawlsian contributors. Hooker and Williams (2012) present the following MILP model for the many-person problem:

$$\text{maximise } \{ z \} \tag{12}$$

s.t.

$$z \leq (n-1)\Delta + \sum_{i=1}^{n} v_i \tag{13}$$

$$u_i - \Delta \leq v_i \leq u_i - \Delta\delta_i \qquad \forall i \tag{14}$$

$$w \leq v_i \leq w + (M - \Delta)\delta_i \qquad \forall i \tag{15}$$

$$u_i, v_i \geq 0, \quad \delta_i \in \{0,1\} \qquad \forall i \tag{16}$$

$$w, z \geq 0 \tag{17}$$

In the optimal solution, w will take the value $u_{\min}$ and

$$(\delta_i, v_i) = \begin{cases} (0, w) & \text{if } u_i - u_{\min} < \Delta \\ (1, u_i - \Delta) & \text{otherwise} \end{cases} \tag{18}$$

5.3 Case study: an EGP-based diet model

Mathematical modelling of diets can be defined as the use of mathematical techniques to formulate and optimise diets (Buttriss et al., 2014). Commonly, the decision variables are defined as: x_j = (proposed) daily intake of food j (j = 1...J), $x_j \geq 0$. Acceptability for the consumer is assured by palatability constraints, which define upper and lower bounds on intake of various foods (e.g. "no more than 5 slices of bread", "at least one sweet snack"), and link intakes of foods (e.g. "3-7g spread per slice of bread"). In order to be nutritious, the generated diets should meet restrictions on the amounts of nutrients (for example carbohydrates, iron). Gerdessen and De Vries (2015) present a EGP-based diet model in which the nutritional constraints are formulated via nutrient utility curves[2], see Figure 5.4.

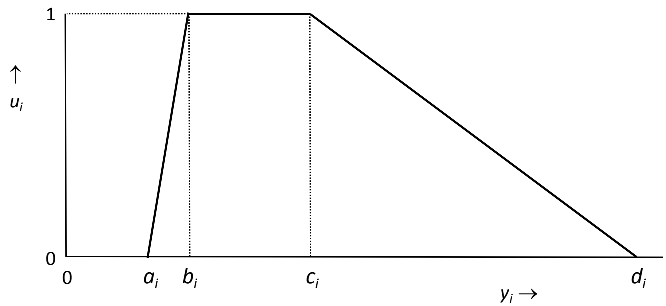

Figure 5.4 Utility curve for intake y_i of nutrient i. Intakes between b_i and c_i are considered as fully adequate: $u_i = 1$. Intake a_i and intake d_i are considered as fully inadequate: $u_i = 0$. Intakes lower than a_i or higher than d_i are not allowed.

In the utility curve of Figure 5.4, intakes between b_i and c_i are considered as fully adequate: the utility for nutrient i is $u_i = 1$. Intakes a_i and d_i are considered as fully inadequate: $u_i = 0$. The model does not allow intakes lower than a_i or higher than d_i. Then, diets are searched that maximise

(i) a Rawlsian criterion: $\max(u_{\min})$ with $u_{\min} = \min_i\{u_i\}$,

(ii) a utilitarian criterion: $\max\{\Sigma_i u_i\}$,

(iii) a convex combination of (i) and (ii): $\max\left((1-\lambda)\cdot\sum_i u_i + \lambda\cdot u_{\min}\right)$, $0 \leq \lambda \leq 1$, thus balancing between equity and utilitarianism.

The decisions x_j (j = 1, …, J) result in an allocation of utilities $u_1, u_2, \ldots, u_n$ to nutrients i (i = 1, …, n). So, the nutrients in the diet model equal the persons of the CEU models. In order to apply CEU, the diet model is extended with the MILP

[2] Note that in Chapter 4 these are called 'nutrient adequacy curves'.

formulation of the many-person problem, i.e. (12) – (17). From here onward, it will be referred to as the *many-nutrient* problem.

5.4 Experiment and observations

EGP and CEU are applied on a diet model for a case of women aged 19-30 yrs (Gerdessen and De Vries, 2015). The model contains 144 foods and optimises 21 utilities. Figure 5.5 shows the Pareto front of the total utility (horizontal axis) and the lowest utility (vertical axis) plus all solutions that are found by:

- *EGP* with achievement function: $\max\left((1-\lambda)\sum_i u_i + \lambda u_{\min} + \varepsilon_1 \sum_i u_i + \varepsilon_2 u_{\min}\right)$, denoted by (●). The tie-breaking terms $\varepsilon_1 \sum_i u_i$ and $\varepsilon_2 u_{\min}$ (with ε_1, ε_2 very small positive numbers) are added to prevent that non-efficient solutions are generated.

- *the many-nutrient CEU model*, with a tie-breaking term $\varepsilon_3 \sum_i u_i$ (with ε_3 a very small positive number) added to the right-hand side of inequality (13), denoted with (×).

Tables 5.1 and 5.2 list the utilities of the 31 solutions generated by EGP and the 7 solutions generated by CEU. Associated food intakes x_j are given in Appendix 5A.

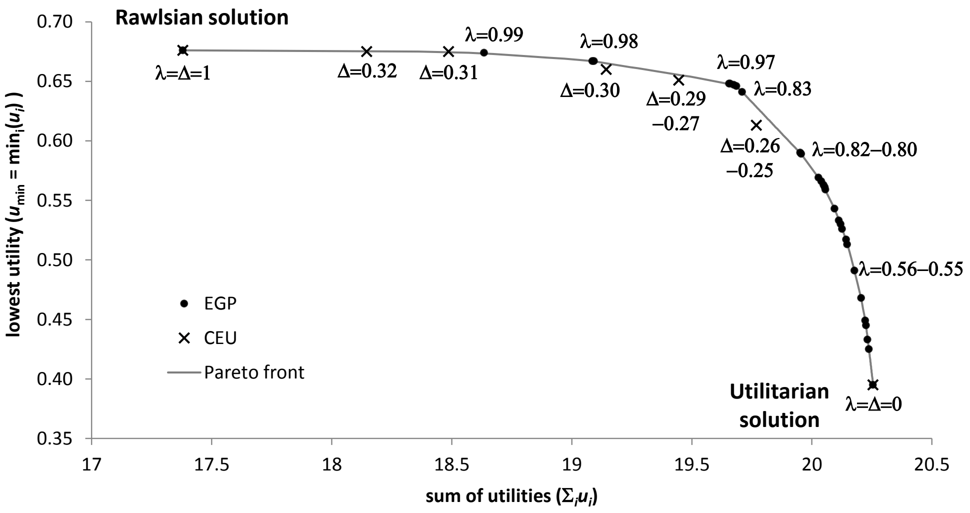

Figure 5.5 Total utility and lowest utility of all solutions found with EGP and CEU.

Table 5.1 Nutrient utilities u_i ($i = 1..21$) of all diets found with EGP. For instance, for $0.84 \leq \lambda \leq 0.91$ a diet with $u_{13} = 0.936$ is found. The shaded column for $0 \leq \lambda \leq 0.36$ is the utilitarian solution and the shaded column for $\lambda = 1$ is the Rawlsian solution.

	λ																														
i	0 0.36	0.37	0.38	0.39 0.42	0.43 0.47	0.48 0.54	0.55 0.56	0.57 0.61	0.62 0.63	0.64	0.65	0.66	0.67 0.68	0.69 0.72	0.73	0.74	0.75 0.77	0.78	0.79	0.80 0.81	0.82	0.83	0.84 0.91	0.92	0.93 0.94	0.95	0.96	0.97	0.98	0.99	1
1.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.674	0.676
2.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.864	0.676
3.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.967
4.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.915
6.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.676
7.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.676
8.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.676
9.	0.395	0.425	0.433	0.445	0.449	0.468	0.491	0.513	0.517	0.526	0.530	0.533	0.543	0.559	0.561	0.563	0.566	0.569	0.589	0.590	0.590	0.641	0.646	0.646	0.647	0.648	0.648	0.667	0.667	0.674	0.676
10.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.993	0.995	0.992
12.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.667	0.667	0.674	0.676
13.	1	1	1	1	1	1	1	1	1	1	1	1	1	0.980	0.980	0.978	0.973	0.965	0.938	0.936	0.936	0.939	0.936	0.935	0.929	0.916	0.912	0.943	0.943	0.919	0.931
14.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.998
15.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16.	1	1	1	1	1	1	1	1	0.999	0.999	0.999	0.999	0.998	0.993	0.993	0.992	0.989	0.987	0.982	0.981	0.981	0.981	0.98	0.98	0.977	0.975	0.975	0.978	0.978	0.973	0.975
17.	0.951	0.913	0.900	0.884	0.876	0.864	0.812	0.739	0.729	0.703	0.693	0.683	0.656	0.633	0.629	0.627	0.626	0.632	0.640	0.639	0.639	0.641	0.646	0.646	0.647	0.648	0.648	0.667	0.667	0.674	0.676
18.	0.909	0.900	0.899	0.897	0.898	0.874	0.875	0.895	0.898	0.898	0.898	0.898	0.898	0.892	0.891	0.889	0.886	0.875	0.807	0.806	0.806	0.830	0.832	0.832	0.828	0.825	0.825	0.837	0.837	0.840	0.842
19.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.667	0.667	0.674	0.676
21.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.678	0.646	0.646	0.647	0.648	0.648	0.667	0.667	0.674	0.676
$u_{\min}$	0.395	0.425	0.433	0.445	0.449	0.468	0.491	0.513	0.517	0.526	0.530	0.533	0.543	0.559	0.561	0.563	0.566	0.569	0.589	0.590	0.590	0.641	0.646	0.646	0.647	0.648	0.648	0.667	0.667	0.674	0.676
u_{sum}	20.26	20.24	20.23	20.23	20.22	20.21	20.18	20.15	20.14	20.13	20.12	20.11	20.10	20.06	20.05	20.05	20.04	20.03	19.96	19.95	19.95	19.71	19.69	19.69	19.68	19.66	19.66	19.09	19.09	18.64	17.38

Table 5.2 Nutrient utilities u_i (i = 1..21) of all diets found with CEU. Rawlsian contributors are shown in bold font. For example, for Δ = 0.28 we have u_{min} = 0.651, so all utilities with u_i < 0.651 + 0.280 = 0.931 are Rawlsian contributors.

	Utilitarian solution			**Δ**			**Rawlsian solution**
Nutrients *i*	**0-0.24**	**0.25-0.26**	**0.27-0.29**	**0.30**	**0.31**	**0.32**	**0.33-1**
1. Protein	1	1	1	1	**0.691**	**0.675**	**0.676**
2. Total fat	1	1	1	1	**0.675**	**0.675**	**0.676**
3. SFA[1]	1	1	1	1	1	1	**0.967**
4. MUFA[2]	1	1	1	1	1	1	**1**
5. PUFA[3]	1	1	1	1	1	1	**0.915**
6. Linol	1	1	1	1	1	1	**0.676**
7. EPA+DHA[4]	1	1	1	1	1	**0.675**	**0.676**
8. Cholesterol	1	1	1	1	1	1	**0.676**
9. Mono-/disacch.	**0.395**	**0.613**	**0.651**	**0.660**	**0.675**	**0.675**	**0.676**
10. Dietary fiber	1	1	1	1	1	1	**1**
11. Calcium	1	1	1	1	1	1	**0.992**
12. Iron	1	1	**0.651**	**0.660**	**0.675**	**0.675**	**0.676**
13. Potassium	1	0.999	1	1	**0.929**	**0.929**	**0.931**
14. Vitamin B_1	1	1	1	1	1	1	**0.998**
15. Vitamin B_2	1	1	1	1	1	1	**1**
16. Vitamin B_6	1	0.997	0.991	0.992	**0.975**	**0.975**	**0.975**
17. Vitamin B_{12}	0.951	**0.613**	**0.651**	**0.660**	**0.675**	**0.675**	**0.676**
18. Vitamin C	0.909	0.934	**0.851**	**0.852**	**0.842**	**0.842**	**0.842**
19. Folate	1	1	1	1	1	1	**1**
20. Vegetables	1	1	1	**0.660**	**0.675**	**0.675**	**0.676**
21. Fruits	1	**0.613**	**0.651**	**0.660**	**0.675**	**0.675**	**0.676**
u_{min}	0.395	0.613	0.651	0.660	0.675	0.675	0.676
$\Sigma_i u_i$	20.255	19.769	19.446	19.144	18.487	18.146	17.380

[1]Saturated fatty acids; [1]Mono-unsaturated fatty acids; [2]Poly-unsaturated fatty acids;
[3]EPA, eicosapentaenoic acid; DHA, docosahexaenoic acid.

Figure 5.5 and Tables 5.1 and 5.2 show that EGP and CEU differ with respect to:

- *Location of the solutions*
  For λ = Δ = 0, both EGP and CEU find the utilitarian solution. For λ = Δ = 1, both approaches find the Rawlsian solution. Apart from these extremes, EGP and CEU find different solutions, which are located rather differently.

- *Differences between neighbouring solutions*
  Neighbouring CEU solutions are more different than neighbouring EGP solutions.

- *Distance to the Pareto front of the bi-criteria problem max*$\{\sum_i u_i\ ;\ u_{\min}\}$
  All EGP solutions are on the Pareto front of the bi-criteria problem max$\{\sum_i u_i\ ;\ u_{\min}\}$. The CEU solutions are located at a small distance from this front.

The following section provides explanations for these observed differences. Also, it discusses other differences between CEU and EGP.

5.5 Analysis and discussion

This section analyses and discusses differences between EGP and CEU. First, the observations from the results are discussed. Next, several other aspects will be discussed. Differences are summarised in Table 5.4.

Location of the solutions

Figure 5.5 shows that in the neighbourhood of the utilitarian solution the EGP solutions are very close together. The first CEU solution, however, is located at a relatively large distance from the utilitarian solution.

The utilitarian solution has $u_{\min} = u_9 = 0.395$. The difference between $u_{\min}$ and the next lowest utility ($u_{18} = 0.909$) is relatively large. Eighteen nutrient intakes y_i are within their optimal range (that is, $b_i \leq y_i \leq c_i$) (see Figure 5.4), and thus have $u_i = 1$. For nutrients with $b_i < y_i < c_i$ (such as protein in Table 5.2), a small change in y_i will not change the utility u_i; the nutrient intake y_i may move towards b_i or c_i, but the resulting utility u_i remains equal to 1. Therefore, moving to vertices with higher $u_{\min}$ will have a minor impact on $\Sigma_i u_i$. As a result, the EGP solutions at the right-hand side of the graph in Figure 5.5 are relatively close together. When moving towards the Rawlsian solution, more and more nutrients get an intake outside $[b_i; c_i]$, which implies they have $u_i < 1$ (see for instance vitamin B_6 in Table 5.2). For those nutrients, any change in y_i will have an immediate impact on u_i. As a result, for this diet model the EGP solutions near the Rawlsian solution are wider spaced than those near the utilitarian solution. We performed a test in which $b_i = c_i$ (for i in 1, …, n). The outcome confirms the analysis, that is, the vertices near the utilitarian solution are no longer close together, see Figure 5.6.

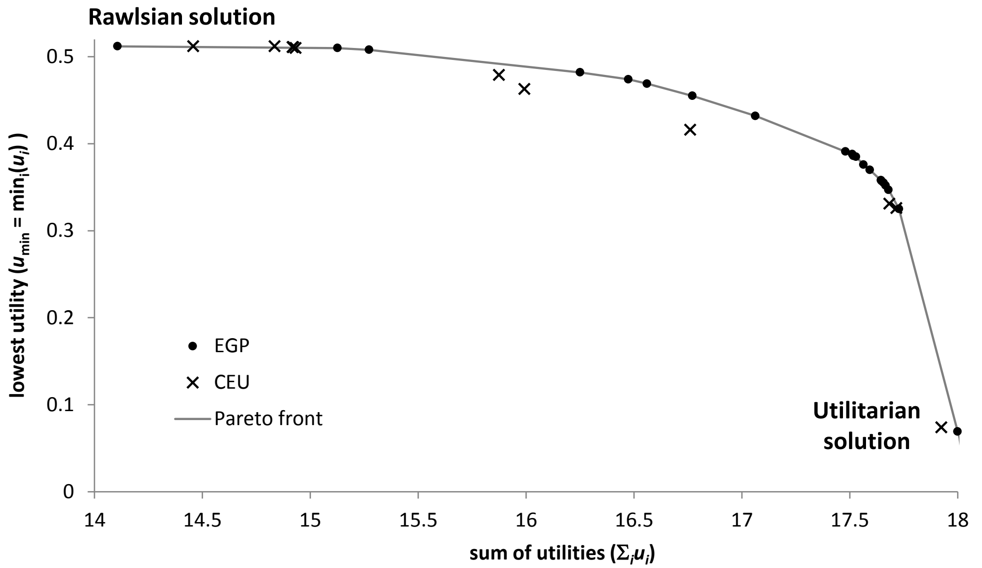

Figure 5.6 Total utility and lowest utility of all solutions found with EGP and CEU for a (fictitious) diet model with triangular utility curves. For mono-di-saccharides: $b_i = c_i = 5\%$. For dietary fibre: $b_i = c_i = 3$ gr/MJ. For all other nutrients, the average of b_i and c_i (see Table 4.5) was used. The curves do not reflect a realistic diet model, but are used to investigate the impact of the shape of the utility curves on the number of EGP solutions near the utilitarian solution.

The CEU is an MILP model: binary variables δ_i indicate whether nutrient i makes a utilitarian contribution to the objective function value ($\delta_i = 1$ if $u_i - u_{\min} > \Delta$) or not ($\delta_i = 0$ if $u_i - u_{\min} \leq \Delta$). In the utilitarian solution ($\Delta = 0$), all nutrients except the worst off $(u_{\min} = u_9)$ make a utilitarian contribution. Moving into the direction of the Rawlsian solution (i.e. putting more emphasis on maximising $u_{\min}$ by increasing Δ) implies that at least one nutrient will change from utilitarian contributor ($\delta_i = 1$) into Rawlsian contributor ($\delta_i = 0$). Nutrient i can only become a Rawlsian contributor if $u_i - u_{\min} \leq \Delta$. Because in the utilitarian solution the lowest utility ($u_{\min} = u_9 = 0.395$) is substantially lower than all other utilities, it takes a relatively large increase in $u_{\min}$ (which implies a relatively large decrease in $\Sigma_i u_i$) and a substantial value of Δ to achieve $u_i - u_{\min} \leq \Delta$ and move away from the utilitarian solution. As a result, for this dataset no CEU solutions are found at the right-hand side of the graph. For the Rawlsian solution, the lowest utility is $u_{\min} = 0.676$, so the largest possible difference between any u_i and $u_{\min}$ is $1 - 0.676 = 0.324$. Therefore, for all $\Delta \geq 0.324$ the Rawlsian solution is found.

We conclude that the difference between EGP and CEU in the location of the solutions can be explained by the shape of the utility curves and by the discrete nature of CEU.

Differences between neighbouring solutions

Tables 5.1 and 5.2 show that in any pair of neighbouring CEU solutions one or more utilities are very different. Neighbouring EGP solutions, however, commonly hardly differ.

CEU is a discrete approach: every utility value is either a utilitarian ($\delta_i = 1$) or a Rawlsian contributor ($\delta_i = 0$). When moving to the next solution (by increasing Δ), one or more u_is will change from utilitarian contributor into Rawlsian contributor. As soon as a nutrient becomes a Rawlsian contributor ($u_i - u_{\min} \leq \Delta$), the value of maximise$\{0,\ u_i - u_{\min} - \Delta\}$ in expression (11) becomes zero, irrespective of the actual value of u_i. As a result, the model tends to give many Rawlsian contributors an utility value that is equal to the utility of the worst off ($u_i = u_{\min}$), which means that one or more nutrients has a utility that is very different from the utility of its neighbour. This can be observed when Δ increases to $\Delta = 0.25$. For $\Delta = 0.25$, utilities u_{17} and u_{21} become Rawlsian contributors and both utility values decrease to $u_{17} = u_{21} = u_{\min} = 0.613$. For $\Delta = 0.30$, utility u_{20} becomes a Rawlsian contributor and its value decreases to $u_{20} = 0.660$. For $\Delta = 0.32$, utility u_7 becomes a Rawlsian contributor and its value decreases to $u_7 = 0.675$.
EGP is a continuous approach. Moving from one solution to the next (by increasing λ) implies moving from one vertex to a neighbouring vertex, which often causes only minor changes in u_i. Three exceptions are $\lambda = 0.83$, $\lambda = 0.97$, $\lambda = 0.99$, where u_{21} resp u_{20} resp u_1 decrease substantially. In the large set of EGP solutions, these could be among the most interesting alternatives for the decision-maker, because at least one of their utilities truly differs from the neighbouring solution.

In conclusion, CEU shows larger differences between neighbouring solutions than EGP because CEU is a discrete approach and EGP is a continuous approach.

Distance to the Pareto front

EGP searches efficient solutions to the n-criteria problem maximise$\{u_1;\ u_2;\ \ldots\ ;\ u_n\}$ by maximising a weighted sum of the aggregate criteria $\sum_i u_i$ and $u_{\min}$. Therefore, all solutions found with EGP are on the Pareto front of the bi-criteria problem maximise$\{\sum_i u_i \,;\, u_{\min}\}$.

For $0 < \Delta < 1$, the solutions found with CEU are not on the Pareto front of the bi-criteria problem maximise$\{\sum_i u_i \,;\, u_{\min}\}$. This is caused by the fact that CEU's welfare function can be maximised by a solution that is dominated in terms of $\sum_i u_i$ and $u_{\min}$. This is elaborated in Appendix 5B. From the point of view of the decision-maker, it might not be strictly necessary that every generated solution is on the Pareto front of the bi-criteria problem. After all, the decision-maker faces the *n*-criteria problem from which the aggregate criteria $\sum_i u_i$ and $u_{\min}$ are derived, and every CEU solution is on the Pareto front of this *n*-criteria problem. Moreover, it is balanced. In order to support the decision-maker in judging whether or not the achieved balance outweighs the distance to the Pareto front of the bi-criteria problem, we calculate the relative distance of each CEU solution to its radial projection on the Pareto front, see Table 5.3. No CEU solution has a relative distance of more than 2%.

In conclusion, solutions generated by CEU are on the Pareto front of the decision-maker's *n*-criteria problem, but not necessarily on the Pareto front of the associated bi-criteria problem. The decision-maker might not perceive a small distance as problematic.

Table 5.3 Relative distance of CEU solutions to their radial projection on the Pareto front of the bi-criteria problem maximise$\{\sum_i u_i \,;\, u_{\min}\}$.

	Solutions found with CEU		**Radial projection* on Pareto front**		**Relative distance**
Δ	**$\Sigma_i u_i$**	**$u_{\min}$**	**$\Sigma_i u_i$**	**$u_{\min}$**	
0.25 - 0.26	19.770	0.613	19.820	0.618	2.0%
0.27 - 0.29	19.445	0.651	19.476	0.655	1.5%
0.30	19.144	0.660	19.172	0.664	1.6%
0.31	18.484	0.675	18.484	0.675	0.0%
0.32	18.149	0.675	18.149	0.675	0.1%

*from point (17.380; 0.395), which combines the lowest values for the Rawlsian and utilitarian criterion.

Number of solutions

In the investigated diet model, CEU finds substantially fewer solutions than EGP, and the solutions are commonly wider spaced. This is explained by the discrete nature of CEU. In practice, a decision-maker might prefer a relatively small set of reasonably different solutions to a large set of rather similar solutions.

Parameters of EGP and CEU

A basic difference between EGP and CEU is in the associated parameters. EGP's parameter λ is a dimensionless control parameter expressing the relative importance of the Rawlsian criterion and the utilitarian criterion. The convex combination in the achievement function of EGP expresses that one unit decrease in $u_{\min}$ can be compensated by $(1-\lambda)/\lambda$ units of increase in $\Sigma_i u_i$, irrespective of the current level of $u_{\min}$ or $\Sigma_i u_i$. Determination of the right value of λ is not easy. Jones and Jimenez (2013) advise to carry out a sensitivity or parametric analysis on λ.

CEU's parameter Δ is measured in the same units as the Rawlsian and the utilitarian criterion. In a two-person problem, it does have an intuitive meaning for the decision-maker in practice, that is, it quantifies maximum sacrifice of total utility one is willing to make in order to maximise the utility of the worst off, see (10). Expression (11) generalises the two-person problem to a many-person problem. In the many-person problem, parameter Δ still quantifies the maximum acceptable sacrifice for maximising the utility of the worst off for a two-person comparison, namely the comparison (u_i, $u_{\min}$), for all persons i. The decision-maker may perceive this two-person problem as hypothetical, because he does not know *a priori* which person will have the lowest utility, and because the actual decision problem may concern (many) more than two persons. However, the decision-maker does not need to provide a justified value for Δ *a priori*. Running the many-person model for a range of Δ will contribute to the decision process: the observed breakpoints in Δ will help the decision-maker to become aware of his trade-offs and preferences. The value of Δ that results in the solution that the decision-maker prefers, can be selected for further use, thus ensuring that the same policy is applied consistently (Hooker and Williams, 2012).

We conclude that both for EGP and CEU it seems advisable to run the model for a range of values of λ and Δ, respectively. This gives the decision-maker insight into the trade-offs between the nutrient utilities and the trade-offs between the utilitarian and the Rawlsian criterion, and in the impact of parameter choice on the solutions that are found. Moreover, it will provide him with a set of promising solutions.

Computational effort

Using the many-nutrient CEU model requires adding a set of binary variables. Hooker and Williams (2012) prove that their basic model (12) – (17) is sharp. By adding additional constraints, the sharpness may easily be lost. As a result, CEU will generally require a larger computational effort than the continuous EGP. In the diet modelling case, the computation times for the CEU model were still very small. However, in other application areas the computational complexity of the CEU model might become prohibitive.

Table 5.4 Summary of differences between EGP and CEU.

Aspect	EGP	CEU
Parameter	Dimensionless	Same dimension as criteria
Solutions on Pareto front of *n*-criteria problem	Yes	Yes
Solutions on Pareto front of bi-criteria problem	Yes	At a small distance
Nature of the method	Continuous	Discrete
Number of proposed solutions	Large	Small
Differences between solutions	Commonly small	Larger differences
Computational effort	Small	Could become large

Applicability for diet modelling

With respect to the applicability of CEU's welfare function for diet modelling, two issues have to be taken into account: additivity of utility values, and impact of increasing low utilities on diet quality.

Additivity of utility values

CEU's welfare function requires, for a 2-person case, that the decision-maker can quantify how much total utility he is willing to sacrifice in order to maximise the utility of the worst-off. This implies that summing utilities is meaningful for the decision-maker. In the context of Hooker and Williams (2012)'s case, where medical treatments are assigned to groups of patients, this additivity is imaginable: the decision-maker could decide that QALYs gained in one group are – from his position – indistinguishable from QALYs gained in another group, and can therefore be summed. In the context of a diet model, however, the utilities for various nutrients may not be additive, see also Section 6.4.2.

Impact of increasing low utilities on diet quality

CEU's welfare function assigns the value of the lowest utility to each Rawlsian contributor, that is, each person that has a utility within a prespecified bound (Δ) from this lowest utility, see expressions (10) and (11). In the context of a diet model, this might not be appropriate: small increments of low nutrient utilities are likely to have a positive impact on overall quality of a diet. For instance, in column $\Delta = 0.27$ of Table 5.2, vitamin C is a Rawlsian contributor, because its utility ($u_{18} = 0.851$) differs less than Δ from the lowest utility ($u_{\min} = 0.651$). As a result, vitamin C's contribution to the welfare function equals $u_{\min} = 0.651$. So, the welfare function does not reflect that a diet with $u_{18} = 0.851$ may have a larger overall quality than a diet with $u_{18} = 0.651$ (all other u_i's being unchanged).

These two issues limit the applicability of CEU's welfare function for diet modelling. The insights gained in this study, however, are valuable, because they are not limited to a diet modelling case. They may be particularly useful for any researcher who considers using CEU, or who struggles with the explanation and interpretation of its results. With respect to the applicability of CEU, we conclude that the investigated decision problem should have the following two characteristics: (i) summing individual utility values is meaningful, (ii) small increments of utilities within a predefined (commonly small) range from the lowest utility do not affect the overall quality of a solution.

5.6 Conclusions

CEU's way of combining the conflicting criteria of equity and utilitarianism in a single model is basically different from the widely used convex combination of a Rawlsian and a utilitarian criterion in EGP. Moreover, Hooker and Williams (2012) provide an intuitive interpretation for the associated parameter Δ. Because of its discrete nature, CEU generates a set of solutions that is smaller and more widely spaced than that of EGP. This is an advantage for any decision-maker who prefers a relatively small set of reasonably different solutions to a large set of rather similar solutions. CEU generates balanced solutions that are not found with EGP. Achieving the balance may incur a (small) distance to the Pareto front of the bi-criteria problem maximise$\{\sum_i u_i\ ;\ u_{\min}\}$. However, all CEU solutions are on the Pareto front of the n-criteria problem maximise$\{u_1;\ u_2;\ \ldots\ ;\ u_n\}$. We suggest that for use in practical cases this distance is reported. This supports the decision-maker to assess whether the achieved balance is worth its price. Because of the presence of binary variables, CEU may require a larger computational effort than EGP. Future research will show whether current observations hold to the same extent for other cases.

Appendix 5A Food intakes x_j

Food intakes x_j in several solutions generated with EGP and CEU, expressed in grams per day (j in 1, …, J) in diets generated with EGP (λ) and CEU (Δ).

	Rawls. sol.	**CEU**	**EGP**	**CEU**	**CEU**	**EGP**	**EGP**	**CEU**	**EGP**	**Util. sol.**
λ	**1**	**–**	**0.99**	**–**	**–**	**0.97**	**0.83**	**–**	**0.55**	**0**
Δ	**1**	**0.32**	**–**	**0.30**	**0.27**	**–**	**–**	**0.25**	**–**	**0**
Coffee prepared	0	0	0	0	0	0	0	600	600	600
Tea prepared	600	600	600	600	600	600	600	0	0	0
Tomato juice	300	300	300	300	300	300	300	0	118	0
Beer alcohol free <0,1 vol%	0	0	0	0	0	0	0	300	182	300
Water 50-100 mg calcium per litre	600	600	600	600	600	600	600	600	600	600
Bread wholemeal average	105	105	105	105	105	105	105	118	200	210
Pasta wholemeal boiled	78	81	72	107	100	71	107	107	107	107
Rice brown boiled	0	0	0	0	0	0	0	71	71	71
Rice white boiled	101	98	107	71	79	107	71	0	0	0
Cheese 48+ less salt average	0	3	0	2	3	3	8	25	0	0
Cheese 30+ average	23	20	23	21	20	19	15	0	19	0
Eggs chicken boiled average	28	18	19	20	21	20	23	21	32	37
Margarine liq 80% fat <17g sat	0	0	0	0	0	10	0	19	0	0
Salad dressing vinaigrette	0	0	0	5	9	0	0	14	0	0
Oil olive	19	20	17	22	23	8	16	11	17	0
Oil sunflower seed	11	10	13	8	7	13	14	0	9	7
Low fat margarine 40% fat <17g sat	8	3	21	0	0	0	19	0	40	1
Low fat margarine 35% fat <10g sat	0	0	0	0	0	12	0	24	0	17
Margarine product 60% fat <17g sat	0	13	0	21	21	0	0	0	0	0
Margarine product 70% fat >17g sat	13	5	0	0	0	0	0	0	0	0
Fish fat > 10 g fat raw	13	13	14	14	14	14	14	14	14	14
Fruit fresh citrus average	134	134	134	133	133	133	134	131	150	150
Lentils boiled	71	71	71	71	71	71	71	71	71	71
Beef <10% fat prepared	102	109	107	110	116	105	128	128	51	0
Pork <10% fat prepared	60	57	57	1	12	24	0	0	0	32
Pork 10-19% fat prepared	0	0	0	18	0	0	0	0	0	0
Processed meat products 10-20g fat	15	15	15	15	15	15	15	19	29	30
Yoghurt full fat	150	150	150	150	150	150	150	150	150	150
Nuts mixed unsalted	0	0	0	0	0	0	0	0	0	32
Crisps potato average	62	63	70	28	24	67	68	35	25	26
Japanese rice cracker wo peanuts	0	0	0	38	38	5	16	25	0	0
Snack sausage roll puff pastry	0	0	0	14	18	28	0	2	0	0
Peanut butter	0	0	0	0	0	0	0	0	16	30
Vegetables average boiled	184	184	184	166	171	183	200	155	200	200
Vegetables mixture raw	0	0	0	17	29	0	0	45	0	0

Appendix 5B Behaviour of CEU's welfare function

This appendix presents a small-scale numerical example that demonstrates that CEU's welfare function can be maximised by a solution that is not on the Pareto front of the bi-criteria problem maximise $\{\Sigma_i u_i \,;\, u_{\min}\}$.

1. Numerical example – analysis without using the welfare function

Consider the feasible set U defined by

$$\begin{array}{rcrcrcr} 5u_1 & + & 2u_2 & + & 3u_3 & \leq & 25 \\ u_1 & & & & & \leq & 1 \\ u_1 & + & u_2 & & & \leq & 5 \\ & & & & u_3 & \leq & 6 \end{array}$$

$$u_1, u_2, u_3 \geq 0$$

Without using CEU's welfare function, the following can be observed:

- Maximising total utility ($u_1 + u_2 + u_3$) over U provides the utilitarian solution $\underline{u}^{\mathrm{U}} = (0,5,5)$ with total utility $\Sigma_i u_i = 10$.

- For the Rawlsian objective $\max\{u_{\min}\}$, alternative optimal solutions exist: $u_{\min} = 1$ for any solution with $u_1 = 1$. This is a well-known weakness of the maximin criterion. Using a hierarchical method or adding a tie-breaking term $\varepsilon \cdot \Sigma_i u_i$ helps to steer away from non-efficient solutions and select the (only) efficient solution within the set of alternative optimal solutions:

  $\max\left\{u_{\min} + \varepsilon \cdot \sum_i u_i\right\}$ with $\varepsilon > 0$ and small (e.g. $\varepsilon = 0.001$)

  provides the Rawlsian solution: $\underline{u}^{\mathrm{R}} = (1,4,4)$, $u_{\min} = 1$ and total utility $\Sigma_i u_i = 9$.

- Solution (1,1,6) has $u_{\min} = 1$ and $\Sigma_i u_i = 8$. It is therefore dominated by the Rawlsian solution. We will refer to it as the dominated solution $\underline{u}^{\mathrm{D}} = (1,1,6)$.

2. Behaviour of CEU's welfare function

For analysing the behaviour of CEU's welfare function z, we take inequality (11) as a starting point. Welfare function z has to be maximised subject to:

$$z \leq (n-1)\Delta + n u_{\min} + \sum_{i=1}^{n} \max\{0, u_i - u_{\min} - \Delta\} \tag{11}$$

There are no other restrictions to z. In the right-hand side of (11), the term $(n-1)\Delta$ is constant. Via the second term, each 'person' i contributes $u_{\min}$ to z. Via the third term, person i makes a utilitarian contribution if u_i differs from $u_{\min}$ more than Δ, that is if $u_i - u_{\min} > \Delta$. This utilitarian contribution equals the amount φ_i by which u_i exceeds $(u_{\min} + \Delta)$:

$$\varphi_i = \max\{0, u_i - u_{\min} - \Delta\} \tag{19}$$

so

$$z = (n-1)\Delta + nu_{\min} + \sum_{i=1}^{n} \varphi_i \tag{20}$$

Table 5B.1 shows the behaviour of welfare function z for Δ = 1, 2, …, 5.

Table 5B.1 Behaviour of welfare function z for Δ = 1, 2, …, 5.

		u_1	u_2	u_3	$u_{\min}$	$\Sigma_i u_i$	$(n-1)\Delta$	$nu_{\min}$	φ_1	φ_2	φ_3	$\Sigma_i \varphi_i$	z
Δ = 1	$\underline{u}^U$	0	5	5	0	10	2	0	0	4	4	8	10
	$\underline{u}^R$	1	4	4	1	9	2	3	0	2	2	4	9
	$\underline{u}^D$	1	1	6	1	8	2	3	0	0	4	4	9
Δ = 2	$\underline{u}^U$	0	5	5	0	10	4	0	0	3	3	6	10
	$\underline{u}^R$	1	4	4	1	9	4	3	0	1	1	2	9
	$\underline{u}^D$	1	1	6	1	8	4	3	0	0	3	3	10
Δ = 3	$\underline{u}^U$	0	5	5	0	10	6	0	0	2	2	4	10
	$\underline{u}^R$	1	4	4	1	9	6	3	0	0	0	0	9
	$\underline{u}^D$	1	1	6	1	8	6	3	0	0	2	2	11
Δ = 4	$\underline{u}^U$	0	5	5	0	10	8	0	0	1	1	2	10
	$\underline{u}^R$	1	4	4	1	9	8	3	0	0	0	0	11
	$\underline{u}^D$	1	1	6	1	8	8	3	0	0	1	1	12
Δ = 5	$\underline{u}^U$	0	5	5	0	10	10	0	0	0	0	0	10
	$\underline{u}^R$	1	4	4	1	9	10	3	0	0	0	0	13
	$\underline{u}^D$	1	1	6	1	8	10	3	0	0	0	0	13

Examples of entries in Table 5B.1

- Utilitarian solution $\underline{u}^U = (0, 5, 5)$ has $u_{\min} = 0$.
  For Δ = 1, every person that differs from $u_{\min}$ more than Δ makes a utilitarian contribution $\varphi_i = u_i - u_{\min} - \Delta = u_i - 0 - 1$. This holds for persons 2 and 3, so $\varphi_2 = \varphi_3 = 5 - 1 = 4$ and $z = 2\cdot1 + 3\cdot0 + (0+4+4) = 10$.
- Dominated solution $\underline{u}^D = (1, 1, 6)$ has $u_{\min} = 1$.
  For Δ = 3, only person 3 differs more from $u_{\min}$ than Δ, so $\varphi_3 = 6 - 1 - 3 = 2$. Now $z = 2\cdot3 + 3\cdot1 + (0+0+2) = 11$.

Observations in Table 5D.1

- For $\Delta = 1$, the utilitarian solution $\underline{u}^{\mathrm{U}}$ is the welfare maximising solution.
- For $\Delta = 2$, alternative optimal solutions exist: utilitarian solution $\underline{u}^{\mathrm{U}}$ and dominated solution $\underline{u}^{\mathrm{D}}$ both have $z = 10$. Adding a tie-breaking term $\varepsilon \cdot \Sigma_i u_i$ to the right-hand side of inequality (20) will make $\underline{u}^{\mathrm{U}}$ the unique welfare maximising solution. However, adding a tie-breaking term $\varepsilon \cdot u_{\min}$ will make $\underline{u}^{\mathrm{D}}$ the unique welfare maximising solution.
- For $\Delta = 3, 4$, the dominated solution $\underline{u}^{\mathrm{D}}$ is the welfare maximising solution.
- For $\Delta = 5$, alternative optimal solutions exist: both Rawlsian solution $\underline{u}^{\mathrm{R}}$ and dominated solution $\underline{u}^{\mathrm{D}}$ have $z = 13$. Adding a tie-breaking term $\varepsilon \cdot \Sigma_i u_i$ to the right-hand side of inequality (20) makes $\underline{u}^{\mathrm{R}}$ the unique welfare maximising solution.

Conclusion: CEU's welfare function can be maximised by a solution that is not on the Pareto front of the bi-criteria problem maximise$\{\sum_i u_i\ ;\ u_{\min}\}$. The issue is not caused by the existence of alternative optimal solutions.

Chapter 6

Conclusions and discussion

6.1 Introduction

As stated in Chapter 1, unhealthy diets contribute substantially to the worldwide burden of non-communicable diseases (NCDs). Globally, NCDs are the leading cause of death, and numbers are still rising. NCDs have a negative impact on the quality of life and well-being of the individual and of society as a whole, and put a high burden on health systems and economy. They are regarded as an important barrier to poverty reduction and sustainable development in low- and middle-income countries. In order to fight NCDs, the WHO (2013a) formulated a voluntary global target of “25% relative reduction in risk of premature mortality from cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases” for prevention and control of NCDs to be attained by 2025. As unhealthy diets are an important risk factor for NCDs, healthy diets can contribute to achieving this overall target. Formulating healthy diets requires target group-specific dietary assessment and advice. Both comprise complex decision problems that commonly have multiple objectives. Therefore, there is a need for multi-objective decision-making (MODM) approaches for dietary assessment and advice.

As stated in Section 1.3 (Figure 1.2), the scientific contribution of this thesis is two-fold: it provides opportunities for better decision-making in dietary assessment and advice and it contributes to model building and solving in OR.

This chapter is organised as follows. First, in Section 6.2 the conclusions with respect to the four individual research questions are considered. Section 6.3 gives the integrated findings. Section 6.4 discusses methodological choices, modelling choices, case-based choices, and challenges for implementation, and points out directions for future research. Section 6.5 provides a concluding remark.

6.2 Conclusions

Based on the research challenges defined in Table 1.2, Table 6.1 summarises the key results of this thesis for nutrition research and for OR. Then, the individual research questions are considered in more detail.

Table 6.1 Key results for Nutrition Research and Operations Research.

Research challenge		Nutrition Research	Operations Research
1. MODM for FFQ development	RQ1/Ch2	MILP-generated food lists can either be shorter or provide more information	New application area
2. Solve fractional programming problem	RQ2/Ch3	Indirect items can be modelled	Solution approach for a general 0-1 fractional programming problem
3. Explore preference structures in diet models	RQ3/Ch4	Awareness among nutritionists of impact of modelling choices on generated solutions	Nutritionists become aware of added value of MODM for diet modelling
4. Combine total and lowest utility	RQ4/Ch5	Approach not suitable for diet modelling	Methodological insights

6.2.1 RQ1 – How can MODM support selection of items for FFQs?

Nutritionists face a challenge to develop short FFQs that provide sufficient information for each of a potentially large set of nutrients of interest. The common procedure for selecting food items for the food list of FFQs, usually based on stepwise regression (Molag et al., 2010, Willett, 1998), is time-consuming. Moreover, it is hard to select items in such a way that all nutrients of interest are sufficiently covered. As a result, the selection of items strongly depends on intuition and (different) personal experiences of domain experts.

Chapter 2 of this thesis models selection of the most informative items to be included in the food list of FFQs as a mixed-integer linear programming (MILP) model. The MILP model chooses the most informative combination of food items for different aggregation levels fast and objectively. Also, food lists of various lengths may be generated and the increase of coverage and variation explained by adding more or other items to the food list may be investigated. The model provides objective information that can help to judge whether the extra information obtained by adding more food items justifies the additional burden for respondents and the additional research cost. With the MILP model, multiple nutrients can be optimised simultaneously. In contrast with a manual selection procedure, the number of nutrients has no impact on the complexity of the model. The results of the MILP-based selection procedure are highly reproducible. In addition, the MILP model can be included in a computer system.

The generated food lists have good performance in terms of length, coverage and variation explained of all nutrients of interest. The results suggest that the MILP model makes the selection process faster, more standardised and transparent, and is especially helpful in coping with multiple nutrients. The generated food lists appear either shorter or provide more information than a food list generated without the MILP model.

6.2.2 RQ2 – How to solve general 0-1 fractional programming problems with conditional fractional terms?

Food lists generated with the FFQ model of Chapter 2 can contain single items such as APPLES and ORANGES and aggregated items such as CITRUS or FRUITS. The composition of such items follows directly from the dataset, so it is known before optimisation. Therefore, we refer to them as *direct items.* In FFQs, also *indirect items* like OTHER CITRUS are used. The interpretation of an indirect item depends on the direct items that are included in the food list, and is therefore only known after optimisation, see Figure 1.3. Chapter 3 shows that this extension to the FFQ problem can be modelled as a general 0-1 fractional programming problem with more than 200 fractional terms. All fractional terms are conditional, that is, in every feasible solution only a subset of the fractional terms is actually defined. Existing literature does not provide a solution method for such problems. Chapter 3 describes how classical transformation principles can be combined and extended in order to eliminate the undefined fractional terms from the objective function. The resulting MILP model can be solved with standard software. Practical instances were solved fast.

6.2.3 RQ3 – What is the impact of achievement functions in diet models?

Existing diet models commonly use weighted sum achievement functions, which presume that the decision-maker has a preference structure in which the trade-offs between objectives are known and constant. Weighted sum achievement functions may generate solutions in which some objectives completely meet their targets, whereas others are (very) far off. Moreover, these achievement functions can be very sensitive to preferential weights. Chapter 4 of this thesis investigates the weighted sum achievement function and two other achievement functions, which generate more balanced solutions that tend to be less sensitive to preferential weights: MaxMin and Extended Goal Programming (EGP). These achievement functions are demonstrated on a diet model in which no solution exists that satisfies all nutritional constraints.

In order to find a 'best possible' diet, the preferences of the nutritionist with respect to food and nutrient intakes are modelled via utility curves[1]. The various achievement functions represent ways of aggregating single food and nutrient utilities into one indicator for overall diet quality. For a further discussion of nutrient utility curves and diet adequacy we refer to Section 6.4.2 and Figure 6.2.

Chapter 4 demonstrates that a weighted sum achievement function generates a solution in which the difference between the lowest and the highest utility is relatively large: the lowest utility is much lower than the other utilities. The MaxMin achievement function generates a solution in which the utilities are as 'equal' as possible: the lowest utility is high, and the differences with the other utilities are small. This, however, incurs a substantial decrease of total utility. The EGP achievement function is a convex combination of the weighted sum and MaxMin achievement functions, and generates compromises: it provides solutions with higher total utility than the MaxMin achievement function, and smaller differences between lowest and highest utility than the weighted sum achievement function. Offering multiple solutions allows the decision-maker choice of a solution that is most suitable for a specific decision problem, and that best meets non-quantifiable goals and preferences.

6.2.4 RQ4 – What is the added value of a novel method for finding a compromise between total utility and lowest utility in the context of diet models?

As discussed in Chapter 4, the preference structure incorporated in an MCDM model is of crucial importance for the quality of decision-making. It is therefore worthwhile to keep an open mind for novel approaches, investigate their behaviour, and assess their added value for practice. Chapter 5 explores such a novel approach, proposed by Hooker and Williams (2012), which we refer to as CEU. It provides new insights into CEU and assesses its added value for practice by comparing it with the extended goal programming (EGP) approach of Chapter 4.
Chapter 5 shows that, because of its discrete nature, CEU generates a set of solutions that is smaller and more widely spaced than that of EGP. This is an

[1] Note that in Chapter 4 these are called 'adequacy curves'.

advantage for any decision-maker who prefers a relatively small set of reasonably different solutions to a large set of rather similar solutions. CEU generates solutions that are not found with EGP. The preference structure modelled in CEU may incur a (small) distance to the Pareto front of the bi-objective problem on the aggregated objectives (viz. total utility and lowest utility). However, all CEU solutions are on the Pareto front of the decision-maker's original *n*-objective problem, which aims to maximise *n* individual utilities at the same time. We suggest that for use in practical cases this distance is reported. This supports the decision-maker to assess whether the achieved balance is worth its price. Because of the presence of binary variables, CEU may require a larger computational effort than EGP.

Chapter 5 concludes that for applying CEU the investigated decision problem should have the following two characteristics: (i) summing individual utility values is meaningful, (ii) small increments of utilities within a predefined (commonly small) range from the lowest utility do not affect the quality of a solution. In diet models this would translate to: (i) diet quality can be calculated as a sum of individual food and nutrient utilities, (ii) small increments of low utilities (except the lowest utility) do not improve the overall diet quality. The diet problems addressed in this thesis will often not have these characteristics, which limits the applicability of CEU for diet modelling. The insights gained in this study, however, are valuable, because they extend beyond diet modelling. They may be particularly useful for any researcher who considers using CEU, or who struggles with the explanation and interpretation of its results.

6.3 Integrated findings

This thesis connects the disciplines of nutrition and OR in order to contribute to formulating healthier diets. Its scientific contribution is two-fold: it provides opportunities for better decision-making in research on dietary assessment and advice, and it contributes to model building and solving in OR, see Figure 6.1. This section elaborates the added value that extends beyond merely answering the research questions that were formulated in Section 1.4. It first reflects on the added value for nutrition research and then on the added value for OR. Throughout this section, the expert on dietary assessment and advice will be referred to as 'the decision-maker' or 'the nutritionist'. The developer of the OR model will be referred to as 'the analyst'.

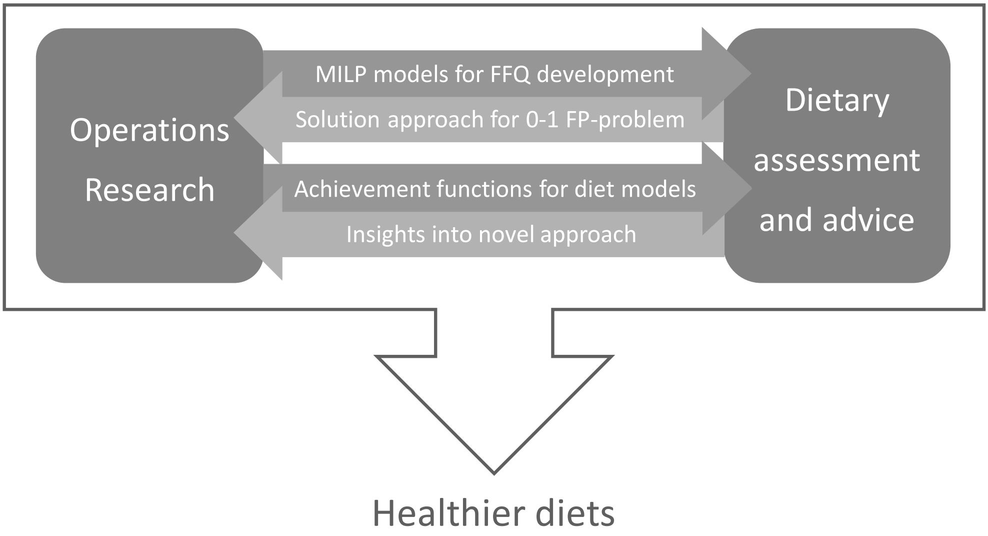

Figure 6.1 This thesis contributes to formulating healthier diets by proposing MILP models for FFQ development (Chapters 2 and 3) and achievement functions for diet models (Chapters 4 and 5). Furthermore, for OR this not only provides models and a (stronger) foothold in nutrition (Chapters 2, 3, 4), but also a solution approach for a 0-1 fractional programming problem (Chapter 3) and new insights into a novel approach for combining equity and utilitarianism in a single model (Chapter 5).

Added value for Nutrition

For the nutritionist, developing MODM approaches contributes to structuring the problem and becoming aware of implicit and explicit assumptions that are made along the way. In this thesis research, developing MODM approaches challenged the nutritionists to explicitly articulate the expertise, assumptions, priorities, and procedures they use for dietary assessment and advice. Many of these procedures strongly depend on the acquired scientific evidence base in nutrition and health research. Developing quantitative models forced the nutritionists to specify their own assumptions and procedures. In FFQ development, it encouraged debate on issues like 'why and/or how much do we prefer CITRUS to ORANGES in a food list targeted for diabetes?'. In diet modelling, the need to specify preference structures encourages debate on quantifying the contribution of a diet to individual and public health.

Being introduced to the basics of quantitative decision-making contributed to the nutritionists' awareness of the possibilities and limitations of currently used tools for dietary assessment and advice. For instance, it was known that the current sequential way of adding items to food lists of FFQs might lead to lists that are unnecessarily long, especially when many nutrients are involved. Also, it was experienced that aiming for multiple nutrients increased complexity of composing a compact and informative food list substantially. In MILP-supported decision-making, however, all direct and indirect items are selected at the same time, and the complexity is hardly influenced by the number of nutrients. This was an eye-opener for nutritionists with respect to the potential added value of MILP models compared to usual methods such as those based on stepwise regression. Concerning diet modelling, the research increased awareness of the implicit assumptions of additive linear achievement functions, and their consequences. This helps to critically assess the outcomes of existing models, and indicates opportunities for improvement.

The developed MODM approaches generate, in a transparent and reproducible way, suggestions that either confirm or challenge the intuition, and foster out-of-the-box thinking (Claassen, 2014). Moreover, they facilitate exploring various what-if scenarios, such as “What is the maximal covered intake of vitamin C in an FFQ with 30 items?” and “What are feasible lower bounds on intake of total mono- and disaccharides if we impose a food-based dietary guideline of eating at least 200g of fruit per day?”.

Developing MODM approaches for dietary assessment and advice increased awareness of methodologies that are complementary to currently used approaches. This encourages searching for MODM approaches for other nutritional issues such as nutrient profiling and prioritising intervention strategies.

Added value for Operations Research

This thesis contributes to model building and solving in OR. As such, it demonstrates the usefulness of venturing out into new application areas, even when the models developed for the initial problem may – from OR perspective – seem straight-forward at first. For the FFQ model of Chapter 2, for instance, the initial challenge was not to build the FFQ model itself, but to define sensible (nutritional) data. Next, there was room for developing the more complicated, fractional model as presented in Chapter 3. Although a vast body of literature exists on fractional programming (e.g. reviews of Stancu-Minasian (1999, 2006)), no solution approach was found for this specific fractional problem. Therefore, a new approach was designed. This demonstrates that real-life problems can in a natural way evolve to OR challenges that lead to extensions of the OR methodology.

Section 5.5 concludes that the added value of CEU for diet modelling is limited. The study itself, however, is valuable, because it provides methodological insights in a novel method that has not yet been investigated further. The MILP models presented by Hooker and Williams (2012) for CEU are elegant and succinct. Only upon further inspection, it becomes clear that the interaction between CEU's welfare function and the problem data can be very subtle and complex. The insights gained with respect to the number, spacing and location of the generated solutions may thus be particularly useful for any researcher who considers using CEU, or who struggles with the explanation of its results.

These contributions to model building and solving in OR are examples of what is referred to as 'applications-driven theory', which means that "the research is initiated in an actual application, which is first successfully solved and then generalised for publication in the literature where it can serve as a basis for both further use and research" (Cooper and McAlister, 1999, Cooper, 2005). As such, they demonstrate the added value of applications-driven research.

Furthermore, this thesis has created awareness among nutritionists of the added value of OR for nutrition research, and it has increased nutritionists' sensitivity for recognising nutritional decision problems that could benefit from an OR approach, thus strengthening the foothold of OR within nutrition research and creating opportunities for development of applications-driven theory.

6.4 Discussion and opportunities for future research

This thesis demonstrates how MODM approaches can make decision-making for dietary assessment and advice more objective, transparent, and reproducible, and how this can contribute to model building and solving in OR. This section discusses methodological choices, modelling choices, case-based choices, and challenges for implementation, followed by opportunities for future research.

6.4.1 Methodological choices

As mentioned in Section 1.1, OR employs techniques such as mathematical modelling to arrive at optimal or near-optimal solutions to complex decision problems. Considering that the FFQ problem and the diet modelling problem aim to optimise multiple conflicting objectives, for instance referring to multiple nutrients, both are classified as MCDM problems. In the FFQ problem as well as in the diet modelling problem, the set of solutions is implicitly described via a set of constraints, so within MCDM both are classified as MODM problems. The aim is therefore to generate Pareto-optimal solutions.

The FFQ problem aims to optimise one objective for every nutrient of interest, while minimising the length of the food list. The model presented in Section 2.2.5 is based on the ε-constraint method: it maximises a weighted sum of objectives, while putting constraints on all objectives. Appendix 2B provides more examples of such models. Moreover, it demonstrates the use of an augmented Chebyshev criterion, which maximises the worst among the set of nutrient-based objectives, while putting an upper bound on the length of the food list. Both the ε-constraint method and the use of the augmented Chebyshev criterion will generate Pareto-optimal solutions.

The diet model has aspiration levels for every nutrient, and aims to minimise the deviations from these levels. Goal programming is a common approach for generating Pareto-optimal solution for such a problem. Chapter 4 investigates several achievement functions, representing various preference structures, for the diet model. Pareto-optimality of the generated solutions is ensured by using sufficiently high target levels, and by using an augmented Chebyshev criterion for the MaxMin optimisation. Chapter 5 explores CEU, a novel approach for balancing between aggregated criteria on total utility and lowest utility. It concludes that CEU's solutions are not necessarily Pareto-optimal in terms of the aggregated objectives (viz. total utility and lowest utility). They are, however, Pareto-optimal with respect to the original set of individual objective functions.

6.4.2 Modelling choices

The modelling choices encountered when using MODM for optimising diet quality are related to open issues in nutritional sciences, which we will refer to as 'nutritional assumptions'. The main nutritional assumptions faced in the context of this thesis are related to the aggregation of individual nutrient adequacies and compliance with food-based dietary guidelines (FBDGs) into one indicator for overall diet quality, see Figure 6.2.

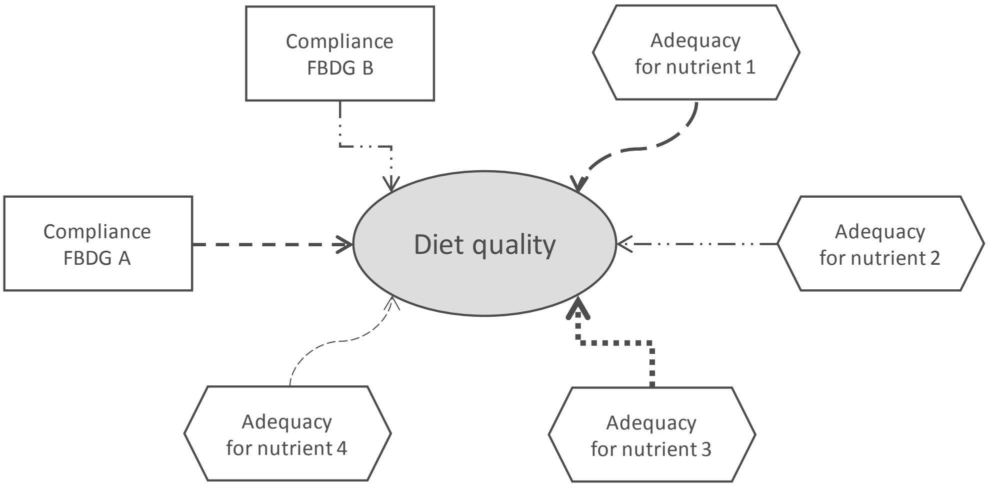

Figure 6.2 Aggregating individual nutrient adequacies and compliance with FBDGs into overall diet quality. The diversity of arrows symbolises that various nutrients and foods contribute to diet quality in different ways.

Nutrient adequacy – Utility curves

The diet models in this thesis aim to generate diets with a high overall nutritional quality. The context is that of a nutritionist who designs a diet for a participant in an intervention study. Based on characteristics such as age, sex, physical activity, it is known to which target group the participant belongs. The probability distribution of nutrient requirements of persons in this target group can be expressed via dietary reference values (DRVs). The personal requirements of the specific participant, however, are not known. Therefore, the nutritionist aims to compose the diet in such a way that for all nutrients it is highly likely that the intake level is sufficient for the participant. This is modelled via nutrient utility curves, which aim to model the preferences of the nutritionist, guided by the question “What intake would the nutritionist aim for?”.

In designing the nutrient utility curves, it is postulated that for ‘beneficial’ nutrients, such as vitamins and minerals, the nutritionist would aim for an intake between the EAR (estimated average requirement) and the RNI (recommended nutrient intake), that is, an intake that would be sufficient for 50% to 97.5% of the target group that the participant belongs to. Therefore, these intakes are assigned a utility value of 1. Intakes lower than the EAR incur a substantial risk of being insufficient, which is expressed via a utility value lower than 1. Intakes lower than the lower intake level (below which an intake could lead to risk in most individuals) are considered insufficient, which is represented by utility value 0.

Intakes exceeding the RNI probably hardly contribute to the nutritional value of the diet, and – for much higher intakes – might even induce adverse effects. In order to express that increasing intake to a higher level than the RNI is not considered to be very useful, such intakes are assigned utility values lower than 1, decreasing to 0 for intakes equal to the upper intake level that is unlikely to pose risk of adverse health effects. For 'non-beneficial' nutrients such as SFA, cholesterol, mono- and disaccharides, the nutritionist would aim for an intake that would for 97.5% of the population not be harmful, so all intakes lower than that are assigned a utility value of 1. As a result, the calculated nutrient utility values should not be interpreted as the percentage of the target group (to which the participant belongs) for which the intake would be sufficient, or as the probability that the intake will be sufficient for the specific participant, but as the extent to which the intake complies with the nutritionist's preferences. The thus defined nutrient utility curves model that increasing intakes that are adequate for only a few people in the participant's target group is more likely to increase the overall nutritional quality of the diet (for the participant) than increasing intakes that are already sufficient for the large majority of the target group to which the participant belongs. A drawback of this modelling choice, however, is that a utility value lower than 1 can either be caused by an intake that is too low or an intake that is unnecessarily high. So, the achievement function of the optimisation model cannot distinguish between inadequacies that are caused by intakes that are potentially too low and potentially too high.

The use of utility curves is not limited to nutrients; also for foods, such as vegetables and fruits, utility curves can be defined, see Table 4.5. The actual shape of such food utility curves will have to reflect scientific evidence on, for instance, sensory and satiety aspects of foods, and on the relationship between FBDG compliance and health outcomes.

Utility curves are not limited to designing diets for intervention studies, but can be tailored to the context of the dietary problem at hand. In developing guidelines for policy makers, for instance, one might use curves that approximate the distribution of nutrient requirements over the population as a whole.

Diet quality – Preference structure in MODM model

It remains unknown how a diet's adequacy with respect to a set of individual nutrients and compliance with FBDGs can be aggregated into an indicator for overall diet quality, see Figure 6.2. Nutritionists deal with this in various ways, depending on the problem context. For supporting consumer decision-making, for instance, nutrient profiling models such as the NRF (Fulgoni Iii et al., 2009) express overall quality of a food via a weighted sum of nutrient-specific indicators. Alternatively, nutritional scoring system NuVal calculates a quality index by dividing a weighted sum of indicators for 'beneficial' nutrients by a weighted sum of indicators for 'non-beneficial' nutrients. On the level of both foods and diets, the Nutrient Balance Concept (Fern et al., 2015) quantifies quality via two indicators that are weighted sums of nutrient-specific indicators, and a third indicator that takes into account that it's not useful to increase nutrient intake further than to a sufficient level. For supporting nutritionists' decision-making, a diet's probability of being adequate for a set of micronutrients can be evaluated via its probabilities of being inadequate for the individual nutrients in the set: P(diet is adequate) = 1 – Π_j P(intake for nutrient j is inadequate), provided that the nutrient set covers all relevant nutrients, and that all probabilities are independent. For supporting health policy makers, nutritionists aim to express health impacts of food or nutrient intake in terms of DALYs or QALYs lost or gained, thus trying to make effects comparable.

In the presented diet models, the aggregation of individual nutrient adequacies and FBDG compliance is modelled via the preference structure (achievement function) in the MODM model. This preference structure is a key consideration in MCDM (Romero, 2004). Weighted sum achievement functions reflect the assumption that diet quality is a weighted sum of nutrient adequacies and FBDG compliance, whereas MaxMin achievement functions reflect the assumption that diet quality is mainly determined by the nutrient or food with lowest adequacy or FBDG compliance, respectively. EGP's convex combination of weighted sum and MaxMin provides a compromise between both. Literature offers several other achievement functions that are worth exploring.

A so-called lexicographic achievement function can be considered when the decision-maker can classify the nutrients into strict priority classes between which no finite trade-offs exist (Tamiz et al., 1998, Romero, 2004). This can be relevant in the context of individual dietary advice, with priority classes for instance: 1. energy, 2. energy-yielding nutrients, 3. micro-nutrients and trace elements for which EARs and RNIs exist, 4. micro-nutrients and trace elements for which no EARs and RNIs exist. In that case, the model would first be run with an equality constraint on energy intake (indicating that energy requirements should be exactly fulfilled) and upper and lower bounds on energy-yielding nutrients. The next iteration then aims to search a diet that fulfils the requirements for the nutrients in class 3 as well as possible, without deteriorating the performance with respect to the nutrients in class 2, and so on.

Another option would be to minimise the number of unmet targets (Jones and Jimenez, 2013). Such an achievement function might be useful in the context of public health policy making, where unmet nutritional constraints incur costs, for instance due to necessary fortification programs.

No diet model can solve nutritional dilemmas on aggregation of food and nutrient utilities into a single indicator for diet quality. However, they can contribute to the debate by providing a fast and transparent way of investigating the impact on advised food intakes and their associated nutrient intakes of various DRVs and various ways of aggregating food and nutrient adequacies. Furthermore, diet models are very useful for identifying ‘problematic’ nutrients, recommendations and consumer preference constraints. For instance, in all our diet modelling results, mono- and disaccharides had lowest utility, irrespective of the achievement function and target group. This demonstrates the consequences of applying very strict DRVs for mono- and disaccharides, combined with lower bounds on fruit intake, and on intake of snacks.

Also in FFQ development, the problem context should be taken into consideration when defining an MODM model. For instance, in aetiologic studies, all nutrients of interest might be of equal importance, because the study needs to assess health effects for all of them. In public health applications, however, some nutrient intakes might be considered more important than others, because they are associated with a larger effect on public health. Experimenting with various weights will provide the nutritionist with a set of solutions with various properties, thus allowing him/her to select the solution that is most suitable for the specific decision problem, and that best meets non-quantifiable goals and preferences.

6.4.3 Case-based choices

In constructing cases for experimenting with the proposed decision support models, several choices are made, for instance on food composition, nutrient sets, dietary reference values, and target groups.

Food composition

Food composition is crucial for nutritional studies: dietary assessment uses it to translate foods into nutrients, and dietary advice uses it to translate nutrients into foods, see Figure 1.1. All studies in this thesis use the NEVO food composition database (RIVM, 2012). With the proposed MILP models, the impact of other data sets or changes in food composition (for instance by fortification or food reformulation) or food use (for instance by using new foods such as meat replacers) can be investigated in a transparent and reproducible way.

Nutrient sets

The proposed models are demonstrated on cases for specific sets of nutrients. The choice of the nutrient set reflects the nutritionist's priorities; nutrients that (s)he does not consider important (enough) will not be included. Besides relevance for the target group, the choice of nutrient sets can be based on data availability. In FFQ development, the selected nutrient set affects the generated food list. For instance, the number of food items in FFQs depends on the dispersion of the selected nutrients through the available foods. For instance, for assessing calcium intake probably a shorter food list would suffice than for assessing protein or energy. Also in diet models, the selected nutrient set affects the generated solutions. For instance, diet models that include calcium are more likely to propose dairy products than diet models that do not take calcium into consideration. Of course, the selected nutrient sets affect the generated solutions. They do, however, not affect the methodologies described in this thesis in terms of speed, transparency, and reproducibility.

Dietary reference values

Nutritionists use dietary reference values (DRVs) for formulating recommendations and goals regarding food and nutrient intake. The adequacy associated with a certain nutrient intake is commonly expressed as percentage of the population for which the intake is sufficient. DRVs are defined for specific populations, depending on age, sex, energy requirements, body size, and physical activity. The exact values of DRVs are an issue of much debate among nutritionists. It is not uncommon that, based on the same information and on sound assumptions, researchers formulate different DRVs, see for instance the study of Brouwer-Brolsma et al. (2016) on DRVs for vitamin D intake. Diet models cannot solve the nutritional dilemmas associated with setting DRVs, but they can contribute to the debate by providing a fast and transparent way of investigating the impact of various DRVs on advised food intakes. The used values of DRVs affect the generated solutions. They do, however, not affect the methodologies described in this thesis in terms of speed, transparency, and reproducibility.

Target group

The proposed models for supporting decision-making in dietary assessment and advice are demonstrated on cases for specific subgroups of the Dutch population. Other subgroups and – especially – other populations may have different foods, which requires different food lists and results in different diets. For instance, when assessing sugar intake or designing a diet for a Dutch target group, including apple sauce might be very useful, whereas for a Mediterranean target group, apple sauce is hardly relevant. With the proposed models, food lists and diets can be generated in a transparent and reproducible way when the user changes the input data or constraints that represent the characteristics of the target group. Nevertheless, the need for independent validation with well-accepted reference methods remains relevant.

6.4.4 Challenges for acceptance of the MODM approaches

The MODM approaches described in this thesis face two important challenges for acceptance by nutritionists: they are presented as pilot versions, and their outputs are starting solutions that need post-processing by interaction between nutritionist and analyst.

Pilot versions

The decision support models described in this thesis were presented to the nutritionist in the form of pilot models; they are not built as user-friendly, user-proof systems that the nutritionist can freely experiment with. Instead, they have to be operated by the analyst. This limits the nutritionist's opportunities to acquire confidence on the models' outcomes and possibilities, to familiarise with the models' properties, and to aim for fast implementation. This is an obstacle for acceptance and adoption of the new approaches.

Post-processing

Solutions proposed by decision support models can often not be used immediately by the decision-maker, but require post-processing. For instance, commonly the food lists generated by the MILP model do not comply with the nutritionist's standards on face validity, because they may contain 'stand-alone' items. For example, the solution contains BLACKBERRY but no similar (soft fruit) items, which might tempt a respondent who has consumed blueberries to report his blueberry consumption under BLACKBERRY (which is unwanted from the perspective of dietary assessment). This may be an obstacle for acceptance of MILP-supported development of FFQs. The need for post-processing should, however, not be a reason for discarding the methodology. In typical OR-supported decision-making, the decision support model is used to propose an initial solution, such as a food list for an FFQ. This solution will commonly first be scrutinised by the nutritionist. Based on expert knowledge, (s)he will point out strengths and weaknesses of the proposed solution to the analyst. For instance, (s)he will point out that stand-alone use of food item BLACKBERRY will confuse the respondent if no reference is made to similar soft fruits like BLUEBERRY. Based on this feedback, the analyst will modify the objectives and/or constraints of the model, for instance exclude BLACKBERRY or include OTHER SOFT FRUITS as well. The model is then re-run in order to generate a solution that takes into account the expert knowledge. This loop is repeated until the nutritionist is satisfied. In this iterative procedure, the decision-maker is supported by the MILP model. It combines the strong points of human insight and experience on the one hand and the efficiency and accuracy of quantitative optimisation techniques on the other hand (Claassen et al., 2007). In the case of the FFQ model, the observed difference in length between the ValNed food list and its MILP alternative is so large that there is ample opportunity for adding (or exchanging) items to the MILP list and still ending up with a shorter list than ValNed.

New technology

In this context, it is worth mentioning the new opportunities offered by development of apps for smartphones. One could, for instance, design an app that asks the respondent a very limited set of questions per day. In such a setting, being asked for BLACKBERRY consumption will not confuse the respondent, because he is used that every day only a very limited part of his consumption pattern is explored.

6.4.5 Opportunities for future research

Several opportunities for further research arise as direct leads resulting from this thesis. Also, opportunities are indicated for extensions to sustainable diets and food supply chains, and for communication to the consumer. Furthermore, opportunities with respect to MCDM are identified.

Direct leads

Based upon the issues discussed in Sections 6.4.2 and 6.4.3, several opportunities for further research arise as direct leads resulting from this thesis, especially in the context of nutrient adequacy and compliance with FBDGs, diet quality, and CEU, as summarised below.

Nutrient adequacy and compliance with FBDGs – Utility curves

- Investigate the impact of other types of utility curves on the proposed food and nutrient intake of the diet. For instance, what is the impact of using utility curves that reflect the percentage probability that a nutrient intake is adequate for the target group.
- Investigate for various diet modelling contexts which type of utility curve is most suitable. For instance, which type of utility curve is suitable for modelling diets on the level of a whole population.

Diet quality – Preference structure in MODM model

- Investigate how sensible weights can be set in achievement functions for diet models. For instance, how to select weights such that actually the most critical nutrients are sufficiently prioritised without paying too little attention to the other nutrients.
- Investigate for various diet modelling contexts which type of preference structure is most suitable. For instance, what is suitable for individual dietary advice, and what is suitable on the level of health policy making.

Challenges for acceptance of the MODM approaches

- Implement the FFQ models of Chapters 2 and 3 in order to assess their added value in a real-life setting by developing FFQs targeted for many nutrients. Develop a procedure for interaction between the nutritionist and the analyst. Investigate how much current lists can be shortened without loss of information and/or how much extra information can be obtained without increasing the length.

CEU

- Investigate for which context CEU is a suitable approach.
- Investigate if CEU can be extended in order to differentiate between Rawlsian contributors.
- Investigate if it is possible to include a lexicographic criterion for the Rawlsian contributors.

Extensions to sustainable diets and supply chains

The global food system faces the challenge to feed a population which may rise to nearly 10.5 billion in 2100 (Fresco and Poppe, 2016). Part of the solution may come from increasing sustainability of diets and food supply chains.

Environmental sustainability

There is an increasing awareness of the environmental impact of our food consumption pattern. For instance, consuming meat has a much larger ecological footprint than consuming legumes and grains (Notarnicola et al., 2017). As a consequence, it might be much more sustainable to satisfy protein requirements via legumes and grains than via meat. Nevertheless, a certain amount of meat might be very useful to satisfy requirements for protein or micronutrients (like iron or vitamin B_{12}). The Health Council of The Netherlands (2011) states that regarding food "there are hardly any guidelines that combine health and ecological perspectives", and that "more research is needed into the further development of guidelines for a healthy and eco-friendly diet". In order to support policy makers, there is a need for diet models that integrate aspects of health and sustainability in a transparent and reproducible way. This thesis research has contributed to the formulation of the so-called SHARP model in EU project SUSFANS (H2020, 2017) and the parallel research project SHARP-BASIC (TIFN, 2017). These projects aim to provide a scientifically underpinned knowledge and data platform that can be used to build models for deriving SHARP diets for European citizens, i.e. diets that are environmentally Sustainable (S), Healthy (H), Affordable (A), Reliable (R) and Preferred from the consumer's perspective (P) (Mertens et al., 2017).

Food supply chains

Not only diet composition affects costs and sustainability of our food pattern, but also the design of food supply chains. For the Netherlands, for instance, tomatoes can be supplied via Dutch greenhouses or be imported from Spain. The trade-off between production and transportation costs (expressed in monetary values and in environmental impact factors such as energy use) may vary with the season, and with choices made in the food supply chain. In order to support policy makers and food industry, there is a need for models that integrate diet and supply chain considerations. This thesis research has contributed to the formulation of NWO research project GREENDISH, which aims to (i) define affordable diets for Dutch citizens that are both healthy and sustainable, as well as (ii) indicate the consequences of alternative healthy and sustainable diets for the design of food supply chains (NWO, 2017). The project addresses issues such as multi-objective network design for the food system under consideration of nutritional demands, with a focus on studying a shift from a meat-based to a plant-based diet on the configuration of food supply chains and the resulting environmental impact.

Communication to the consumer – Nutrient Profiling

This thesis demonstrates how MODM can be used to support decision-making in dietary assessment and advice. As a next step, the findings arising from dietary assessment and advice have to be communicated to the consumer such that (s)he is guided towards a healthy dietary pattern.

Lobstein and Davies (2009) provide a review on the use of systematic methods for categorising foods according to their nutritional quality (nutrient profiling) as a strategy for promoting public health through better dietary choices. The authors state that a nutrient profiling approach should be able to summarise and synthesise key nutritional characteristics (such as sugar, fat and salt content, energy density and portion size) in a way that is easily applied across a variety of products, is understandable to users, and can be strictly defined for regulatory purposes. In short, nutrient profiling should answer the question “how healthy is this food?”, based on multiple attributes, viz. its contents of multiple nutrients. Multi-attribute decision-making (MADM) may contribute to the development of adequate methods for profiling foods and diets.

Multi-criteria decision-making

Identifying a representative subset of Pareto-optimal solutions

The MODM approaches in this thesis aim to provide the decision-maker with a set of Pareto-optimal solutions. Ideally, this set is representative for the whole set of Pareto-optimal solutions. Shao and Ehrgott (2016) propose a new method for obtaining such a set, which is tested for problems with up to eight objectives. The problems in diet modelling, however, typically have many more objectives, which poses substantial computational challenges. So, there is a need to investigate how a representative subset of the Pareto-set can be obtained for problems with many objectives within an acceptable computational effort.

Weight selection in MODM

As mentioned in Section 1.4, a problem in MODM is how to justify and interpret weights, for instance in formulating weighted sum achievement functions and in combining achievement functions. Further research is needed to find out how to set sensible weights in MODM models for dietary assessment and advice.

Looping back: MADM and nutritional assumptions

After finding a representative subset of the Pareto-front, it is up to the decision-maker to rank these alternatives based on their performance with respect to multiple criteria and his/her preferences, or to point out a preferred alternative. This is a typical MADM problem, so it is worth investigating whether the decision-maker can be supported via MADM approaches such as Analytic Hierarchy Process and Electre (Belton and Stewart, 2002). Within context of diet modelling, however, expressing preferences based on diet properties coincides with the aforementioned nutritional assumptions on aggregating FBDG compliance and nutrient adequacy into one indicator for overall diet quality.

6.5 Concluding remark

This thesis explores the use of MODM approaches to improve decision-making for dietary assessment and advice. Considering the added value for nutrition research and the new models and solutions generated, we conclude that the combination of both fields has resulted in synergy between nutrition research and OR. As issues related to diet and to food supply tend to become more and more global and complex, we foresee a lasting need for OR-supported decision-making on levels ranging from consumer decision-making to designing food supply chains and effective and efficient governmental policies, thus contributing to improving global health via healthier diets.

Summary

Unhealthy diets contribute substantially to the worldwide burden of non-communicable diseases (NCDs), such as cardiovascular diseases, cancers, and diabetes. Globally, NCDs are the leading cause of death, and numbers are still rising. NCDs have a negative impact on the quality of life and well-being of the individual and of society as a whole, and put a high burden on health systems and economy. They are regarded as an important barrier to poverty reduction and sustainable development in low- and middle-income countries.

As unhealthy diets are an important risk factor for NCDs, healthy diets are a global priority to reduce NCDs. In the context of healthy diets, two nutritional research fields are particularly relevant: dietary assessment and dietary advice. *Dietary assessment* can contribute to NCD reduction by assessing the food and nutrient intake of target groups and individuals in order to investigate the relation between diet and disease. It helps to point out which foods and nutrients critically contribute to the health status of consumers, and to formulate food and nutrient recommendations. Studies in nutritional epidemiology, for instance, ask respondents to fill in a questionnaire on their use of certain foods products. The challenge for dietary experts is to compose a questionnaire that is short enough to be acceptable for the respondent, and yet sufficiently covers all nutrients that the dietary expert wants to assess. *Dietary advice* contributes to NCD reduction by translating food and nutrient recommendations into realistic food choices. Here, dietary experts face another challenge: from the range of thousands of products that contain multiple nutrients, how to compose a dietary pattern that complies with all nutritional constraints, and is acceptable for the consumer? So, both dietary assessment and dietary advice give rise to complex decision problems that commonly have multiple objectives: which foods to include in dietary assessment or advice to pursue the multiple objectives of the researcher or fulfil the requirements of the consumer?

A discipline that can help to solve these complex decision problems is Operations Research (OR), which deals with the application of advanced analytical methods to help make better decisions. It employs techniques such as mathematical modelling and mathematical optimisation to arrive at optimal or near-optimal solutions to complex decision problems. It is often concerned with determining a minimum (for example, in this thesis, the length of a questionnaire) or a maximum (for example, nutritional adequacy of a diet). Widely-used techniques in OR are linear programming and mixed-integer linear programming (MILP). Within OR, multi-criteria decision-making (MCDM) approaches can be used to support the decision-maker in situations where one wants to pursue multiple maxima or minima at the same time, for instance in situations where multiple nutrients are relevant. Commonly, there is a level of conflict between these maxima and minima. Within MCDM, multi-objective decision-making (MODM) focusses on situations in which the set of solutions is implicitly described via a set of constraints, and often is very large. MODM approaches aim to find a ‘best possible’ solution, that is, they aim to find a feasible solution that has the best possible performance with respect to the decision-maker’s objectives and preferences. Therefore, MODM approaches may contribute to decision-making in dietary assessment and advice.

The aim of this thesis is to investigate MODM approaches for dietary assessment and advice, thus contributing to formulating healthier diets. This is relevant not only to nutrition research as such, but also contributes to model building and solving in OR.

For assessing habitual nutrient intake of subjects in a population, often food frequency questionnaires (FFQs) are used. Nutritionists face the challenge to develop short FFQs that provide sufficient information for each nutrient of interest. In Chapter 2 of this thesis, the selection of the most informative items to be included in the food list of FFQs is modelled as a multi-objective MILP model. With this model, the most informative combination of food items for different aggregation levels can be determined in a standardised and reproducible way. Also, food lists of various lengths may be generated in order to investigate if adding more items justifies the additional burden for respondents and the additional research cost. With the MILP model, the food list can be optimised for multiple nutrients simultaneously. In contrast with a manual selection procedure, the number of nutrients has no impact on the complexity of the model. The generated food lists have good performance in terms of length, coverage and variation explained of all nutrients of interest. The results suggest that the MILP model speeds up the process and increases standardisation, transparency, and reproducibility. Moreover, it is especially helpful in coping with multiple nutrients. The generated food lists appear either shorter or provide more information than a food list generated without the MILP model.

Food lists generated with the MILP models of Chapter 2 can contain single items such as APPLES and ORANGES and aggregated items such as CITRUS or FRUITS.

For all these items, their interpretation follows directly from their names and thus is known *before* optimisation. Therefore, we refer to them as *direct items.* In FFQs, also *indirect items* like OTHER CITRUS are used. The interpretation of such an indirect item depends on the direct items that are included in the food list, for instance, if only ORANGE and LEMON are included in the FFQ as direct items, then OTHER CITRUS comprises all citrus in the dataset (such as tangerine, grapefruit), except ORANGE and LEMON. The composition of OTHER CITRUS will therefore only be known after optimisation. Although there is debate, nutritionists assume that the contribution of indirect items to the quality of an FFQ can be approximated by the average contribution of its constituent items. Chapter 3 shows that this can be modelled by extending the FFQ model of Chapter 2 with fractional terms for the indirect items: the contribution of an indirect item is calculated as the summed contributions of its constituent items divided by the number of constituent items. In case an indirect item is not used, both the summed contribution and the number of items are zero, which makes the resulting fractional term "zero divided by zero", and thus undefined. Chapter 3 shows how classical transformation principles can be combined and extended in order to eliminate these undefined fractional terms from the objective function. The resulting MILP model can be solved with standard software, thus supporting nutritionists to select both direct and indirect items for FFQs with one model. This demonstrates that real-life problems can in a natural way evolve to OR challenges that lead to extensions of the OR methodology.

Supporting decision-making related to dietary advice requires translating nutrient recommendations into realistic food choices. For this, linear programming-based diet models are a robust and flexible tool. Existing diet models commonly use weighted sum achievement functions, which presume that the decision-maker has a preference structure in which the trade-offs between objectives are known and constant. Weighted sum achievement functions may generate solutions in which some objectives completely meet their targets, whereas others are (very) far off. Moreover, they can be very sensitive to preferential weights. Chapter 4 investigates the weighted sum achievement function and two other achievement functions, which generate more balanced solutions that tend to be less sensitive to preferential weights: MaxMin and Extended Goal Programming (EGP). These achievement functions are demonstrated on a diet model in which no solution exists that satisfies all nutritional constraints. In order to find a 'best possible' diet, the preferences of the nutritionist with respect to food and nutrient intakes are modelled via utility curves. The various achievement functions represent ways of aggregating single food and nutrient utilities into one indicator for overall diet quality.

Chapter 4 demonstrates that a weighted sum achievement function generates a solution in which the difference between the lowest and the highest utility is relatively large: the lowest utility is much lower than the other utilities. The MaxMin achievement function generates a solution in which the utilities are as 'equal' as possible: the lowest utility is high, and the differences with the other utilities are small. This, however, incurs a substantial decrease of total utility. The EGP achievement function is a convex combination of weighted sum and MaxMin, and generates compromises: it provides solutions with higher total utility than MaxMin, and smaller differences between lowest and highest utility than MaxSum. This allows the decision-maker the choice of a solution that is most suitable for a specific decision problem, and that best meets non-quantifiable goals and preferences.

As the preference structure incorporated in an MODM model is of crucial importance for the quality of decision-making, it is worthwhile to keep an open mind for novel approaches, investigate their behaviour, and assess their added value for practice. Chapter 5 explores such a novel approach from literature, which we refer to as CEU (Combining Equity and Utilitarianism). It provides new insights into CEU and assesses its added value for practice by comparing it with EGP for the case of the diet model of Chapter 4.

Chapter 5 shows that, because of its discrete nature, CEU generates a set of solutions that is smaller and more widely spaced than that of EGP. This is an advantage for any decision-maker who prefers a relatively small set of reasonably different solutions to a large set of rather similar solutions. CEU generates solutions that are not found with EGP. The preference structure modelled in CEU may incur a (small) distance to the Pareto front of the bi-objective problem on the aggregated objectives (viz. total utility and lowest utility). However, all CEU solutions are on the Pareto front of the decision-maker's original n-objective problem, which aims to maximise n individual utilities. We suggest that for use in practical cases this distance is reported, because it supports the decision-maker to assess whether the balance achieved by CEU is worth its price.

Chapter 5 concludes that for applying CEU the investigated decision problem should have the following two characteristics: (i) summing individual utility values is meaningful, (ii) increments of utilities within a predefined (commonly small) range from the lowest utility do not affect the decision-maker's perceived quality of a solution. The diet problems addressed in this thesis will often not have these characteristics, which limits the applicability of CEU for diet modelling. The insights gained in this study, however, are valuable, because they extend beyond diet modelling. They may be particularly useful for any researcher who considers using CEU, or who struggles with the explanation and interpretation of its results.

Chapter 6 presents the conclusions of the research and a general discussion. This thesis connects the disciplines of nutrition and OR in order to contribute to formulating healthier diets, and its scientific contribution is two-fold: (i) it provides opportunities for better decision-making in research on dietary assessment and advice by proposing MILP models for FFQ development and achievement functions for diet models, and (ii) it contributes to model building and solving in OR by providing a solution approach for a specific 0-1 fractional programming problem and new insights into a novel approach for combining equity and utilitarianism in a single model. Moreover, this thesis has created awareness among nutritionists of the added value of OR for nutrition research, thus strengthening the foothold of OR within nutrition research and creating opportunities for development of applications-driven OR theory. Amongst others, the chapter discusses the main modelling choices encountered when using MODM for optimising diet quality. These are related to nutritional assumptions on nutrient adequacy and compliance with food-based dietary guidelines (modelled via utility curves), and on aggregating individual nutrient adequacies and compliance with food-based dietary guidelines into one indicator for overall diet quality (modelled via various preference structures in the MODM models).

This thesis explores the use of MODM approaches to improve decision-making for dietary assessment and advice. Considering the added value for nutrition research and the new models and solutions generated, we conclude that the combination of both fields has resulted in synergy between nutrition research and OR. As issues related to diet and to food supply tend to become more and more global and complex, we foresee a lasting need for OR-supported decision-making on levels ranging from consumer decision-making to designing sustainable food supply chains and effective and efficient governmental policies, thus contributing to improving global health via healthier diets.

References

Anderson, AM and Earle, MD (1983). Diet planning in the third world by linear and goal programming. *Journal of the Operational Research Society,* 34**:** 9-16.

Barros, AI (1998). Discrete and Fractional Programming Techniques for Location Models. *Discrete and Fractional Programming Techniques for Location Models*.

Bazaraa, MS, Sherali, HD and Shetty, CM (1993). *Nonlinear programming : theory and algorithms,* New York [etc.]: Wiley.

Belton, V and Stewart, TJ (2002). Multiple criteria decision analysis: an integrated approach, Boston [etc.]: Kluwer Academic Publishers.

Bharathi, AV, Kurpad, AV, Thomas, T, Yusuf, S, Saraswathi, G and Vaz, M (2008). Development of food frequency questionnaires and a nutrient database for the prospective urban and rural epidemiological (PURE) pilot study in South India: Methodological issues. *Asia Pacific journal of clinical nutrition,* 17**:** 178-185.

Briend, A, Darmon, N, Ferguson, E and Erhardt, JG (2003). Linear programming: A mathematical tool for analyzing and optimizing children's diets during the complementary feeding period. *Journal of Pediatric Gastroenterology and Nutrition,* 36**:** 12-22.

Brill Jr, ED (1979). The Use of Optimization Models in Public-Sector Planning. *Management Science,* 25**:** 413-422.

Brouwer-Brolsma, EM, Berendsen, AA, Vaes, AM, Dullemeijer, C, Groot, LC and Feskens, EJ (2016). Collection and analysis of published scientific information as preparatory work for the setting of Dietary Reference Values for Vitamin D. *EFSA Supporting Publications,* 13.

Buttriss, JL, Briend, A, Darmon, N, Ferguson, EL, Maillot, M and Lluch, A (2014). Diet modelling: How it can inform the development of dietary recommendations and public health policy. *Nutrition Bulletin,* 39**:** 115-125.

Caballero, R, Gómez, T and Ruiz, F (2009). Goal Programming: realistic targets for the near future. *Journal of Multi-Criteria Decision Analysis,* 16**:** 79-110.

Cade, J, Thompson, R, Burley, V and Warm, D (2002). Development, validation and utilisation of food-frequency questionnaires - A review. *Public Health Nutrition,* 5**:** 567-587.

Carlson, A, Lino, M and Fungwe, T (2007). The Low-Cost, Moderate-Cost, and Liberal Food Plans http://www.cnpp.usda.gov/Publications/FoodPlans/MiscPubs/FoodPlans2007AdminReport.pdf; accessed 25 April 2014: CNPP-20.

Carroll, RJ, Midthune, D, Subar, AF, Shumakovich, M, Freedman, LS, Thompson, FE and Kipnis, V (2012). Taking Advantage of the Strengths of 2 Different Dietary Assessment Instruments to Improve Intake Estimates for Nutritional Epidemiology. *American Journal of Epidemiology,* 175**:** 340-347.

Charnes, A and Cooper, WW (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly,* 9**:** 181-186.

Charnes, A and Cooper, WW (1977). Goal programming and multiple objective optimizations: Part 1. *European Journal of Operational Research,* 1**:** 39-54.

Claassen, GDH (2014a). Mixed integer (0-1) fractional programming for decision support in paper production industry. *Omega (United Kingdom),* 43**:** 21-29.

Claassen, GDH (2014b). Optimization-based decision support systems for planning problems in processing industries. Wageningen: Wageningen University.

Claassen, GDH, Hendriks, THB and Hendrix, EMT (2007). *Decision Science: theory and applications,* Wageningen: Wageningen Academic Publishers.

Cleveland, LE, Escobar, AJ, Lutz, SM and Welsh, SO (1993). Method for identifying differences between existing food intake patterns and patterns that meet nutrition recommendations. *Journal of the American Dietetic Association,* 93**:** 556-563.

Cooper, WW (2005). Origins, uses of, and relations between goal programming and data envelopment analysis. *Journal of Multi-Criteria Decision Analysis,* 13**:** 3-11.

Cooper, WW and McAlister, L (1999). Can research be basic and applied? You bet. It better be for B-schools! *Socio-Economic Planning Sciences,* 33**:** 257-276.

Dantzig, GB (1960). On the Significance of Solving Linear Programming Problems with Some Integer Variables. *Econometrica,* 28**:** 30-44.

Darmon, N, Darmon, M and Ferguson, E (2006a). Identification of nutritionally adequate mixtures of vegetable oils by linear programming. *Journal of Human Nutrition and Dietetics,* 19**:** 59-69.

Darmon, N, Ferguson, E and Briend, A (2002a). Linear and nonlinear programming to optimize the nutrient density of a population's diet: An example based on diets of preschool children in rural Malawi. *American Journal of Clinical Nutrition,* 75**:** 245-253.

Darmon, N, Ferguson, EL and Briend, A (2002b). A cost constraint alone has adverse effects on food selection and nutrient density: An analysis of human diets by linear programming. *Journal of Nutrition,* 132**:** 3764-3771.

Darmon, N, Ferguson, EL and Briend, A (2006b). Impact of a Cost Constraint on Nutritionally Adequate Food Choices for French Women: An Analysis by Linear Programming. *Journal of Nutrition Education and Behavior,* 38**:** 82-90.

Ferguson, EL, Darmon, N, Fahmida, U, Fitriyanti, S, Harper, TB and Premachandra, IM (2006). Design of optimal food-based complementary feeding recommendations and identification of key "problem nutrients" using goal programming. *Journal of Nutrition,* 136**:** 2399-2404.

Fern, EB, Watzke, H, Barclay, DV, Roulin, A and Drewnowski, A (2015). The nutrient balance concept: A new quality metric for composite meals and diets. *PLoS ONE,* 10.

Fletcher, LR, Soden, PM and Zinober, ASI (1994). Linear programming techniques for the construction of palatable human diets. *Journal of the Operational Research Society,* 45**:** 489-496.

Fresco, LO and Poppe, KJ (2016). Towards a Common Agricultural and Food Policy: Wageningen University & Research.

Fulgoni Iii, VL, Keast, DR and Drewnowski, A (2009). Development and validation of the nutrient-rich foods index: A tool to measure nutritional quality of foods. *Journal of Nutrition,* 139**:** 1549-1554.

Gao, X, Wilde, PE, Lichtenstein, AH and Tucker, KL (2006). The 2005 USDA food guide pyramid is associated with more adequate nutrient intakes within energy constraints than the 1992 pyramid. *Journal of Nutrition,* 136**:** 1341-1346.

Gedrich, K, Hensel, A, Binder, I and Karg, G (1999). How optimal are computer-calculated optimal diets? *European Journal of Clinical Nutrition,* 53**:** 309-318.

Gerdessen, JC and De Vries, JHM (2015). Diet models with linear goal programming: impact of achievement functions. *European Journal of Clinical Nutrition,* 69**:** 1272-1278.

Glover, F (1975). Improved linear integer programming formulations of nonlinear integer problems. *Management Science,* 22**:** 455-460.

Gonzalez, CA and Riboli, E (2010). Diet and cancer prevention: Contributions from the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *European Journal of Cancer,* 46**:** 2555-2562.

H2020 (website). SUSFANS: Metrics, Models and Foresight for European SUStainable Food And Nutrition Security. *http://www.susfans.org/,* visited 2017-04-07.

Health Council of the Netherlands (2006). Guidelines for a healthy diet 2006. The Hague: Health Council of the Netherlands.

Health Council of The Netherlands (2011). Guidelines for a healthy diet: the ecological perspective. The Hague: Health Council of the Netherlands.

Hooker, JN and Williams, HP (2012). Combining equity and utilitarianism in a mathematical programming model. *Management Science,* 58**:** 1682-1693.

Imamura, F, Micha, R, Khatibzadeh, S, Fahimi, S, Shi, P, Powles, J and Mozaffarian, D (2015). Dietary quality among men and women in 187 countries in 1990 and 2010: A systematic assessment. *The Lancet Global Health,* 3**:** e132-e142.

INFORMS (2017). What is Operations Research? : https://www.informs.org/About-INFORMS/What-is-Operations-Research, visited 11 Feb 2017.

International Food Policy Research Institute (2016). Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030. Washington, DC.

Jones, D (2011). A practical weight sensitivity algorithm for goal and multiple objective programming. *European Journal of Operational Research,* 213**:** 238-245.

Jones, D and Jimenez, M (2013). Incorporating additional meta-objectives into the extended lexicographic goal programming framework. *European Journal of Operational Research,* 227**:** 343-349.

Jones, D and Tamiz, M (2010). *Practical Goal Programming,* Boston, MA: Springer Science+Business Media, LLC.

Korhonen, PJ, Silvennoinen, K, Wallenius, J and Öörni, A (2013). A careful look at the importance of criteria and weights. *Annals of Operations Research,* 211**:** 565-578.

Lambe, J, Kearney, J, Leclercq, C, Zunft, HFJ, De Henauw, S, Lamberg-Allardt, CJE, Dunne, A and Gibney, MJ (2000). The influence of survey duration on estimates of food intakes and its relevance for public health nutrition and food safety issues. *European Journal of Clinical Nutrition,* 54**:** 166-173.

Li, HL (1994). A global approach for general 0-1 fractional programming. *European Journal of Operational Research,* 73**:** 590-596.

Lim, SS, Vos, T, Flaxman, AD, Danaei, G, Shibuya, K, Adair-Rohani, H, Amann, M, Anderson, HR, Andrews, KG, Aryee, M, Atkinson, C, Bacchus, LJ, Bahalim, AN, Balakrishnan, K, Balmes, J, Barker-Collo, S, Baxter, A, Bell, ML, Blore, JD, Blyth, F, Bonner, C, Borges, G, Bourne, R, Boussinesq, M, Brauer, M, Brooks, P, Bruce, NG, Brunekreef, B, Bryan-Hancock, C, Bucello, C, Buchbinder, R, Bull, F, Burnett, RT, Byers, TE, Calabria, B, Carapetis, J, Carnahan, E, Chafe, Z, Charlson, F, Chen, H, Chen, JS, Cheng, ATA, Child, JC, Cohen, A, Colson, KE, Cowie, BC, Darby, S, Darling, S, Davis, A, Degenhardt, L, Dentener, F, Des Jarlais, DC, Devries, K, Dherani, M, Ding, EL, Dorsey, ER, Driscoll, T, Edmond, K, Ali, SE, Engell, RE, Erwin, PJ, Fahimi, S, Falder, G, Farzadfar, F, Ferrari, A, Finucane, MM, Flaxman, S, Fowkes, FGR, Freedman, G, Freeman, MK, Gakidou, E, Ghosh, S, Giovannucci, E, Gmel, G, Graham, K, Grainger, R, Grant, B, Gunnell, D, Gutierrez, HR, Hall, W, Hoek, HW, Hogan, A, Hosgood Iii, HD, Hoy, D, Hu, H, Hubbell, BJ, Hutchings, SJ, Ibeanusi, SE, Jacklyn, GL, Jasrasaria, R, Jonas, JB, Kan, H, Kanis, JA, Kassebaum, N, Kawakami, N, Khang, YH, Khatibzadeh, S, Khoo, JP, Kok, C, Laden, F, et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet,* 380**:** 2224-2260.

Lobstein, T and Davies, S (2009). Defining and labelling 'healthy' and 'unhealthy' food. *Public Health Nutrition,* 12**:** 331-340.

Maillot, M, Vieux, F, Amiot, MJ and Darmon, N (2010). Individual diet modeling translates nutrient recommendations into realistic and individual-specific food choices. *American Journal of Clinical Nutrition,* 91**:** 421-430.

Makowski, D, Hendrix, EMT, Van Ittersum, MK and Rossing, WAH (2000). A framework to study nearly optimal solutions of linear programming models developed for agricultural land use exploration. *Ecological Modelling,* 131**:** 65-77.

Makowski, D, Hendrix, EMT, Van Ittersum, MK and Rossing, WAH (2001). Generation and presentation of nearly optimal solutions for mixed-integer linear programming, applied to a case in farming system design. *European Journal of Operational Research,* 132**:** 425-438.

Mark, SD, Thomas, DG and Decarli, A (1996). Measurement of exposure to nutrients: An approach to the selection of informative foods. *American Journal of Epidemiology,* 143**:** 514-521.

Masset, G, Monsivais, P, Maillot, M, Darmon, N and Drewnowski, A (2009). Diet optimization methods can help translate dietary guidelines into a cancer prevention food plan. *Journal of Nutrition,* 139**:** 1541-1548.

Mathers, CD and Loncar, D (2006). Projections of global mortality and burden of disease from 2002 to 2030. *Plos med,* 3**:** e442.

Mertens, E, Van'T Veer, P, Hiddink, GJ, Steijns, JM and Kuijsten, A (2017). Operationalising the health aspects of sustainable diets: A review. *Public Health Nutrition,* 20**:** 739-757.

Miettinen, K (2008). Introduction to Multiobjective Optimization: Noninteractive Approaches. in: J Branke, K Deb, K Miettinen and R Slowinski (eds.) *Multiobjective Optimization : Interactive and Evolutionary Approaches.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Molag, ML (2010). Towards transparent development of food frequency questionnaires : scientific basis of the Dutch FFQ-TOOL: a computer system to generate, apply and process FFQs. [S.l.: s.n.].

Molag, ML, de Vries, JHM, Duif, N, Ocké, MC, Dagnelie, PC, Goldbohm, RA and van 't Veer, P (2010). Selecting informative food items for compiling food-frequency questionnaires: Comparison of procedures. *British Journal of Nutrition,* 104**:** 446-456.

Molag, ML, de Vries, JHM, Ocké, MC, Dagnelie, PC, van den Brandt, PA, Jansen, MCJF, van Staveren, WA and van 't Veer, P (2007). Design characteristics of food frequency questionnaires in relation to their validity. *American Journal of Epidemiology,* 166**:** 1468-1478.

NEVO (1996). *Nederlands Voedingsmiddelentabel (Dutch Food Composition Table).* Den Haag: De Commissie Nederlandse Voedingsmiddelentabel van de Voedingsraad (in Dutch).

Nordic council of Ministers (2014). *Nordic Recommendations 2012. Integrating nutrition and physical activity. 5th Edition,* Copenhagen: Nordisk Ministerråd.

Notarnicola, B, Tassielli, G, Renzulli, PA, Castellani, V and Sala, S (2017). Environmental impacts of food consumption in Europe. *Journal of Cleaner Production,* 140, Part 2**:** 753-765.

NWO (2017). GREENDISH: Food supply chains delivering sustainable and healthy diets. *http://www.nwo.nl/en/research-and-results/research-projects/i/68/13068.html,* visited 2017-03-24.

Petersen and Clifford, C (1971). A Note on Transforming the Product of Variables to Linear Form in Linear Programs. *Working Paper, Purdue University*.

Rawls, J (1971). *A theory of justice,* Cambridge, MA: Harvard University Press.

RIVM (2012). Dutch food composition database NEVO 2011. Bilthoven.

Romero, C (2001). Extended lexicographic goal programming: A unifying approach. *Omega,* 29**:** 63-71.

Romero, C (2004). A general structure of achievement function for a goal programming model. *European Journal of Operational Research,* 153**:** 675-686.

Romero, C and Rehman, T (2003). *Multiple criteria analysis for agricultural decisions,* Amsterdam [etc.]: Elsevier.

Romero, C, Tamiz, M and Jones, DF (1998). Goal Programming, Compromise Programming and Reference Point Method Formulations: Linkages and Utility Interpretations. *The Journal of the Operational Research Society,* 49**:** 986-991.

Schaible, S (1995). Fractional Programming. in: R Horst and PM Pardalos (eds.) *Handbook of Global Optimization.* Dordrecht: Kluwer Academic Publishers.

Schaible, S and Shi, J (2004). Recent developments in fractional programming: single-ratio and max-min case. in: W Takahashi and T T (eds.) *Proceedings of the 3rd International Conference in Nonlinear Analysis and Convex Analysis, Aug.* Yokohama Publishing.

Shao, L and Ehrgott, M (2016). Discrete representation of non-dominated sets in multi-objective linear programming. *European Journal of Operational Research,* 255**:** 687-698.

Stancu-Minasian, IM (1999). A fifth bibliography of fractional programming. *Optimization,* 45**:** 343-367.

Stancu-Minasian, IM (2006). A sixth bibliography of fractional programming. *Optimization,* 55**:** 405-428.

Tamiz, M, Jones, D and Romero, C (1998). Goal programming for decision making: An overview of the current state-of-the-art. *European Journal of Operational Research,* 111**:** 569-581.

The Dutch Nutrition Centre (1998). *Zo eet Nederland: Resultaten van de Voedselconsumptiepeiling 1997–1998 (Results of the Dutch Food Consumption Survey 1997/1998),* Den Haag: Voedingscentrum (in Dutch).

Thompson, FE and Byers, T (1994). Dietary Assessment Resource Manual. *Journal of Nutrition,* 124**:** S2245-S2317.

Thompson, S, Gower, R and Darmon, N (2013). LiveWell for Life. A balance of healthy and sustainable food choices for France, Spain, and Sweden. http://livewellforlife.eu/livewell-plate/reports; accessed 25 April 2014 WWF Report.

TIFN (2017). SHARP-BASIC. *http://www.tifn.nl/project/sharp/,* visited 2017-03-27.

Wakai, K (2009). A review of food frequency questionnaires developed and validated in Japan. *Journal of Epidemiology,* 19**:** 1-11.

WHO (2013a). Global action plan for the prevention and control of noncommunicable diseases 2013-2020.

WHO (2013b). Vienna declaration on nutrition and noncommunicable diseases in the context of health 2020. *WHO Ministerial Conference on Nutrition and Noncommunicable Diseases in the Context of Health.*

WHO (2014). *Global status report on noncommunicable diseases 2014*: World Health Organization.

WHO (2017). Fact sheet noncommunicable diseases.

Willett, W (1998). *Nutritional epidemiology,* New York [etc.]: Oxford University Press.

Wirsam, B, Hahn, A, Uthus, EO and Leitzmann, C (1997). Fuzzy sets and fuzzy decision making in nutrition. *European Journal of Clinical Nutrition,* 51**:** 286-296.

Wirsam, B and Uthus, EO (1996). The use of fuzzy logic in nutrition. *Journal of Nutrition,* 126**:** 2337S-2341S.

Wu, TH (1997). A note on a global approach for general 0-1 fractional programming. *European Journal of Operational Research,* 101**:** 220-223.

Yaghoobi, MA and Tamiz, M (2007). A method for solving fuzzy goal programming problems based on MINMAX approach. *European Journal of Operational Research,* 177**:** 1580-1590.

Acknowledgements

The cover of this thesis aims to express that its contents deal with comparing incomparable items like apples and oranges (in Dutch: 'appels en peren vergelijken'). In practice, not all aspects of apples and oranges can be described exactly, which is why not real apples and oranges are depicted, but simplified representations. Yet, these representations are immediately recognisable as apples and oranges, which is also what I aimed for when approximating real-life nutritional problems via mathematical models and techniques. The models in this thesis aim to pursue multiple objectives at the same time. Commonly, these objectives are conflicting, which necessitates us to seek some sort of compromise. This holds in real life as well. To this thesis, multiple people have contributed in incomparable ways, that often cannot be described exactly. That is why also this section of acknowledgements is a simplified representation of reality.

Jack, thank you for the many years of pleasant cooperation. When this project started, you were quite unfamiliar with nutrition research. Yet, you recognised potential, and a few years later you were highly successful in obtaining funds for extending *OR in Nutrition* from "Joke's peculiar little niche" into a substantial research stream within the ORL group. You were constructive and supportive whenever I asked your input or advice, having a keen eye for the big picture while I was toying with the details. I appreciate your effort for finishing this project after you moved up to the first floor, and I will miss the "excuses" for popping into your office every now and then.

Pieter, this thesis research started with a cup of coffee, during which we explained to each other what the disciplines of nutrition research and operations research encompass. Within five minutes, we identified several opportunities for decision support in nutrition research, and soon after that we set off optimising FFQs. I greatly appreciate your scientific curiosity, your enthusiasm, and the way in which you structure things and put them into perspective. I am looking forward to many more cups of coffee to be enjoyed!

Frits, you stood by me when I was fuming with adrenaline, either from euphoria or from deception. You patiently endured the frustration that I had utter fun toying with the content while all you could do was sit and watch and hope that you would find time to keep up with the updates that kept flowing in. You sharpened my mind and my output whenever we were sparring about elevating the bar, tending the garden, or tackling any other real-life multi-criteria challenge.

Jeanne, you were always curious what mathematical models could contribute to nutrition research. Between your very many pressing duties you managed to squeeze in time for seeking sensible nutritional interpretation for my latest findings and ideas, and ensuring that our work met solid nutritional standards.

Aleksander and Argyris, you are the best roommates I could wish for. Aleksander, your well-intended contribution surely set me off into the right direction! Guys, your smiling and teasing presence provides superior force :)

Karin and Frits, thank you for being not only good colleagues but also very dear friends, being in the same boat, sharing chats, laughs, tears and hikes, and providing a stationary supply of supportive jokes.

Argyris and Karin, I am glad to have you on my side.

Natasja, thank you for pushing the right buttons at the right times. Ilona, Jeanette, Claudia, Anne, Sylvia, Jacqueline, Didi, thank you for managing me and my promo-team, arranging meetings and answering many questions. Colleagues from ORL and INF, your friendly chats and banter contribute greatly to the pleasant working atmosphere at our corridor.

Matthijs and Timon, after my first paper was published, you were shocked that all these months of hard work had resulted in only four pages of printed text. Fortunately, the page count of the final booklet exceeds the magical bound of one hundred, which hopefully repairs my in-home reputation. Wim, my loving anchor and foundation, you supported me in every possible way. The three of you endured my absent-mindedness, my ever-increasing amnesia, the kitchen table stacked with paper work and mice roaming the premises. Things might be on the mend … :)

This thesis illustrates both the value and the limitations of operations research. It can offer decision support, but it cannot fully take the pain out of nutritional and other decision-making; it merely makes us aware where our biggest challenges are. Fortunately, real-life apples and oranges can be fully enjoyed without the slightest necessity for making any comparison or decision!

Joke C. van Lemmen-Gerdessen
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan

Wageningen School
of Social Sciences

Name of the learning activity	**Department/Institute**	**Year**	**ECTS***
A) Project related competences			
Agricultural and natural resources management: a multi-criteria approach	Mansholt Graduate School, Carlos Romero	2010	1.5
Writing research proposal	WUR	2012-2013	6
B) General research related competences			
Techniques for Writing and Presenting a Scientific Paper, Michael Grossman	WGS	2010	1.2
Participate in writing grant proposals	GreenDish, Sharp-Basic	2014	2
C) Career related competences/personal development			
'MCDM in Nutritional Epidemiology. Multicriteria Decision Making'	MCDM Conference, Malaga, Spain	2013	1
'Diet models with linear Goal Programming: impact of achievement functions'	27th Euro Conference, Glasgow, UK	2015	1
'Combining Equity and Utilitarianism - Comparison of Two Approaches in Diet Modelling Context'	28th Euro Conference, Paznań, Poland	2016	1
"Recent Advances in Multi-Objective Optimization", Nantes	Workshop, Nantes, France	2015	1
Multiobjective Optimization	Summerschool ATOM, Lille, France	2017	1
Participation in supervision thesis students	WUR	throughout	4
D) Accreditation of prior learning			12
Didactic courses (Onderwijs BasisKwalificatie en vervolgcursussen)			
Two chapters in a scientific book (Claassen et al, 2007, Decision Science : theory and applications: Wageningen Academic Publishers, Chapters 7 & 10)			
Coordinator and developer of courses for ORL: Kwantitatieve Modellen voor de Beleidsvoorbereiding, Decision Science 2 (a.o. the topic MCDM) since 2003			
Write an article in a peer reviewed journal: Gerdessen JC (1996) Vehicle routing problem with trailers. Eur J Oper Res, 93, 135-147			
Review for European Journal of Operations Research			
Total			**31.7**

*One credit according to ECTS is on average equivalent to 28 hours of study load

Cover: fruits by Katrinaku / Shutterstock, design by Ruben Stoel, Joke van Lemmen
Printed by Digiforce / Proefschriftmaken.nl, Vianen